

INVESTIGAÇÃO E AVALIAÇÃO EM PSICOLOGIA E SAÚDE

JOSÉ LUÍS PAIS RIBEIRO



JOSÉ LUÍS PAIS RIBEIRO

**INVESTIGAÇÃO E AVALIAÇÃO
EM PSICOLOGIA E SAÚDE**

Titulo: *Investigação e Avaliação em Psicologia e Saúde*. 2ª Edição. Outubro 2010.

ISBN: 978-989-8463-01-2.

Capa: Ricardo Romão

Lisboa: Placebo, Editora LDA.

ÍNDICE

PRÓLOGO	1
PRIMEIRA PARTE – INVESTIGAÇÃO	
INICIAR E IMPLEMENTAR UMA INVESTIGAÇÃO	2
Introdução	2
<i>O que diferencia o método científico do senso-comum</i>	2
<i>Raciocínio dedutivo ou indutivo</i>	4
<i>História do controlo em psicologia</i>	6
As fases dos projectos de investigação	8
<i>Iniciar a investigação</i>	8
<i>Recolha de dados</i>	13
<i>Apresentar os resultados</i>	14
O que deve incluir um relatório de investigação	16
<i>Página de título</i>	16
<i>Resumo</i>	17
<i>Introdução</i>	19
<i>Método</i>	21
<i>Resultados</i>	25
<i>Discussão</i>	27
<i>Referências</i>	27

<i>Tipos de trabalhos escritos</i>	28
ASPECTOS METODOLÓGICOS DE UMA INVESTIGAÇÃO	30
Desenhos de investigação	31
<i>Experimental</i>	30
<i>Observacional</i>	30
Métodos de investigação	32
<i>Métodos epidemiológicos</i>	32
<i>Métodos de sondagem</i>	33
<i>Métodos experimentais</i>	34
<i>Métodos clínicos</i>	35
Validade dos desenhos de investigação	37
<i>Validade interna</i>	37
<i>Validade externa</i>	39
População ou universo e amostra	41
<i>Porquê da amostragem</i>	42
<i>Tipos de amostra</i>	42
<i>Amostras não probabilísticas ou intencionais</i>	43
<i>Amostras probabilísticas</i>	45
<i>Quantos indivíduos deve ter uma amostra</i>	46
Análise estatística	49
<i>Estatística paramétrica versus não paramétrica</i>	49
<i>Análise correlacional versus experimental</i>	50
<i>Análise inferencial versus descritiva</i>	50

<i>Análise univariada versus multivariada</i>	51
<i>Análise estatística</i>	51
<i>Análise bivariada</i>	53
<i>Erro tipo I e II</i>	55
<i>Testes unicaudais versus bicaudais</i>	56
 ASPECTOS ÉTICOS DA INVESTIGAÇÃO	 57
Ética na investigação	57
<i>Aspectos gerais da ética na investigação no campo da saúde e doenças</i>	57
<i>Medidas relativas à autoria dos resultados</i>	61
 SEGUNDA PARTE – AVALIAÇÃO	
O QUE AVALIA A AVALIAÇÃO	63
Introdução	63
Definição de avaliação psicológica	64
Tipo de técnicas utilizadas na avaliação psicológica	64
<i>Questionários</i>	65
<i>Testes</i>	66
<i>Inventários</i>	67
<i>Escalas</i>	67
<i>Entrevista</i>	68
<i>Observação naturalista</i>	68
<i>Listas de verificação</i>	68
<i>Listas de classificação</i>	69

Literácia e avaliação psicológica	69
Tipos de testes psicológicos	70
<i>Testes individuais versus colectivos</i>	71
<i>Testes objectivos e não objectivos</i>	71
<i>Testes de papel e lápis</i>	71
<i>Testes de cultura equilibrada</i>	71
<i>Testes livres de cultura</i>	72
<i>Testes de referência a critério ou a conteúdo</i>	72
Tipo de preenchimento	72
<i>Testes preenchidos pelo próprio</i>	73
<i>Testes preenchidos pelo testador</i>	73
Preocupações relacionadas com o preenchimento	74
<i>Desejabilidade social</i>	74
<i>Resposta ao calha</i>	74
<i>Falsificação</i>	74
<i>Estilo de resposta</i>	75
O que avalia a avaliação psicológica	75
<i>Inteligência</i>	75
<i>Personalidade</i>	79
<i>O que é específico na utilização dos teste em saúde</i>	87
<i>Quando é que a avaliação em saúde é avaliação psicológica</i>	89
INTERPRETAÇÃO DAS RESPOSTAS	90
Tipos de escalas de medida	91

<i>Medidas nominais/categoriais</i>	93
<i>Medidas ordinais</i>	93
<i>Escalas de intervalo</i>	93
<i>Medidas de rácio</i>	94
Técnicas de medição/tipo de escala	94
<i>Métodos de classificação categorial</i>	94
<i>Escala de Likert</i>	95
<i>Escala análoga visual</i>	96
<i>Escala de Thurstone</i>	96
<i>Escala de Gutman</i>	98
<i>Diferencial semântico</i>	98
<i>Escolha forçada</i>	98
<i>Dicotómicas</i>	99
Propriedades psicométricas	99
<i>Propriedades psicométricas dos testes</i>	99
<i>Propriedades psicométricas da avaliação comportamental</i>	104
Validade dos itens	104
<i>Índice de dificuldade do item</i>	105
<i>Índice de discriminação do item</i>	105
<i>Distribuição das respostas</i>	106
<i>Correlação do item com a nota total do teste</i>	106
<i>Consistência interna do item</i>	106
<i>Validade convergente-discriminante do item</i>	107
<i>Itens dos testes de referência a critério (ou de referência a conteúdos)</i>	107

Técnicas Psicométricas ou Clinimétricas	107
Outras características recomendadas na avaliação em saúde	109
<i>Apropriabilidade</i>	<i>109</i>
<i>Economia (burden)</i>	<i>109</i>
<i>Adequabilidade</i>	<i>111</i>
<i>Interpretabilidade</i>	<i>111</i>
<i>Conformidade</i>	<i>112</i>
<i>Aceitabilidade</i>	<i>112</i>
<i>Utilidade</i>	<i>112</i>
CONSTRUÇÃO DE TESTES	113
Introdução	113
<i>Com base no conteúdo dos itens</i>	<i>113</i>
<i>Com base em critérios empíricos</i>	<i>113</i>
<i>Com base na análise factorial</i>	<i>114</i>
<i>Com base em teorias da personalidade</i>	<i>114</i>
Estudos em culturas e línguas diferentes	114
<i>A construção de técnicas de avaliação transculturais</i>	<i>115</i>
<i>Como se costuma fazer a adaptação de instrumentos</i>	<i>126</i>
<i>Hipótese psicolexical</i>	<i>127</i>
<i>Abordagem absolutista versus abordagem universalista</i>	<i>118</i>
<i>Aspectos técnicos na tradução</i>	<i>118</i>
<i>Aspectos a considerar quando se opta pela tradução</i>	<i>119</i>
Como se constrói uma escala	123

Ética na avaliação	128
Direitos de autor dos testes	128
REFERÊNCIAS	131

PRÓLOGO

Investigação e avaliação em Psicologia e Saúde

O presente livro constitui uma cópia do que foi publicado na década de 90 e responde a diversas solicitações tanto em Portugal como noutros países de língua Portuguesa, sobre como lhe aceder.

A reedição do texto de 1999 propõe-se disponibilizar a versão original do livro entretanto deixada de publicar. Os aspectos ligados à metodologia – enquanto, por definição, ciência que estuda os métodos – e ao desenvolvimento de técnicas ou instrumentos de avaliação psicológica continuam no centro do conhecimento psicológico. Os métodos porque estão por detrás de toda a investigação em psicologia, e a avaliação psicológica porque suporta a investigação e a intervenção.

Hoje a utilização das técnicas de avaliação que foram desenvolvidas no seio da psicologia, e que constituíram a sua fundamentação científica nos primeiros decénios, são partilhadas por muitas outras disciplinas quer na aplicação profissional quer na investigação. Tal é lisonjeiro para nós psicólogos, mas traz a responsabilidade de chamar a atenção para a importância do rigor no desenvolvimento e uso das técnicas de avaliação.

A iniciativa da editora Placebo merece assim todo o nosso apoio e reconhecimento, por facilitar o acesso a um texto que por alguns é reconhecido como um clássico escrito em Português.

José Luís Pais Ribeiro

PRIMEIRA PARTE – INVESTIGAÇÃO

INICIAR E IMPLEMENTAR UMA INVESTIGAÇÃO

Introdução

Ciência é entendida como um conjunto de conhecimentos humanos a respeito da natureza, da sociedade e do pensamento, adquiridos pela descoberta e explicação das leis objectivas que regem os fenómenos. Salienta-se que por definição se impõe que a ciência é o conhecimento deriva das leis objectivas que regem os fenómenos e a sua explicação.

Ciência tanto se refere ao conjunto de informação acumulada de forma sistemática, como aos meios que foram utilizados para obter esse conhecimento. O meio utilizado para obter o conhecimento científico é o método científico, e o método científico refere-se a um conjunto de procedimentos objectivos. Esta “objectividade” impõe limites ao que é “científico”. A ciência consiste num jogo muito estruturado que reflecte a história, a cultura, a filosofia da cultura ocidental.

Se a ciência reflecte a filosofia, história e cultura, facilmente se compreende porque muitas práticas provenientes de outras culturas são consideradas não científicas na medida em que não decorrem nem têm subjacentes leis objectivas. Não são científicas porque foram estabelecidas à margem das regras deste jogo chamado “ciência”.

O que diferencia o método científico do senso-comum

Uma das características dos seres humanos é a sua capacidade para interrogarem o que observam e de arranjar teoria para o explicarem. Fazem-no a partir da observação e fundamentados no conhecimento que têm. No entanto tal não cai na definição de ciência. O que diferencia, então, o método do senso comum do método científico? A diferença fundamental reside em que o cientista se baseia

num conjunto de regras alicerçadas e fundamentadas por uma filosofia que é aceite pela ciência. Bem e de Jong (1998) explicam que há um continuo entre o senso comum e a ciência. A ciência emergiu da necessidade das pessoas resolverem os problemas do dia a dia. A forma como se entende hoje a ciência nasce no final da idade média. O que distingue a ciência tal como começa a ser entendida é o carácter sistemático de perguntar e responder às perguntas. Bem e de Jong listam as seguintes características do método científico: é sistemático no sentido que tem de se aplicar a todos os fenómenos com as mesmas características; recorre a métodos claramente definidos; é redutor no sentido em que ignora aspectos da realidade que são distractores, e em que reduz os fenómenos a princípios gerais; é objectivo, no sentido em que é controlável e em que diferentes observadores que utilizassem os mesmos procedimentos chegariam aos mesmos resultados; é claro porque os conceitos utilizados são definidos sem ambiguidade; nunca termina no sentido em que no futuro será aperfeiçoado.

O método científico consiste num sistema de regras e de processos na qual se baseia a investigação (Bowling, 1998). A ciência suporta a, e suporta-se na, investigação. É esta que permite afirmar aquela. O jogo regrado da ciência inclui o jogo regrado da investigação.

O objectivo do método científico é fazer afirmações sem erro. O que é afirmado é “verdade” dentro das regras que a ciência estipula, deixando de parte os aspectos irrelevantes. Sendo a investigação um processo delicado e complexo inclui várias fases e só é investigação se o relatório final for tão detalhado que permita a outro investigador, ou outra equipa de investigação, seguir, exactamente, os mesmos passos e chegar, ou não, aos mesmos resultados. Trata-se, pois, de um exercício técnico detalhado e com regras precisas.

Um aspecto que caracteriza uma investigação é “controlo”. Com efeito, numa investigação, todos os passos são meticulosamente controlados. Perder um passo, saltar um passo, deixa de se poder falar em ciência, inviabiliza-se a investigação. A fronteira entre o que é ciência e o que não é ciência é muito ténue e baseia-se fundamentalmente no processo como foi conduzida a investigação e produzido o conhecimento.

Uma investigação começa por uma grande questão. O investigador, na sua prática diária, reparou em algo para que não conhece explicação, ou então os modelos ou teorias disponíveis sugerem relações entre variáveis que interessa confirmar ou desafiar. Desta questão emergem variáveis e relações entre variáveis. Num momento seguinte, o investigador procede a pesquisa em base de dados de modo a verificar se encontra estudos que respondam à sua pergunta. Hoje, as bases de dados são cada vez maiores e mais acessíveis pelo que este exercício fundamental é simples. O problema que se coloca agora, aliás, é o

excesso de informação que provém de uma base de dados, e o de saber se as próprias bases, ou as palavras seleccionadas para proceder à pesquisa na base de dados são as mais adequadas.

Num terceiro momento, com base na informação recolhida, desenha-se o projecto de investigação que deve, incluir, para além da questão de investigação já depurada pela pesquisa, os objectivos, as variáveis as técnicas de recolha de dados ou de intervenção, a população alvo da investigação e o modo como se selecciona a amostra, e os momentos de recolha de dados ou de intervenção.

Num quarto momento procede-se à implementação do projecto, introduzem-se os dados em base de dados, analisam-se e escrevem-se os resultados.

Finalmente, num quinto momento, atinge-se a fase mais criativa da investigação, que é a de discutir os resultados. Com efeito, neste momento, a pergunta inicial, clarificada com as leituras ou consultas que se fizeram acerca de resultados de outras investigações, e os resultados encontrados, permitem responder à pergunta inicial do investigador e formular novas questões para outros projectos de investigação. Para além das respostas encontradas e, talvez mais importante, são as questões que os resultados levantam.

Por último resta divulgar o estudo, o que deve ser feito numa revista científica (*peer reviewed*) que garanta que chega aos cientistas interessados no tema. Esta última parte é importante porque evita que outros iniciem investigações que se proponham responder à mesma questão, e sugere questões para outras investigações. A publicação do relatório da investigação contribui para o avanço do conhecimento

Qual é a parte mais importante da investigação? Todas são importantes. Ocorrem em momentos diferentes mas se alguma for mal realizada o esforço total do projecto é prejudicado.

Raciocínio dedutivo ou indutivo

A ciência pratica-se de muitas maneiras, desde a observação de acontecimentos que ocorrem de modo natural, até a acontecimentos que ocorrem em consequência de procedimentos experimentais mais ou menos complexos. Os cientistas podem fazer ciência construindo teorias, podem realizar experiências, podem observar a natureza à margem de qualquer teoria descrevendo somente o que ocorre, etc. O que constrói a teoria não necessita de se envolver em experimentação e o que

implementa uma experiência pode não mergulhar no mar profundo da teoria. O observador não manipula as variáveis enquanto o experimentador manipula fazendo variar as variáveis. O avanço da ciência depende da teia construída por todos estes tipos de cientistas, e todos eles têm mérito.

A observação é investigação de fenómenos naturais, é naturalística, é orientada pelos factos, e indutiva. A experimentação tende a ser artificial, derivada do raciocínio lógico, orientada pelos problemas e pelas questões de investigação, e dedutiva. Ambos são orientados pelo princípio da “objectividade” e da “verdade”.

O desenvolvimento do raciocínio dedutivo ou do raciocínio indutivo constituem uma parte importante do conhecimento e do pensamento científico. Segundo Pinto (1990) o raciocínio indutivo foi proposto inicialmente no século XVII e o raciocínio dedutivo foi defendido já no presente século.

Segundo o *raciocínio dedutivo* o investigador começa com uma teoria e com hipóteses verificáveis a partir de ideias gerais, gerando dados e verificando, de seguida, as hipóteses.

O *método indutivo* propõe que se parta da observação para a teoria. Começava-se por efectuar inúmeras observações sem nenhuns pressupostos, de modo inocente. Os dados seriam analisados de modo a clarificar a sua associação das variáveis em estudo, para, por fim, se produzirem enunciados gerais que depois constituiriam hipótese para outras investigações realizadas com outras condições. Na psicologia John Locke popularizou o método indutivo ajudando a estabelecer o empirismo que salienta a importância de se fazerem observações, ao invés de afirmações teóricas (Bowling, 1998).

Popper (1959) argumenta que o método indutivo está errado, e que toda a observação pressupõe uma teoria: não há observação científica inocente. A observação parte de um conjunto de postulados que impõem que a observação seja feita de um modo e não de outro, e que a interpretação do que é observado seja feita à luz da teoria que está por detrás da observação, e não de outra. Neste método o primeiro passo da investigação será de, com base na teoria, elaborar a questão de investigação e, de seguida desenhar um procedimento que lhe permita responder à questão.

Se uma teoria, um modelo permite deduzir que as relações entre determinadas variáveis serão de um tipo, então tal será verdade até que a investigação permita refutar a teoria. Por outro lado, também na perspectiva do método dedutivo, uma teoria, qualquer teoria, só será verdadeira se tiver sido submetida ao teste da refutação e lhe sobreviver.

Na perspectiva do método dedutivo resulta, ainda, uma dimensão prática que afirma que uma teoria para ser científica tem de ser susceptível de se submeter à prova da investigação. Quanto mais fácil de verificar for uma teoria melhor será essa teoria.

Parece haver um dilema sobre a natureza do método científico, nomeadamente sobre o papel da indução e da dedução. Poderá dizer-se que o método dedutivo suplantou o método indutivo e que este está ultrapassado? Não. Consoante a investigação que se está a fazer, assim o método utilizado se pode aproximar do método dedutivo ou indutivo. O método científico envolve ambos os procedimentos: o método indutivo salienta a observação cuidada e consequente generalização; o método dedutivo salienta o teste das hipóteses que emergiram das teorias.

História do controlo em psicologia

A investigação impõe um rigoroso controlo das variáveis em estudo. Só com este controlo rigoroso das variáveis se pode falar em ciência. A investigação em psicologia afirmou-se como ciência pela adopção de procedimentos importados da física, procedimentos esses que garantiam a “boa” investigação. Este padrão de controlo rigoroso de variáveis provém da física, que em condições laboratoriais rigorosamente concebidas e controladas consegue isolar as variáveis mais complexas. No entanto, reconhece-se que em ciências humanas um controlo semelhante ao da física é impossível.

Ora, em psicologia da saúde, as questões de investigação debruçam-se normalmente sobre problemas, populações, contextos que são praticamente impossíveis de reproduzir em laboratório, entre outras coisas, porque envolvem procedimentos muito stressantes, interacções com outros significativos não reproduzíveis, e/ou problemas que decorrem ao longo de um tempo prolongado. Ou seja, os planos de investigação com um controlo rigoroso do que constitui ameaça à validade da investigação não se pode aplicar à maioria de situações concretas com que se depara a psicologia que se pratica em contexto de saúde. Os procedimentos de investigação muitas vezes não são implementados como se planeou, a distribuição aleatória dos sujeitos nem sempre é possível, e, frequentemente, os sujeitos não chegam ao fim da intervenção.

Segundo West, (1985), o controlo de variáveis tem, historicamente, três origens diferentes. A primeira origem, são as ciências físicas, envolvendo o isolamento do fenómeno que está a ser estudado de factores que a investigação mostrou serem susceptíveis de influenciar o fenómeno estudado. Nas ciências físicas em geral, em que os objectos em estudo são inertes, não mudam com o tempo, e são

uniformes, este procedimento é fácil de implementar. Ora, em psicologia em geral, e em psicologia aplicada em particular, este método não é fácil de aplicar, se não mesmo impossível.

A segunda abordagem sobre o controlo de variáveis que influenciam a investigação em ciências humanas provém da agricultura. Envolve a distribuição aleatória dos sujeitos por diferentes tratamentos, mas, ao invés dos físicos os investigadores em agricultura estão interessados em estudar os fenómenos em condições naturais e não em laboratório. Ao invés de, para verificarem se uma nova estripe de planta é mais produtiva se se controlar a chuva os ventos, as condições do solo, ou a quantidade de sol, interessam-se por mostrar que um determinado tratamento produz efeitos *apesar* das variações *normais* daqueles elementos. As condições do meio variam, mas o sujeito (a planta) mantém as mesmas condições.

Este procedimento é adequado quando se pode assumir que as unidades experimentais são recipientes relativamente passivos. Ora, em ciências humanas e em circunstâncias reais, os sujeitos são activos, influenciam os tratamentos porque têm conhecimentos diferentes, motivações diferentes, alterando com a sua acção ou inacção, a igualdade das condições de tratamento. Por isso a distribuição aleatória só por si pode não ser suficiente.

Uma terceira abordagem ao controlo das variáveis foi implementado pela economia, pela ciência política, pela sociologia, onde a manipulação das condições de experiência visando controlar as variáveis não são, frequentemente, possíveis. Nestas áreas os investigadores tentam eliminar por via da estatística aqueles factores externos susceptíveis de influenciar os fenómenos em estudo. West exemplifica com um tipo de estudo epidemiológico, como por exemplo o estudo do impacto da dieta no desenvolvimento da doença coronária, tentando remover via estatística os efeitos de factores de risco que a investigação demonstrou influenciarem a doenças, tais como a idade, o sexo, e o consumo de tabaco. Este tipo de abordagem permite um controlo eficaz dos efeitos de variáveis externas. Isto no entanto só é verdadeiro quando (a) se conhecem todas as variáveis externas capazes de influenciar os resultados, (b) elas são medidas de modo fiel e válido, (c) se conhecem as relações entre a variável em estudo e as variáveis externas, (d) se conhece de que modo as interacções entre as variáveis externas contribuem para o resultado da variável em estudo. Embora as técnicas estatísticas disponíveis facilitem o controlo das variáveis externas, as condições referidas anteriormente não são fáceis de satisfazer em psicologia da saúde.

Verifica-se assim que a exigência de controlo das variáveis de investigação de um modo semelhante ao da física é inadequado em psicologia e que, o mais apropriado será eliminar mediante o controlo estatístico das variáveis, o efeito daquelas que se sabe contribuem para o resultado.

As fases dos projectos de investigação

Fazer investigação é inerente a qualquer prática profissional embora haja profissões que se dediquem exclusivamente à investigação. Iniciar um projecto de investigação exige sempre medidas de rigor inerentes ao acto de investigar. Consoante se decida fazer investigação como actividade complementar à sua prática diária, ou fazer investigação como investigador profissional, há sempre uns passos e questões a que se deve responder antes de iniciar a investigação. Não sendo eliminatórias é no entanto aconselhável que a resposta a grande parte delas seja positiva. A caixa 1 lista algumas dessas questões:

CAIXA 1

1. A motivação é sólida?
3. Os empregadores facilitam a realização da investigação?
4. Conhecem-se bem as teias burocráticas que permitem o acesso a subsídios para a investigação?
5. Os recursos existentes permitem que a investigação seja levada até ao fim?
 - 5.1. Como serão tratados os dados?
 - 5.2. A produção do relatório de investigação, ou tese está garantida?
6. O suporte social dos amigos, família, e próximos em geral é boa?
7. Se for caso disso, tem um orientador ou orientadores que o supervisionem?
 - 7.1. O campo de investigação do supervisor é o que lhe interessa?

Podem-se considerar três grandes fases de um projecto de investigação:

- a) uma fase inicial que inclui a decisão de iniciar o projecto e inclui a definição da questão de investigação;
- b) uma segunda fase sensível que abrange a recolha de dados;
- c) uma terceira fase que abrange a redacção do relatório de investigação, e a apresentação de resultados.

Iniciar a investigação

Um projecto inicia-se pela questão de investigação. Em psicologia da saúde esta surge da experiência do psicólogo e pode começar por ser muito genérica. O passo seguinte consiste em esclarecer se ela já foi respondida.

Procurar as respostas que foram dadas à questão genérica está muito facilitado com a crescente disponibilidade de bases de dados onde toda a investigação publicada e reconhecida como possuindo mérito científico está registada e com acesso fácil e rápido. Para além das bases de dados como, por exemplo, o PsycLIT ou o MEDLINE, facilmente se encontram sites na internet em que os assuntos são discutidos. Para além disto os próprios autores estão acessíveis através do correio electrónico. Todos estes recursos permitem estabelecer redes de comunicação e informação que facilita este primeiro passo.

Deve-se, assim, aperfeiçoar a questão de investigação tomando em consideração o que já foi investigado e afirmado, o que está publicado e os modelos e teorias disponíveis. Estando a questão de investigação mais clara o passo seguinte consiste em avaliar as possibilidades de implementação da investigação. Algumas questões têm de ser colocadas nesta fase. A caixa 2 lista algumas dessas questões:

CAIXA 2

1. A questão de investigação tem mérito?
 - 1.1. não viola aspectos éticos básicos?
 - 1.2. a autorização para implementar a investigação é viável?
 - 1.3. as variáveis a estudar são exequíveis?
5. Conhece bem as teias burocráticas que facilitam ou dificultam o acesso à amostra e ao contexto de investigação?
6. A recolha de dados é viável com garantia de validade interna e externa do desenho de investigação?
 - 6.1. As variáveis são operacionalizáveis?
 - 6.2. As condições consideradas adequadas para recolher os dados existem e podem ser disponibilizadas?
 - 6.3. A recolha de dados pode ser realizada num período de tempo adequado?
7. Os procedimentos éticos exigidos pela profissão e pelas instituições são exequíveis?

Se a generalidade das questões for respondida positivamente então pode passar-se à fase seguinte da investigação.

O protocolo de investigação

O caminho da investigação continua pela elaboração do projecto de investigação. Projecto (que etimologicamente significa lançar antes) ou programa (que significa

escrever antes) ou protocolo (que significa o que vai na frente) são termos utilizados para nomear o documento que se escreve antes de se começar a investigação. Este documento serve, normalmente, de compromisso entre o investigador e as entidades que estarão ligadas à investigação. Estas podem ser a instituição onde se vai realizar a investigação, a comissão de ética que se deverá pronunciar sobre se é adequado submeter a população que tutela ao procedimento de investigação, a instituição a quem se pede subsídio, etc.

Este documento é, também, um compromisso entre o investigador e a sociedade. Este projecto, protocolo, programa, deverá incluir as intenções do investigador, aquilo que ele pensa que vai realizar, como o vai fazer e quando. Descreve o que o investigador fará se tudo correr nas condições que ele supunha e propunha no início. Ora, no decorrer de uma investigação algumas coisas não correrão como era suposto, pelo que o investigador deverá fazer adaptações sem desvirtuar o seu projecto de investigação. Com efeito, o investigador deverá garantir que não introduzirá alterações que sejam susceptíveis de mudar a investigação. De outro modo tratar-se-ia de outra investigação e não da que estava prevista.

O protocolo de investigação em saúde deve incluir um conjunto de informações que esclareça as intenções do investigador e que estabeleça balizas que não ultrapassará. A caixa 3 apresenta alguns dos aspectos que devem ser incluídos no projecto:

CAIXA 3

1. O título da investigação
2. O nome do ou dos responsáveis pela investigação
3. O nome da ou das instituições a que os investigadores pertencem
4. Justificação do projecto
5. Quem são os participantes
6. Qual é a amostra
7. Qual é o material que vai ser utilizado
8. Qual o desenho de investigação
9. O procedimento
10. Os recursos necessários
11. A calendarização

Passando a explicar: O *título da investigação* que já deverá incluir o nome das variáveis, a população que se propõe estudar, e permitir adivinhar a questão de investigação;

O nome do ou dos responsáveis pela investigação e os da instituição ou instituições a que pertencem;

Justificação do projecto que deverá explicar porque é que aquela investigação é importante, o estado actual da investigação naquela área, porque é que ela deve ser feita, deixar claro que não põe em causa aspectos éticos essenciais nomeadamente a saúde dos participantes, mostrar que os resultados eventuais serão úteis para a sociedade, e, finalmente, qual é ou quais são os objectivos da investigação;

Quem são os participantes, que deverá explicar *quem é a população* do estudo, quantos são, onde é que ela está, e se é acessível para efeitos do estudo; *Qual é a amostra* e se o seu número é adequado para o tipo de estudo que se pretende implementar;

Qual é o material que vai ser utilizado, ou seja, que técnicas de recolha de dados vão ser utilizadas e a sua pertinência para o estudo;

Qual o desenho de investigação que vai ser implementado;

O procedimento, ou seja, os passos que se propõe implementar para escolher e contactar os participantes, o que fará para recolher os dados, os cuidados no manuseamento dos dados;

Os recursos necessários para implementar a investigação, ou seja que material é necessário, que instalações, que ajuda técnica ou quantas pessoas e com que qualificação, que tipo de treino necessitam receber para participarem no projecto;

A calendarização, ou seja, o que se propõe fazer em cada momento e em que datas.

Consoante a instituição a que se apresenta o protocolo assim será necessário apresentar mais dados que sejam específicos dessa instituição.

Questões de investigação e objectivos

A questão de investigação constitui o elemento fundamental do início de uma investigação. À questão de investigação segue-se a formulação dos objectivos da investigação. Estes representam aquilo que o investigador se propõe fazer para responder à questão de investigação.

Teoria, hipóteses e estudos exploratórios – A teoria tem um papel importante na investigação embora a investigação não esteja inevitavelmente associada à produção ou à verificação de teorias (Breakwell, 1995). Uma teoria representa

um conjunto organizado de explicações para que as coisas ocorram desse modo. A teoria postula um conjunto de proposições que clarificam a natureza das relações entre construtos ou entre variáveis. Deste modo, a investigação que parta de um modelo teórico deve supor, com base na teoria, que as variáveis da investigação mantêm algum tipo de relação.

No entanto é frequente os investigadores recorrerem à investigação para descrever determinados acontecimentos por razões meramente práticas, sem tentarem relacionar as variáveis com qualquer modelo teórico, embora se possa argumentar que existem, pelo menos, teorias implícitas, dado que todas as pessoas constroem constantemente teorias. Em ciências sociais é comum desenvolverem-se teorias baseadas em dados acumulados (*grounded theory*). Esta refere-se a um processo de descoberta de teoria a partir de dados que foram acumulados sistematicamente e analisados (Bowling, 1998) e aproxima-se do método indutivo. Nos casos em que se gera teoria a partir dos dados decorre que muita investigação se faça sem partir de uma teoria.

Se, pelo contrário, há um modelo ou uma teoria que explica as relações possíveis entre as variáveis deve-se, então, gerar um conjunto de *hipóteses* que emergem do que o modelo ou a teoria postulam. As hipóteses propõe-se clarificar o que é que o investigador espera encontrar no final do estudo e porquê. Os modelos e as teorias dão as ideias enquanto as hipóteses fornecem evidência de que as teorias predizem os acontecimentos do mundo real. As *hipóteses derivam de modelos, da teoria, de investigações anteriores*, ou de uma combinação de várias.

Deve haver sempre hipóteses? – Não. A maioria das vezes não devem existir hipóteses. Com se referiu antes, há dois grandes tipos de praticas de investigação: a observação dos factos tal como eles existem na natureza, e a experimentação que implica a manipulação das variáveis de investigação. A observação do que existe na natureza aproxima-se do método indutivo. Queremos saber se na natureza a variável “a” varia quando a variável “b” varia? Trata-se de observar o que acontece na natureza sem manipulação de variáveis: estamos a explorar a realidade. Então não se justifica a formulação de hipóteses.

Se a investigação tem propósitos meramente práticos a existência de uma teoria que postule relações entre variáveis é crucial; se as investigações anteriores ou não se debruçaram sobre o problema ou as evidências encontradas da relação entre variáveis são contraditórias, ou ainda, há razões para questionar se o que a investigação tem mostrado para outras culturas, se aplicará do mesmo modo na nossa. Nestes casos o investigador deverá desenhar um *estudo exploratório* para recolher informação que lhe permita formular hipóteses.

Um tipo de estudo frequente são os estudos descritivos. Neste tipo de estudo descritivo não se pretende explicar porque é que ocorrem determinados fenómenos. O investigador somente apresenta o que encontrou (Aday, 1989). Os estudos descritivos são normalmente estudos exploratórios que decorrem do facto do investigador não ter necessariamente um conjunto de suposições bem desenvolvidas para formular hipóteses.

Em estudos exploratórios, qualquer resultado é um bom resultado e pode ser discutido com o mesmo mérito. Se tiverem sido estabelecidas hipóteses com base na teoria e nos dados da investigação, no caso de não se verificar o resultado esperado a discussão não terá o mesmo mérito do que se os resultados estivessem de acordo com as hipóteses. No caso de as hipóteses não se confirmarem tal pode dever-se a inúmeras razões que têm a ver com a validade de toda a investigação, desde a concepção até à implementação não se podendo inferir que as hipóteses estavam erradas.

Variáveis na investigação – As variáveis da investigação constituem um elemento central dado ser à volta delas que se estrutura a investigação. Como o nome diz, variável é uma característica que varia, que se distribui por diferentes níveis, ou que é de diferentes tipos, e é oposto a uma constante. A escolha de uma variável deve garantir que ela varia.

Recolha de dados

A recolha de dados pertence a uma segunda fase do projecto de investigação e que é diferente da anterior. É mais prática, que é realizada no terreno. Com efeito há um momento em que se torna necessário passar ao terreno para recolha de dados. Esta parte pode ser feita pelo investigador ou por assistentes treinados. Esta parte inclui o que se vai fazer com os dados. A caixa 4 sugere algumas acções a considerar nesta fase:

CAIXA 4

1. Obter aprovação escrita da instituição onde vai decorrer a investigação
2. Definir como se vai seleccionar os participantes
3. A população a estudar existe?
4. A população a estudar é acessível?
5. A relação entre variáveis em estudo e o número de participantes a incluir na amostra é adequado?

6. Seleccionar os participantes
7. Escolher o material de recolha de dados
8. Existe uma forma aferida em português? Se não proceder à adaptação da técnica de recolha de dados (ver na última parte do livro como proceder)
9. Elaborar a folha de consentimento informado que o respondente deve assinar (se for caso disso)
10. Elaborar o protocolo de recolha de dados
11. O consentimento informado está previsto
12. Foi pedida autorização à comissão de ética
13. Garantir que as pessoas da instituição que devem apoiar a recolha de dados estão disponíveis e estão a par da sdatas e do que se espera delas
14. Se for caso disso iniciar o recrutamento de assistentes para a recolha de dados
15. Treinar os assistentes para recolher dados de modo uniforme
16. Marcar reuniões com os membros da equipa
17. Iniciar o estudo piloto
18. Proceder às adaptações necessárias em função do estudo piloto
19. Calendarizar a recolha de dados
20. Arranjar um espaço para atender o respondente na recolha de dados
21. Elaborar um manual de cotação dos dados
22. Definir os procedimentos de tratamento de dados (desde o final da recolha como onde se arquiva, quem o faz, onde se guarda o material, garantir que o material está fechado e seguro)
23. Definir os procedimentos de introdução de dados em base de dados computadorizada (nomeadamente escolher o programa de tratamento estatístico)
24. Definir os procedimentos de verificação
25. Garantir ficheiros de segurança

Apresentar os resultados

A última fase de uma investigação consiste no tratamento dos dados e na apresentação dos resultados. Uma investigação pode ter vários destinos. Normalmente os resultados devem ser apresentados sob forma escrita, e o destino deste relatório pode ser uma academia e visar a defesa perante um júri tal como acontece nas pós- graduações (mestrado ou doutoramento); pode visar a apresentação à entidade que subsidiou o estudo, ou pode visar a publicação em

jornais científicos ou outros. Pode ainda ser apresentada oralmente, na defesa perante o júri se se tratar de pós-graduação académica, ou em reuniões científicas.

A redacção do relatório de investigação é uma fase decisiva que tem uma função de comunicação com a comunidade científica. Se esta fase não decorrer, se o trabalho não for submetido ao escrutínio dos pares, se ninguém souber que a investigação decorreu, como decorreu, quem participou, como os dados foram recolhidos e quais os resultados, pode questionar-se se houve investigação.

Esta fase possui características completamente diferentes das anteriores, e no entanto decisivas para dar expressão a elas. Uma ideia a reter é que visa comunicar com a comunidade científica e que, enquanto ciência, o deve ser feito com critérios que exibam e garantam a cientificidade do estudo. A caixa 5 propõe algumas questões que devem estar presentes num relatório:

1. CAIXA 5

1. Os agradecimentos estão presentes?
2. A revisão da literatura foca documentos recentes (se for caso disso)?
3. As questões de investigação estão claras?
4. O texto vai direito ao assunto a investigar sem se perder em aspectos supérfluos?
5. Os documentos consultados são apresentados de acordo com as regras estabelecidas?
6. Não há choque entre os modelos teóricos defendidos e os procedimentos utilizados?
7. A análise dos dados está de acordo com os objectivos da investigação?
8. A estatística utilizada é adequada?
9. Os resultados estatísticos são apresentados de modo claro?
10. A forma como os resultados são apresentados é a mais adequada?
11. Responde à questão de investigação?
12. Os resultados estão de acordo com outras investigações na área?
13. Discute os resultados que encontrou?

Como o relatório tem funções de comunicação, para além do rigor da apresentação dos dados é necessário que esteja bem escrito, que seja claro. Nomeadamente a discussão é um exercício de criatividade e de expressão de inteligência.

O que deve incluir um relatório de investigação

Um relatório de investigação apresenta as razões, os objectivos, os procedimentos, os resultados e a discussão de uma investigação. Deve fazê-lo de modo tal que qualquer outro investigador possa repetir todos os passos da investigação para verificar os resultados.

Um relatório deste tipo é tanto mais minucioso quanto a inexperiência do investigador. Por exemplo, um relatório de fim de curso (mestrado ou doutoramento) é muito minucioso enquanto um artigo para publicação num periódico é menos. Tal deriva do primeiro poder espalhar-se por mais páginas do que o segundo, e por o primeiro colocar quem o apresenta numa situação de avaliação iniciática em que o avaliado tem de mostrar que conhece todos os detalhes da investigação.

Um investigador deve dominar perícias de comunicação de tal modo que as suas ideias e investigações alcancem outros investigadores ou especialistas da sua área. Todas as áreas de conhecimento partilham modos de comunicação mais ou menos comuns. No entanto, de domínio para domínio mantêm-se detalhes de apresentação que são específicos dessa área: por exemplo, um texto da área médica difere de um da psicologia, pelo menos, pelo modo como as referências bibliográficas são apresentadas. Todos os periódicos apresentam numa página regras de publicação que devem reger os textos que lhes são submetidos: Estas regras variam entre periódicos quanto aos detalhes, embora partilhem em comum as regras da sua área de conhecimento.

Na área da psicologia são normalmente seguidas as regras de publicação da *American Psychological Association* – e serão essas que guiarão as nossas recomendações.

Um manuscrito na área da psicologia, segundo a APA, deve incluir as seguintes secções: página de título; resumo; introdução; método; resultados; discussão; referências.

Página de título

O título deve sumariar as principais ideias da investigação em 10 – 12 palavras. Uma boa receita, segundo a APA, para um relatório de investigação experimental é: (variável dependente) em função da (variável independente) ou,

os efeitos da (variável independente) na (variável dependente). Com outros tipos de investigação deverão ser incluídas no título as variáveis principais do estudo (de tal modo que não implique causalidade). Deve também incluir referência à população que está a ser estudada.

Os periódicos incluem normalmente um *running head* que consiste num título curto que surge no cimo das páginas ímpares do artigo. Este deve ser um resumo do título.

A caixa 6 apresenta como exemplo o título de um artigo do autor publicado na *Acta Neurologica Scandinavica*¹:

CAIXA 6

Título

Impact of Epilepsy on QOL in a Portuguese Population: Exploratory Study

Running head

QOL in Portuguese Epileptic Patients

Resumo

O objectivo desta secção é fornecer, em cerca de 250 palavras um sumário compreensivo do estudo. É um elemento fundamental da investigação porque é a parte mais lida, frequentemente a única, de uma investigação. Há actualmente revistas que publicam os sumários seleccionados e comentados de artigos publicados. Os sumários são seleccionados por investigadores de entre artigos relevantes. O sumário do artigo é reproduzido, acrescentam-se os comentários, e tudo junto não excede uma página da revista, como é o caso, por exemplo, da *Evidence Based Mental Health*.

Um resumo deve ser apresentado, também, em Inglês. A versão em Inglês é publicada nas bases de dados como o PsychLit, o MedLine, o ERIC, ou outra.

¹ Ribeiro, J. L., Mendonça, D., & Martins da Silva, A. (1998). Impact of epilepsy on QOL in a Portuguese population: Exploratory study. *Acta Neurol Scand.*, 97, 287-294).

O resumo é um texto com o máximo de informação no mínimo de palavras. Em geral começa por uma descrição breve do problema investigado, ou mais especificamente pelo objectivo do estudo, do método (participantes, tipo de amostra, material), dos resultados e suas implicações. Não se devem citar referências no resumo.

Na caixa 7 apresenta-se um exemplo de um resumo de 250 palavras de acordo com as exigências da organização que o solicita²:

CAIXA 7

The objective of the study was to inspect the relationship between satisfaction, expectations, and the evolution of pain in patients submitted to general surgery at a general hospital – Hospital Geral de Santo António, in Porto. Participants were 81 patients submitted to seven different types of surgery, mean age 56,35 years, range 17-86, of which 59,3% were female. We used measures recommended by the *American Pain Society Quality of Care Committee* (APSQCC) namely, satisfaction with treatment (sat1) measured with one item, satisfaction with the way medical doctors gave support if patients had pain (sat2), satisfaction with the way nurses gave support if patients had pain (sat3). Based on APSQCC recommendations we measured pain with one analogue visual scale about 24 hours after surgery, and three pain numeric scales about 24 hours before discharge from the hospital. Correlation between pain after surgery and before discharge shows statistically significant but modest correlations ($Rho=.30$, $p<.02$), suggesting that the two types of pain can be different. We used a scale with two items (Cronbach alfa= .62) to measure congruence with expectancy about pain. Preliminary results show no statistically significant correlation between satisfaction or expectations and post- surgery pain; however, a statistically significant correlation was found between pain at discharge from the hospital and sat1 ($rho=-0,42$, $p<.001$) (more satisfaction less pain); a statistically significant correlation was found between pain and expectations ($rho=0,26$, $p<.03$) (higher expectation about a good control on pain less pain). No statistically significant correlations were found between support by medical doctors or nurses and pain.

Escrever um resumo para uma revista científica, ou para uma reunião científica é idêntico. Como as bases de dados são consultadas com facilidade e, frequentemente, o investigador se fica somente pela consulta dessas bases, a tendência é para

² Ribeiro, J., Couto, P., Teixeira, C., Mendonca, D., & Bastos, R. (1999). Satisfaction with treatment and acute pain in general surgery: The 13th Conference of the European Health Psychology Society.

que os resumos nelas contidos se constituam como pequenos artigos. Deste modo alguns dos principais jornais científicos pedem um resumo organizado segundo regras precisas, respeitando domínios específicos. A caixa 8 apresenta um exemplo de artigo científico publicado segundo as regras de um Jornal Europeu³:

CAIXA 8

Objectives – To develop an outcome measure of QOL (Quality Of Life) for the Portuguese population; describe Qol in epileptics; analyse how QOL is affected by epilepsy and analyse how QOL is affected by patients' characteristics.

Patient and methods – study of a consecutive sample of 92 epileptics from an outpatient clinic, men and women, aged 15-65, having the basic education level. Firm diagnosis of epilepsy (with or without seizures) without other illness. A questionnaire based on ESI-55 and Health Insurance Experiment Short-Forms was used. Principal component analysis (orthogonal varimax) showed a solution explaining 55.6 of total variance. After the inspection, the 33 remaining itens were grouped into 5 sub-scales (dimensions).

Results – Results showed that QOL was poor for people with partial seizures. The best scores were found on patients now free from seizures.

Conclusion – The QOL measure is sensitive to cultural differences, demographic and disease variables which could explain the results found.

O *The Journal of the American Medical Association* (JAMA) recomenda, para os artigos que expõem resultados de investigação, a seguinte estrutura de organização do abstract: Context; Objectives; Design; Patients; Setting; Intervention; Main Outcome Measures; Results; Conclusions.

Com o aumento das publicações científicas a tendência será para que a apresentação dos resultados seja cada vez mais económica em termos de espaço.

Introdução

O principal objectivo desta secção é explicar ao leitor porque é que se realizou o estudo. Deve explicar qual é a questão de investigação, indicar porque é que ela é importante, e o que é que é inovador relativamente aos estudos anteriores.

³ Ribeiro, J. L., Mendonça, D., & Martins da Silva, A. (1998). Impact of Epilepsy on QOL in a Portuguese population: Exploratory study. *Acta Neurol. Scand.*, 97, 287-294.

Esta parte deve iniciar-se por aspectos genéricos que se tornam cada vez mais específicos. Deve incluir uma revisão actual da literatura, mas evitar uma revisão histórica exaustiva, e deve relacionar a investigação que é apresentada com a que tem sido feita. Deve incluir as hipóteses, se as houver, assim como o raciocínio que lhes deu origem.

A introdução só deve incluir assuntos que sejam relevantes para a investigação. Não se deve espalhar por assuntos que não sejam considerados no desenho de investigação, que não serão abordados nos resultados, e que não serão discutidos.

O último parágrafo desta secção deve incluir uma afirmação que explicita, claramente, o objectivo do estudo. No caso de se apresentarem hipóteses elas serão apresentadas aqui.

Se uma investigação deve ou não incluir a formulação de hipóteses dependerá, como já se referiu, da filosofia que o investigador adopta e do tipo de investigação. Definir hipóteses não é expor expectativas baseadas no senso comum do investigador. Uma hipótese é uma afirmação declarativa que se propõe esclarecer as relações entre as variáveis. A hipótese deriva directamente do problema. É uma tentativa de adivinhar a solução do problema. Para que a “adivinha” seja inteligente segundo Mason e Bramble (1978), ela deve basear-se no que se sabe acerca do problema. A revisão da literatura, a investigação anterior e a experiência pessoal, contribuem para esse conhecimento. “As hipóteses baseiam-se geralmente numa teoria ou modelo” (Mason & Bramble, 1978, p. 54).

Se o problema estiver claro definir hipóteses é simples uma vez que a formulação do problema sugere normalmente a natureza da hipótese. Mason e Bramble (1978) defendem que uma boa hipótese dev possuir cinco características:

- devem ser colocadas numa forma declarativa;
- devem descrever as relações entre duas ou mais variáveis;
- devem ser testáveis;
- devem ser operacionais, de tal modo que não haja ambiguidades acerca das variáveis ou das relações entre elas;
- devem reflectir a adivinha da solução ou resultado do problema, a qual se deve basear no conhecimento, e na investigação anterior.

Se não se definirem hipóteses então a formulação dos objectivos da investigação que derivam do problema deverão respeitar os cinco princípios referidos acima.

Método

O objectivo desta parte é descrever em detalhe como se realizou o estudo, de tal modo que qualquer um possa o replicar com base na informação fornecida.

Numa investigação este sector divide-se nas seguintes partes: participantes; material, planificação; procedimento. A planificação e procedimento podem trocar a ordem de aparecimento e por vezes fundem-se numa só.

Participantes

Indica quem participa no estudo (a população ou universo), quantos são, e como foram seleccionados (a amostra). Inclui todos os detalhes considerados relevantes para o estudo, nomeadamente, por exemplo, sexo, idade, profissão, número de filhos, etc. Devem ser acrescentadas as condições que os participantes devem ter para pertencerem à população e as que, se as, tiverem, determinam que são excluídos.

Se o estudo for realizado com animais em vez de “participantes” deve usar o termo “sujeitos”. A caixa 9 exemplifica como se pode apresentar esta secção ⁴:

CAIXA 9

Participantes

A população do estudo é constituída por estudantes entre o 11º ano de escolaridade e o último ano da universidade frequentando escolas da cidade do Porto. Os participantes constituem uma amostra por cotas com 609 estudantes saudáveis, 53% do sexo feminino, pertencentes a três escolas secundárias de zonas diferentes da cidade do Porto, e a nove escolas da Universidade do Porto, com idades entre os 16 e 30 anos, $M=19,88$ anos. Os estudantes do 11º ano de escolaridade foram escolhidos de entre os que, pelo seu comportamento escolar, tal como era percebido pelos professores, tinham alta probabilidade de aceder à universidade.

Material

Descreve em detalhe o material utilizado (origem, características). Deve apresentar exemplos dos itens de cada escala e do modo de lhes responder, as propriedades psicométricas e os construtos que mede. Deve ser claro porque é

⁴ Ribeiro, J.. (1999) mental health inventory: Um estudo de adaptação (submitted).

que este material foi utilizado. Não deve descrever o procedimento (não utilizar verbos de acção)⁵. A caixa 10 propõe um exemplo:

CAIXA 10

Material

A avaliação foi feita através de questionários, anónimos, de auto-resposta, que incluíam: questionário demográfico que incluía idade, sexo, formação, número de anos de trabalho, estado civil, número de filhos;

Questionário de avaliação do stress no trabalho desenvolvido para o estudo, que focava a satisfação com áreas do trabalho. O questionário possui 39 itens, distribuídos por seis dimensões, em que a resposta é dada segundo uma escala tipo Likert entre “discorda totalmente”, discorda bastante”, “discorda parcialmente”, “concorda parcialmente”, “concorda bastante”, e “concorda totalmente”. Recorrendo à análise de componentes principais com rotação varimax verificou-se que os itens com uma carga factorial no factor a que pertencem superior a 0,45 explicam 68% da variância. As dimensões encontradas foram as a seguir descritas:”condições económicas, com seis itens itens que apresentavam um Alfa de Cronbach =0,76, e de que é exemplo “a minha profissão é uma profissão com segurança para o futuro”; “apoio da hierarquia” com oito itens que apresentavam um Alfa de Cronbach =0,88 de que é exemplo “o meu chefe dá-me sugestões para eu fazer melhor o meu trabalho”; “reconhecimento pelo trabalho realizado” com cinco itens que apresentavam um Alfa de Cronbach = 0,76, de que é exemplo “as pessoas reconhecem quando faço bem as coisas”; “interesse da profissão” com oito itens que apresentavam um Alfa de Cronbach =0,88, de que é exemplo “o meu trabalho é muito rotineiro”; “relação com colegas” com oito itens que apresentavam um Alfa de Cronbach =0,87, de que é exemplo “a minha relação com os colegas é boa”; “condições físicas do local de trabalho” com quatro itens que apresentavam um Alfa de Cronbach =0,87, de que é exemplo “as condições de trabalho do meu serviço são confortáveis”.

(...)

Consentimento informado – Um aspecto importante na realização de uma investigação consiste em obter o consentimento informado dos participantes. Uma das maneiras de o fazer é através da folha de rosto do questionário. De modo geral esta página de rosto contém um conjunto de informações visando esclarecer o que se pretende do participante. A folha de rosto do questionário

⁵ Ribeiro, J., Maia, P., & Paes Cardoso (1998). Fontes de Stress, Saúde, e Mal-Estar, em Profissionais e o SCI. V Simpósio Internacional de Cuidados Intensivos.

deve incluir (a) a carta de apresentação da investigação e do investigador e, (b) o pedido de assinatura que concorda em participar no estudo.

Há recomendações precisas sobre o que deve ser incluído nesta folha de rosto. Segundo Dillman (1978), a carta que constitui a face do questionário e o apresenta, é um elemento fundamental que vai influenciar a motivação do respondente para responder. Dillman explica que a carta de apresentação deve caber numa página e incluir as diversas partes que simultaneamente informem e motivem. Nesta carta explica-se para que serve o questionário, o que o indivíduo vai fazer e solicita-se a sua participação. A caixa 11 descreve o que deve ser incluído na folha de rosto:

CAIXA 11

O primeiro parágrafo, (a) apresenta o tema da investigação, (b) explica o que a investigação se propõe estudar e, (c) defende a utilidade do estudo;

O segundo parágrafo (a) procura convencer o respondente que a sua resposta, a sua participação, é fundamental e não pode ser substituída, (b) que a participação é voluntária, (c) que os dados são confidenciais, (d) que não há consequências pela recusa em participar;

O terceiro parágrafo (a) coloca ênfase na motivação, salientando a importância social e académica da investigação, (b) explica como o respondente poderá ter acesso aos resultados da investigação quando forem publicados, (c) informa quem são os responsáveis pela supervisão do projecto.

A parte final da folha de rosto deverá incluir um espaço destacado para o participante assinar a sua concordância em participar na investigação. A caixa 12 apresenta uma sugestão do que deve ser incluído no texto que o participante em investigações que não sejam anónimas deve assinar:

CAIXA 12

Se concordar em participar neste estudo por favor assine no espaço abaixo e obrigado por aceitar dar a sua importante contribuição para este estudo.

Eu, _____ tomei conhecimento do objectivo do estudo e do que tenho que fazer para participar no estudo. Fui esclarecido sobre todos os aspectos que considero importantes e as perguntas que coloquei foram respondidas. Fui informado que tenho direito a recusar participar e que a minha recusa em participar não terá consequências para mim. Assim declaro aceito participar na investigação

Assinatura _____ data _____

Planificação

Descreve o desenho e o método e, se for caso disso, o modo como os participantes foram distribuídos pelos grupos. Descreve as variáveis independentes e dependentes. Descreve quais os processos de controlo utilizados.

A caixa 13 mostra como se pode apresentar o desenho e o método:

CAIXA 13

O desenho do estudo é descritivo de comparação entre grupos. O método é experimental natural em que a variável manipulada é o transplante renal (transplantados *versus* não transplantados).

Dependendo do contexto de investigação podem ser designadas de diferentes modos; Sampaio e Ferreira (1997), no âmbito de ensaios clínicos controlados, designam por *variáveis primárias*, aquelas “que podem responder à questão inicialmente formulada e *variáveis secundárias*, aquelas que “completam a resposta mas cuja análise é desprovida de sentido sem a resposta obtida através das variáveis primárias “ (p. 751).

As ciências sociais costumam referir dois tipos básicos de variáveis: as *variáveis independentes* e as *variáveis dependentes*. Estas variáveis são, como se referiu no parágrafo anterior as variáveis primárias de uma investigação. Uma questão que se coloca com frequência é se todas as investigações devem Ter variáveis independentes e variáveis dependentes. Se o desenho é experimental, dado que, por definição, uma investigação deste tipo manipula a variável independente para verificar o que ocorre na variável dependente, esta terminologia tem de ser utilizada. Em investigações que se propõem observar o que ocorre – por exemplo verificar se há diferenças entre grupos, ou se uma variável varia quando outra varia – sem manipular as variáveis, é muitas vezes, difícil definir qual é a variável dependente ou independente. Nestes casos é preferível denominá-las de variáveis principais.

Outro grupo de variáveis importantes em investigação são as *variáveis de controlo*, ou seja, aquelas variáveis que a investigação ou a experiência tem demonstrado que se relacionam com a variável dependente e que é necessário controlar para poder afirmar que a variável independente influencia a dependente. Estas variáveis são, normalmente, incluídas no grupo das variáveis secundárias.

As variáveis independentes são aquelas que o investigador manipula, ou que designa como variável central no estudo. A variável dependente é aquela que varia em consequência da variação da variável independente.

Repare-se que a variável que num desenho é considerada independente pode ser, noutra investigação, com outra formulação do problema, com outro desenho, a variável dependente. Ser dependente ou independente não é uma propriedade da variável é, antes, uma consequência do desenho da investigação. Todas as variáveis podem ser independentes ou dependentes. As variáveis podem ser *quantitativas* se a variação for na quantidade, ou *categoriais* se variarem no tipo.

Há algumas variáveis que são de difícil manipulação, por estarem fora do controlo do investigador, e, por isso, nem são candidatas a variáveis independentes nem a variáveis dependentes dado dificilmente influenciarem ou serem influenciáveis por outra variável, como seja o sexo e a raça. Chamam-se a estas *variáveis fixas* (Davis, 1995). A caixa 14 exemplifica como se podem apresentar as variáveis:

CAIXA 14

As variáveis principais do presente estudo são o tempo de doença diagnosticada e a exaustão vital. A primeira será a variável independente do estudo, e a segunda a variável dependente. Como variáveis de controlo incluiremos as variáveis demográficas sexo, idade, e nível sócio económico, e as variáveis de doença, estimativa dos dias que faltou ao trabalho no último ano, estimativa de dias que passou na cama por doença, e percepção geral de saúde.

Procedimento

Esta parte do estudo sumaria cuidadosamente todos os passos da execução do estudo. Descreve todas as fases do estudo e todas as instruções que os participantes receberam.

Resultados

Antes de redigir esta parte observe cuidadosamente todos os números produzidos na análise estatística e pense na melhor maneira de os sumariar e de melhor exibir o que é saliente. Sugere-se que faça primeiro os quadros ou figuras que pretende utilizar.

Quando apresentar os resultados saliente o significado das estatísticas. Ou seja, descreva o que está a testar e qual o significado que o resultado tem para as variáveis envolvidas. Não discuta as implicações dos resultados nesta secção.

Não explique aspectos elementares da estatística (por exemplo “como o valor de p é $<0,01$, logo inferior a $0,05$, o valor é significativo”): a comunidade científica que vai ler o relatório sabe estatística.

Para que o cientista que lê o relatório de investigação compreenda o significado da estatística deve-se indicar um conjunto de detalhes que são decisivos para essa compreensão. Não chega, por exemplo, dizer que a correlação é $r=0,30$. Deve dizer-se que a correlação, para determinado número de sujeitos, é igual a tal valor que, por sua vez, tem um nível de significância estatística determinado, ou, por exemplo, $r(120)=0,30$, $p<0,01$.

Para além da explicação detalhada em termos estatístico acrescente uma descrição geral que sumarie, num parágrafo, a estatística que foi esmiuçada nos parágrafos anteriores.

Apresente as estatísticas descritivas antes da estatística inferencial, ou seja apresente as médias ou percentagens (eventualmente numa tabela ou numa figura), antes de apresentar os resultados dos testes estatísticos.

Na caixa 15 apresenta-se o modo como as principais estatísticas devem ser apresentadas:

CAIXA 15

(...) manifestam-se diferenças estatisticamente significativas entre sexos com as mulheres a adoptar comportamentos e atitudes relacionadas com a saúde mais positivos ($M=99,20$) do que os homens ($M=93,20$), $t(539)=4,43$, $p<0,0001$. (...)

(...) Para a sub-escala "eficácia social", a média mais elevada (mais eficácia social) é de $M=20,52$ correspondendo aos alunos do 11º ano e a média mais baixa, de $M=18,87$, correspondendo ao segundo ano da universidade. A análise de variância univariada indica que a diferença entre grupos é significativa, $F(2,596)=8,87$, $p<0,0003$. O teste de Scheffé assinala diferenças significativas entre o grupo de alunos do 11º ano do secundário e, quer os do segundo, quer os dos últimos anos da universidade. (...)

A análise do número de horas passadas a ver televisão e a frequência de actos agressivos de cada uma das 20 crianças revela uma correlação directa entre horas a ver televisão e indícios de comportamento agressivo. A análise com recurso ao coeficiente de correlação de Pearson apoia esta conclusão, $r(59)=0,63$, $p<0,05$.

Enquanto 60% dos homens concordaram que as suas habilidades de leitura eram boas, somente 35% das mulheres o fizeram. Uma análise de Qui Quadrado 2×2 mostra diferenças significativas, $\chi^2(1, N=119)=10,51$, $p<0,025$, sugerindo que os homens têm mais confiança nas suas habilidades de leitura.

De salientar que “ M ”, “ r ”, “ t ”, “ F ”, N e “ p ”, ou seja todas as letras, se apresentam sobressaídas em sublinhado (ou bold ou outro). Note-se que o valor de p não se apresenta “=” – igual a – tal como sai impresso no computador, mas sim “<” inferior a. Note-se ainda que em português os valores se representam com zero vírgula valor e não ponto valor como nos países anglófonos, ou seja, por exemplo no terceiro parágrafo da caixa 15, a correlação não é “.63” mas sim “0,63”.

Discussão

O objectivo desta secção é discutir e interpretar os resultados, sem perder de vista a questão de investigação. Esta parte é a parte criativa do relatório. É nesta parte que se faz a síntese entre a teoria, a investigação e os resultados.

Cone e Foster (1993) sugerem que a discussão deve começar por um resumo dos resultados, mas não em termos de apresentação dos valores da estatística que já foram expostos em detalhe nos resultados. Este resumo deve ser sumariado quase em linguagem comum.

Não se devem comentar os resultados que não são estatisticamente significativos, a não ser que tal tenha algum outro significado especial que não estatístico. Deve-se interpretar criativamente, sempre dentro do quadro da ciência, os resultados. Porque é que foram encontrados estes resultados? Está de acordo com a investigação? É contrário ao esperado? Quais são as consequências dos resultados da investigação?

Referências

A última parte do trabalho inclui as referências que foram utilizadas no decorrer do relatório. A forma como as referências se devem apresentar estão tipificadas e variam com o contexto onde se pretende publicar o trabalho que se está a escrever. Em psicologia é costume recorrer-se às normas da *American Psychological Association*.

Como a psicologia da saúde tanto publica em revistas de psicologia como noutros tipos de revistas deve-se consultar as recomendações aos autores que cada revista inclui numa das páginas iniciais ou finais.

Tipos de trabalhos escritos

Actualmente há mais informação disponível do que a que se pode consultar. Daí resulta que a economia de espaço seja fundamental. Consoante o modo como tem de se descrever o estudo e os objectivos, assim se tem de decidir o número de páginas. Os trabalhos escritos têm vários objectivos: uma tese de doutoramento, um mestrado, um artigo científico, um relatório para um departamento.

Se o relatório for uma tese de doutoramento tudo tem de estar detalhado com minúcia e, simultaneamente, não pode haver texto desnecessário (não esquecendo que os dados devem estar disponíveis à parte para serem enviados aos cientistas que os pretendam consultar).

Uma tese define-se como uma dissertação sobre um assunto particular, em que é feita uma investigação original visando a apresentação para um diploma ou um grau académico (Cone & Foster, 1993). Por sua vez, ainda segundo estes autores, uma dissertação é o tratamento extensivo de um assunto, normalmente sob forma escrita. Estes autores referem que nos Estados Unidos da América (EUA) é costume distinguir uma dissertação de uma tese, em que a primeira se destina a doutoramento e a tese a mestrado, embora tal distinção não seja universal.

O objectivo de uma tese ou dissertação é evidenciar a capacidade do estudante realizar investigação independente sobre um tema que contribua de forma importante e de modo original para o conhecimento nessa área de conhecimento. Ela pode incluir, ou não, trabalho empírico embora normalmente o faça.

Quantas páginas deve ter uma tese

Uma tese ou dissertação é um relato detalhado e exaustivo de uma investigação, que deve ser apresentado de acordo com regras da comunidade científica de modo a facilitar a comunicação entre os cientistas. Enquanto mapa detalhado de tudo o que ocorreu na investigação uma tese tende a incluir muitas páginas. As variações sobre o número de páginas são grandes e, em Portugal, este tipo de trabalho tende a ter muitas páginas. Enquanto algumas escolas fazem exigências específicas sobre o máximo de páginas a incluir, a maioria não o faz.

Cone e Foster (1993) referindo-se a uma pesquisa sobre dissertações em psicologia publicadas no *Dissertation Abstracts International* em 1990

encontraram variações entre 59 e 679 páginas com uma média de 174. Ou seja, parece razoável considerar que uma tese de doutoramento se deve conter em cerca 200 páginas, incluindo referências e anexos. Estes autores recomendam que as teses de mestrado deverão ter menos um terço do que as de doutoramento, ou seja, limitar-se a cerca de 130 páginas.

As escolas deveriam ter definida uma política referente a estes aspectos: quantas páginas incluir, os aspectos que devem ser contemplados no trabalho, etc. Na ausência de orientações claras compreende-se que os estudantes tenham dificuldade em decidir sobre o que excluir do seu trabalho, resultando relatórios com muita informação desnecessária.

Se em vez de uma tese se tratar de um artigo que visa ser publicado numa revista científica então o que se escreve é um sumário detalhado da investigação. Há uma grande variedade de revistas científicas. Se há revistas com um grau de exigência muito elevado que rejeita cerca de 80% dos artigos submetidos, outras há que publicam quase todos com muito pouca crítica.

O número de páginas adequado para um artigo a publicar numa revista científica dependerá da revista e, as recomendações aos autores explicarão eventuais limitações. Revistas há em que os artigos não excedem uma página, o que corresponderá cerca de duas páginas A4 a dois espaços. Outras, normalmente de pendor mais teórico, incluem um grande número de páginas. Pode dizer-se que um artigo não deverá exceder 15 páginas, lembrando que a tendência é para reduzir.

Algumas revistas científicas incluem uma secção que designam por *short communications* que são apresentadas em uma ou duas páginas. Convém ter em mente que a função de um artigo científico é explicar sumariamente a investigação que foi feita. Num artigo nunca se pode dizer tudo. Aqueles que estão interessados no trabalho têm sempre a possibilidade de contactar os autores a pedir esclarecimentos adicionais.

ASPECTOS METODOLÓGICOS DE UMA INVESTIGAÇÃO

Desenhos de investigação

Desenho de investigação refere-se à estrutura geral ou plano de investigação de um estudo como seja se o estudo é experimental ou descritivo, e qual o tipo de população. Definido o desenho torna-se necessário especificar o método do estudo e de recolha de dados. Por *método de investigação* entende-se as técnicas e práticas utilizadas para recolher, processar e analisar os dados, com seja, por exemplo, investigação ou sondagem (Bowling, 1998).

Diferentes autores recorrem a termos diferentes para caracterizar os estes aspectos. Aday (1989) propõe que se considerem os seguintes desenhos de investigação: Experimental ou observacional.

Experimental

Num estudo que adopte um desenho experimental o investigador actua sobre a variável independente para identificar se esse tipo de intervenção produz alterações na variável dependente. Definem-se dois grupos, em que os sujeitos de investigação são distribuídos de modo aleatório por cada grupo, intervém-se sobre um, e não sobre o outro, e depois comparam-se as modificações que se verificaram após a intervenção. Num estudo deste tipo conclui-se por uma relação de causalidade, ou seja, as mudanças ocorridas na variável dependente são *por causa* das mudanças que o investigador introduziu na variável independente.

Observacional

Num estudo que adopte um desenho observacional o investigador não intervém. Antes, desenvolve procedimentos para descrever os acontecimentos que

ocorrem, naturalmente, sem a sua intervenção, e quais os efeitos nos sujeitos em estudo. Estes podem ser *descritivos* ou *analíticos*.

- *Descritivos* – os estudos observacionais-descritivos, basicamente, fornecem informação acerca da população em estudo e podem ser: *transversais*; *de comparação entre grupos*; ou *longitudinais*.
 - a) *transversais* – os estudos observacionais-descritivos transversais focam geralmente um único grupo representativo da população em estudo, e os dados são recolhidos num único momento;
 - b) *comparação entre grupos* – os estudos observacionais-descritivos de comparação entre grupos focam dois ou mais grupos, escolhidos com base no critério de um grupo possuir uma característica de interesse para o estudo e o outro não. Como no anterior, os dados são recolhidos num único período de tempo;
 - c) *longitudinais* – os estudos observacionais-descritivos longitudinais focam mais de um grupo que no decorrer do tempo serão submetidos a determinados acontecimentos. Os dados são recolhidos em vários momentos do período de estudo.
- *Analíticos* – Os estudos observacionais analíticos permitem responder à questão de porque é que os sujeitos têm aquelas características. Estes estudos podem ser: *transversais*; *controlo de caso*; e *prospectivos*.
 - a) *transversais* – os estudos analítico-transversais procuram explicar os resultados através do exame das relações estatísticas (correlações) entre variáveis num único momento;
 - b) *controlo de caso* – estudos analíticos de controlo de caso, também chamados estudos retrospectivos, são estudos típicos em saúde onde grupos com e sem doença são comparados nas variáveis em estudo de um modo retrospectivo. Neste tipo de desenho os sujeitos são identificados depois do acontecimento (por exemplo doença) ter ocorrido;
 - c) *prospectivo* – estudos analítico-prospectivos são estudos longitudinais em que o investigador procura o que acontece ao longo do tempo (prospectivamente), numa amostra que está exposta a diferentes factores.

Os contextos de investigação desenvolvimental, propõe-se analisar aspectos ligados ao tempo, tanto as mudanças que ocorrem nos indivíduos devido ao processo de desenvolvimento humano, como as mudanças que decorrem das alterações históricas, culturais ou geracionais. Este tipo de investigações impõe

desenhos típicos. Podem-se identificar: (a) os desenhos *desenvolvimentais transversais* que comparam grupos diferentes de idade num mesmo momento; (b) os *desenvolvimentais longitudinais* que seguem as mesmas pessoas através de diferentes idades com observações repetidas; (c) os de *seguimento temporal (time-lag)* que comparam pessoas da mesma idade de diferentes gerações (Baltes, Reese, & Nesselroade, 1977). Estes autores apresentam desenhos mais complexos do que os referidos, que visam identificar ou exaltar aspectos do desenvolvimento associados à idade, à geração e ao momento da avaliação que designam por: (a) *sequências transversais*; e, (b) *sequências longitudinais*. Nos desenhos designados de *sequências transversais* fazem-se observações independentes de todas as coortes e de todas as idades. Nos desenhos designados de *sequências longitudinais* fazem-se observações repetidas da mesma coorte.

Métodos de investigação

Os métodos de investigação em psicologia da saúde, segundo Bishop (1994) podem subdividir-se do seguinte modo: *métodos epidemiológicos*; *métodos de sondagem*; *métodos experimentais*; e, *métodos clínicos*.

Métodos epidemiológicos

Os métodos epidemiológicos permitem identificar a distribuição das doenças e dos factores que lhes estão asso-ciados. Fornecem indicadores tais como a prevalência, incidência e risco relativo do aparecimento de uma doença.

Bishop refere as seguintes estratégias de investigação em epidemiologia: *relato de casos*; *estudos correlacionais*; *estudos de caso com grupo de controlo*; *estudos de coorte*:

- a) *relatos de casos* – é um método epidemiológico que se caracteriza por descrições escritas e detalhadas, feitas por vários clínicos acerca de doentes individuais que apresentam problemas interessantes e únicos. Este tipo de casos, dado as suas idiossincrasias, fornecem indicações importantes acerca de factores de riscos potenciais.
- b) *estudos correlacionais* – são métodos epidemiológicos em que se inspecciona se a ocorrência de determinadas doenças se relaciona com aspectos que se

suspeita são factores de risco para essa doença, em diferentes populações, ou na mesma população, em vários momentos. Correlação forte entre o presumível factor de risco e a doença sugere possível associação entre o factor de risco e doença.

- c) *estudos de caso com grupo de controlo* – são métodos epidemiológicos que se propõe identificar factores de risco para determinadas doenças, através de estudos de observação desenhados especialmente para observar essas relações. Neste tipo de estudos compara-se um grupo de doentes que têm essa doença com um grupo de pessoas que não têm a doença. Assim emergem possíveis factores de risco que estão presentes no grupo doente e não no saudável.
- d) *estudos de coorte* – são métodos epidemiológicos que se propõem comparar os indivíduos que possuem o que se suspeita seja o factor de risco, com outros que não o têm, de modo a poder verificar quem desenvolve a doença ou as doenças em foco. Os estudos de coorte podem ser *prospectivos*, ou *retrospectivos*.

No primeiro caso a observação é feita para diante. Por exemplo, Bishop (1994) cita um estudo em que mulheres que tiveram rubéola na gravidez foram seguidas durante dois anos após o nascimento do filho, para verificar se as alterações congénitas eram, ou não, superiores ao que acontecia com mulheres sem rubéola o que se verificou ser verdade. Trata-se de um exemplo de estudo prospectivo: o factor de risco é identificado antes de terem ocorrido as alterações congénitas e do início do estudo. Nos estudos retrospectivos, ou seja que investigam o que já sucedeu, o investigador define o nível de exposição a que os participantes estiveram sujeitos e, então, examina os de modo a identificar quantos em cada grupo de exposição ao risco tinha a doença ou doenças em estudo. Bishop refere um estudo em que pessoas que no passado estiveram sujeitas a radiação nuclear anormal foram estudadas para verificar o risco de falecerem por leucemia.

Métodos de sondagem

Os estudos epidemiológicos focam essencialmente os aspectos relacionados com doenças que atingem um número elevado de pessoas, que estão em franco crescimento, ou que são particularmente graves, ou seja focam as epidemias. Importa, segundo as novas perspectivas e os novos interesses de saúde, identificar aspectos tais como as atitudes e comportamentos relacionados com a saúde, de pessoas que não estão doentes, e mesmo de pessoas que são particularmente saudáveis. Ou seja, também é importante perguntar “o que é que as pessoas fazem para ser saudáveis?”, ao contrário dos estudos epidemiológicos que perguntam “o que é que aconteceu para as pessoas terem esta doença?”. Os métodos de sondagem

vão para além da morbilidade e mortalidade. Como focam a população não doente o seu espaço de pesquisa é a população em geral que não está ligada ao campo da saúde, debruçando-se tipicamente sobre grandes grupos populacionais instalados na comunidade. Exemplos destes estudos são o Estudo Alameda em que o total da população de *Alameda County* foi definida como a população do estudo, e que visava identificar comportamentos e atitudes e relacioná-los com a saúde e doença.

Tanto os estudos epidemiológicos como os de sondagem são essencialmente estudos descritivos que se baseiam em estatística correlacional que permitem estabelecer ligações entre dois eventos e não uma relação causal embora ela possa ser suspeita.

Métodos experimentais

Ao contrário dos métodos anteriores que apenas permitem estabelecer relações entre as variáveis sem determinação de causalidade, os métodos experimentais são desenhados para verificar hipóteses de causalidade. Um termo chave dos métodos experimentais é *controlo*. Nos métodos experimentais, e ao contrário dos anteriores, a variável independente é manipulada com vista a verificar o impacto que tal manipulação tem sobre a variável dependente. Numa investigação típica deste tipo definem-se vários grupos, em que os participantes, com características idênticas, são distribuídos pelos grupos de modo aleatório. Cada grupo é submetido a níveis diferentes da variável independente, e o impacto dos diferentes níveis da variável independente na variável dependente é avaliado. Outro aspecto que é controlado é o contexto de intervenção, ou seja, para se poder garantir que o resultado na variável dependente tem a ver com as diferentes condições da variável independente, tem de se controlar rigorosamente o contexto de intervenção para que não seja possível a existência de variáveis parasitas que, por si só, possam, eventualmente, explicar os resultados da variável dependente.

Tendo em conta o grau de controlo das variáveis, podem-se considerar as investigações experimentais em; *verdadeiramente experimentais* (true experiments); *quase experimentais*; e *experimental – naturais*.

- a) *estudos verdadeiramente experimentais* – Os estudos verdadeiramente experimentais são aqueles em que o investigador tem o máximo controlo. O local onde é mais provável realizar investigação deste tipo é em laboratório. Ora, dado que, ou por razões éticas, ou porque não é possível distribuir os participantes pelos grupos de maneira aleatória, ou por outras razões, os estudos verdadeiramente experimentais são reduzidos;

- b) *estudos quase experimentais* – Os estudos quase experimentais, são uma variação dos estudos verdadeiramente experimentais em que o controlo é menor. A variável independente é controlada, assim como muitos outros aspectos da investigação mas os participantes não são distribuídos de modo aleatório pelos grupos;
- c) *estudos experimentais – naturais* – Este tipo de investigações recorre a desenhos que alguns autores designam de *Ex pos facto designs* (Brannon & Feist, 1992), dado que a variável independente não é manipulada mas, ao invés disso, ela é seleccionada pelo investigador *após* o facto (a mudança nessa variável) ter ocorrido: o investigador controla-a mas não é responsável por essa manipulação. Por exemplo, uma investigação sobre o impacto do transplante renal na sua qualidade de vida das pessoas, num desenho em que há um grupo de comparação constituído por transplantados, o investigador não pratica o transplante. Limita-se a seleccionar as pessoas que foram transplantadas. O controlo do investigador sobre a variável independente é reduzido mas é a única forma de realizar estudos com variáveis independentes deste tipo. Uma das maiores limitações reside no grupo de comparação que não é, de facto, um grupo equivalente (num estudo experimental um grupo inicial deve ser aleatoriamente dividido em sub grupos e, à partida todos os participantes iniciais têm igual probabilidade de pertencer a qualquer dos sub-grupos). Assim, as diferenças encontradas não podem, em segurança ser atribuídas à manipulação da variável independente (o transplante). Por esta limitação do controlo, com frequência este tipo de desenho não é considerado um desenho experimental.

Métodos clínicos

A lógica subjacente aos métodos clínicos é idêntica à subjacente aos métodos experimentais, pelo que o que foi dito para os métodos experimentais se aplica aos métodos clínicos. A diferença provém de que o que se designa por investigação clínica introduz problemas únicos. Nomeadamente, há aspectos que caracterizam o que aqui se designa por investigação clínica: (a) os participantes, geralmente, são pessoas doentes, (b) que acorreram ao sistema de saúde para pedir ajuda para o seu sofrimento, (c) e as variáveis independentes são procedimentos que visam aliviar tal sofrimento.

Neste tipo de investigação alguns aspectos devem ser tomados em consideração e controlados. Bishop (1994) sugere os seguintes:

- a) *os critérios de inclusão* do doente no grupo seleccionado devem ser especialmente cuidados;

- b) a *co-morbilidade* é outro aspecto que é importante considerar. Com efeito, neste tipo de investigação, muitos doentes juntam mais do que uma doença. Frequentemente são doenças que decorrem da primeira doença, como é o caso dos doentes diabéticos que têm problemas de circulação, renais, ou outros;
- c) os *tratamentos simultâneos* são comuns em pessoas que têm uma doença. As pessoas com determinada doença podem estar a realizar outro tratamento, ou com outros medicamentos, ou com doses diferentes do medicamento, ou a tomar medicamentos para outras doenças. Estes são susceptíveis de influenciar os resultados da variável dependente;
- d) a *participação do doente* ou como a psicologia se lhe refere, a adesão dos doentes aos tratamentos é, geralmente, baixa. Se num grupo de investigação clínica em que é suposto o doente realizar tarefas fora do encontro com o investigador, alguns dos participantes não as fazem, ou as fazem em quantidades variadas, os resultados da investigação são falsificados.
- e) o *efeito placebo* deriva da crença que o doente tem, acerca do tratamento, ou de quem o trata. Reflecte, também a vontade do doente se tratar, e é fonte de interferência numa investigação.

Um dos modos que a investigação farmacológica tem de ultrapassar a influência do efeito placebo é com *estudos cegos* ou de *ocultação*, ou com estudos *duplamente cegos* ou de *dupla ocultação*. Num estudo *cego* ou de *ocultação*, o investigador dá a um grupo de doentes medicação de dois tipos: Um dos sub-grupos de doentes recebe a medicação em estudo, e o outro sub-grupo recebe medicação de aparência idêntica, mas inerte do ponto de vista químico como seja o caso de comprimidos de açúcar, em que a cor e aspecto seja exactamente igual ao real. A diferença no resultado do tratamento entre os dois grupos – o que recebe a medicação e o que recebe a imitação ou placebo – deve-se ao efeito do composto químico.

Como tem sido demonstrado que, pelo facto de o investigador que administra o comprimido saber qual é o falso e o verdadeiro pode de algum modo ser sinalizado ao doente, recorre-se com frequência a estudos *duplamente cegos* ou de *dupla ocultação*. Neste caso, não só o doente não sabe qual o comprimido verdadeiro, como quem o administra também o ignora, sendo a investigação controlada por um outro investigador.

Em psicologia da saúde estes procedimentos não são aplicáveis. O que mais se aproxima são as situações em que de dois grupos em experiência, um é submetido a uma terapia psicológica convencional enquanto o outro somente

conversa com o psicólogo. Outra variante será um dos grupos ficar em espera para se verificar se a intervenção realizada foi eficaz.

Validade dos desenhos de investigação

Uma investigação deve apresentar dados precisos sobre a relação entre variáveis, a partir dos quais se podem confirmar ou formular teorias, tomar decisões, propor novas investigações. É suposto que os dados sejam precisos, ou seja, que os resultados a que se chega sejam, com alguma garantia, os resultados exactos que é possível alcançar com tal desenho de investigação, e não outros, e que a sua aplicação a esta população seja, também, verdadeira. Toda e qualquer investigação deve garantir estes aspectos.

O primeiro aspecto diz respeito à *validade interna* do plano de investigação. Baltes, et al. (1977) definem-na como a garantia que a relação que se observa entre as variáveis é interpretada ou identificada com precisão e com validade. O segundo aspecto diz respeito à *validade externa*, ou seja, segundo estes autores, garante que as relações observadas entre as variáveis, neste conjunto de dados, se pode aplicar a outro conjunto de dados.

Validade interna

Baltes et al. (1977) referem alguns aspectos que devem ser controlados para garantir a validade interna de uma investigação, nomeadamente:

História

Quando se realiza uma investigação a vida continua. Há sempre a possibilidade de o resultado que encontramos ser influenciado por qualquer acontecimento que corre ou ocorre paralelamente à investigação.

Maturação

É semelhante à história, mas enquanto esta se refere a acontecimentos externos a maturação refere-se a mudanças que ocorrem internamente. Podem ocorrer mudanças internas por causas maturacionais mais ou menos subtis. Alguns

momentos do desenvolvimento dos indivíduos são mais sensíveis a mudanças deste tipo do que outros.

Efeito de teste

Parece hoje claro que fazer um teste⁶ hoje influencia a realização de uma segunda passagem do teste. Quer em testes de resposta habitual quer de resposta máxima, o conhecimento anterior das perguntas é susceptível de influenciar os resultados.

Instrumentação

A investigação recorre sempre a qualquer técnica de recolha de dados, seja um cronómetro para recolher o tempo, um polígrafo, uma contagem de glóbulos do sangue, a observação do comportamento, uma entrevista, um questionário de auto-preenchimento, um teste de realização máxima, ou outra. A utilização em dois momentos diferentes destas técnicas é susceptível de produzir resultados diferentes. Se a técnica de recolha de dados é mecânica as diferenças podem dever-se a problemas de calibração, às mudanças na humidade do ar ou a flutuações da electricidade; se é baseada na observação ou em questionários, o observador pode estar cansado, o respondente pode estar preocupado com algum aspecto, tão pueril com a preocupação com o horário do transporte que o pode levar a casa, ou a hora do dia em que os dados são recolhidos.

Regressão estatística

Este fenómeno também conhecido por regressão no sentido da média, refere-se a que os indivíduos que produzam resultados extremos (altos ou baixos) numa primeira avaliação obterão resultados menos extremados numa segunda avaliação. Note-se que neste caso a média de resultados pode manter-se o que muda é a variância. Esta mudança é susceptível de alterar a significância estatística das diferenças de resultados recolhidos entre dois momentos.

Seleccção

Quando se realiza uma investigação, e quando se fazem estudos que envolvem a comparação entre grupos, pode haver a necessidade de seleccionar os

⁶ Por teste entende-se qualquer recolha de dados realizada de modo padronizado, e abrange todas as técnicas de avaliação psicológica.

participantes para pertença dos grupos. O modo como os indivíduos são seleccionados para o grupo de comparação é fundamental, para se poder concluir que a diferença entre os grupos se deve a alguma característica que existia *há priori* e não à variável em estudo.

Mortalidade experimental

Quando se formam grupos de comparação é possível que os grupos percam elementos por razões tão variadas como morte ou mudança de cidade. Estas perdas geralmente não afectam do mesmo modo todos os grupos de investigação.

Efeitos combinados

É possível que dois ou mais aspectos se combinem para prejudicar a validade interna, ou seja, para produzir resultados que não se distinguem dos que seriam produzidos se todos os aspectos estivessem controlados.

Validade externa

Como já se referiu, a *validade externa* garante que as relações entre um conjunto de dados que se observaram numa investigação podem ser generalizadas a outro conjunto de dados que não foram observados mas poderiam ter sido. Os dados observados numa investigação constituem, geralmente, apenas uma pequena parte dos dados que poderiam ter sido incluídos na investigação mas que não o foram. Campbell e Stanley (1963) afirma que a validade externa responde à seguinte questão: a que populações, contextos, tratamentos e avaliações pode ser generalizado o que foi observado? Planos de investigação com boa validade externa permitem amplas generalizações. Se o plano de investigação é pobre deverá afirmar-se: os resultados da presente investigação devem ser generalizados com cuidado para além deste estudo.

Inúmeros aspectos podem ameaçar a validade externa, tais como: unidades experimentais; avaliação; efeitos reactivos da avaliação; interacção com o tipo de tratamento; reacção aos arranjos experimentais; interferência de múltiplos tratamentos (Baltes, Reese, & Nesselroade, 1977).

Unidades experimentais

Refere-se a uma dimensão importante que é a amostragem. Na investigação em psicologia é frequente este aspecto ser negligenciado. O modo como a amostra é seleccionada deve ser claramente explicitado. Amostras seleccionadas de modo inadequado comprometem a validade do estudo.

Tipo de avaliação

Os dados recolhidos com um sistema de avaliação seriam iguais se o sistema utilizado fosse outro? Este assunto é particularmente importante dado que a falta de instrumentos de avaliação adaptados para a nossa população não garantem o rigor métrico das avaliações feitas.

Efeitos reactivos da avaliação

É frequente em investigações proceder-se a avaliação pré experimental. Se o grupo de controlo não é submetido à mesma avaliação pré experimental então não se pode afirmar que o resultado obtido não é influenciado por essa avaliação experimental. Baltes, Reese, e Nesselroade (1977) dão o seguinte exemplo. Imagine-se que a investigação visa experimentar o efeito de um medicamento para emagrecer. Se o peso do grupo experimental é medido em grupo antes da experiência, poderá esse acto de avaliar o peso ou das conversas que ocorrem nesse momento aumentar a consciência do peso que se tem e alterar o padrão de alimentação, a par da tomada de medicação e da alteração da dieta, que não era prevista no plano de investigação? Se sim, então não se saberá se uma possível alteração do peso será consequência da alteração da dieta em vez da medicação. O mesmo efeito poderia ocorrer em pessoas que comprassem o medicamento na farmácia sem passarem antes pela situação de avaliação em grupo.

Interacção com o tipo de tratamento

Imagine-se uma situação em que se pretende aplicar um programa para deixar de fumar a um grupo de pessoas. Suponha-se que os anúncios para pedir voluntários para participar no programa são colocados em lugares onde é proibido fumar. Como os grandes fumadores, provavelmente, evitam os locais onde é proibido fumar, tais anúncios pode ter como consequência que as pessoas que os lêem os e se ofereçam como voluntários sejam fumadores moderados. Se

o tratamento for eficaz a generalização do tratamento só pode ser feita a fumadores moderados (Baltes, Reese & Nesselroade, 1977).

Reacção aos arranjos experimentais

Por vezes o ambiente em que decorre a investigação pode produzir efeitos separados do plano de investigação e influenciar os resultados de tal modo que os resultados não são devidos a essas características do ambiente onde decorre a investigação.

Interferência de múltiplos tratamentos

Diz-se que há este tipo de efeito quando a aplicação simultânea de múltiplos tratamentos produzem resultados desconhecidos ou indesejados. Imagine-se um programa para fazer as pessoas ganharem peso em que são prescritos diferentes menus, tomados em diferentes frequências, petiscar entre refeições e suplementos vitamínicos. O ganho de peso é devido à regularidade das refeições, aos petiscos ou aos suplementos?

População ou universo e amostra

As pessoas ou objectos acerca das quais se pretende produzir conclusões designam-se por população (Norusis, 1991). Os habitantes de Lisboa; ou as mulheres que habitam no bairro de Alvalade; ou os trabalhadores do sexo masculino do metropolitano de Lisboa; ou os estudantes do quinto ano de escolaridade; ou os estudantes do sexo feminino que têm 17 anos; etc. cada um destes grupos é uma população. A *população* ou *universo* é a *totalidade* das observações pertinentes que podem ser feitas num dado problema (Miaoulis & Michener, 1976).

Smith (1975) designa por *universo geral* a população abstracta ou teórica à qual o investigador pretende generalizar os resultados da investigação. Designa por *universo de trabalho* a operacionalização concreta desse universo. É importante distinguir estes universos dado que os investigadores em ciência sociais e comportamentais raramente estudam universos gerais.

Uma *amostra* é um subgrupo da população (ou universo) seleccionado para obter informações relativas às características dessa população (ou universo) (Miaoulis & Michener, 1976). O processo de definição da amostra designa-se por *amostragem*.

Quando se pretende recolher informação acerca de uma população duas estratégias podem ser seguidas: recolher a informação de toda a população, ou seja fazer um censo, ou recolher informação de uma amostra que represente essa população e fazer uma *sondagem*. Em psicologia raramente se fazem censos.

Porquê da amostragem

Segundo Finsterbusch, Liewellyn e Wolf (1983), a menos que uma população alvo tenha menos de 2000 pessoas deverá, sempre, ser decidido fazer um estudo por amostragem, e tal princípio pode aplicar-se a muitas populações com menos de 2000 indivíduos. Estes autores defendem, que a selecção do tamanho da amostra é uma questão de julgamento, e que o número de sujeitos que constituirá a amostra não é importante: referem que, até amostras de 40 ou 80 indivíduos podem fornecer características sensíveis da população, e que uma amostra de 100 é suficiente para ser tomada a sério. Apresentam uma lista de amostras, variando entre 50 e 1000 indivíduos, sendo a moda os 200, consideradas em vários estudos representativos de populações regionais.

A amostragem permite, geralmente, ao investigador reduzir custos, reduzir a necessidade de mão-de-obra, recolher informação mais rapidamente, e obter dados mais compreensivos (Smith, 1975). Paradoxalmente, segundo vários autores (Miaoulis & Michener, 1976; Smith, 1975), amostras bem seleccionadas fornecem melhores estimativas da população do que os censos, principalmente porque se podem controlar melhor certos tipos de erros.

Tipos de amostra

As amostras caem em dois tipos (Dommermuth, 1975): não probabilística e probabilística. Numa amostra *não probabilística*, também chamada *intencional* (Smith, 1975), a probabilidade relativa de um qualquer elemento ser incluído na amostra é desconhecida. O problema principal da amostra intencional ou não probabilística é que se desconhece quanto representativa ela é de uma população ou universo.

Numa amostra *probabilística*, a probabilidade relativa de que um qualquer elemento da população ser incluído na amostra é conhecida, e uma vez que o procedimento de amostragem esteja determinado, a selecção de unidades específicas a ser incluídas na amostra será feita de modo aleatório.

Amostras não probabilísticas ou intencionais

Smith (1975) propõe o seguinte tipo de amostras não probabilísticas: acidentais, homogéneas, heterogéneas, estruturais, de informantes estratégicos. Pode-se acrescentar ainda a amostra sequencial que é uma amostra típica do sistema de saúde.

- *Amostras acidentais* – são amostras simples constituídas por indivíduos que acidentalmente participaram no estudo.
- *Amostras homogéneas* – constituídas por um leque relativamente estreito ou uniforme da variável em estudo. Podem ser de dois tipos:
 - a) *amostra de casos extremos* – é uma amostra homogénea em que só participam indivíduos que possuam características extremas de uma variável, como por exemplo estudar a busca de excitação em indivíduos que subiram ao Monte Everest, ou em indivíduos que mergulham em apneia mais de 100 metros;
 - b) *amostras de casos raros* – é uma amostra homogénea em que os indivíduos seleccionados são pouco frequentes, como seria o caso de seleccionar indivíduos que receberam transplante de um órgão.
- *Amostras intencionais heterogéneas* – constituídas por um leque de características da variável. Podem ser de dois tipos:
 - a) *amostras representativas* – são amostras intencionais heterogéneas em que um conjunto de elementos da variável são intencionalmente escolhidos para garantirem a amplitude da representação da variável. Em oposição ao exemplo dado acima acerca da busca de excitação, uma amostra deste tipo juntaria grupos de esquiadores, estudantes antes dos exames, praticantes de yoga, etc.;
 - b) *amostra por quotas* – são amostras intencionais heterogéneas em que a variável representada seria proporcional ao universo de trabalho. Por exemplo se se pretende estudar uma variável de estudantes do sexo masculino da cidade de Lisboa, se sabe que 10% dos estudantes são afro-portugueses, então a amostra teria 10% de afro-portugueses.

- *Amostra estrutural* – inclui os indivíduos que são seleccionados devido a uma propriedade relacional determinada tal como a posição de dominância na hierarquia, rede sociométrica, ou cadeia de comunicação. Uma amostra deste tipo utiliza como unidade de selecção elementos ligados por uma relação especial. Como as listas de elementos relacionais raramente existem enquanto universo de trabalho é necessário desenhar este tipo de amostras.
- *Amostras de informantes estratégicos* – é uma variante das amostras estruturais e homogéneas. Para melhor estudar certos aspectos da organização social, comunitária e da sociedade em geral; assumindo que o conhecimento não está igualmente distribuído, por vezes o investigador pretende localizar pessoas que detenham informação pertinente acerca do sistema social ou de um dos seus componentes. Para tal procuram-se pessoas que ocupem posições de liderança na organização social. Este tipo de amostra pode subdividir-se em dois sub tipos:
 - a) *amostra de bola de neve* – é um tipo de amostra em que o investigador constrói a amostra de população especial perguntando a um conjunto de informadores iniciais que forneça nomes de outros potenciais membros. É o caso, por exemplo, de uma amostra de mulheres que recorreram ao aborto. Como o aborto é ilegal e não há listas de pessoas que recorreram ao aborto, uma maneira de investigar esta população é perguntando a um grupo inicial se conhecem outras pessoas que recorreram ao aborto, repetindo a pergunta a este grupo, e assim sucessivamente;
 - b) *amostra escolhida por especialistas* – é um tipo de amostra em que se pede a um especialista que escolha os indivíduos típicos representativos das características em estudo.
- *Amostra sequencial* – Uma amostra comum em contexto de saúde é a amostra sequencial. Designa a amostra em que as unidades constituintes são escolhidas estritamente na base “primeiro que aparece, primeiro escolhido”. Todos os indivíduos que são elegíveis para participar no estudo deverão ser incluídos conforme vão aparecendo (JAMA, 1996).

Diferentes autores nomeiam de modo diferente as amostras não probabilísticas. Por exemplo, Dommermuth (1975) chama *amostra de conveniência* a uma amostra escolhida por conveniência do investigador. Suponha-se que se pretende estudar a adesão ao tratamento de doentes de centros de saúde. Para tal é necessário obter informação de 200 centros de saúde. O investigador pode decidir que, para esta característica a população é relativamente homogénea e que, como tal, quaisquer centros de saúde servem para a investigação. Neste caso quem recolhe os dados pode escolher, com base na sua conveniência quais os centros de saúde onde recolher dados. Chama *amostra de opinião* em que o investigador decide, com base

na sua opinião, fundamentada, que determinados centros de saúde fornecem informação mais adequada do que outros. Polgar e Thomas (1988) chamam amostra accidental a uma amostra constituída pelos indivíduos que estão mais a jeito para a investigação. Por exemplo, uma amostra que é frequentemente utilizada são os alunos da universidade onde o investigador está sediado. É uma mostra fácil e que satisfaz muitos dos objectivos de investigação.

Amostras probabilísticas

Aday (1989) propõe os seguintes tipos de amostras probabilísticas.

- *Amostra aleatória simples* – aquela em que todos os elementos da população têm uma probabilidade igual, conhecida e diferente de zero, de ser incluídos na amostra. Os métodos mais conhecidos de seleccionar os elementos para esta amostra são a lotaria ou procedimentos baseados nas tabelas de números aleatórios.
- *Amostra aleatória sistemática* – selecciona-se um número ao calha para começar entre o primeiro e o último dos elementos da população, com 1 a n elementos de intervalo e seleccionam-se todos os elementos que caem no intervalo escolhido. Por exemplo se queremos escolher 500 indivíduos de uma população de 20000 alunos da universidade do Porto, dividimos 20000 por 500 (que daria o valor de 40), ordenamos os alunos por ordem alfabética, e escolhemos todos os indivíduos que caem no lugar múltiplo de 40 (o 40, o 80, o 120, o 160, etc) até alcançar os 500 participantes.
- *Amostra estratificada* – garante que determinados grupos da população são incluídos. Assim, são definidos todos os grupos de interesse e são seleccionados elementos de cada grupo. Por exemplo, definindo-se que da população de estudantes da universidade do Porto deveriam estar representados elementos dos dois sexos e de todos os anos de escolaridade. Neste caso seleccionavam-se, com um dos métodos probabilísticos anteriores, os elementos de cada sub-grupo.
- *Amostra de cacho* – este método é mais utilizado em estudos nacionais ou outros que envolvam grandes áreas geográficas, e também envolve a divisão da população em grupos. O seu grande mérito consiste em maximizar a dispersão da amostra pela comunidade de modo a que o máximo de diversidade esteja representada minimizando os custos. Identificam-se todos os cachos de unidades de habitação, por exemplo, quarteirões com mais de 50 habitações. Amostra destes cachos são seleccionadas, e depois, todos ou unidades de habitação (entre 7 a 10) seleccionadas para participar.

Quantos indivíduos deve ter uma amostra

O número de indivíduos a incluir numa amostra é uma questão importante e nem sempre compreendida. Como explicam Polgar e Thomas (1988) não há um número mágico de participantes que possa ser considerado o número ótimo. Também não se pode dizer qual a percentagem da população que deve ser utilizada.

O número ótimo de participantes depende das características da investigação e do contexto onde a amostra é recolhida. É suposto que quanto maior o número de participantes na amostra menor é o *erro de amostragem*. Tal é verdadeiro em parte mas o aumento do número de participantes não constitui o elemento mais crítico para tornar a amostra mais representativa.

Com efeito o erro de amostragem está relacionado com o número de participantes pela relação seguinte: O erro de amostragem é igual a $1/\sqrt{n}$. Ou seja, o erro de amostragem é inversamente proporcional à raiz quadrada do tamanho da amostra. Esta equação mostra que quanto maior é a amostra (n) menor é o erro de amostragem. No entanto, da equação, verifica-se que se se duplicar o tamanho da amostra só se reduz o erro na raiz quadrada de dois (1,414). Dito de outro modo, para reduzir o erro de amostragem pouco se ganha com o aumento de amostras acima de 250. Por exemplo, enquanto o erro de amostragem com 250 participantes é de 0,063, se duplicássemos esse número para 500 o erro reduzia-se para 0,044. Pequeno ganho para o custo que significaria esse aumento de participantes.

O tamanho da amostra depende do grau de precisão desejado, da variância da amostra e do tipo de amostra. Quando se pretende fazer uma interpretação estatística dos resultados de uma amostra o erro de amostragem pode ser calculado se se utilizar uma amostra probabilística. Se não for probabilística o cálculo do erro é impossível.

Smith (1975) apresenta os tamanhos da amostra para uma tolerância de 1 e 2% e limites de confiança de 99 para 1 ou 997 para 3 (Quadro 1).

Verifica-se que quando cresce o universo o número de elementos na amostra cresce menos, e para uma precisão de 98% em 99 em 100 amostra o número de sujeitos necessários cresce muito pouco ou permanece constante

Quando a amostra não é probabilística os cálculos anteriores acerca do erro de amostragem não podem ser feitos. As amostras em psicologia são frequentemente

não probabilísticas. O número de participantes a incluir será ditado normalmente pelo número de variáveis a tratar.

QUADRO 1

Número de participante consoante o tamanho do universo, nível de tolerância e limite de confiança, numa amostra estratificada simples

Tamanho do universo	Tamanho da amostra necessária para uma precisão igual ou superior a 99%, 997 amostras em 1000		Tamanho da amostra necessária para uma precisão igual ou superior a 98%, 99 amostras em 100	
	Número de unidades da amostra	Porcentagem do universo	Número de unidades da amostra	Porcentagem do universo
200	171	85,5	105	52,5
500	352	70,4	152	30,4
1.000	543	54,3	179	17,9
2.000	745	37,2	197	9,8
5.000	960	19,2	209	4,2
10.000	1.061	10,6	213	2,1
20.000	1.121	5,6	216	1,1
50.000	1.160	2,3	217	0,4
100.000	1.173	1,2	217	0,2

In Smith (1975), p. 126.

Na investigação psicológica normalmente estão em estudo mais de uma variável. Tem, então, de se garantir que a amostra inclui o número suficiente de participantes para que os resultados estatísticos da análise sejam seguros. A análise estatística multivariada exige (um valor confortável) que do cruzamento das variáveis resulte, em cada, cela 20 sujeitos (número de participantes a dividir pelas categorias do conjunto das variáveis). Referindo-se à análise multivariada, Tinsley e Tinsley (1987) sugerem que uma regra a seguir é considerar 10 sujeitos, ou 5 sujeitos como o mínimo requerido, por variável em análise.

Segundo estes autores, quando o tamanho da amostra aumenta, o erro de medida tende a estabilizar e é menos importante aumentar a amostra, de tal modo que acima de 300 sujeitos a relação entre variáveis e número de sujeitos é menos crucial.

Assim, se um estudo englobar duas variáveis (estado civil e nível de escolaridade), em que cada uma tem 4 categorias (solteiro/a, casado/a, viúvo/a e divorciado/a ou separado/a), (menos que o primeiro ciclo do ensino básico, segundo e terceiro ciclo do ensino básico, ensino secundário, universidade) então o número de participantes necessário seria 320. A quadro 2 mostra os valores necessários consoante o número de variáveis e o número de categorias por variável.

QUADRO 2

*Número mínimo de participantes para 10 indivíduos por cela
(20 participantes entre parêntesis)*

Número de variáveis	Número de categorias por variável		
	2	3	4
1	20 (40)	30 (60)	40 (80)
2	40 (80)	90 (180)	160 (320)
3	80 (160)	270 (540)	640 (1280)
4	160 (320)	810 (1620)	2560 (5120)

In Smith (1975), p. 128.

A partir do número de variáveis incluído, a necessidade de participantes é muito maior do que se não se considerarem as variáveis. Em psicologia a investigação procura, normalmente, relações entre duas ou mais variáveis. Pelo que o cálculo para as necessidades do número de participantes para um tratamento estatístico adequado exige muito mais participantes do que se a preocupação fosse a representação da população ou universo. Tinsley e Tinsley (1987) afirmam que, no âmbito da análise multivariada, uma amostra de 100 sujeitos é pobre, de 200 é razoável, de 300 é boa, de 500 é muito boa e de 1000 é excelente. Ou seja, sumariando o que tem sido dito acerca do número de participantes, recomendam-se valores acima de 300.

Ora, se no domínio da saúde pública conseguir este número de participantes é fácil, no da doença estes números são raramente alcançáveis: quer porque nas instituições de saúde de uma zona geográfica não há tantos doentes, como, em algumas doenças (por exemplo, transplantados renais, cirurgia cardioráxica, etc.) estes números seriam difíceis de atingir.

Análise estatística

O objectivo do investigador quando se debruça sobre um conjunto de dados é verificar a existência e a natureza das relações entre variáveis. Os procedimentos de análise de dados podem ser classificados de inúmeras maneiras para salientar os aspectos a que se dá mais relevância na investigação e que não são métodos alternativos. Estes esquemas de classificação são, normalmente dicotómicos e serão referidos três deles: (a) análise correlacional *versus* experimental; (b) análise inferencial *versus* descritiva; (c) análise univariada *versus* multivariada.

Antes de se passar ao tipo de análise há que fazer uma escolha fundamental para a credibilidade da investigação, acerca do tipo de testes estatísticos a utilizar. Estes variam consoante a natureza da distribuição dos resultados, ou das técnicas de recolhas de dados. Esta escolha diz respeito às técnicas estatísticas de análise de dados *paramétricas* ou *não paramétricas*.

Estatística paramétrica versus não paramétrica

A escolha depende da natureza da distribuição dos dados. A *estatística paramétrica* assume que os valores se distribuem normalmente, ou seja, segundo uma curva tipo sino com uma altura apropriada e simétrica, e que a amostra foi escolhida aleatoriamente. Se a distribuição dos valores não se fizer desta forma deve-se recorrer à estatística *não-paramétrica*.

A estatística paramétrica é adequada quando se trata de variáveis intervalares. Nos casos em que as variáveis são nominais ou ordinais a estatística adequada é não paramétrica. No entanto se observarmos alguns dos relatórios de investigação mais conceituados publicados em periódicos conhecidos pelo seu rigor verifica-se com frequência que recorrem a estatística paramétrica com variáveis ordinais.

Muitos autores defendem que, caso se esteja em presença de medidas ordinais de boa qualidade chegam-se aos mesmos resultados quer se utilize estatística paramétrica ou não paramétrica (Fife-Schaw, 1995; Miller, 1984; Reckase, 1990). Comrey em 1973 refere que para se obviar à questão da distribuição dos resultados não seguir padrões estatísticos de normalidade, as respostas às técnicas utilizadas na avaliação devem distribui-se por 12 pontos. Em 1988 o mesmo autor afirma que esta distribuição deve ter pelo menos 5 pontos.

Em congruência com o tipo de estatística escolhido deve dar-se atenção ao conjunto de procedimentos estatísticos a que se submetem os dados. Por exemplo, não há congruência quando se recorre à estatística não paramétrica para a análise de dados uni e bivariada e, de seguida, se recorre à estatística paramétrica – tal como a análise factorial ou à regressão múltipla – para outras análises do mesmo estudo.

Análise correlacional versus experimental

Na *investigação correlacional* analisam-se as relações entre variáveis sem que se manipulem as variáveis em estudo. Os dados recolhidos expressam as relações entre variáveis tal como elas existem naturalmente. A análise correlacional não permite fazer inferências acerca de causalidade, do tipo “a” provoca “b”. Somente permite afirmar que, provavelmente, quando “a” aumenta (ou diminui) “b” aumenta (ou diminui).

Nos desenhos *experimentais* há manipulação da variável independente. Neste tipo de estudo o desenho de investigação preocupa-se com a eliminação de todas as variáveis susceptíveis de influenciar os resultados para além da variável que é manipulada. Assim é possível concluir se o tratamento utilizado tem algum efeito na variável dependente, ou seja, ao contrário da investigação correlacional, se tal ocorre por causa do uso de tal procedimento. A investigação experimental é utilizada para identificar a causa. Inferência causal deste tipo justifica-se por duas razões (Baltes, Reese, & Nesselroade, 1977): (a) porque o investigador controlou a manipulação da variável independente e (b) porque há um lapso de tempo breve entre o tratamento e o resultado.

Análise inferencial versus descritiva

A *estatística inferencial* propõe-se, com base na teoria da probabilidade, decidir o nível de confiança a partir do qual se pode generalizar resultados obtidos à

população em geral. Este valor é dado pelo valor da probabilidade p , que se refere à probabilidade de o resultado que foi encontrado naquela amostra ser o resultado verdadeiro, ou o resultado que se encontraria se se avaliasse toda a população de onde saiu a amostra. O valor normalmente aceite é de $p < 0,05$. A *estatística descritiva* permite caracterizar os dados de determinada amostra. É um procedimento que se propõe apresentar a fotografia do grupo estudado relativamente a determinado número de variável. É um procedimento básico que, por isso, é incluído, normalmente, em todas as investigações.

Análise univariada versus multivariada

Se os investigadores focarem as relações entre variáveis diz-se que a estatística utilizada é univariada ou multivariada (Baltes, Reese, & Nesselroade, 1977). *Estatística univariada* designa procedimentos para analisar a distribuição de notas representando uma única variável dependente. É o caso, por exemplo, de uma investigação que pretenda analisar o impacto de diversos tratamentos (intervenção em vários grupos com tipos de intervenção diferentes, mais um de controlo, numa variável psicológica, por exemplo, no auto-conceito).

O termo estatística multivariada utiliza-se quando se analisa a distribuição conjunta de duas ou mais variáveis. Por exemplo, uma investigação que procure avaliar se quando o auto-conceito se altera, se verifica, igualmente, alteração da ansiedade, do bem-estar e da auto-eficácia. Alguns autores usam a expressão estatística bivariada para se referirem ao estudo das relações entre duas ou mais variáveis, enquanto se estiverem em jogo mais de duas variáveis utilizam, então, a expressão estatística multivariada (Aday, 1989).

Análise estatística

A análise estatística faz-se consoante o foco dos estudos referidos acima. O primeiro passo da análise estatística consiste na descrição dos dados – médias, variância, etc. –, ou seja, na caracterização dos dados da amostra. Como se referiu é um procedimento básico que descreve os resultados de todas as variáveis do estudo. Cada variável considerada, variáveis dependentes e independentes, variáveis de controlo, variáveis demográficas, variáveis de comparação, etc., deve ser detalhadamente descrita de modo a que fique claro as

medidas de tendência central – moda, mediana e média – e as medidas de variabilidade – valores extremos, variância e desvio padrão – que tipifiquem os resultados da amostra naquelas variáveis.

Medidas de tendência central

A *moda*, a *mediana* e a *média* são medidas de tendência central utilizadas para descrever dados. A *moda* é a medida mais pobre de tendência central: sendo o valor que ocorre mais frequentemente numa distribuição ele é adequado quando se trata de variáveis nominais. Nestas não se pode utilizar a mediana nem a média.

A *mediana* é o valor em que a amostra se divide em duas partes iguais. É a medida de tendência central que deve ser utilizada quando a variável é ordinal. É um descritor mais rico do que a moda, menos rico que a média, mas o adequado para as variáveis ordinais.

A *média* é o descritor mais rico dos três referidos que só se deve utilizar com variáveis intervalares. A moda e a mediana também podem ser utilizados com variáveis intervalares.

Medidas de variabilidade

A par das medidas de tendência central devem ser consideradas as medidas de variabilidade. Com efeito, as medidas de tendência central não nos dão informações sobre a variabilidade das medidas. A medida mais adequada a incluir num relatório, quando se trata de variáveis ordinais ou intervalares, são os valores limites ou os valores extremos (*range*), ou seja, os valores máximos e mínimos que se verificaram nos resultados, para aquela variável. Os limites são o descritor de eleição quando se trata de variáveis ordinais. São um descritor pobre e, quando se está em presença de uma variável intervalar, devem ser considerados outros descritores como é o caso da variância e do desvio padrão. A variância e o desvio padrão descrevem não só os limites como a distribuição de todos os valores. O quadro 3 mostra o tipo de estatística descritiva a utilizar consoante a natureza das variáveis.

QUADRO 3

Resumo da estatística descritiva adequada consoante a natureza das variáveis

Tipo de escala	Estatística descritiva			
	Distribuição de frequências	Medida de tendência central	Variabilidade	Posição individual
Nominal	Gráfico de barras	Moda		
Ordinal	Histograma	Mediana	Valores extremos (<i>range</i>)	Ordem ou ordem percentílica
Intervalar	Histograma, polígono de frequências	Média	Variância e desvio padrão	Notas padrão

Análise bivariada

Como se referiu acima neste tipo de estatística inspeccionam-se as relações entre duas variáveis. A relação entre duas variáveis principais, ou entre uma variável independente e outra dependente, pode ser feita pela inspeção do modo como elas variam em conjunto, como é o caso da análise correlacional que nos dá o grau de associação entre elas, e indica como uma variável muda quando a outra muda, ou ainda, a proporção da variância de uma variável que é prevista pela outra variável a que está associada. Se as notas (scores) respeitam os critérios de uma distribuição normal, como ocorre quando as variáveis são intervalares, os testes são adequados e o número de participantes é elevado, recorre-se à estatística paramétrica, à correlação de Pearson. No caso daqueles aspectos serem violados e a distribuição não for normal, então recorre-se à estatística não-paramétrica e a outras medidas de associação entre variáveis. O quadro 4 mostra as estatísticas a utilizar consoante a natureza das variáveis.

QUADRO 4

Medidas de associação consoante a natureza das variáveis

Tipo de variável	
Dicotômica ou nominal	Coeficiente ϕ (ϕ_i) de correlação
	Coeficiente de contingência de Cramer
	Lambda de Goodman & Kruskal's
Ordinal	Correlação r_s de spearman
	Tau de Kendal (tau a, tau b, tau c)
Intervalar ou racio	Correlação r de Pearson

Correlação e causalidade

Em ciências humanas o estabelecimento de uma relação de causalidade entre variáveis é cientificamente inadequado. Com efeito o que se pode estabelecer é a variância conjunta ou independente de duas ou mais variáveis. Nunca se pode afirmar que não há outras variáveis que correm paralelas, ou que antecedem uma delas. No entanto há um conjunto de critérios que, se usados com cuidado, poderão permitir explorar a hipótese de causalidade. Hill (1965) sugere os seguintes critérios para atribuição de causalidade entre variáveis:

- poder*, que se refere à magnitude da associação entre as duas variáveis;
- consistência*, que se refere à probabilidade elevada de aquela relação tornar a ocorrer;
- especificidade*, que se refere a que a causa dá origem a um efeito simples e não a múltiplos efeitos;
- temporalidade*, referindo-se a que a causa precede o efeito;
- gradiente biológico*, para referir à presença de um efeito de resposta associada à magnitude do estímulo, ou a um efeito de curva;
- plausibilidade da hipótese*, ou seja, à coerência da relação com outras fontes de informação.

A ausência de alguns destes critérios não inviabiliza a hipótese de causalidade, no, entanto é prudente garantir a presença de todos.

Se a análise bivariada em vez de se propor identificar a associação entre variáveis pretender identificar a probabilidade de dois grupos ou duas amostras pertencerem à mesma população, então a estatística que se deve utilizar é a comparação entre médias. O Quadro 5 mostra a estatística que deve ser utilizada em função das características das medidas.

QUADRO 5

Comparação ente médias consoante a natureza das variáveis

Tipo de variável	
Dicotómica ou nominal	Diferença de proporções: qui quadrado
	Teste de Fisher
	Teste de Macnemar
Ordinal	Teste de sinais
	Teste de Mann-Witney
	Teste de Kolmogorof-Smirnof
	Teste de Fisher
	Teste de Wilcoxon
	Teste de McNmar
Intervalar ou racio	Teste t

Quando se pretende identificar o grau de associação entre mais de duas medidas intervalares existem inúmeras outras técnicas. Se a relação for entre uma variável dependente e um grupo de variáveis independentes recorre-se à análise de regressão múltipla. Se a relação for entre um grupo de variáveis independentes e um número de variáveis dependentes, então recorre-se à correlação canónica.

Erro tipo I e II

Diz-se que estamos na presença de erro Tipo I quando concluímos que a variável independente afecta a variável dependente e na realidade tal não acontece. Isto pode ocorrer quando, por exemplo, na definição de dois grupos que se pretendem comparar se seleccionam participantes de tal modo diferentes

que num grupo permanecem os que tendem a ter nota baixa na variável e no outro grupo se incluem os que têm nota alta na variável. Quando se encontram diferenças entre os dois grupos, que se supunha espelhar a relação com a variável independente e tal é falso, ou quando num estudo correlacional se conclui que há correlação entre variáveis quando de facto não há, diz-se que se está em presença de erro Tipo I.

O erro Tipo II ocorre quando se conclui que a variável independente não tem impacto ou não se relaciona com a variável dependente quando de facto tem. Tal ocorre por vezes quando o número de participantes é muito pequeno e não espelha diferenças estatisticamente significativas.

Testes unicaudais versus bicaudais

Escolhe-se testes unicaudais (one tailed) quando a direcção da relação entre as variáveis pode ser definida *a priori*. Se um sujeito tiver um ataque epiléptico do lobo temporal e de seguida se avaliar a realização num teste de memória, como os dados da investigação existentes apontam para o facto de que o ataque epiléptico afecta a memória então podemos assumir a ideia de causalidade, ou seja, que o resultado do teste é afectado pelo ataque. Nestes casos os testes estatísticos utilizados devem ser unicaudais. Na maior parte da investigação em psicologia não é possível concluir pela direcção da relação entre variáveis pelo que é mais apropriado recorrer a testes estatísticos bicaudais.

ASPECTOS ÉTICOS DA INVESTIGAÇÃO

Ética na investigação

Os aspectos éticos são decisivos em investigação. Sem um código de ética que aponte limites oriente os passos da investigação é a própria investigação que fica em causa.

Se os limites que governam a investigação em geral estão razoavelmente definidos os da investigação em psicologia não estão claramente delimitados ao que não é alheia a relativa juventude da formação e da profissão. É necessário que uma organização da profissão estabeleça de forma mandatória para protecção da própria psicologia.

Outras organizações mais antigas já se debruçaram e definiram estes aspectos pelo que a nossa discussão tomará como referência o código de ética de uma organização centenária – a APA – salvaguardando no entanto a possível existência de aspectos culturais que possam impor diferenças relativamente a um código de ética português ou europeu.

A psicologia que se exerce no campo da saúde acrescenta uma nova dimensão que deriva do facto de ser uma prática interdisciplinar e multiprofissional por excelência. Swencionis e Hall (1987), explicam que a colaboração íntima entre profissões inerente à psicologia da saúde acarreta problemas éticos que derivam da existência de códigos diferentes. Os psicólogos necessitam conhecer as diferenças e estar conscientes acerca do modo de lidar com elas. Dois aspectos que estes autores referenciam dizem respeito às técnicas de avaliação e à investigação com pessoas.

Aspectos gerais da ética na investigação no campo da saúde e doenças

No campo da saúde encontra-se um grupo importante para a investigação que são as pessoas doentes que faz toda a diferença relativamente a outras

psicologias aplicadas. Swencionis e Hall (1987), explicam que uma das maiores diferenças relativamente a outras práticas psicológicas deriva do facto dos participantes na avaliação estarem fragilizados pela doença.

As organizações profissionais que definem regras de funcionamento em investigação partilham um conjunto de preocupações e regras comuns. Recorreremos às regras das associações de psicologia que, acerca de cada um dos pontos referidos, recomendam:

- *Planificação da investigação* – os investigadores desenham, conduzem e reportam investigação de acordo com padrões éticos e de competência científica reconhecidos; planificam, a investigação de modo a reduzir a possibilidade de os resultados serem equívocos; na planificação da investigação consideram a concordância da investigação segundo o código de ética. Se um aspecto ético é pouco claro deverão esclarecer a dúvida consultando os órgãos competentes da instituição, consulta a colegas ou outros mecanismos adequados; fazem todos os possíveis para proteger os direitos e o bem estar dos participantes e de outras pessoas que possam ser afectadas pela investigação.
- *Responsabilidade* – os investigadores conduzem a investigação de modo competente e com preocupações pela dignidade e bem-estar dos participantes; são responsáveis pela conduta ética da investigação conduzida por eles ou por outros sob a sua supervisão ou controlo; Aos investigadores só é permitido realizar as tarefas para que estão preparados e treinados; enquanto parte do processo de implementação e desenvolvimento dos projectos de investigação, devem consultar especialistas nas populações especiais que vão ser investigadas ou que têm elevada probabilidade de ser afectadas.
- *Lei e regras* – os investigadores planificam e conduzem investigação de acordo com as leis e regras em vigor no local, com as regras que regulam a prática da profissão, particularmente as que regulam a investigação com pessoas.
- *Aprovação pela instituição* – os investigadores deverão obter a devida aprovação para realizar a investigação por parte da instituição ou organização de acolhimento, ou seja aquela onde se vai realizar a investigação. Deverão fornecer à organização informação cuidada e precisa sobre a sua proposta de investigação. A investigação será conduzida de acordo com o protocolo de investigação aceite pela instituição.
- *Responsabilidade pela investigação* – antes da iniciação da investigação (excepto na investigação anónima, observação naturalista, ou outra

semelhante), os investigadores deverão chegar a acordo com os participantes de modo a clarificar a natureza da investigação e a responsabilidade de cada parte.

- *Consentimento informado* – para obterem o consentimento informado, os investigadores utilizam uma linguagem que seja razoavelmente compreendida pelos participantes (excepto nos casos em que por motivos do próprio desenho se dispense o consentimento informado, e que se referirá abaixo). O consentimento informado deverá estar devidamente documentado; recorrendo a uma linguagem que seja entendida pelos participantes. Aos participantes deverão ser dadas informações sobre a natureza da investigação; que são livres de participar ou de declinar de participar, ou de abandonar a investigação; será explicado aos participantes quais as consequências previsíveis de declinarem participar ou de abandonarem a participação; Aos participantes deverão ser dadas informações sobre os factores susceptíveis de influenciar a sua disponibilidade para participar (tais como riscos, desconforto, efeitos adversos, ou limitações da confidencialidade, excepto para o referido no ponto “decepção na investigação”); deverão, ainda, explicar outros aspectos acerca dos quais os candidatos a participar questionem; quando os investigadores fazem investigação com pessoas que sejam seus alunos ou subordinados, terão especial cuidado para proteger esses participantes de consequências adversas de declinarem participar ou de abandonarem a participação. Se a participação na investigação for condição para frequentar o curso, ou tal fornecer créditos de algum tipo ao candidato à participação, este deverá ter alternativas equitativas ao seu desejo de não participar; com pessoas que estão legalmente incapacitadas de darem o consentimento informado, os psicólogos, para além de (1) fornecerem explicação apropriada. (2) obterem o seu assentimento, (3) obtêm permissão apropriada das pessoas legalmente autorizadas, se tal consentimento substitutivo for permitido por lei.

Os problemas da necessidade do consentimento informado continuam a ser discutidos nalgumas circunstâncias que levantam problemas éticos e de filosofia em geral, embora seja indiscutível a necessidade deste procedimento. Ver por exemplo Truog, Robinson, Randolph e Morris (1999).

- *Dispensa do consentimento informado* – antes de determinar que a investigação não requer consentimento informado dos participantes (tal como a investigação anónima, observação naturalista, ou no caso de investigação realizada com arquivos), os investigadores deverão tomar em consideração os requisitos das instituições que tutelam a instituição ou da própria instituição, e deverão consultar colegas se tal for apropriado.

- *Consentimento informado quando se recorrer a registos em imagem ou som* – Os investigadores obterão o consentimento informado prévio dos participantes para fazer registos de imagem ou som de qualquer modo, a menos que a investigação recorra somente a observação naturalista em lugares públicos e não for antecipado que o registo será usado de modo que possa levar a identificação pessoal ou provocar danos.
- *Oferta de incentivos à participação* – se oferecerem serviços profissionais em troca da obtenção de participantes para a investigação, os psicólogos tornam clara a natureza dos serviços, tal como dos riscos, obrigações, e limitações: os psicólogos não oferecem incentivos excessivos ou inapropriados, principalmente se tal tiver probabilidade de ter efeitos coercivos.
- *Decepção na investigação* – Os investigadores não conduzirão um estudo que envolva decepção, a menos que se tenha determinado que o recurso a técnicas de decepção seja justificado pela natureza prospectiva do próprio estudo e tal tenha valor científico, e se formas alternativas que não impliquem decepção não forem possíveis de utilizar; investigadores nunca minimizarão aspectos significativos susceptíveis de afectar a vontade das pessoas participarem, tais como os riscos físicos, desconforto, ou experiências emocionais desagradáveis: qualquer outra decepção que seja parte do desenho ou processo de investigação, deverá ser explicada aos participantes tão cedo quanto possível, de preferência no final da sua participação, mas nunca depois da investigação ter terminado.
- *Partilha e utilização de dados* – os investigadores informam os participantes na investigação de previsível futura partilha dos dados provenientes da sua participação, do uso dos dados de modo que o participante possa ser identificado, e da possibilidade de utilização futura dos dados:
- *Minimização da invasibilidade* – na condução de investigação, os investigadores interferirão na vida dos participantes ou do meio onde os dados são recolhidos de modo garantido por um desenho de investigação apropriado e que seja consistente com o papel do psicólogo.
- *Fornecer aos participantes informações acerca do estudo* – os investigadores fornecem aos participantes oportunidade para obterem informação adequada acerca da natureza, resultados e conclusões da investigação, e providenciarão para corrigir ideias erradas que os participantes tenham; se valores científicos ou humanos justificarem o atraso ou o não fornecimento de informação, os psicólogos adoptarão medidas razoáveis para reduzir os riscos de danos.
- *Honrar compromissos* – os investigadores adoptam medidas razoáveis para honrar os seus compromissos que assumiram com os participantes na investigação.

Medidas relativas à autoria dos resultados

Diz Azeredo (1997) “a autoria injustificada é aquela que inclui pessoas que não contribuíram ou contribuíram de uma forma insignificante para o artigo” (p813). Explica o consenso internacional sobre a matéria que “a autoria deve, unicamente ser atribuída a quem tenha contribuído substancialmente e intelectualmente para um artigo, bem como esteja disposto a assumir a responsabilidade do conteúdo do mesmo” (p813). Considera que se trata de autoria incompleta quando há omissão de nomes que contribuíram de modo importante para o artigo.

Cone e Foster (1993) explicam que segundo os critérios de ética da APA se deve considerar como contribuição importante actividades tais como a sugestão do problema, das hipóteses, escrever grandes partes do artigo, realizar análises estatísticas, sugerir o desenho de investigação e outros aspectos do género. Contribuições intelectuais menores tais como preparar o material de avaliação, recolher dados, fazer sugestões acerca da análise estatística não deve fazer parte da autoria mas ser objecto de agradecimento em nota de pé de página ou de fim de texto. Ou seja, a autoria do artigo depende da participação conceptual na realização do trabalho e não da realização instrumental. Se o director do departamento se limitar a fornecer encorajamento, fornecer críticas, ou a disponibilizar espaço para o trabalho, ou verbas, tal não é, segundo as regras da APA motivo para autoria do artigo devendo ser somente objecto de agradecimento.

A ordem dos nomes depende igualmente da contribuição científica de cada uma das pessoas.

Uma dimensão importante das investigações diz respeito à autoria dos artigos e relatórios genéricos de investigação. Acerca desse assunto, e com base, ainda, no código de ética da APA (1992) recomendam-se os seguintes aspectos:

- *Registo dos resultados* – Os investigadores não inventam dados, ou falsificam resultados nas suas publicações; Se os investigadores descobrirem erros significativos nos seus dados que foram publicados, tomarão medidas razoáveis para corrigir tais erros, numa correcção, anulação, errata, ou outra medida de publicação adequada
- *Plágio* – os investigadores não apresentam porções substanciais, ou elementos de trabalhos de outros, ou de dados que não os seus próprios, mesmo que ocasionalmente citem o outro trabalho ou fonte dos dados.
- *Crédito pela publicação* – os investigadores só assumem a responsabilidade e o crédito, incluindo o crédito pela autoria, por trabalhos que eles

realizaram ou para os quais contribuíram; A autoria principal e outros créditos pela publicação deverão reflectir com precisão a contribuição científica ou profissional relativas do indivíduo envolvido, independentemente do seu estatuto relativo. A mera posseção de uma posição institucional tal como director do departamento ou director do serviço, não justifica créditos de autoria. Contribuições menores para a investigação ou para a escrita da publicação deverão ser objecto de agradecimento em nota de pé de página, na introdução ou na parte final consoante o tipo de relatório escrito; Um estudante é geralmente o primeiro autor ou autor principal em qualquer artigo com autores múltiplos, que seja substancialmente baseado na tese de dissertação do estudante.

- *Duplicação de dados publicados* – Os investigadores não publicarão, como dados originais, dados que tenham sido anteriormente publicados. Tal não invalida a reprodução de dados quando tal for informado adequadamente.
- *Partilha de dados* – após os resultados da investigação terem sido publicados, os psicólogos disponibilizam os dados a outros profissionais competentes que procurem re-analisar as conclusões do estudo, e que pretendam utilizar os dados somente para esse propósito, garantindo que a confidencialidade dos participantes seja protegida, e que os direitos legais concernentes à propriedade dos dados sejam garantidos.

SEGUNDA PARTE – AVALIAÇÃO

O QUE AVALIA A AVALIAÇÃO

Introdução

A Avaliação Psicológica constitui uma matriz base de identidade para os Psicólogos. Com efeito, foi através da construção de técnicas de avaliação de dimensões psicológicas que, no virar do século, a psicologia se estabeleceu como ciência.

Paralelamente ao desenvolvimento das técnicas de avaliação psicológica assistiu-se à evolução dos procedimentos estatísticos. Com efeito, são inúmeros os procedimentos estatísticos que foram desenvolvidos para aperfeiçoar o estudo das técnicas de avaliação psicológica. Uma das primeiras foi a correlação de Pearson. Karl Pearson era aluno de Galton e, este último é, justamente, considerado um dos pioneiros da avaliação psicológica. Juntos desenvolveram a técnica de análise de regressão (Anastasi, 1990; Jerison, 1982). O desenvolvimento dos procedimentos estatísticos estão na base do estabelecimento e do desenvolvimento da avaliação psicológica como prática científica. Durante os primeiros 50 anos de evolução da avaliação psicológica aqueles dois desenvolvimentos andaram a par.

Historicamente a avaliação psicológica pode ser organizada em várias etapas. Com base na proposta de Murphy e Davidshofer (1998) poderiam indicar-se as seguintes:

- a) até 1900 experimentação da avaliação psicológica;
- b) até 1920 desenvolvimento dos testes de avaliação intelectual individuais e de grupo, e desenvolvimento da teoria psicométrica;
- c) até 1940 desenvolvimento da análise factorial, desenvolvimento dos testes projectivos e de inventários padronizados de avaliação da personalidade;
- d) até 1960 desenvolvimento de medidas de interesse vocacional, e de medidas padronizadas de psicopatologia;

- e) até 1980 desenvolvimento da teoria de resposta aos itens e da avaliação neuropsicológica; desenvolvimento de técnicas de avaliação psicológica em pessoas com doenças específicas; desenvolvimento e declínio da avaliação de inspiração comportamental;
- f) de 1980 em diante adaptação dos testes à aplicação por computador. Adaptação dos testes existentes a contextos muito diversos nomeadamente o desenvolvimento de versões reduzidas de testes clássicos, ou a construção de testes mais pequenos de avaliação da personalidade.

Definição de avaliação psicológica

Toda a avaliação visa a produção de informação que seja útil para tomar decisões. A avaliação psicológica visa, portanto produzir informação psicológica adequada para tomar decisões. Nunca se avalia só para avaliar, principalmente em psicologia: a seguir à avaliação algo deve ocorrer.

As decisões podem agrupar-se nas seguintes categorias: decisões de selecção se a avaliação visa seleccionar um indivíduo para alguma função; decisões de monitorização da intervenção que visam verificar se a acção (por exemplo o tratamento, ou o ensino) está a conseguir os resultados que eram supostos; decisões de investigação que fornecem informação acerca das hipóteses que tinham sido formuladas, ou dos objectivos da investigação e permitem escrever o relatório de investigação; decisão de diagnóstico que visa determinar a acção que se vai seguir. Com efeito uma das funções importantes do testes é permitir o diagnóstico de uma perturbação do comportamento. O diagnóstico implica valor preditivo da técnica de avaliação (Anastasi, 1990), ou seja, a capacidade de, a partir do resultado, se poder inferir que o avaliado, a prazo, provavelmente, terá um dado ajustamento. Tal capacidade diagnóstica implica que a avaliação tenha validade preditiva.

Tipo de técnicas utilizadas na avaliação psicológica

Os instrumentos ou técnicas de avaliação psicológica recorrem a uma grande variedade de termos nos seus nomes. “Teste” é o termo mais conhecido e mais conotado com a psicologia, embora muitas outras profissões recorram também a

ele: um teste escolar, um teste de condução, etc. De um modo geral o termo “teste” dá um sentido de seriedade e desafio à actividade, o sentido de que tem de responder ao teste o melhor que é capaz.

Em psicologia, no entanto, nem sempre se utiliza a palavra teste. Por exemplo nos testes de personalidade, assume-se genericamente que não utilizam o termo “teste” no seu título devido à impressão que a palavra teste pode causar no avaliado (Cronbach, 1970). Como os instrumentos que avaliam a personalidade questionam o indivíduo acerca de si próprio, lhes pede respostas que descrevam aspectos íntimos e a forma como habitualmente se sente, a palavra “teste”, pela conotação que tem com “fazer o melhor possível” pode levar o respondente a falsificar as respostas de acordo com o que ele pensa que será a descrição adequada. Geralmente o termo utilizado para os testes de personalidade é “inventário” ou “questionário”. São excepção os testes projectivos de personalidade e tal deve-se, provavelmente, ao facto de as “questões” colocadas (que não são realmente questões) pedirem respostas que não é possível para um leigo saber qual a adequada. As respostas deste tipo de técnicas de avaliação são indícios de personalidade. A seguir apresentam-se e discutem-se vários termos que são utilizados para qualificar nas técnicas de avaliação psicológica, a saber: questionários, testes, inventários, escalas, entrevistas, observação naturalista, listas de verificação, listas de classificação.

Questionários

O termo “questionário” é utilizado por leigos e a expressão significa, neste caso, que se trata de um conjunto de questões, ou como diz o dicionário de língua portuguesa, “série de questões ou perguntas ordenadas sobre um assunto” (Porto Editora, 6ª ed. p. 1375). Todos os procedimentos de avaliação psicológica são basicamente constituídos por questões. Esse conjunto de questões pode, então designar-se por teste, inventário, escala, etc., de acordo com a relação que se assume existir entre os itens, ou questões, que compõem o questionário. Mas, se todos os testes, inventários, escalas, são questionários, nem todos os questionários são testes, inventários ou escalas. Para o serem, como se explicará ao longo deste texto, têm de possuir um conjunto de propriedades métricas, éticas, entre outras, que garantam que o são.

Com efeito, o modo como as técnicas de avaliação são designadas têm, geralmente, um significado técnico: por exemplo, “escala” significará que os itens estão relacionados e pertencem à mesma família. Enquanto o termo “lista” significará que os itens não estão agrupados em escalas; num teste de realização máxima os

itens estão ordenados por nível de dificuldade, enquanto nos de realização habitual estão distribuídos aleatoriamente. No entanto o que foi dito não constitui regra e pode haver diferenças na nomenclatura das técnicas de avaliação.

Testes

“Teste” é o termo nobre da avaliação psicológica. O termo “teste” foi introduzido por Cattell num artigo de 1890. Este termo difundiu-se associado à avaliação da inteligência, embora o primeiro teste de inteligência que, aliás, viria a constituir matriz para a avaliação deste construto – a “Escala Métrica de Inteligência” de Binet-Simon, de 1905 – não utilizasse o termo “teste” no seu título.

Não há uma definição completamente satisfatória de teste embora tenda a haver um conjunto de aspectos que são inerentes a um teste psicológico. Anastasi (1990) define teste psicológico como uma *medida objectiva e padronizada de uma amostra de comportamento*. Murphy e Davidshofer (1998) explicam que um teste psicológico tem três características identificatórias:

- 1) um teste psicológico é uma amostra de comportamento – ou seja, nunca é uma avaliação exaustiva de todos os comportamentos possíveis que poderiam ser utilizados na medição ou avaliação de determinado atributo;
- 2) a amostra de comportamento é recolhida segundo condições padronizadas – ou seja, as respostas só são válidas se forem recolhidas segundo as condições descritas no protocolo do teste. Se forem recolhidas noutras condições tratar-se-à de outro teste. Anastasi (1990) acrescenta outro aspecto à padronização, a saber, o teste tem normas de comparação, ou seja, a realização do indivíduo é comparada com a realização de uma amostra de outros que são considerados como *referência* e da qual este se afasta ou aproxima, afastamento que, se acima de determinado valor pode definir deficiência do comportamento.
- 3) há regras claras para avaliar as respostas, ou para produzir informação quantitativa da amostra de comportamento – ou seja, se as respostas ao teste fossem avaliadas por qualquer outro técnico chegaria a resultados muito parecidos, senão iguais, acerca da realização do sujeito.

Basicamente, e de um ponto de vista científico, todas as técnicas de avaliação psicológica são testes embora, como já se disse, não utilizem esse termo no seu título.

Inventários

“Inventário” é o termo que designa a maioria das técnicas de avaliação da personalidade. Como já se referiu, este termo pretende evitar o uso da palavra “teste”, susceptível de influenciar as respostas dos sujeitos avaliados, no sentido em que “teste” sugere que o indivíduo tem de fazer o melhor possível. Ora, na avaliação da personalidade, não há respostas certas. Todas as respostas que é possível dar são, em princípio, correctas. Uma classificação quanto ao tipo de resposta diria que são provas de resposta habitual, ao contrário dos testes de inteligência e de capacidade em geral, que são provas de realização máxima, e exigem ao respondente que faça o melhor ou o mais rápido que é capaz. Ao sujeito é pedido que descreva os seus sentimentos, pensamentos, atitudes, interesses, sintomas, bem-estar, ou seja, que descreva sensações ou sentimentos que só ele conhece.

Este tipo de testes são, geralmente de papel e lápis (termo que hoje se pode generalizar ao uso de computadores), de auto preenchimento, ou seja o sujeito pode responder autonomamente sem que o psicólogo esteja a dar instruções, ao contrário do que acontece com os testes de realização máxima.

Bowling (1991) afirma que este tipo de medidas assume grande variedade de formas nomeadamente medidas com um único item, bateria, escala. Há questões auto respondidas que recorrem unicamente a um item. Quando recorre a várias questões para avaliar um conceito estamos na presença de uma bateria. Por sua vez uma bateria pode ser uma escala.

Escala

“Escala” é outra designação para os testes que se propõem avaliar a personalidade (Jackson, 1970). Dawis (1987) define escala como uma colecção de itens, cujas respostas são classificadas e combinadas para produzir uma nota da escala. Dito de outro modo, quando um grupo grande de questões ou itens são criados para avaliar a mesma dimensão, atributo, traço, factor, etc., que é suposto terem uma correlação elevada entre si, e que devem ser somadas para fornecer uma nota dessa dimensão atributo, traço, factor, etc., estamos na presença de uma escala. Idealmente uma escala deve ter, pelo menos 20 itens, em que metade são fraseados pela negativa e a outra metade pela positiva (Reckase, 1990). Há no entanto autores que aceitam três itens, e hoje é frequente encontrar dimensões definidas por um ou dois itens.

Entrevista

Tipicamente a entrevista representa o contacto inicial entre o psicólogo e o sujeito (Turk & Kerns, 1985). Consiste num método de recolher informação através de conversa ou de questões colocadas directamente, e a maioria da informação obtida através da entrevista dificilmente poderá ser conseguida de outra maneira, como sejam os contactos que o psicólogo estabelece com as pessoas em contexto de saúde e doenças, por exemplo com doentes acamado (Kaplan & Saccuzzo, 1993). Para estes autores “em muitos aspectos a entrevista é como um teste” (p226), nomeadamente nos seguintes aspectos: (a) é um método de recolher dados, (b) é utilizada para fazer previsões, (c) pode ser avaliada em termos de fidelidade, (d) pode ser avaliada em termos de validade, (e) pode ser feita em grupo ou individual, (f) pode ser estruturada ou não. Caracteriza-se por o entrevistador ou testador, registarem os aspectos que definiram como alvos de avaliação, e as estratégias de registo são muito variadas.

Observação naturalista

Segundo Anastasi (1990), trata-se de técnicas de observação directa de comportamento espontâneo em contexto natural. A observação naturalista recorre a inúmeras técnicas, a longo termo como o método de diário, a métodos mais breves como a amostra temporal. Pode consistir na observação da totalidade do comportamento ou de um aspecto particular como a linguagem, a motricidade, ou a relação interpessoal. Como qualquer teste é um método de observação rigoroso que difere da situação de teste descrita acima porque na observação naturalista, a) não se controla a situação estímulo e, b) geralmente, abrange uma amostra de comportamento mais ampla.

Listas de verificação

Listas de verificação – *check list* no original – são técnicas de observação preenchidas por alguém que observa outro. Consiste, normalmente, num conjunto de itens que não têm de se relacionar uns com os outros como nas escalas, ou cuja relação é accidental (não tem de haver consistência interna ou homogeneidade). Uma lista de verificação propõe-se verificar se determinados comportamentos, atitudes, pensamentos, etc., existiram ou não. A resposta é dada em termos de “sim” ou “não” e cada um vale por si, ou seja, não se somam as respostas (podendo no entanto ser somadas).

Listas de classificação

Listas de classificação – rating scales no original – são listas como as anteriores, em que um observador treinado, através de observação não-sistemática ao longo do tempo, acumula casualmente e informalmente dados. Estas observações são feitas em condições naturais em que o observador não é considerado um observador, e abrangem, normalmente, um período longo. É uma observação retrospectiva em que o observador sumaria as observações acumuladas ao longo de um período largo de tempo. Normalmente envolve um sumário avaliativo de experiências presentes ou passadas em que as impressões do observador são processadas de modo complexo e não especificado para chegar ao julgamento final (Thorndike & Hagen, 1977). É uma técnica muito utilizada em contextos educativos e industriais. Segundo Thorndike e Hagen (1977) o procedimento mais comum com listas de classificação consiste em o observador com um conjunto de atributos, cada um com uma lista de números, de adjectivos, ou de descrições que representam níveis ou graus de existência desses atributos. O observador regista (classifica) nessa lista uma ou mais pessoas em cada atributo, atribuindo-lhe um número, letra, adjectivo, ou descrição que julga melhor o descrever. Na função pública o chefe de serviço classifica no final do ano os funcionários que trabalham com ele em atributos como, por exemplo, “pontualidade” classificando-o, ainda como exemplo hipotético, em “falta muito”, “raramente falta”, “pouco pontual”, “raramente se atrasa”, “pontual”. Para além deste atributo inúmeros outros são susceptíveis de integrar uma lista de classificação.

Literacia e avaliação psicológica

A avaliação psicológica recorre usualmente a material escrito em questionários de vários tipos que impõem uma capacidade razoável de leitura. Em Portugal, tanto mais quando cresce a idade do respondente, cresce a probabilidade de a literacia do respondente ser baixa. Esta baixa literacia, acrescentada à pouca experiência da população na utilização destas técnicas, pede aos técnicos cuidados redobrados na utilização de técnicas de avaliação psicológica.

A utilização por populações pouco letradas impõe cuidados no texto escrito das questões e, se os procedimentos de passagem forem cuidados, a utilidade desta avaliação não é prejudicada. A nossa experiência prática com populações deste género confirma-o (Pais-Ribeiro, Mendonça, & Martins-da-Silva, 1997, 1998).

Outros investigadores têm apresentado procedimentos que apoiam a utilização destas técnicas em populações com literacia limitada (Bernal, Wooley, & Schensul, 1997, Orley, 1994). Orley (1994) no âmbito de um projecto para ser aplicado em diversos países com níveis de literacia média diferentes sugere a utilização de vários procedimentos de passagem do questionário consoante a literacia do respondente.

Tipos de testes psicológicos

Há inúmeras maneiras de classificar os testes psicológicos. Murphy e Davidshofer (1998) afirmam que a maioria dos testes psicológicos se podem distribuir por três categorias gerais:

- testes de realização, que são aqueles em que se dá aos respondentes uma tarefa bem definida que eles devem realizar o melhor que são capazes. Um exemplo deste tipo de testes são os testes de inteligência, e podem incluir-se nesta categoria todos os testes de realização;
- observações comportamentais envolvem a observação do comportamento do sujeito em situações particulares;
- medidas de auto registo em que o sujeito descreve o que sente, suas atitudes, crenças, interesses, etc.

Cronbach (1970) classifica os testes em duas grandes categorias: os que se propõem medir a *realização máxima* do sujeito, que são utilizadas quando se pretende saber quanto bem o indivíduo consegue fazer no seu melhor, e podem ser consideradas testes de capacidade. A característica básica de um teste de capacidade é que se pede ao sujeito para fazer o melhor que é capaz e, o objectivo, é identificar o que a pessoa é capaz de fazer ao seu melhor nível.

A outra categoria abrange as técnicas que visam determinar a *realização típica ou habitual* do sujeito, ou seja, Como é que o sujeito responde numa dada situação ou em situações semelhantes. Os testes de personalidade, hábitos, interesses, etc., caem nesta categoria. Estes testes não se propõem avaliar o que o sujeito é capaz de fazer mas sim o que ele faz habitualmente (ou normalmente).

Cronbach subdivide os testes de realização típica ou habitual em duas categorias: *observação do comportamento* e instrumentos ou *técnicas de auto-preenchimento (self-reports)*. Os primeiros propõem-se estudar o sujeito quando actua de modo natural. As observações são feitas em *situação padronizada* e em

condições naturais. As técnicas de auto-preenchimento são as mais adequadas para avaliar muitos aspectos inerentes ao sujeito que, ou são muito morosas de observar de outro modo, ou não são observáveis de todo, como, por exemplo, o que é que o sujeito pensa em determinada situação.

Testes individuais versus colectivos

Consoante o tipo de teste, seus objectivos, e normas de aplicação os testes podem ser classificados como técnicas de avaliação individual *versus* colectiva. Os grandes testes são de aplicação individual dado que, para além dos estímulos que elicitam as respostas há todo um outro conjunto de dados que o psicólogo pode recolher relativos ao processo de preenchimento e que complementam o conhecimento acerca do respondente. No entanto muitas passagens de testes (caso, por exemplo da investigação) dispensam a informação adicional que a recolha em situação clínica não dispensa.

Testes objectivos e não objectivos

Sundberg (1977) designa dois grandes tipos de testes. De um lado estão as técnicas *objectivas* ou *psicométricas*, do outro, as técnicas *comportamentais* e *projectivas*. As técnicas objectivas são aquelas em que não há uma intervenção de interpretação entre o comportamento do respondente e uma terceira pessoa. É objectiva quando os dados são directamente do sujeito a outros que então a podem estudar. Os testes objectivos não necessitam de interpretação por parte do passador. A consequência disso é que qualquer um pode passar o teste que depois é interpretado pelo psicólogo.

Testes de papel e lápis

Designam-se assim os testes em que o material consiste em papel e lápis e em que as respostas são dadas por escrito, quer seja por resposta aberta quer seja por resposta fechada ou forçada.

Testes de cultura equilibrada

Este tipo de testes normalmente designados por *culture fair tests* são técnicas de avaliação psicológica concebidas para *minimizar* os efeitos das diversas

experiências culturais na realização dos testes. Propunham-se comparar de forma justa, indivíduos provenientes de culturas diferentes, assumindo que os antecedentes culturais não influenciavam a realização desses testes.

Testes livres de cultura

Este tipo de testes normalmente designados por *culture free tests* são técnicas de avaliação psicológica concebidas para *anular* os efeitos das diversas experiências culturais na realização do teste.

Testes de referência a critério ou a conteúdo

Os testes de referência a critérios comparam a realização do indivíduo avaliado com um critério que define o que se espera que o indivíduo realize, ao invés de o comparar com a realização normal de um grupo de referência. Visam avaliar os indivíduos antes e depois de submetidos a uma aprendizagem. O objectivo destas técnicas não é comparar a realização do observado com um valor médio que provém de uma amostra normativa, mas sim de comparar o indivíduo consigo mesmo, em dois momentos diferentes, antes e depois da aprendizagem, nomeadamente para identificar se esta foi eficiente.

O foco é o que o respondente é capaz de fazer, ao invés de qual é a sua posição comparativamente a outros. Com uma avaliação deste tipo sabe-se o que o indivíduo é capaz de fazer (pode-se descrever tal), ao invés de como nas técnicas de avaliação tradicionais em que sabemos a posição do indivíduo relativamente aos seus iguais mas não se conhecem os conteúdos da sua realização (o que faz bem e o que faz mal).

Tipo de preenchimento

A forma como os testes são preenchidos são muito variadas podendo, no entanto ser agrupadas, nomeadamente a) os que são preenchidos pelo próprio e b) os que são preenchidos pelo testador.

Testes preenchidos pelo próprio

Quando é o próprio a preencher pode fazê-lo (a) perante o testador ou, (b) na ausência do testador.

- *Perante o testador* – trata-se da forma clássica de utilização dos testes. O sujeito é colocado numa situação padrão, uma sala apropriada com luz, som, conforto, adequados, o testador apresenta-se, cria um bom ambiente, apresenta o teste, explica as tarefas a realizar normalmente em forma de instruções que fazem parte do protocolo de muitos testes, o indivíduo responde às questões do teste ou realiza as tarefas supostas, no tempo dado.
- *Auto preenchimento (self-report)* – significa que é requerida inferência mínima entre a resposta dada e a variável/ construto alvo (Baker & Brandon, 1990). O que tipifica este tipo de resposta é ela permitir ao respondente identificar o construto que está em avaliação.

O auto preenchimento pode ser feito sem a presença do avaliador ou testador. Em saúde, muitas investigações clássicas, principalmente se exigem grande quantidade de respondentes, são realizadas com os questionários enviados por correio – conhecido por método *mail type*. Pais-Ribeiro (1995b) verificou que este tipo de preenchimento não produzia resultados diferentes em amostras da mesma população que responderam na presença do psicólogo ou em casa com devolução posterior.

Testes preenchidos pelo testador

Outra situação típica consiste no preenchimento pelo testador e podem ser de vários tipos tais como (a) o testador regista directamente as respostas do sujeito como ocorre em algumas formas de entrevista estruturada, (b) quando o testador regista, assinala, indirectamente, as respostas ou sinais do sujeito como ocorre, por exemplo com a escala de Hamilton, ou do teste de Rorschach, (c) quando em situação de observação estruturada como ocorre quando o testador assinala o comportamento do observado segundo fracções de tempo e ou categorias pré-estabelecidas como, por exemplo com a escala de Flanders, e (d) finalmente, através de observação não estruturada como ocorre quando o testador observa em situação natural e posteriormente regista o que observou. A entrevista pode ainda ser realizada à distância, via telefone, por exemplo.

Preocupações relacionadas com o preenchimento

Quando se avaliam aspectos pessoais ou íntimos, como ocorre nos testes de personalidade, muitos autores receiam que a resposta não seja verdadeira. Como os respondentes são de algum modo coagidos a responder a todas as questões, uma possibilidade que se coloca é que os indivíduos não respondam com precisão. Alguns testes clássicos como, por exemplo, o Minnesota Multiphasic Personality Inventory (MMPI) incluem escalas de mentira, validade e correcção, que visam avaliar a veracidade do padrão de respostas do sujeito. Alguns aspectos devem, então, ser considerados, nomeadamente:

Desejabilidade social

Tem sido demonstrado que pessoas de diferentes níveis escolares, culturas e níveis sócio económico tendem a concordar com as respostas que são mais adequadas socialmente. No âmbito dos testes de personalidade tem sido designado como uma característica de personalidade que pode ser avaliada. Há testes que são utilizados em conjunto para avaliar a desejabilidade social como é o caso do Escala de Desejabilidade Social de Edwards (1957), ou formas reduzidas como a Marlow-Crowne Social Desirability Scale (Strahan & Gerbasi, 1972).

Resposta ao calha

Respostas ao calha ocorrem quando o respondente não está motivado ou quando não é capaz de responder. Neste caso ele assinala a resposta de modo quase aleatório, sem qualquer critério.

Falsificação

Consiste em responder de modo a dar, intencionalmente, uma imagem oposta à realidade, em responder de modo a distorcer intencionalmente a realidade.

Estilo de resposta

O estilo de resposta consiste na tendência para concordar (aquiescência) ou para discordar (crítico) com a resposta. Uma das maneiras de combater esta orientação consiste em incluir na escala o mesmo número de itens formulados pela negativa e pela positiva.

O que avalia a avaliação psicológica

Grosso modo poderia dizer-se que os testes de avaliação psicológica são técnicas concebidas para avaliar construtos psicológicos. Esses construtos são muito amplos e nem sempre é fácil afirmar que tal construto é ou não psicológico. Há no entanto um conjunto de construtos que são clássicos em psicologia nomeadamente inteligência e personalidade.

Inteligência

O primeiro autor que se propôs avaliar a capacidade mental (ou inteligência termos que são equivalentes) foi Galton em meados do século passado que sugeriu que a inteligência era um traço que existia nas famílias. Foi no entanto James McKeen Cattell que introduziu em 1890 o termo “teste mental” (Carror, 1982). Ainda segundo este autor é a Binet que pode ser creditada a descoberta de uma maneira apropriada de avaliar a inteligência ou capacidade mental.

O que é a inteligência ou capacidade mental? Inúmeras definições têm sido produzidas e Sternberg e Salter (1982) resumiam uma definição de inteligência como “comportamento adaptativo dirigido para um fim” (p. 3). Binet (1890) considerava que uma definição estreita de inteligência incluía dois processos: primeiro perceber o mundo externo e, segundo, trazer os acontecimentos à memória para os trabalhar ou pensar neles. Perron-Borelli e Perron (1970), com base em Binet propõem que perante um problema um indivíduo, revela-se inteligente se: (a) compreende a natureza e os dados do problema, e que se trata realmente de um problema; (b) inventa uma ou mais soluções para atingir o objectivo; (c) durante as operações mantém a atenção focada no objectivo a seguir, e a direcção da acção; (d) mantém a crítica sobre o processo mental e os resultados da acção.

O conceito de inteligência é complexo e controverso, e tem sido estudado a partir de inúmeras perspectivas. Por exemplo, Siegal e Richards (1982) referem três abordagens principais ao estudo da inteligência: (a) abordagem psicométrica; (b) abordagem desenvolvimental; e (c) abordagem do processamento de informação.

Abordagem psicométrica – esta abordagem defende que a inteligência pode ser quantificada, e que as pessoas podem ser ordenadas ao longo de um contínuo, do indivíduo menos inteligente ao indivíduo mais inteligente. A avaliação da inteligência é feita com questões ou itens que são escolhidas pela sua capacidade de diferenciar as pessoas. Por exemplo os itens dos testes de inteligência na linha dos de Binet e Simon são escolhidos por distinguirem bem indivíduos de idades diferentes enquanto os testes tipo dos de Wechsler são escolhidos por se correlacionarem bem com critérios externos de inteligência e por diferenciarem bem indivíduos da mesma idade. As estratégias de diferenciação adoptadas pelos diferentes testes derivam da concepção de inteligência subjacente à medida.

Ou seja, cada teste de inteligência ou de capacidade mental é susceptível de medir uma inteligência que pode ser diferente da que é medida por outro testes de inteligência. Historicamente as formas de avaliar a inteligência podem agrupar-se de vários modos. Os testes de Binet e Simon concebiam a inteligência como compósita ou seja, para se avaliar a inteligência tinha de se avaliar a realização do indivíduo em tarefas muito diferentes e a inteligência medida resultava da realização no conjunto de tarefas propostas pelos itens. Os teste desenvolvidos por Wechsler adoptam a mesma ideia de inteligência dos de Binet-Simon. Na evolução do estudo da inteligência na perspectiva psicométrica desenvolveram-se várias formas de avaliar este construto, nomeadamente a inteligência expressava-se como: a) compósita, (b) bifactorial, (c) factores de grupo, (d) fluida ou cristalizada, (e) hierárquica, e (f) estrutura do intelecto (Murphy & Davidshofer, 1998). Esta evolução reflecte, também, o desenvolvimento da estatística aplicada, ao estudo dos testes de inteligência.

Inteligência bifactorial – Esta teoria foi desenvolvida por Spearman (1927) que defendia que toda a realização intelectual nos testes se podia explicar por um factor intelectual geral ou nível de factor g. Ou seja, o que explicava a realização inteligente era o factor g, embora a nota total de um teste fosse influenciada (a) por um erro de medida “e”, e (b) cada teste medisse um factor específico “s”, único desse teste.

Inteligência multifactorial – Thurstone (1935) entendia que, ao contrário da linha anterior, a inteligência era multifactorial. Embora admitisse a importância de um factor intelectual geral considerava que a inteligência era melhor compreendida por factores de grupo que eram independentes uns dos outros, apesar de todos se relacionarem com o factor g. Este autor sugeria que a

inteligência podia ser explicada por sete factores de grupo que intitulou de “Aptidões Mentais Primárias” e que eram: *Compreensão verbal* que implicava tarefas de vocabulário, compreensão de leitura, analogias verbais, e equivalentes; *fluência verbal* que abrangia a capacidade para gerar e manipular rapidamente um conjunto amplo de palavras com características específicas; *número* que se referia à capacidade para realizar operações matemáticas rápida e eficientemente; *espaço* que se referia à capacidade para apreender as relações espaciais e operar com figuras espaciais; *memória associativa* ou memória de trabalho; *velocidade de percepção* que se referia à velocidade para perceber os detalhes visuais, discrepâncias e semelhanças; *raciocínio* ou a capacidade para realizar tarefas que implicassem dedução, indução ou raciocínio aritmético.

Inteligência fluida ou cristalizada – Cattell (1963) considerava dois tipos de inteligência distintos embora relacionados. A inteligência fluida era a capacidade para identificar relações e configurava numa capacidade de raciocínio que era basicamente hereditária. A inteligência cristalizada abrangia o conhecimento e perícias aprendidas.

Estas duas inteligências evidenciam elevada correlação dado que as pessoas com elevada inteligência fluida teriam mais conhecimentos por via de possuírem maior capacidade de aprendizagem.

Inteligência hierárquica – Vernon (1960) propõe um modelo integrador que engloba as propostas de Spearman e de Thurstone. Sugere que um factor *g* (como Spearman) se manifesta na realização de qualquer teste de inteligência e que este subdivide em duas grandes categorias ou grupos: grupo verbal-educacional e grupo espacial-motor. Cada um destes grupos, por sua vez, se pode subdividir em grupos de nível mais baixo que, dalgum modo se assemelham às capacidades mentais primárias de Thurstone. Estes grupos, por sua vez, englobam factores menores que são específicos de cada teste e que se assemelham aos factores específicos de Spearman.

Carroll (1993) propõe um modelo de capacidade cognitiva de três estratos semelhante ao anterior. Baseado num largo estudo com base na análise factorial propõe um modelo hierárquico moderno que, no entanto, vai na linha do de Vernon. No nível mais geral identifica um factor *g*. Num segundo nível identifica um conjunto de factores semelhante aos de Thurstone. Finalmente no nível seguinte identificam-se factores específicos.

Estrutura do intelecto – Guilford (1967) propõe um modelo tridimensional, apresentado como um cubo, em que o cruzamento dos três eixos ou dimensões dá origem a 180 diferentes tipos de inteligência. Os três eixos são (a) as *operações* que o indivíduo faz, (b) os *conteúdos* ou materiais sobre os quais se realizam as

operações e, (c) os *produtos* que se refere à forma como a informação é guardada e processada. Há seis tipos de operações, cinco tipos de conteúdos e seis tipos de produtos, de cujas combinações possíveis resultam as 180 inteligências. Este modelo representa uma ruptura com os modelos anteriores no sentido em que não aceita a existência de um factor g.

Resultados da avaliação da inteligência – a inteligência tal como é avaliada na perspectiva psicométrica expressa-se normalmente por um quociente de inteligência (QI) que, no entanto, tem vários significados. No início, QI tal como era medido com a escala métrica de inteligência de Binet representava mesmo um quociente. O sujeito respondia acertadamente a um conjunto de itens, e a soma dos resultados certos convertiam-se numa idade mental. Esta constituía o numerador de uma fracção em que o denominador era a idade real. Ao resultado desta operação chamava-se “Quociente de Inteligência” (QI) do indivíduo. Mais tarde surge uma outra forma de expressar o QI que já não resultava de um quociente mas sim da dispersão dos resultados que se deveria distribuir segundo uma curva normal. Aqui, o QI depende dos resultados dos indivíduos do mesmo grupo de idade em que arbitrariamente se decide que o valor médio é 100 e que o desvio padrão é, por exemplo nos testes de Wechsler de 15. Na realidade não se trata, como se vê, de um QI mas sim de um índice de inteligência. No entanto conservou-se a expressão QI. A vantagem desta forma de expressar o QI sobre a anterior é que o desvio se mantém igual em todas as idades ao contrário do de Binet em que a expressão do QI se modificava consoante a idade do indivíduo (um indivíduo que tivesse um atraso de um ano na idade mental teria um QI mais baixo se o indivíduo tivesse seis anos de idade do que se tivesse 10 anos). Apesar de tudo há diferenças na concepção dos dois tipos de QI. No de Binet o QI expressa o atraso que o indivíduo tem relativamente ao caminho que deveria ter percorrido, ou seja compara-se o indivíduo consigo mesmo, enquanto no QI de dispersão se compara o indivíduo com a realização dos outros indivíduos que constituem o grupo de referência ou normativo.

Dentro da perspectiva psicométrica os resultados podem ainda ser apresentados em percentagem de respostas certas e em que o resultado máximo possível será 100 (de 100%) e a média será de 50%. Por vezes esta forma de expressar a realização intelectual pode ser confundida com o tradicional QI.

Abordagem desenvolvimental – é expressa pela teoria de Piaget. Esta perspectiva interessa-se por aspectos qualitativos da expressão da inteligência nomeadamente pela identificação de padrões universais de realização tal como, por exemplo, a invariância da aquisição de estruturas intelectuais. Enquanto a perspectiva psicométrica se interessa pelas respostas certas que o indivíduo dá num teste, a perspectiva desenvolvimental interessa-se pelos erros que o respondente faz. Enquanto a perspectiva psicométrica se interessa pela soma das

respostas certas a desenvolvimental interessa-se pela qualidade das respostas (responder correctamente a uma pergunta significa que o indivíduo já possui a estrutura cognitiva e nesse caso tanto faz responder correctamente a uma pergunta como a duas ou a três: A resposta certa a uma questão já mostra que a estrutura mais complexa existe). Nesta perspectiva não existe nenhuma medida de expressão de nível intelectual. Na Perspectiva de Piaget, por exemplo, os indivíduos evoluem através de quatro estádios de desenvolvimento que é suposto serem universais para todos os humanos: estágio sensório motor, estágio pré-operacional, estágio operacional e estágio formal. Estes estádios subdividem-se em outros sub-estádios e a terminologia utilizada pelo autor não é sempre a mesma, embora mantenha os mesmos passos de desenvolvimento.

Abordagem do processamento de informação – Esta abordagem concebe o homem como um manipulador de símbolos e visa descrever aos símbolos que são manipulados, ou seja a *representação*, e identificar a maneira como são manipulados, ou seja, o *processamento*. Interessam-se por duas áreas: (a) memória e (b) resolução de problemas. Os primeiros interessam-se pelas características qualitativas e quantitativas do sistema de memória, nomeadamente capacidade de memória, estratégias de memorização, metamemória e conhecimento base. Os que se interessam pela resolução de problemas focam a interface entre a resolução de problemas e o ambiente da tarefa (Siegler & Richards, 1982).

Personalidade

Historicamente desenvolveram-se duas formas sequenciais no tempo, e incompatíveis no método: a avaliação tradicional da personalidade e a avaliação comportamental.

Avaliação tradicional da personalidade – Nesta perspectiva da avaliação da personalidade identificam-se várias orientações que se exprimem nas medidas utilizadas na avaliação, nomeadamente: (a) as que se propõem avaliar a totalidade da personalidade e (b) as que se propõem avaliar aspectos (dimensões, traços, factores, etc.) específicos. No primeiro caso os testes tendem a incluir no título o termo “personalidade” e no segundo recorrem ao nome da dimensão a avaliar (ansiedade, depressão, etc.).

Por outro lado há (a) os testes (inventários) que avaliam dimensões básicas estruturais, dificilmente mutáveis, e (b) os que avaliam a interacção com o meio, (c) e os que avaliam dimensões flutuantes (estados). Exemplo clássico dessa distinção é o *State-Trait Anxiety Inventory (Form Y)* (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983).

Segundo Piéron (1968) na década de 30 tinham-se identificado mais de 50 definições de personalidade. De acordo com este dicionário Personalidade representa essencialmente a noção de *unidade integrativa* de uma pessoa, com todas as *características diferenciais* permanentes – inteligência, carácter, temperamento, constituição, etc. –, e as suas *modalidades únicas de comportamento*. Piéron adopta a seguinte definição de personalidade: organização dinâmica dos aspectos cognitivos, afectivos, conativos, fisiológicos e morfológicos do indivíduo. Trata-se de uma ideia de personalidade dinâmica em constante mutação, dinâmica essa que depende da interacção entre todas aqueles aspectos.

Magnusson e Torestad (1993) afirmam que não existe definição de personalidade. O que existe, dizem, é um organismo vivo, activo, intencional, funcionando e desenvolvendo-se como um ser total e integrado, e, assim, a investigação em personalidade seria definida como o estudo de como os indivíduos pensam, sentem, actuam e reagem, tal como eles o fazem, ou seja, do ponto de vista do indivíduos como um organismo total e integral. Para estes autores os aspectos que outros nomeiam são secundários.

Genericamente os psicólogos adoptam uma perspectiva alargada do conceito de personalidade, ou seja abrangem tanto “os traços intelectuais como os não intelectuais” (Anastasi, 1990, p. 18 confirmar). No entanto, como refere também Anastasi, na terminologia dos testes psicológicos a designação de “teste de personalidade” frequentemente refere-se a medidas de características tais como ajustamento emocional, relações interpessoais, motivação, interesse, atitudes, ou seja, a características que são avaliadas com provas de realização habitual em vez de realização máxima. Deixa, portanto, de fora a avaliação da inteligência.

A avaliação da personalidade é, segundo Ozer e Reise (1994) uma prática científica que visa, (a) determinar aquelas características que constituem diferenças individuais importantes, (b) desenvolver medidas precisas de tais atributos e, (c) explorar amplamente o significado das consequências da identificação e medição de tais características.

Como se referiu antes as medidas de personalidade dividem-se em medidas de traços simples ou de um traço abrangendo medidas de variáveis como a ansiedade ou a depressão, e medidas que se propõem avaliar a totalidade da personalidade. Seria suposto que os traços que emergem dessas diversas teorias que suportam as técnicas de avaliação que se propõem avaliar a totalidade da personalidade fossem idênticas, mas tal não acontece.

A avaliação da personalidade inclui grande variedade de medidas, de técnicas e de variáveis. Estas técnicas são orientadas, quer para o diagnóstico

psicopatológico quer para a avaliação de aspectos positivos da personalidade. Por exemplo, o MMPI foi concebido para diagnosticar indivíduos que têm uma personalidade patológica e inclui uma nota, um valor, fronteira entre o que é normal e o que é patológico. Os indivíduos que têm nota abaixo são normais enquanto os que têm acima são doentes.

No entanto, a maioria dos testes de personalidade não define essa fronteira. Apenas se pode dizer que o indivíduo tem muito, pouco, ou outra classificação, da variável, mas não se sabe se tal valor é patológico. São instrumentos que, como afirma Bowling (1994) permitem medir a gravidade da variável após o diagnóstico estar feito. Acontece com inúmeras técnicas tais como o inventário de depressão de Beck (Beck, 1972) ou a Escala de Depressão de Hamilton (Hamilton, 1967), ou outros equivalentes que avaliam traços simples.

Muitas teorias da personalidade assumem que os traços são características persistente do indivíduo, que sustentam ou espelham a personalidade, e que estão por detrás, do pensamento, dos sentimentos, do comportamento, da forma de agir, nas actividades do dia a dia, sendo suposto que, ao identificar os traços, e assim, a personalidade do indivíduo, seria possível prever o seu modo de actuar agora e no futuro.

Magnusson e Torestad (1993) discutem as teorias que procuram entender como e porquê os indivíduos funcionam de determinado modo num dado momento ou estágio, referindo duas teorias ou modelos complementares: aquelas que discutem o tema de uma perspectiva *corrente* e as que o fazem de uma perspectiva *desenvolvimental*. As que salientam a perspectiva corrente (tais como os modelos cognitivos e psico-biológicos) analisam e explicam o modo como os indivíduos funcionam com base nas suas disposições biológicas e psicológicas. Não consideram a possibilidade de existência de um processo desenvolvimental passível de conduzir o indivíduo ao seu estado actual. Os modelos desenvolvimentais (como o psicanalítico e os genéticos) analisam e explicam o funcionamento actual do indivíduo em função da sua história desenvolvimental. Estes modelos estão atentos ao momento e à expressão de acontecimentos importantes no passado e na actualidade, e no modo como esses aspectos interagiram para produzir o funcionamento actual.

Os traços na personalidade – O traço tem sido uma unidade de análise do comportamento privilegiada desde os primórdios da psicologia. Praticamente todos os teóricos do traço assumem que os traços de personalidade têm ligação ao comportamento (Digman, 1990), embora varie o modo como traços e comportamentos se relacionam.

Lersch (1954) explicava que se podiam distinguir três tipos de traços: Os traços de comportamento que descrevem o aspecto externo do comportamento sem se interessar pelo que o provoca; os traços de realização que descrevem as maneiras características de realizar as tarefas e as realizações objectivas sem se interessar, também, pelas causas; finalmente os traços de essência que se interessam pelos processos e estados intra-psíquicos, ou seja, pela realidade interior.

Numa outra forma de analisar os traços, Murphy e Davidshofer (1998) há três modos diferentes de pensar acerca dos traços: (a) os traços provocam os comportamentos, (b) traços só existem na cabeça do investigador, e (c) os traços são sumários descritivos. No primeiro caso os traços explicam o comportamento. O comportamento honesto da pessoa decorreria de o indivíduo ter um traço de honestidade. Na segunda perspectiva o traço seria uma organização cognitiva do investigador que, ao pretender organizar o comportamento dos outros numa totalidade coerente construía traços; é próxima da anterior. Na terceira perspectiva o comportamento do indivíduo não seria provocado pelos traços, antes os traços seriam descrições do comportamento. Assim, quando uma pessoa fosse descrita como honesta este nome do traço descreveria um conjunto de comportamentos que se relacionavam com a honestidade.

A ideia que os traços são algo inerente ao indivíduo, uma estrutura tal como um esqueleto, e que todo o comportamento estaria dependente desse esqueleto, era apelativa se se tivesse conseguido identificar ou avaliar essa estrutura. Ora o que se verifica é que há uma longa história de discordância acerca de quais os traços a avaliar. Um escrutínio breve por alguns dos mais marcantes teste de personalidade mostra, o seguinte:

O *Califórnia Psychological Inventory* possui 420 itens distribuídos por 20 escalas, a saber: dominância; capacidade para estatuto; sociabilidade; presença social; auto-aceitação; independência; empatia; responsabilidade; socialização; auto-controlo, boa impressão; comunhão; bem-estar; tolerância; realização via conformidade; realização via independência; eficiência intelectual, mente psicológica, flexibilidade; masculinidade/feminilidade.

O *Personality Research form* é uma das mais recentes medidas de personalidade. A versão longa do teste contém 440 itens (a curta 300) distribuídos por 20 escalas de personalidade mais duas de validade, a saber: humildade; realização; pertença; agressão; autonomia; mudança; estrutura cognitiva; defensivo; dominador; resistente; exibicionismo; evitamento da ameaça; impulsividade; sustentador; ordem; jogo; atento; reconhecimento social, procura protecção; compreensão; desejabilidade; infrequente.

O *Sixteen Personality Factor Structure* (16PF) tem 185 itens agrupados em 16 dimensões de personalidade (factores primários) e 5 factores globais, a saber: caloroso; racional; estabilidade emocional; dominância; vivacidade; consciência das regras; arrogante; sensibilidade; vigilância; abstracção; privacidade; apreensão; abertura à mudança; confiança em si; percepcionismo; tensão. Os factores globais incluem: extraversão; ansiedade; resoluto; independência; auto-control.

O *Edward Personal Preference Schedule* inclui 210 pares de afirmações que definem 15 necessidades de personalidade, a saber: realização; deferência; ordem; exibição; autonomia; pertença; introspecção; segurança; dominância; vergonha; sustentação; mudança; resistência; heterossexualidade; agressão.

Uma análise destes testes mostra grande desacordo quanto à consistência dos constituintes da personalidade. Historicamente tem sido teorizada a hipótese de que cinco factores explicariam a personalidade dos indivíduos embora, também neste caso, houvesse divergências acerca desses cinco factores (Digman, 1990). Segundo Digman na década de 30 tinha sido defendido que a personalidade podia ser definida por cinco grandes factores distintos e a análise factorial mostrava tal distribuição.

Na década de 80 pareceu nascer algum consenso de que a personalidade podia ser caracterizada por cinco factores, que são comuns em muitas das técnicas referidas antes – conhecidos por “*big five*” – dado que têm emergido em vários estudos e com vários métodos de avaliação e em diferentes culturas. Desenvolveu-se uma técnica de avaliação para medir esses cinco factores – o *NEO Personality Inventory*.

O *Revised NEO Personality Inventory* constitui uma taxonomia que tem sido aceite (embora não por todos) como uma base para descrever o que os inventários de personalidade deverão medir. Tem cinco factores – neuroticismo; extraversão; abertura à experiência; agradabilidade; consciência, e cada um destes factores inclui seis facetas (total de 30 facetas).

Aplicações da avaliação psicológica tradicional – É normalmente utilizada para uma das seguintes funções: (a) classificar pessoas por comparação com um grupo de referência a que o avaliado pertence; (b) descrever pessoas segundo traços de personalidade; (c) descrever pessoas segundo psicopatologia; (d) verificar a estrutura, a estabilidade ou a magnitude dos traços de personalidade.

Na avaliação tradicional os resultados dos testes podem ser considerados de duas maneiras distintas (a) por comparação com a norma e (b) por comparação com a teoria.

No primeiro caso, os resultados dos testes são apresentados como uma nota (score), e o significado dessa nota decorre da comparação com valores padrão que provém de médias de resultados encontrados na realização do teste por pessoas “normais” ou padrão.

No segundo caso as respostas do sujeito são comparadas com a teoria que governa o teste. A resposta do sujeito é indexada, com base na teoria, a uma estrutura ou dimensão da personalidade e, assim, a personalidade do sujeito é descrita.

A avaliação da personalidade na perspectiva comportamental – As técnicas de avaliação da personalidade foram dominadas até aos anos sessenta pela concepção das teorias de traços. Nos anos sessenta cresceu o interesse pela avaliação em situação específica, com rejeição da perspectiva anterior (Anastasi, 1990). Este modo que emergiu nos anos sessenta surge no âmbito das teorias da aprendizagem, das comportamentais, mantendo-se como uma referência importante até aos anos oitenta (Fernandez-Ballesteros, 1994).

Esta nova forma de avaliar a personalidade foi designada por *comportamental* em oposição à vigente até então designada por *tradicional* (Fernandez-Ballesteros, 1994; Goldfried & Kent, 1972). A avaliação *tradicional* visaria predominantemente a compreensão das características pessoais – estrutura da personalidade – subjacente ao comportamento. As *comportamentais*, por contraste, propõe-se avaliar directamente a realização, ou o modo de responder do indivíduo em situação natural na vida do dia a dia. A assunção básica nesta perspectiva é que a unidade de investigação no estudo da personalidade *é a interacção entre o indivíduo e o seu meio ambiente significativo, e não a personalidade por si*. Nesta perspectiva assume-se que o comportamento em diferentes situações varia enquanto na perspectiva tradicional se assume que é estável em diferentes situações.

A perspectiva tradicional baseava-se na ideia de determinismo psíquico, no sentido em que se assumia que o comportamento do indivíduo é determinado por uma dinâmica estrutural pré existente ao comportamento. O funcionamento individual dependeria dos *traços* e não do meio ambiente.

Ao contrário, a perspectiva comportamental, mais descritiva, coloca a ênfase na situação onde o comportamento ocorre, com a ligação entre o comportamento e os estímulos que o antecedem ou precedem. Uma das características da avaliação comportamental consistia na recusa sistemática dos testes tradicionais, dos conceitos que ele avaliavam, e do diagnóstico psiquiátrico.

A avaliação comportamental que se estabeleceu em oposição aos sistemas de classificação diagnóstica, e focando aspectos opostos aos construtos cognitivos, entra em declínio nos anos oitenta. Por um lado, devido à crescente adopção do *Diagnostic and Statistical Manual of Mental Disorders* e ao sistema de classificação da Organização Mundial de Saúde, o *International Classification of Diseases* que impõem aos psicólogos comportamentais a utilização de categorias diagnósticas. Por outro a própria fusão das perspectivas comportamentais com as cognitivas (Fernandez-Ballesteros, 1994).

Ao contrário da teoria psicométrica tradicional em que a resposta a um teste tem implícita a ideia de que o modo como o sujeito responde ao teste é idêntico ao modo como o sujeito responderia na situação real (ligação resposta-resposta, ou R-R), a avaliação de inspiração comportamental tem implícita a ideia de ligação da resposta com um estímulo antecedente ou consequente (ligação estímulo-resposta -E-R-, ou R-E).

A avaliação na perspectiva comportamental supõem várias características inerentes a esta teoria que Fernandez-Ballesteros, (1994) designa por (a) análise funcional, (b) análise indutiva (em oposição a uma lógica hipotético-dedutiva) e (c) idiografismo.

A análise funcional, segundo Fernandez-Ballesteros (1994), propõe-se determinar através da observação rigorosa, as relações entre o comportamento problema que desencadeou a observação, e os antecedentes e consequentes. A observação deve ser realizada em meio natural (ao invés da avaliação tradicional padronizada que deve ser realizada em condições laboratoriais ideais).

Se originalmente a análise dos antecedentes e consequentes se referia unicamente aos aspectos comportamentais observáveis, num segundo momento a observação passou a incluir igualmente eventos internos ao organismo. A observação visa não somente o comportamento motor observável mas também o comportamento cognitivo-verbal e fisiológico. Os auto-registos acerca de acontecimentos internos (comportamentos cobertos em oposição aos comportamentos abertos, os que são observáveis) constituem, também, um tipo especial de comportamento desde que o avaliador não pretenda inferir deles atributos intrapsíquicos ou estruturas molares (traços).

A análise indutiva, em oposição à hipotético-dedutiva, impõe que não se saia dos dados estritos que o cliente apresenta, ou que se observa, ao invés de incluir o que se observa em grandes construtos indexados a conceitos teóricos: Num primeiro momento recolhe-se dados do sujeito e do seu meio ambiente, por observação ou por informação prestada pelo observado; constata-se quais os

comportamentos que podem ser problemáticos, e quais são os antecedentes ou consequentes que o sustentam.

O idiografismo está estreitamente ligado ao indutismo e refere-se aos procedimentos de recolha de dados descritivos de modo exaustivo em que toda a informação é pertinente, para identificar o comportamento problema. Como se nota a metodologia de avaliação comportamental difere radicalmente da tradicional que, ao invés de identificar o que o indivíduo faz, o que está associado ao comportamento, e o que é único nesse indivíduo, fornece uma nota, (score). Esta deverá ser comparada com uma média, e em que as notas resultam da soma de itens que diluem a informação fornecida por cada um desses itens.

Aspectos positivos da personalidade – Durante muito tempo foi dominante uma avaliação da personalidade que se debruçava predominantemente sobre os aspectos negativos desta. Progressivamente ganhou estatuto uma avaliação que focava outros aspectos que podem ser avaliados no âmbito da personalidade embora frequentemente sejam considerados como independentes desta.

Huber (1977) explicava que a psicologia da personalidade se tinha interessado por diferentes unidades de análise a saber: traços, hábitos, motivos, self, papel, e atitude. Para Anastasi (1990) interesses, atitudes, motivos e valores individuais representam aspectos importantes da personalidade. Sundberg (1977) explica que “aqueles que desejam compreender a personalidade fariam bem se dessem atenção às atitudes, opiniões, crenças, e valores” (p. 192). Diz este autor que estes construtos se referem a ligações importantes entre o indivíduo e a sociedade. As atitudes caíram por acidente histórico no domínio da psicologia social mas teria sido mais razoável se tivessem sido desenvolvidas no seio da teoria da personalidade. Ozer e Reise (1994), explicam que “a avaliação da personalidade é muitas vezes medição de traços. No entanto motivos, intenções, crenças, estilos e estruturas constituem as unidades conceptuais preferidas de alguns investigadores” (p. 358).

Ora, porque a psicologia da saúde é um domínio de aplicação que recorre aos conhecimentos provenientes da psicologia, para actuar no contexto da saúde, neste manual ultrapassar-se-à esta fronteira para considerar todas as dimensões referidas integradas na personalidade e importantes para compreender a acção do indivíduo na relação com a saúde.

As dimensões que serão referidas a seguir constituem elementos positivos da personalidade, no sentido em que, ao contrário das avaliações tradicionais da personalidade, valores baixos destes aspectos não constituem defeito ou patologia.

Interesses são respostas afectivas, aprendidas, para com um objecto ou actividade; coisas nas quais um indivíduo está interessado elicitam sentimentos positivos, coisas em que há pouco interesse elicitam pouco afecto, e coisas que não despertam interesse elicitam apatia ou mesmo sentimentos de aversão (Murphy & Davidshofer, 1998).

Atitude é segundo refere (Sundberg, 1977), a predisposição do indivíduo para avaliar determinado símbolo, objecto ou aspecto do seu mundo de uma maneira favorável ou desfavorável. Para Ajzen (1988), atitude “é a disposição para responder favoravelmente ou desfavoravelmente a um objecto, pessoa, instituição, ou acontecimento” (p. 4). Este autor explica que as atitudes são construtos hipotéticos inacessíveis por observação directa, que têm de ser inferidos por respostas mensuráveis. Continua afirmando que uma das classificações mais populares das respostas consiste no cruzamento de dois eixos, um eixo, referente ao tipo de resposta, abrange respostas não verbais versus verbais, e outro eixo, englobando três categorias de respostas: cognitivas, afectivas e conativas.

Opiniões são muito semelhantes a atitudes de tal modo que podem ser utilizadas como sinónimos, mas o termo opinião tem sido utilizado para referir a expressão de um sentimento para com um item de cada vez em vez de vários itens à vez. Considera-se que as opiniões são mais objectivas e menos contaminadas por aspectos emocionais do que as atitudes, relacionando-se mais de perto com factos e conhecimento do que as atitudes.

Crenças são segundo Sundberg (1977) assunções que nunca são conscientemente examinadas ou verificadas. Ajzen (1988) refere que as crenças são categorias de resposta que expressam atitudes.

Valores são assunções pessoais acerca do que é desejável e que é relevante para a selecção do comportamento Sundberg (1977).

O que é específico na utilização dos teste em saúde

A psicologia clínica evoluiu de uma função diagnóstica, típica na primeira metade do século para uma função cada vez mais interventora (Pais-Ribeiro, 2001; Pais-Ribeiro & Leal, 1996, 1997).

A psicologia da saúde é, grosso modo, a aplicação do conhecimento acumulado pela ciência e prática da psicologia geral na arena da saúde, como defende

Matarazzo (1987, p. 55). Deste modo a avaliação psicológica que se realiza em psicologia da saúde rege-se pelas mesmas regras que a avaliação psicológica em geral. Os testes ou técnicas a que se recorre são as mesmas. Grosso modo, os testes de inteligência e de personalidade. No entanto são necessários ajustamentos aos contextos e problemas próprios deste domínio que se pode exemplificar com a avaliação de sintomas. A psicologia da saúde tem recorrido a técnicas de avaliação que se baseiam na descrição de sintomas que tipificam dimensões tais como a ansiedade ou a depressão. Entre os mais conhecidos encontram-se o inventário de depressão de Beck (Beck, 1972) de Hamilton, de Zung, ou o Symptom Checklist de 90 itens (SCL-90) que agrupa estes sintomas em dimensões de personalidade. Embora estes instrumentos já fossem utilizados na avaliação da personalidade, têm agora mais utilidade, numa época que se interessa pela identificação dos sintomas que as pessoas apresentam.

No modelo biomédico que a psicologia adoptou, a avaliação psicológica tinha dois objectivos principais: (a) a *identificação* e (b) o *tratamento* das perturbações psicológicas. A avaliação visava uma dimensão única do doente, o seu estado mental, sem se interessar pela dimensão corporal (Belar & Deadorff, 1995). As mudanças históricas, sociais, económicas e culturais, impuseram uma nova maneira de pensar a psicologia que se fazia no Campo da Saúde em geral, e no da avaliação psicológica em particular.

Segundo Green (1985), no início, os psicólogos nos serviços de saúde não-mental recorriam aos instrumentos tradicionais de avaliação psicológica úteis na avaliação psicopatológica em contextos de doença mental mas desadequados noutros contextos de doença. Com efeito a necessidade de garantir as propriedades métricas dos testes impõem a existência de instrumentos muito longos que não são adequados para pessoas fragilizadas, acamados, ou em condições semelhantes.

Da panóplia de técnicas de avaliação e diagnóstico que a psicologia desenvolveu durante quase um século, Green refere uma lista de instrumentos clássicos que incluem algumas daquelas técnicas, e que podem ser úteis. A autora agrupa-os em quatro categorias: listas de sintomas; instrumentos que avaliam traços simples; inventários de estilo de vida; inventários de personalidade.

A mudança dos contextos de saúde arrastou a avaliação para domínios e formas menos consideradas até então e que exibem diferenças com a avaliação tradicional.

Johnston (1997) enuncia várias áreas, domínios, aspectos, de avaliação em Psicologia da Saúde, tais como, por exemplo, dor e comportamento de dor,

stress, emoções e acontecimentos de vida, processos de coping, suporte social, estado de saúde, e qualidade de vida relacionada com a saúde, doenças, sintomas, alterações funcionais, reabilitação, expectativas, experiência e avaliação dos cuidados de saúde, diferenças individuais e demográficas, crenças acerca das causas das doenças, crenças e conhecimentos acerca das doenças e da saúde, comportamento relacionado com a saúde e com as doenças, etc.

McDowell e Newell (1987) utilizam as seguintes categorias para avaliação em contextos de saúde: Incapacidade funcional; bem-estar psicológico; saúde social; qualidade de vida e satisfação com a vida; dor, saúde geral.

Bowling (1991) organiza as medidas em saúde nas seguintes categorias: capacidade funcional; estado de saúde; bem-estar psicológico; redes sociais e suporte social, satisfação com a vida e moral.

Quando é que a avaliação em saúde é avaliação psicológica

Tem-se defendido que a avaliação psicológica se mestiçizou com a pragmática da intervenção psicológica em contextos de saúde, tendo-se desenvolvido inúmeras técnicas de avaliação que à primeira vista não parecem provenientes da psicologia. Propusemos os seguintes critérios que permitam contribuir para diferenciar entre o que, eventualmente, poderá ser avaliação psicológica e o que não o será. As funções do Psicólogo nos Serviços de Saúde exigem que o psicólogo: identifique os factores que facilitam a emergência de, ou mantêm a crise; identifique as reacções que são normais à doença física das que são psicopatológicas; avalie a saúde, o bem-estar, a qualidade de vida de pessoas que têm, ou não, doenças. Assim, o que caracteriza a avaliação psicológica seria uma combinação dos seguintes aspectos:

- *Os construtos* – se os construtos utilizados tiverem origens claras na Psicologia pode-se dizer que a variável é psicológica;
- *O tipo de técnicas de avaliação utilizadas* – a psicologia recorre a determinadas técnicas (testes, questionários, escalas, inventários, listas de verificação, listas de classificação, etc.) que avaliam amostras de comportamento ou indícios. Se uma técnica avalia indícios de um atributo, ou comportamentos, ela terá, provavelmente, uma base psicológica forte;
- *Dimensões subjectivas* – o objecto de avaliação deve ter uma base de expressão baseada na percepção pessoal, variando consoante a teoria ou modelo subjacentes à técnica de avaliação utilizada.

INTERPRETAÇÃO DAS RESPOSTAS

Há três maneiras de interpretar as respostas que os respondentes dão nas técnicas de avaliação: a resposta pode ser interpretada como (a) uma *amostra de comportamento*, (b) como *indício de uma característica de personalidade*, ou como (c) *correlato de outro comportamento* (Loevinger, 1957; Sundberg, 1977).

Considerar as respostas como amostras de comportamento significa que a pergunta pede uma resposta que é um comportamento habitual: por exemplo “quantos cigarros fuma?” a resposta que o sujeito dá é o comportamento.

Uma resposta como indício é o tipo de resposta que é pedida, por exemplo, nos testes projectivos. A resposta dada indicia determinada característica da personalidade que só o psicólogo pode conhecer. Aqui há a suposição de causalidade: o indivíduo dá tal resposta porque tem tal característica de personalidade. Como explica Sundberg a resposta expressa a *natureza* da pessoa; é indicadora de uma condição inata, simboliza um processo pessoal abrangente. O psicólogo regista e infere a característica de personalidade de que a resposta é indício.

Uma resposta como correlato de outro comportamento significa que aquela resposta é dada por pessoas que possuem determinada característica. Por exemplo, se um indivíduo faz afirmações que têm sido associadas (correlacionadas) com racistas, a resposta permite dizer que o indivíduo é, provavelmente, racista. Considerar a resposta como correlato não implica assumir a ideia da causalidade: apenas a de correlação, ao contrário da resposta assumida como indício.

As técnicas de avaliação psicológica podem ser classificadas quanto ao modo como as respostas são interpretadas de vários modos, nomeadamente: (a) relativamente a normas, (b) relativamente à teoria e (c) relativamente a critérios.

Na comparação com normas, (ou interpretação com base em normas) as respostas são classificadas numericamente, fornecendo, normalmente uma nota (score) que é comparada com a nota de um grupo que serviu de referência, o grupo normativo. O sujeito em avaliação obteve uma nota que é igual, superior,

ou inferior à média do grupo de referência, e este resultado pode, ainda, ser indicador de patologia. Este tipo de resultados é o único adequado para classificar os indivíduos, para os comparar uns com os outros. É o tipo de classificação das técnicas psicométricas. De modo geral as normas, ou seja o grupo que foi escolhido para representar o grupo normativo, pode ser uma amostra nacional, regional, local, ou pode ser ainda por grupo de idade, por sexo, por nível de escolaridade ou outro.

No segundo tipo de comparação das respostas – interpretação com base na teoria – as respostas dos sujeitos, normalmente qualitativas, são analisadas com base na teoria que está subjacente ao teste, e, com base nesta, as respostas dos sujeitos são indexadas para uma categoria, uma estrutura, dessa teoria. É assim com os testes projectivos, e é assim, também, com a avaliação piagetiana. Esta avaliação recorre àquilo que Piaget chama “Método Clínico” que, perante um questionamento apropriado, feito segundo o método clínico, visa elicitar a resposta do sujeito perante situações padrão. Tal ou tal resposta indicará a existência de uma determinada estrutura cognitiva. A interpretação com base na teoria implica, não só um bom conhecimento do método de questionamento como um bom conhecimento da teoria. Aliás, a aplicação do método de questionamento impõe um bom conhecimento da teoria.

No terceiro tipo de comparação – respostas comparadas com um critério – as respostas do sujeito são comparadas com um critério definido antes. Por exemplo estabelece-se que os sujeitos submetidos a um ensino de matemática devem, no final da aprendizagem, ser capazes de responder acertadamente a 75% dos problemas de determinado tipo que lhe são colocados de um tipo que é conhecido no início. É um tipo de avaliação desenvolvido em contextos de intervenção por objectivos, em que os objectivos são estabelecidos com base, não no que os indivíduos normalmente respondem, mas sim no que, criterialmente, se define que idealmente deveriam ser capazes de responder. O sujeito atinge ou não os objectivos, ou seja, o sucesso é determinado em termos de sim ou não, e não de certo ou errado. Este método não permite classificar os respondentes. Os indivíduos não se comparam entre si, eles são comparados somente com a tarefa a realizar.

Tipos de escalas de medida

Uma máxima da psicologia que permitiu a sua afirmação como ciência a par de outras ciências “duras” é a de que se um construto existe ele existe nalguma quantidade, logo é mensurável. Com efeito a investigação científica depende de

quanto bem os conceitos se podem medir, da possibilidade dos resultados duma qualquer variável se distribuírem por categorias ou valores. Por medição entende-se a atribuição de números a objectos, acontecimentos ou observações de acordo com um conjunto de regras. Por vezes os números significam que uma observação pertence a determinada categoria, enquanto outras vezes significa que a observação tem mais quantidade de uma propriedade do que outra observação que tem um valor mais baixo.

Em geral, em psicologia, as propriedades psicológicas são medidas indirectamente porque não temos maneira de as observar directamente. Por exemplo, no caso da inteligência que se pode exprimir por um QI, podemos inferi-la pelo modo como as pessoas respondem a testes que lhes pedem que resolvam problemas de grau de dificuldade variável. Assume-se que as pessoas que resolvem problemas mais difíceis, ou que os resolvem mais depressa, são mais inteligentes. Não podemos observar a inteligência de um modo mais directo do que este. O mesmo se passa com a maioria dos outros conceitos da psicologia. Por vezes, outras disciplinas chamam a este tipo de medidas “subjectivas” ou “moles”, em oposição a outras medidas “objectivas” ou “duras”. Frequentemente considera-se que, se a pergunta ou a resposta envolver palavras, se está na presença de medidas subjectivas, na medida em que a palavra é mediada por processos de pensamento passíveis de alterar a resposta. Mesmo na psicologia, os testes livres de cultura (sem palavras ou com um mínimo de palavras) eram considerados mais fiáveis do que os que tinham muito texto.

A psicologia teve de defrontar-se sempre com a crítica de que as suas medidas são “subjectivas” e tem sabido lidar bem com ela. Principalmente no sistema de saúde e na perspectiva do modelo biomédico desenvolvido numa perspectiva mecanicista, tende-se a acreditar que as medidas que são dadas directamente por máquinas são melhores medidas. Por exemplo, tende-se a acreditar que a medida de pressão arterial que é representada pela elevação do mercúrio ou por um registo electrónico, ou que o registo da actividade eléctrica do cérebro ou do coração, que são fornecidos por máquinas são menos subjectivas, ou mais precisos, o que não é verdade. As medidas fornecidas por máquinas são afectadas como qualquer outra por um erro de medida que pode ser devido ao material utilizado, ou a qualquer outra característica. O que acontece é que em psicologia o erro de medida é um aspecto que o psicólogo considera sempre que faz a avaliação, e que deve estar presente em todo o processo. O que está em jogo não será subjectividade versus objectividade, mas sim fidelidade e validade da medida versus a falta dela.

De um modo geral as técnicas de avaliação psicológica estão bem estudadas, para todos os fenómenos psicológicos, com validade e fidelidade bem estabelecidas que podem ser usadas com eficácia desde que os manuais sejam utilizados apropriadamente.

Na década de 40 Stevens (1946) propõe que todas as medidas podiam ser agrupadas num de quatro tipos.

Medidas nominais/categoriais

As medidas nominais ou categoriais reflectem diferenças qualitativas em vez de diferenças quantitativas. Exemplos comuns são masculino/feminino, certo/errado ou casado, solteiro, viúvo, divorciado. Números podem substituir as palavras (1 para masculino, 2 para feminino), como por vezes se faz quando se introduzem dados no computador, sem que tal signifique que 2 é mais do que 1. Os sistemas de medida categoriais ou nominais requerem duas propriedades, mutua exclusividade e exaustividade (Fife-Schaw,1995). Mutua exclusividade significa que cada observação só pode caber numa categoria; exaustividade significa que o sistema de categorias deve ter tantas categorias quantas as necessárias para todas as observações. Nalguns casos como o certo/errado é fácil assumir que certo é melhor do que errado, no entanto as medidas nominais/ categoriais não assumem à priori que uma categoria é mais ou melhor do que outra.

Medidas ordinais

Trata-se de um nível de medição mais complexo do que o anterior. Tal como na anterior as assunções de mutua exclusividade e exaustividade aplicam-se, e todos os casos são agrupados numa das categorias. A grande diferença é que as categorias são ordenadas segundo um critério definido à priori, de tal modo que pertencer a uma categoria significa ter mais ou menos do que pertencer a outra categoria. Os resultados dos instrumentos de avaliação em psicologia são normalmente expressos em escalas ordinais. Se numa escala que supostamente mede a ansiedade, por exemplo, determinado indivíduo difere três pontos de outro, sabemos que um tem mais do que o outro, mas *não sabemos quanta mais ansiedade um tem do que o outro*. Uma das consequências desta propriedade é que este tipo de escalas não deve ser tratado estatisticamente com estatística paramétrica, como é frequente observar-se.

Escala de intervalo

Tal como nas escalas ordinais, os números significam que uns têm mais do que outros na dimensão medida. A distinção principal da medida anterior é que com

medidas intervalares os números estão à mesma distância, ou seja, num termómetro 2 graus de diferença entre os 38 e 40 graus significa exactamente o mesmo que 2 graus de diferença entre os 10 e 12 graus. Muitos autores de medidas em psicologia assumem que os resultados se exprimem em medidas intervalares, tal como ocorre, por exemplo com a avaliação da inteligência, ou em escalas que sejam construídas para o fazer como as escalas de Thurstone.

Medidas de rácio

A diferença entre estas medidas e as de intervalo é que as de rácio possuem um zero absoluto. Um bom exemplo de medidas deste tipo são escalas de comprimento, de peso, tempo ou respostas correctas num teste. Em qualquer destes casos é possível haver medida zero. A consequência principal deste tipo de medidas é que se um indivíduo tem 120 kg de peso pesa o dobro de outro que pesa 60 kg, o que não ocorre se não houver “zero” como é, por exemplo, o caso da inteligência.

Técnicas de medição/tipo de escala

Inúmeras técnicas de medição recorrem aos tipos de escala referidos acima, embora tenham propriedades específicas que qualificam a técnica de construção das escalas ou a forma de as utilizar.

Métodos de classificação categorial

Neste procedimento o investigador apresenta um conjunto de itens a um grupo de sujeitos com instruções para classificar os itens por ordem segundo determinado critério. Por exemplo Holmes e Rahe (1967) apresentaram aos respondentes (para classificar por ordem) um conjunto de acontecimentos e pediu-lhes que os classifikassem consoante o ajustamento social que achavam que esses acontecimentos lhes exigiram maior esforço, ou lhes provocou maior desconforto, no ano anterior. A mecânica da ordenação e magnitude dos eventos era assim pedida pelos autores: “Ao primeiro acontecimento, casamento, foi dado um valor arbitrário de 500. Quando responder a cada um dos restantes

eventos pense para si próprio, este acontecimento exige mais ou menos ajustamento do que o casamento? Se decidir que o ajustamento é mais intenso e prolongado, então escolha um número proporcionalmente maior e coloque-o no espaço” (Hormes & Rahe, 1967, p. 213). Com base nos valores arbitrários atribuídos pelos respondentes a cada acontecimento a escala de Holmes e Rahe atribui ao acontecimento “a morte do conjugue” a classificação mais elevada, um valor arbitrário de 100, ao acontecimento seguinte, ao divórcio o valor de 73, ao terceiro, a separação conjugal o valor de 65, ao quarto e quinto acontecimento o mesmo valor, 63, enquanto o último, o 43º acontecimento, uma violação menor da lei recebeu o valor arbitrário de 11.

Este método de classificação tem sido considerado difícil por alguns autores que sugerem variações (Smith, 1975).

Escala de Likert

É a escala mais utilizada em avaliação de atitudes. Consiste numa série de afirmações em que os respondentes devem indicar a sua concordância ou discordância numa escala de intensidade, por exemplo, “Concordo bastante”, “concordo”, “indeciso”, “discordo”, “discordo bastante”, ou “nunca”, “quase nunca”, “indeciso”, “quase sempre”, ou “sempre”. A escala de Likert produz uma escala ordinal que deve ser tratada com estatística não-paramétrica.

Começa-se por produzir afirmações ou itens acerca de um assunto a que os sujeitos respondem assinalando uma posição numa escala do tipo que é indicado acima. Bowling (1998) exemplifica com os seguintes itens ou afirmações: “Sinto sempre dores”, “sinto dores graves”, “sinto dores mas isso não me incomoda”, “as minhas actividades estão limitadas por causa da dor”, e a cada uma das quais os sujeitos devem responder consoante o grau de concordância com a afirmação, de “concordo bastante” até “discordo bastante”.

As escalas de Likert incluem vários itens que devem ser somados para produzir uma nota, e os itens somados por pertencerem à mesma escala devem ter correlação elevada uns com os outros.

Formas mais sofisticadas da escala incluem a selecção dos itens que melhor diferenciam os 25% de respondentes com atitudes mais positivas dos 25% com atitudes mais negativas. Para uma afirmação ser considerada boa deve haver diferenças estatisticamente significativas entre estes dois grupos para cada um dos itens. Se tal não ocorrer então o item deve ser eliminado.

Escala análoga visual

Bowling (1998) considera-a idêntica a uma escala de Likert na medida em que é pedido ao respondente que assinale o grau de concordância ou discordância com uma afirmação numa escala, normalmente colocada na horizontal, constituída por uma linha em que se marcar num extremo significa que “concorda bastante” enquanto se marcar no outro extremo significa que “discorda bastante”. Por exemplo ela é utilizada na avaliação psicológica da dor (Schoenfeld & Costello, 1994), função em que é considerado um dos procedimentos mais usados (Knape, 1994). Willard (1995) explica que numa escala análoga visual para avaliação da dor se apresenta ao respondente uma linha de 10 cm em que num extremo se refere à “ausência de dor” e o outro extremo se refere à “pior dor possível de imaginar”. O indivíduo deve registar nesse contínuo a dor que sente. O resultado consiste na medição em centímetros do ponto assinalado pelo respondente.

Há várias formas idênticas a esta. Por exemplo, uma escala de classificação numérica é idêntica à anterior mas o valor de dor mais baixo descrito tem o valor “0”, no mais alto o valor “10” (ou 100) em que todos os valores de 0 a 10 estão indicados e o respondente deverá escolher um desses valores para indicar a dor que sente (Knape, 1994). Outras formas semelhantes podem ser encontradas tais como a escala de faces em que uma série de faces são apresentados tendo num extremo uma face sorridente e no oposto uma face triste com outras faces intermediárias indicando um contínuo da alegria à tristeza., ou como uma escada que é apresentada verticalmente em que o degrau mais baixo corresponde ao valor mais baixo da escala e o mais alto ao mais elevado.

Escala de Thurstone

Smith (1975) refere três métodos de construção de escalas que foram propostas por Thurstone e que ficaram com o seu nome: (1) *comparações emparelhadas*, (2) *intervalos de aparência igual*, e (3) *intervalos sucessivos*.

Técnica de comparações emparelhadas – Esta escala propõe-se ser uma escala intervalar em vez de uma escala ordinal. Parte da premissa que se existir um número grande de afirmações sobre atitudes e se fizer comparações emparelhadas dessas afirmações ao longo de um contínuo favorável-desfavorável, o respondente pode expressar uma afirmação de cada par como sendo mais favorável ou desfavorável do que a outra.

O modo de construir uma escala deste tipo começa pela produção de uma pool de afirmações ou itens (pelo menos 20 segundo Smith, 1975, mas podendo ir até 100) acerca de um assunto qualquer. Todos os itens dessa pool são emparelhados de modo a que cada item apareça emparelhado com cada um de todos os outros itens (para 20 itens há 190 emparelhamentos). Em segundo lugar, pede-se a um grupo de pessoas que actuam como juízes, que classifiquem em cada par qual a afirmação mais favorável e a menos favorável. De seguida organiza-se uma tabela de dupla entrada em que aparece o número de vezes que cada afirmação é cotada como mais favorável do que cada uma das outras afirmações. Os valores brutos encontrados nesta tabela são transformados em notas z e, através de um procedimento estatístico que o autor define, chega-se à selecção de afirmações que se aproximam de intervalos iguais. Como se nota este procedimento de escolha dos itens é muito dispendioso pelo que se chegou a outro procedimento mais leve.

Técnica dos intervalos de aparência igual – É outra escala de Thurstone em que os itens ocupam uma posição na escala que foi previamente determinada por uma operação de classificação realizada por juízes. Propõe-se ser uma escala de resposta intervalar, ou seja, em que a distância entre quaisquer dois números é de magnitude conhecida, com a qual se pode utilizar estatística paramétrica.

Começa-se por um número elevado de afirmações ou itens como no método anterior. Solicita-se a um número elevado de juizes (entre 50 e 300) segundo Miller (1970), que classifiquem de modo independente cada uma das afirmações em 11 grupos que variam do mais favorável ao menos favorável passando por um ponto neutro. Com base na classificação dos juizes encontra-se uma média para cada afirmação. As afirmações com um desvio padrão muito elevado, e que não tenham uma distribuição normal, são eliminadas porque significa que há grande discordância entre os juizes. Agora cada item possui um valor médio que deriva da classificação dada pelos juizes. Com base nestes valores os itens são ordenados dando origem a um coeficiente moderador na nota do indivíduo. Um item tem índice, por exemplo, de 0,2, enquanto outro terá de 0,6 ou 1,3.

Método de intervalos sucessivos – O procedimento é idêntico ao anterior mas acrescenta-se a identificação das distâncias que cada juiz atribui a cada item de modo a identificar, tal como na técnica de comparações emparelhadas, uma escala intervalar.

A construção de escalas de Thurstone é muito dispendiosa, dado a validação da escala exigir, ou o trabalho de juizes que nem sempre estão disponíveis, ou que se comece com um número de itens muito elevado.

Escala de Gutman

O escalograma de análise de Gutman parte da assunção que (a) um conjunto de itens pode ser ordenado ao longo de um contínuo de dificuldade ou magnitude e (b) este conjunto de itens mede uma variável dimensional (Smith, 1975). É uma técnica de classificação hierárquica (Bowling, 1998) e, portanto os itens têm de possuir a propriedade hierárquica de quem concordar com um item também concorda com os itens que estão na hierarquia mais baixa da escala. Tem sido utilizada em saúde para avaliar o funcionamento físico. Trata-se de uma escala unidimensional que é difícil aplicar a outros domínios do comportamento que, normalmente, não são tão consistentes como o funcionamento físico de modo que possam ser hierarquizados.

Diferencial semântico

Este tipo de escala foca o significado que uma pessoa atribui a uma palavra ou conceito. Refere-se à medição de várias dimensões semânticas, ou a diferentes tipos de significado que se reflecte no adjectivo adoptado como descritor. Uma das amplitudes mais utilizadas é um descritor de sete posições em que a média é neutra e vale “0” e para cada um dos extremos cresce até +3 ou -3. Pede-se ao respondente que assinale em escalas que em cada extremo mostram adjectivos bipolares, por exemplo “bom-mau”, “lento-rápido”, “quente-frio”, “feliz-triste”, “fraco-forte”, “activo-passivo”. A maioria dos adjectivos pode classificar-se em três categorias tais como avaliação (p. ex., bom-mau), potência (p. ex., forte-fraco), e actividade (p. ex., rápido- lento).

Escolha forçada

Numa escala deste tipo pede-se ao respondente que escolha entre duas afirmações que sejam igualmente apelativas. Uma afirmação representa um dos atributos enquanto a outra afirmação representa o oposto, ou não representa o atributo em estudo. Se houver mais de duas alternativas a terceira escolha possível consiste num item de elevada desejabilidade social. A escala de Controlo Interno *versus* Externo de Rotter (1966) é uma escala deste tipo.

Dicotómicas

As escalas de resposta dicotómica são escalas de escolha forçada em que o respondente tem de escolher entre duas respostas claramente antagónicas (ao contrário da anterior em que ambas são plausíveis) do tipo sim não, certo errado, verdadeiro falso, etc.

Propriedades psicométricas

A psicometria desenvolveu-se principalmente a partir dos primeiros esforços para medir a inteligência. Como já foi referido, os testes devem respeitar critérios métricos que são próprios da generalidade das medidas. Entre estas propriedades psicométrica podemos distinguir as que são inerentes ao teste e os que são inerentes a cada item.

Propriedades psicométricas dos testes

Dois grandes tipos de propriedades métricas se têm distinguido nos testes psicológicos: a *fidelidade* e a *validade*. Com efeito, a psicologia desenvolveu técnicas para garantir as propriedades métricas na avaliação de características de personalidade que são únicas, dado debruçarem-se sobre variáveis subjectivas (leia-se variáveis avaliadas por palavras ou baseadas na opinião o respondente). Muita gente considera impossível medir características subjectivas. Para a psicologia se afirmar como ciência teve de demonstrar que isso era possível. Tal foi conseguido na primeira metade do século pelos investigadores com a ajuda de técnicas estatísticas que, entretanto, se foram desenvolvendo, e que evidenciaram que as medidas deste tipo tinham propriedades métricas aceitáveis, produzindo medidas com rigor idêntico às medidas da física.

Para os leigos, a noção de objectividade ou subjectividade da medida está ligada ao modo como a avaliação é feita. Um teste em que o respondente assinala a resposta no papel será uma resposta subjectiva. O mesmo sucede se se tratar de uma entrevista. Se pelo contrário, a pergunta aparecer no ecrã do computador, a resposta for assinalada mediante a pressão de uma tecla, e no final o resultado for impresso pelo computador, já será mais objectiva.

Ora, o essencial para qualquer medição ser precisa é, primeiro, que meça o que se pretende medir e não outro aspecto diferente ou parecido (validade) e, segundo, que se a medição for repetida, nas mesmas condições, com os mesmos respondentes, o resultado encontrado seja idêntico (dentro de um erro aceitável) (fidelidade). Estes aspectos devem ser garantidos por todas as medidas, quer se trate de um registo de pressão sanguínea, de pressão atmosférica, ou outra.

Fidelidade – Fidelidade, de fiel, fiável, significa que as notas obtidas por avaliadores (os mesmos ou outros) quando avaliam as mesmas pessoas em diferentes ocasiões, com os mesmos testes, ou com conjuntos equivalentes de itens, são semelhantes (Anastasi, 1990). Dito de outro modo, ainda segundo a mesma autora, a fidelidade de um teste indica a extensão em que as diferenças individuais na realização do teste são atribuíveis a diferenças “verdadeiras” ou reais, das características avaliadas.

Dado que todos os tipos de fidelidade se interessam pelo grau de consistência ou concordância entre dois conjuntos de notas independentes, eles expressam-se por um coeficiente de correlação.

Os tipos de fidelidade que podem ser identificadas num teste são: teste-reteste, formas alternativas, imediatas ou em tempos diferentes, duas metades, alfa de Cronbach ou Kuder-Richardson, de cotador.

Teste-reteste – Na situação de teste-reteste o teste é passado aos mesmos indivíduos, num momento e de novo passado algum tempo. O tempo de intervalo considerado é uma variável importante por várias razões. Nuns casos devido à aprendizagem, como é o caso de itens dos testes de inteligência, em que os respondentes, em caso de o tempo entre as passagens do teste ser curta, podem aprender a resposta. Noutros casos, porque é esperado que haja alterações na variável, como é o caso, por exemplo, da ansiedade estado, em que as correlações deverão, caso o tempo entre passagens seja elevado, ser moderadas. Ou seja, os valores desejáveis da correlação que expressam a fidelidade do teste devem ser considerados caso a caso. O valor de referência considerado bom é de $r = 0,80$, embora se possam aceitar valores tão baixos como de $r=0,60$.

Formas alternativas – É um procedimento parecido com o anterior. Podem ser imediatas ou em momentos diferentes, e refere-se à passagem de duas formas equivalentes do teste (duas formas que avaliam o mesmo construto, a mesma dimensão). Uma das vantagens deste procedimento reside na eliminação da aprendizagem das respostas que pode ocorrer quando se passa o mesmo teste

duas vezes. Outra das vantagens consiste na possibilidade de passar a forma alternativa em sequência, ou seja, responder aos dois testes de seguida, o que seria impossível na situação teste-reteste.

Duas metades – Trata-se de um procedimento próximo do anterior em que metade do teste é comparado com a outra metade. Os sujeitos respondem ao teste e, depois, o teste é repartido em dois. No caso de um teste ser composto por várias sub-escalas, supõe-se que metade dos itens de cada sub-escala estejam presentes em cada metade. Para evitar a variável cansaço que pode provocar respostas diferentes na parte final do teste, a melhor maneira de proceder é fazer a divisão pelos itens pares e ímpares.

Consistência interna – O coeficiente alfa de Cronbach, que fornece a medida de consistência interna, está matematicamente ligado ao método das duas metades (Murphy & Davidshofer, 1998). Na realidade a consistência interna é encontrada com base na média das intercorrelações entre todos os itens do teste. Uma boa consistência interna deve exceder um alfa de 0,80. São no entanto aceitáveis valores acima de 0,60. Estes baixos valores justificam-se quando as escalas têm um número de itens muito baixo.

Fidelidade do cotador – Consiste na garantia que o teste dá de que há concordância dos resultados quando estes são recolhidos por dois cotadores diferentes na passagem do mesmo teste.

Validade – A validade refere-se ao que o teste mede e a quanto bem o faz. *É a garantia que o teste dá que mede o que se propõe medir.* Não se pode referir à validade de um teste em termos gerais. A validade deve ser determinada com referência ao uso particular para que o teste foi desenvolvido. Os processos de determinação da validade abrangem a determinação da relação entre a realização no teste e outros factos observáveis ligados às características do comportamento em consideração. Estabelecer a validade de um teste é sempre um procedimento que depende do julgamento do investigador perante o valor da evidência que conseguiu produzir. Os juízos para estabelecer a validade de uma medida devem ser baseados em evidência empírica sólida.

Não é o teste, propriamente, que é validado, mas sim as inferências e as conclusões a que se chegam com base na nota do teste. Ou seja, a validade não é uma propriedade do teste mas sim uma função do que a nota do teste mede.

Os métodos de determinação da validade foram definidos em 1954 pela Associação Americana de Psicologia com a intenção de uniformizar a grande

confusão de métodos que até então tinham sido utilizados para a determinar (Murphy & Davidshofer, 1998). As quatro faces da validade, como estes autores lhe chamam, são a validade de conteúdo, a validade de construto, a validade preditiva e a validade concorrente: estas duas últimas agrupam-se na validade de critério. A validade de construto é a validade nobre de qualquer medida. É ela que garante que o teste mede o construto. Foca aspectos mais abstractos, mais vastos e mais persistentes dos construtos do que os outros tipos de validade. A validade de conceito é uma validade de primeira linha que garante que a inspecção do conteúdo do teste por especialistas permite afirmar que o teste avalia o que é suposto. Trata-se de uma validade de pendor mais teórico. Por outro lado, a validade de critério garante, de modo empírico, que o teste mede o que é suposto medir. Ao contrário da validade anterior é mais empírica. A validade de critério expressa os critérios que foram escolhidos para determinar a sua validade e que se relacionam com o teste.

Validade de conteúdo – é estabelecida através da evidência de que os atributos incluídos no teste são uma amostra representativa dos que se pretendem medir. Todos os testes psicológicos são amostras sistemáticas de um domínio particular de comportamento. Uma descrição detalhada do conteúdo do domínio a avaliar fornece a base para avaliar a validade de conteúdo.

O procedimento básico para avaliar a validade de conteúdo consiste nos seguintes passos (Murphy & Davidshofer, 1998): (a) descrever o conteúdo do domínio ou construto; (b) determinar as áreas do conteúdo do domínio que são avaliadas por cada item; (c) comparar a estrutura do teste com a estrutura do conteúdo do domínio.

Como a validade de conteúdo é, basicamente, um julgamento e não um exercício de objectividade, uma das maneiras para identificar o julgamento mais adequado é por análise de vários juizes especialistas no conteúdo do domínio em avaliação. Se, após análise, todos os juizes concordarem que tal item avalia tal conteúdo então, provavelmente, tal pode ser aceite como verdadeiro. Murphy e Davidshofer (1998) referem-se à extensão da concordância dos especialistas como rácio de validade de conteúdo

Validade de construto – Os investigadores debruçam-se frequentemente sobre atributos abstractos, sobre atributos que, em sentido literal não existem, que não têm existência física. Não se trata de um problema exclusivo da psicologia. Frequentemente os físicos medem propriedades da matéria que não são observáveis, como é o caso da massa que é uma propriedade hipotética dos objectos, ou da gravidade. Os construtos são fundamentais em ciência: é com base neles que se constroem as leis científicas.

Os atributos que a psicologia se propõe medir – inteligência, motivação, felicidade, bem-estar, etc. – são ideias construídas pelos cientistas – construtos – para sumariar grupos de fenómenos. Se uma pessoa diz a verdade em situações variadas pode ser rotulada de honesta. Honestidade é um construto que não pode ser observado directamente. Conclui-se que uma pessoa é honesta pela observação de acções que expressam honestidade.

Sendo a psicologia uma ciência que se debruça sobre construtos, ou seja, sobre dimensões que não são observáveis, o modo de determinar a validade de construto é complexa e decisiva para se considerar uma medida como séria.

Todos os construtos possuem duas propriedades (Murphy e Davidshofer, 1998) (a) são sumários abstractos de um fenómeno que ocorre com regularidade e (b) estão relacionados com, ou ligados a acontecimentos observáveis, concretos. Se uma maçã cai da árvore para o solo, o construto de gravidade é desenvolvido para explicar e prever o comportamento dos corpos que caem. A gravidade em si não é observável. O que se observa é a queda dos corpos. O mesmo se passa com a inteligência. O que se observa é o comportamento da pessoa, não a inteligência.

Em psicologia o processo de definir um construto passa por três fases: (a) identificar os comportamentos relacionados com o construto a medir; (b) identificar outros construtos e definir a sua relação com o construto a medir; (c) identificar comportamentos que estão relacionados com estes outros construtos e, com base na relação entre construtos determinar em que condições o comportamento está relacionado (ou não) com o construto a medir.

Anastasi (1990) apresenta os seguintes métodos de determinação da validade de construto. Correlação com outros testes, análise factorial; consistência interna; efeitos de variáveis experimentais na nota do teste; validade convergente e discriminante. A validade convergente refere-se à extensão em que a correlação do instrumento com instrumentos que medem o mesmo construto é maior do que a correlação com os que medem construtos diferentes. (Herdman, Fox-Rushby, & Badia, 1998). A validade divergente refere-se à extensão em que a correlação do instrumento com instrumentos que medem diferentes construtos é menor do que a correlação com os que medem o mesmo construto. (Herdman, Fox-Rushby, & Badia, 1998).

Validade de critério – A validade relacionada com um critério indica a eficiência de um teste em prever o comportamento de um indivíduo numa dada situação (Anastasi, 1990). A realização num teste é comparada com um critério, ou seja, uma medida objectiva e independente do que o teste se propõe medir. Por exemplo, os testes de inteligência recorrem frequentemente como critério

aos resultados escolares. Identificando as notas dos alunos e correlacionando-as com as notas de um teste de inteligência é esperado que tal correlação seja elevada. Os melhores alunos terão mais inteligência. Se essa comparação for feita no mesmo tempo, ou seja, se as notas que os alunos tiraram agora e a realização actual do teste de inteligência tiverem uma correlação elevada, estamos perante uma validade concorrente ou simultânea. Se a correlação elevada for entre o teste feito agora e as notas escolares do fim do ano lectivo, ou daqui a alguns meses, ou daqui a cinco anos, ou daqui a 10 anos, diz-se que a validade é preditiva, ou seja que o teste de inteligência tem capacidade para prever as notas que o indivíduo vai ter, ou, ainda dito de outro modo, um bom resultado actual no teste de inteligência permite afirmar que há uma probabilidade elevada de o indivíduo ter notas elevadas após tal período de tempo.

Propriedades psicométricas da avaliação comportamental

A avaliação comportamental baseia-se, como se referiu, na análise funcional. Logo a psicometria, que na avaliação tradicional tem por base a correlação, tem pouca importância neste tipo de avaliação. No entanto, é necessário garantir, também na avaliação comportamental que as medidas são fiéis e válidas, que são precisas. A validade deste tipo de medidas suporta-se fundamentalmente na inspecção teórica em vez de, como nas técnicas psicométricas, na inspecção estatística. A evolução da avaliação psicológica de inspiração comportamental alterou-se nos últimos anos e a diferença entre estas técnicas, para além de se ter esbatido, já não passa por esta fronteira. No entanto a natureza funcional da avaliação comportamental mantém-se como característica identificatória. Alguns dos aspectos referidos para a avaliação psicométrica são relativizados na avaliação comportamental.

Validade dos itens

Todos os critérios utilizados para validar os são adequados para validar os itens. Cada item pode ser considerado como um teste (Murphy & Davidshofer, 1998). Todos os índices de validade do item são baseados na relação entre a resposta ao item e a realização critério (Anastasi, 1990).

Índice de dificuldade do item

Este indicador é destinado aos itens dos testes de capacidade, ou seja, que, a) pedem ao respondente que façam o melhor possível, e, b) se cota a resposta como certa ou errada. Para a maioria destes testes a dificuldade do item define-se pela percentagem de pessoas que respondem correctamente. Um item que é respondido por 90% das pessoas é um item fácil enquanto um que é respondido por 10% das pessoas é difícil. Os testes de capacidade são concebidos para avaliar o nível de funcionamento das pessoas. Como tal a escolha dos itens que permita uma cuidadosa diferenciação dos respondentes é fundamental. Um bom item, um item capaz de fornecer mais informação acerca da capacidade do indivíduo, será aquele que é respondido por 50% dos respondentes. No entanto poderá ser escolhido outro índice qualquer. Por exemplo, os testes de capacidade são construídos de modo que os itens surgem por ordem de dificuldade. Assim, o primeiro item será de resposta fácil de modo a que o respondente vá ganhando confiança, e a dificuldade vai crescendo de modo que os últimos itens são respondidos por poucos sujeitos. Então, se se tratar de um teste de capacidade o primeiro item pode ter um índice de dificuldade de 0,90 ou próximo, e o último de 0,10.

Índice de discriminação do item

Como o anterior trata-se de um indicador para testes de resposta máxima. O índice de discriminação de um item revela a capacidade do item discriminar entre os sujeitos que realizam melhor no teste e os que realizam pior. O primeiro passo para encontrar o índice de discriminação é seleccionar os 27% dos sujeitos que tiveram melhor nota no teste e os 27% que tiveram pior nota. Este valor de 27% é o que, no caso dos resultados se distribuírem segundo a curva normal, permite substituir a fórmula estatística que deveria ser utilizada. De qualquer modo, para se encontrar um índice de discriminação aceitável, a comparação pode ser feita entre os 25% a 35% melhores resultados no teste e os 25% e 35% piores resultados no teste (Murphy & Davidshofer, 1998). Imaginemos o item sete de um teste de capacidade qualquer. Para identificar o índice de discriminação do item sete seleccionavam-se os 27% que tinham tido melhor nota no total do teste e os 27% que tinham pior nota. O índice de discriminação do item corresponderia à diferença entre os que tinham respondido correctamente ao item 7 nos 27% superiores e os que tinham respondido correctamente nos 27% inferiores. Quanto maior for este valor melhor capacidade de discriminação tem o item.

Distribuição das respostas

Nos testes de resposta habitual um bom item deve garantir, (a) que há respostas em todas as alternativas de resposta que cada item oferece e (b) que a distribuição de respostas se faz de modo que a hipótese nula (a hipótese da distribuição não ser normal) seja rejeitada.

No primeiro caso se, por exemplo, a resposta exigida for numa escala tipo Likert com sete alternativas de resposta em que num extremo a resposta é “concordo totalmente” e no outro extremo “discordo totalmente” passando por uma resposta intermédia “indeciso”, todas as sete alternativas devem ter sido assinaladas por algum dos respondentes e a maioria das respostas devem estar nas alternativas intermédias.

No segundo caso, a distribuição das respostas deve ser inspeccionada com recurso a um teste estatístico. Os programas estatísticos contêm testes que permitem esta inspeção como é o caso do Kolmogorov-Smirnov.

Correlação do item com a nota total do teste

Quanto mais elevado o valor da correlação do item com a nota total do teste melhor o item discrimina entre os que têm bom resultado no teste e os que têm pior resultado. Uma correlação elevada indica também que o item mede o mesmo construto que a nota total do teste.

Consistência interna do item

Uma boa consistência interna do item implicaria uma de duas coisas: ou o item tem uma elevada correlação com o critério externo utilizado para validar o teste e uma baixa correlação com a nota do teste; ou o item tem uma boa correlação com o resultado do teste a que pertence (Anastasi, 1990). No segundo caso a homogeneidade do teste é maior, enquanto no primeiro caso se garante que os itens que compõem o teste têm maior amplitude de conteúdo e maior validade de critério. Geralmente, caso não seja especificado o contrário, a consistência interna do item representa a correlação do item com a escala a que pertence.

Validade convergente-discriminante do item

É normalmente exibida como um bom indicador de que o item mede o mesmo construto da escala a que pertence e não outro. Imagine-se um teste composto por várias escalas, cada uma avaliando um traço ou característica de personalidade. Uma boa validade convergente-discriminante do item mostrará que a correlação do item com a escala a que pertence é substancialmente mais elevada do que a correlação do item com as escalas a que não pertence.

Itens dos testes de referência a critério (ou de referência a conteúdos)

A resposta aos itens destes testes são avaliadas em termos de certas ou erradas, como ocorre com os testes de capacidade. No entanto, um bom item para estes testes não é o que tem uma correlação elevada com a nota total da escala a que pertence. Um bom item em testes de referência a critério é aquele que discrimina bem um sujeito antes e depois da aprendizagem. Ou seja antes da aprendizagem o indivíduo não respondia ao item e depois da aprendizagem responde correctamente então o item é bom. O item deste teste é sensível à aprendizagem e não às diferenças individuais.

Portanto, se 100 indivíduos antes de se submeterem a uma aprendizagem não respondem correctamente a uma questão e depois da aprendizagem todos respondem correctamente, estaríamos na presença de um item ideal para um teste deste tipo. Tal como nos outros testes um valor de 100%, se é considerado ideal raramente é alcançado.

Técnicas Psicométricas ou Clinimétricas

Já discutimos noutro estudo este aspecto (Pais-Ribeiro, 1998). Com efeito, a intervenção psicológica em saúde e doença visa provocar alterações nos sinais ou nos sintomas do indivíduo, ou nas alterações nos comportamentos. Torna-se importante salientar que nem todas as alterações que são estatisticamente significativas são clinicamente satisfatórias. Imaginemos um grupo de indivíduos que tenham uma pressão sistólica de 210 e que, após intervenção a reduzem para 180. Esta redução é, provavelmente, estatisticamente significativa, mas o valor alcançado não é clinicamente satisfatório. Blanchard e Schwartz

(1988) discutem o problema explicando a importância dessas diferenças. Tal significa que a consideração da significância dos valores varia entre o clínico e o investigador, ou seja, estatisticamente significativa e clinicamente significativa não quer dizer o mesmo.

Evans, Margison e Barkham (1998) explicam que os investigadores tendem a comparar médias entre grupos, recorrendo a procedimentos estatísticos que se baseiam na distribuição normal dos resultados, enquanto os clínicos se interessam pelas mudanças que ocorrem em determinados sujeitos, em que “sucesso” e “falha” são definidos arbitrariamente. Para ultrapassar esta dificuldade na identificação da significância clínica dos resultados apresentam um método descrito por Jacobson et al. Conhecido por mudanças clínicas fiéis e significativas, que deve ser utilizado para avaliar se as mudanças verificadas são significativas e fiéis (de confiança).

As técnicas de avaliação utilizadas pela psicologia têm-se suportado em técnicas estatísticas robustas. Uma boa técnica de avaliação, como já foi referido, deve possuir propriedades psicométricas rigorosas que, frequentemente se reportam a normas. Ora, em clínica privilegia-se o que há de único no sujeito. Como consequência, é frequente as técnicas que satisfazem do ponto de vista psicométrico não serem satisfatórias para as decisões clínicas a tomar. No contexto desta discussão emergiu, em contexto de saúde, a expressão “clínimética”.

Wijsman et al. (1991) definem clínimética como “a ciência de quantificação dos fenómenos clínicos com particular atenção para a validação das variáveis de resultado” (p. 182).

A diferença entre clínimétrica e psicométrica não parece ser grande. Feinstein (1994) discute os dois conceitos considerando que eles se diferenciam em vários aspectos, nomeadamente nos seguintes:

- 1) Os instrumentos psicométricos contêm inúmeros itens, agregados ou não em escalas, que mascaram ou obscurecem sintomas particulares que são significativos do ponto de vista clínico, como é o caso da dispneia ou da dor numa articulação que deve ser o foco da intervenção.
- 2) Em segundo lugar – continua Feinstein – a psicometria suporta-se em procedimentos, ou em conceitos, que não são familiares para os profissionais que trabalham em contexto de cuidados de saúde, tais como, validade, fidelidade, consistência interna, etc.
- 3) Um terceiro problema é que um instrumento sensível num contexto clínico pode não o ser noutro contexto.

- 4) Um quarto problema é que a agregação dos itens em escalas produzem indicadores que são pouco sensíveis às mudanças.
- 5) Finalmente, segundo Feinstein, embora na avaliação baseada em considerandos psicométricos os doentes expressem os seus sentimentos, valores e crenças, o resultado final é tratado por procedimentos matemáticos que devem ser robustos.

Em resultado destas discussões parece que o mais adequado é que os clínicos desenvolvam dados e taxonomias com base na matemática, na intuição, e numa concordância crítica entre os especialistas, considerando também a opinião dos doentes, de modo a produzir indicadores que sejam úteis em contexto de saúde e de doenças.

Outras características recomendadas na avaliação em saúde

Um teste para o ser do ponto de vista métrico, deve possuir características psicométricas e clinimétricas tal como foi referido. Quando aplicados nalguns domínios, nomeadamente em cuidados de saúde, há características que não sendo psicométricas são fundamentais para a tomada de decisão acerca da utilização do teste com pessoas que estão fragilizadas. Estas propriedades ou características são.

Apropriabilidade

Trata-se de uma propriedade dos testes decisiva na utilização com doentes. Não é aceitável por razões éticas violar esta propriedade. Não havendo um consenso internacional para esta definição, Bowling (1998) expõe a seguinte: a consideração que os benefícios esperados na saúde, dos procedimentos utilizados excedem as consequências negativas por uma margem suficientemente grande que justifique a implementação do procedimento, excluindo-se considerações de ordem económica.

Economia (burden)

O SAC (1995) define economia como as exigências em termos de tempo, energia e outras, que são feitas àqueles a quem os testes são administrados. A

avaliação psicológica foi, durante muito tempo uma prática privilegiada pelos psicólogos. A partir da segunda metade do século XX, o desenvolvimento da psicologia clínica enviou a avaliação para o lugar actual, ou seja, o primeiro momento da intervenção psicológica. Desta evolução decorre a importância da economia, que se torna um aspecto especialmente importante em psicologia e saúde.

O *aparatus* da avaliação psicológica é frequentemente difícil de utilizar, de utilização impraticável, e fornece resultados que são difíceis de interpretar e de integrar pela equipa terapêutica nas decisões clínicas. Os instrumentos para utilização rotineira em cuidados de saúde devem ser breves e simples de utilizar dizem Fitzpatrick, Fletcher, Gore, Jones Spiegelhalter e Cox (1992). Estes autores salientam, no entanto, que a brevidade pode significar que se perde informação importante acerca da experiência do avaliado e que a validade destes instrumentos breves tem de ser cuidadosamente verificada.

É possível e adequado utilizar técnicas clássicas de avaliação psicológica – como por exemplo a *Wechsler Adult Intelligence Scale-Revised (WAIS-R)*, – na avaliação da deterioração mental. No entanto, técnicas deste tipo são pouco utilizadas, nomeadamente porque são muito dispendiosas no tempo que é necessário para as passar. Para este fim específico outras técnicas têm sido desenvolvidas como, por exemplo, o *Short Test of Mental Status* (Kokmen, Smith, Petersen, Tangalos, & Ivnik, 1991), que é um teste de rastreio (*screening*) cuja administração demora cinco minutos, e exhibe uma correlação de 0,78, 0,71 e 0,78 (numa amostra de 268 participantes), respectivamente com o QI verbal, QI de realização e QI total da *WAIS*. Ou seja, cerca de dois terços da variância explicada é comum aos dois testes, um que demora mais de uma hora e outro que demora cinco minutos. O mesmo se verifica com o Inventário de Saúde Mental de cinco itens, conhecido por MHI-5, que tem, na adaptação portuguesa (Pais-Ribeiro, 2001), uma correlação de 0,92 (de 0,95 na versão original), com a versão longa de 38 itens. Ou seja, cerca de 90% da variância do teste longo é explicada pela versão curta.

Numa investigação acerca da qualidade de testes com um único item Gardner, Cummings, Dunham, e Pierce, (1998) numa investigação em que se propõe comparar a eficiência métrica da medida de um construto psicológico realizada com um item *versus* uma medida do mesmo construto realizada com vários itens, concluem que um método não é superior ao outro. Concluem que um bom item pode ser mais eficaz do que escalas que agrupam itens menos bons.

No que foi dito parece resultar um paradoxo para a psicologia. Com efeito, uma das máximas na psicometria é que quanto mais itens uma medida de auto-relato tiver acerca de um construto psicológico melhor (Gardner, Cummings, Dunham,

& Pierce, 1998). Sendo imprescindível considerar a economia dos testes, e embora as versões curtas expliquem uma proporção elevada da variância das versões longas, o que se perde em variância explicada, sendo aparentemente pouco, é suficiente para que o teste deixe de ter validade e, por isso, deixe de ser utilizável. Por esta razão, é necessário tomar cuidado quando se decide pela utilização de versões curtas ou reduzidas, no sentido em que, sendo a versão longa uma técnica de avaliação psicológica, a versão curta, na medida em que deixa de ter propriedades psicométricas deixa de o ser.

Todos os testes, como todas as medidas, são afectadas por um erro esperado que deve cair dentro de uma margem curta razoável. Dito de outra maneira, o resultado de qualquer teste é verdadeiro numa margem de erro que oscila entre mais ou menos determinado valor que é explicitado no manual do teste. Este erro inerente à nota está ligado à validade e fidelidade que, por sua vez, dependem, entre outros aspectos do número de itens, do tamanho do teste. Ora quando o teste é reduzido para ser económico compromete-se a validade e a fidelidade que, acima de determinada magnitude do compromisso pode torná-lo inutilizável.

Adequabilidade

Significa que os aspectos a avaliar são adequados para a decisão a tomar. Não se trata de aspectos de validade. Enquanto esta é a garantia que o teste mede o que se pretende medir, a adequabilidade é uma característica que deve ser inspeccionada antes da escolha do instrumento, ou seja, o instrumento escolhido deve garantir que mede aspectos que *são úteis* para a decisão a tomar. Pode acontecer que não haja disponível, nenhum instrumento para avaliar o aspecto de interesse. Devido a isto, e porque inúmeras propriedades são exigidas aos instrumentos de avaliação, o psicólogo pode decidir recorrer a instrumentos que avaliem aspectos próximos do que interessa.

Interpretabilidade

O SAC (1995) define interpretabilidade como o grau em que é possível atribuir um significado qualitativo aos valores quantitativos de um instrumento. Murphy e Davidshofer (1998) descrevem a interpretabilidade como uma característica que os testes devem possuir que garanta que a informação que fornecem acerca do indivíduo é interpretada de modo semelhante por vários utilizadores. Explica

dizendo que testes que sejam específicos acerca do que se propõem medir, têm maiores probabilidades de ser úteis do que aqueles que são vagos.

Conformidade

De *Responsiveness* no original refere-se à capacidade do instrumento para detectar mudanças mínimas. A utilização de medidas em psicologia e saúde, mais do que classificar os indivíduos por referência a uma norma deverá ser sensível a mudanças que ocorram no indivíduo ao longo do tempo, quer se trate de sensibilidade psicométrica *versus* sensibilidade clínicométrica (ver noutro local).

Por definição, conformidade, que também se pode designar por sensibilidade, abrange mudanças mínimas que são consideradas importantes. Considera-se *sensibilidade da razão do sinal* a mudança real que ocorre ao longo do tempo, e *sensibilidade de ruído* as mudanças nas notas que ocorrem ao longo do tempo e que não estão associadas a mudanças do estado SAC (1995).

Aceitabilidade

Refere-se à propriedade que a técnica de avaliação tem que garante que é bem recebida, que é recebida sem desconfiança, pelo respondente

Utilidade

É uma propriedade que garante que a técnica utilizada é útil para a investigação em causa. Por vezes recolhe-se informação supérflua que depois não é utilizada na investigação.

CONSTRUÇÃO DE TESTES

Introdução

O desenvolvimento de técnicas para avaliar características de personalidade utilizou várias abordagens para agrupar os itens que Anastasi (1990) classifica de com base: (a) no conteúdo do item; (b) em critérios empíricos; (c) na análise factorial; (d) nas teorias da personalidade.

Com base no conteúdo dos itens

Recolhe-se informação relativamente ao construto que se pretende avaliar, através da revisão da literatura e na entrevista a especialistas. Uma vez clarificado o construto escrevem-se itens que de modo claro se refiram a esse construto. A relação dos itens com o construto deve basear-se na concordância de juizes especialistas no construto, que devem concordar que os itens o avaliam. Por essa razão Sundberg chama-lhe estratégia de julgamento, teórico-racional. A validade subjacente é, basicamente, uma validade facial (ou validade de fé) e de conteúdo, porque depende do julgamento subjectivo dos juizes e, portanto, não se pode dizer que tal teste possui validade empírica.

Com base em critérios empíricos

Anastasi (1990) denomina assim as escalas que são construídas com base num qualquer critério externo. Loevinger (1957) explicava que o termo critério empírico se refere normalmente à selecção de itens com base na sua correlação com o critério externo. Sundberg (1977) chama a este método de contraste entre grupos. Como este autor explica, um conjunto de itens é passado a dois ou mais grupos de indivíduos que se sabe diferirem nalguma característica (o critério ou construto a avaliar). Um item que diferencie os grupos de modo, estatisticamente significativo,

é candidato, a ser incluído no teste. Com frequência os itens assim escolhidos são difíceis de interpretar teoricamente. Um exemplo bem sucedido desta estratégia é o *Minnesota Multiphasic Personality Inventory*, um dos testes de personalidade mais conhecidos e mais utilizados. Um número elevado de itens foi definido sem nenhum critério teórico. Posteriormente foram passados a uma população normal e a uma população com um diagnóstico claro de psicopatologia. Os bons itens eram aqueles cujas respostas eram estatisticamente diferentes nos dois grupos.

Com base na análise factorial

A análise factorial permite agrupar os itens num conjunto homogéneo. Com base na análise factorial ou na consistência interna identifica-se a relação entre itens, que se exprime por valores de relação estatística elevada, quer na análise factorial (relação de um grupo de itens com um factor) quer na consistência interna. O facto de haver um agrupamento factorial consistente e/ou consistência interna elevada, não garante a existência de uma relação conceptual satisfatória com o critério externo e, frequentemente a fundamentação teórica de tal agrupamento torna-se difícil.

Com base em teorias da personalidade

Uma outra estratégia referida por Anastasi (1990) é a baseada em teorias da personalidade. Neste caso no princípio está uma teoria, para responder à qual se constroem itens ou estratégias de avaliação. Note-se que, ao contrário da estratégia com base no conteúdo dos itens nesta não tem de existir relação evidente entre o item e o construto: deve existir, sim entre o item e a teoria. Um exemplo deste tipo de itens são os que integram os testes projectivos, ou a avaliação clínica Piagetiana.

Cada uma das estratégias referidas acima é susceptível de conduzir a testes diferentes. Quem desejar construir um teste terá vantagens em recorrer, de modo sequencial e organizado, a várias das estratégias descritas.

Estudos em culturas e línguas diferentes

O rápido desenvolvimento e a facilidade das comunicações, o fenómeno da globalização, tornou moda, nos anos 90, os estudos transculturais. Este tipo de

estudos propõe-se estudar as mesmas variáveis, os mesmos construtos, com as mesmas técnicas de avaliação, em povos e culturas diferentes com línguas diferentes. Parece, por vezes, aceitar-se que as palavras, as frases, quando literalmente traduzidas de uma língua para outra constituem, exactamente, o mesmo tipo de estímulo susceptível de produzir em pessoas com as mesmas características, as mesmas respostas.

No auge da utilização dos testes de inteligência, a discussão acerca do impacto da compreensão da pergunta no resultado do teste levou à construção de testes que foram intitulados de *cultura equilibrada* ou *livres de cultura*, assumindo-se que a cultura do avaliado jogava um papel fundamental na realização do teste. Os mesmos pressupostos são actualmente verdadeiros. A investigação em saúde salienta a importância de “os investigadores serem prudentes na utilização de um instrumento num contexto cultural diferente daquele para que foi concebido. Para além da validade de conteúdo e facial, outros problemas incluem a validade da tradução e a importância relativa dos itens do instrumento” (Fletcher, Gore, Jones, Fitzpatrick, Spiegelhalter, & Cox, 1992, pp. 145-146).

A construção de técnicas de avaliação transculturais

Anastasi (1990) diferencia a avaliação *dentro da, ou intra cultura* (crosscultural), da avaliação *entre culturas* (transcultural). Dentro de uma mesma cultura há grandes variações entre grupos com base na educação ou, noutras dimensões, susceptível de influenciar a realização nos testes. Os aspectos ligados à literacia têm sido dos mais discutidos e considerados. Refere Anastasi que se podem identificar três abordagens no desenvolvimento de testes destinados a pessoas de culturas diversas.

Uma envolve a escolha de itens que sejam comuns às diversas culturas e a validação do resultado do teste perante critérios locais em culturas diferentes.

Outra abordagem consiste em desenvolver um teste numa cultura e passá-lo a indivíduos de outras culturas para identificar se o teste mede os mesmos aspectos.

Uma terceira abordagem consiste em desenvolver testes para cada cultura, e cada teste seria aplicado somente na cultura para que foi desenvolvido. Anastasi afirma que quando a língua difere, a simples tradução do teste raramente é suficiente, exigindo adaptação e revisão do conteúdo do teste. Quanto ao procedimento a utilizar para desenvolver testes a aplicar em diferentes culturas com diferentes línguas sugere dois passos: (a) o primeiro consiste na selecção de

um conjunto de itens básicos que sejam igualmente apropriados para as duas culturas; (b) estes itens são administrados aos participantes das duas línguas; (c) a análise dos itens deverá permitir a eliminação dos itens que têm significados diferentes nas diferentes línguas

Por exemplo, Orley (1994) explica os seguintes passos na construção de um instrumento para avaliar a qualidade de vida em diferentes países no âmbito de um estudo da OMS: primeiro, com base na revisão da literatura identificaram-se um conjunto de domínios do construto “qualidade de vida” que incluíam, cada um, inúmeras facetas. Cada um dos centros participantes (um de cada país) avaliou e aceitou a definição da faceta assim como a tradução proposta, pronunciou-se sobre cada uma das facetas e redigiu itens – que avaliassem aspectos objectivos e subjectivos – que se relacionassem com cada faceta. De seguida era-lhes pedido que atribuíssem a cada faceta um valor que expressasse quanto importante ele era para a qualidade de vida. Estas questões eram analisadas por equipas de especialistas de cada um dos países participantes que acrescentavam outros itens que considerassem oportunos. Estes itens eram traduzidos para inglês e enviados para Genebra para uma análise mais geral. Desta análise resultou a retirada de alguns itens ou a aglutinação de outros. Chegou-se então a uma pool de 300 itens incluindo os que foram produzidos para línguas ou culturas específicas. Traduziram-se as questões para cada uma das 10 línguas participantes, que devolvidos a cada centro e passada a uma amostra de 300 participantes que fossem representativos da população geral, permitiram a selecção dos itens com base nas propriedades psicométricas. No final resultarão 10 versões do questionário, com, provavelmente, questões específicas para cada centro participante.

Neste exemplo apresentado por Orley, a universalidade do questionário não dependerá da tradução cuidadosa de um conjunto de questões desenvolvidas em inglês, mas sim (a) de um conjunto de facetas comuns, (b) de uma metodologia comum de desenvolvimento e selecção dos itens e, (c) da existência de um conjunto de itens adequado para cada língua. O que é comum no questionário da OMS são os domínios e as facetas. As questões são adaptadas (se necessário) à especificidade de cada cultura de modo a garantir que cada faceta é adequadamente contemplada. Este é o procedimento considerado adequado para o desenvolvimento de técnicas de avaliação comparáveis em culturas diferentes.

Como se costuma fazer a adaptação de instrumentos

Embora estes princípios sejam aceites e reconhecidos, é frequente deparar-se com estratégias muito mais leves de adopção de técnicas de avaliação. A mais frequente consiste na tradução simples de técnicas disponíveis em língua inglesa.

Parece claro que testes escritos em inglês e traduzidos segundo regras lexicais não mantêm sentido idêntico. De Raad (1998) é da mesma opinião e afirma: “as línguas diferem muito umas das outras: traços descritos numa língua não têm necessariamente tradução correspondente noutra língua” (p. 120). Este autor, fazendo uma revisão crítica sobre um conhecido teste de personalidade com versões em vários idiomas confirmava a variedade de estruturas que esse teste apresentava nos vários idiomas. A nossa experiência no estudo de testes originariamente escritos em inglês aponta no mesmo sentido (Pais-Ribeiro, 1994a,b; 1995a).

Um processo de adaptação adequado consiste num procedimento muito semelhante ao de construção. Como afirma Bradley (1994) uma nova tradução necessita ser submetida aos mesmos procedimentos psicométricos, que, é suposto, a versão original ter sofrido.

Como um item traduzido com respeito pelas regras lexicais tem probabilidade reduzida de ter o mesmo significado na língua original e na traduzida, por vezes, para manter o mesmo sentido, a solução é ajustar o texto. Recorrer a itens traduzidos unicamente com respeito pelas regras lexicais e sem respeitar a validade de conteúdo pode ser mais ou menos grave consoante a validade do teste, ou dos itens, é garantida com base na validade de conteúdo (desde que a tradução respeite a mesma validade de conteúdo) ou da validade empírica (desde que o estudo empírico seja feito para a versão traduzida).

Hipótese psicolexical

Segundo De Raad (1998) a hipótese lexical defende que todas as diferenças individuais com alguma importância têm expressão na linguagem. Esta leva à hipótese psicolexical da personalidade que assume que há um conjunto de descritores universais dos traços de personalidade que seriam comuns a todos os humanos. Os traços, entendidos à imagem de peças de anatomia que tipificavam a espécie.

Guyatt (1993) apresenta uma crítica à filosofia da tradução de questionários de avaliação da qualidade de vida, salientando as vantagens e limitações de tal procedimento, e sugere uma abordagem no desenvolvimento de questionários que se aproxima da que foi apresentada acima como exemplo. Assume que a simples tradução com respeito rigoroso pela versão original dará origem a questionários diferentes.

Basicamente sugere que o questionário na língua original seja utilizado como guia mas não mais. Sugere três estratégias de desenvolvimento de questionários que designa por estratégia Rolls-Royce, estratégia Volkswagen e estratégia Volvo, consoante os recursos disponíveis. A primeira é muito semelhante ao processo de construção do questionário. A segunda utilizaria uma versão simplesmente traduzida. A terceira, provavelmente a mais plausível, começaria por assumir um conjunto de assunções: (a) os domínios do instrumento original também são domínios importantes na cultura e língua para que se pretende traduzir; (b) a estratégia para formular as questões também funciona bem nesta cultura e língua; (c) a estratégia para responder também é adequada. A terceira estratégia não se limitaria a uma tradução rigorosa com base em dimensões lexicais.

Abordagem absolutista versus abordagem universalista

Herdman, Fox-Rushby e Badia (1998) referem que a abordagem absolutista da adaptação de técnicas de avaliação a outras culturas tem tendência a prevalecer. A abordagem absolutista assume que as diferenças de conteúdo, e a organização de conceitos entre culturas são negligenciáveis, e que, se se der uma atenção cuidadosa aos elementos linguísticos, um questionário desenvolvido para uso numa cultura será aceitável noutra cultura.

Uma abordagem universalista, pelo contrário, não assume que os construtos são iguais em todas as culturas e, por consequência, defende a necessidade de estabelecer se o conceito existe e, se existir, se é interpretado de modo semelhante nas duas culturas e, caso o seja, em que grau a interpretação é semelhante. A abordagem universalista, dizem Herdman, Fox-Rushby e Badia (1998), propõe-se salientar ou elicitar aqueles aspectos do conceito que são genuinamente universais ao longo de diferentes culturas e utilizar somente os aspectos que o são para medir esses conceitos em diferentes culturas.

Aspectos técnicos na tradução

Um procedimento, frequentemente defendido por autores americanos para garantir que a versão traduzida é equivalente à original é proposto por Brislin, Lonner e Thordike (1973). Consiste em três fases. (a) tradução e retro-tradução, independentes, por vários juizes que dominem a língua, o conteúdo e a amostra do estudo; (b) utilização de classificadores que examinem as versões, original, a tradução, e a retroversão para analisar os erros, e (c) um pré-teste da versão original

e da versão traduzida com sujeitos bilingues que não estejam familiarizados com o instrumento de modo a garantir a equivalência entre as duas versões. Este método tem sido utilizado nos Estados Unidos da América, país de emigrantes, onde é frequente a existência de comunidades que, para além da língua do novo país (o inglês dos Estados Unidos da América) dominam a língua do seu país de origem, podendo, assim, preencher o mesmo questionário em duas línguas e, então, comparar as respostas. Ora, como se imagina a disponibilidade de populações bilingues não é idêntica em Portugal, pelo que este procedimento não é exequível.

Bradley (1994) refere que é essencial que o tradutor de um questionário seja fluente nas línguas em jogo, embora tal, só por si, não chegue. É necessário, igualmente, que o tradutor compreenda o objectivo do questionário e as intenções subjacentes à concepção de cada item. A autora discute estratégias a adoptar na tradução de um instrumento, recomendando a abordagem por um comité, que deveria incluir, na sua versão mais económica, um tradutor, um especialista em linguística e um especialista em psicometria. A tradução, retroversão e retradução seria o primeiro passo destinado a verificar se a versão atingida após tradução seria idêntica à original. Tal deve ser aplicado, segundo Bradley (1994), tanto no caso da tradução ser realizada por uma equipa de tradutores, na sua versão mais dispendiosa, quer por um único tradutor. A retroversão deve ser realizada por tradutores que não tenham conhecimento da versão original do instrumento. Se após comparação dos itens da retroversão se identificarem discrepâncias, o processo repetir-se-á tantas vezes quanto as necessárias até se atingir uma retroversão idêntica à versão a partir da qual se iniciou a tradução.

Todd e Bradley (1994) referem a Técnica Delphi que é basicamente uma sondagem realizada com um painel de especialistas. Primeiro, os membros do painel, (que não sabem quem são os restantes membros do painel) individualmente, e por escrito, analisam e dão informações sobre o teste a traduzir que devolvem ao grupo para nova análise, com a intenção de alcançar consenso. Um aspecto importante da Técnica Delphi é que o painel é anónimo mesmo para os outros membros que o constituem.

Aspectos a considerar quando se opta pela tradução

O SAC (1995) recomenda que a adaptação transcultural de um instrumento deve tomar em consideração três aspectos distintos: (a) equivalência linguística, (b) equivalência conceptual e (c) equivalência psicométrica. Outros autores fazem propostas mais detalhadas.

Hays, Anderson e Revicki (1993) referem quatro equivalências que devem ser considerados na tradução de modo a garantir que as medidas são idênticas: Equivalência da tradução do item; equivalência operacional; equivalência da escala; e equivalência métrica.

- 1) *Equivalência da tradução do item* – que proporciona a equivalência funcional, responde a uma questão importante que é a de saber se a tradução noutra língua mantém o mesmo significado que versão original, que intitula de equivalência funcional. O critério mais rigoroso é o já referido, proposto por Brislin, Lonner e Thordike (1973).
- 2) *Equivalência operacional do item* – pretende verificar se as assumpções referentes à administração do teste na língua original se podem aplicar, igualmente, no idioma em que se pretende utilizar a versão traduzida. Aspectos que devem ser particularmente cuidados são o tempo, ou os custos (não só temporais), com a passagem do teste, a exposição de aspectos pessoais íntimos que o teste pode exigir e que não são sentidos da mesma maneira em diferentes culturas, e a literacia.
- 3) *Equivalência da escala* – garante que o teste mede as mesmas dimensões nas duas culturas. As dimensões avaliadas com técnicas deste tipo estão embebidas numa matriz cultural e num sistema de crenças que se desenvolveu nessa cultura. Os itens deste tipo de instrumentos podem ter o mesmo significado mas ser valorizados e conceptualizados de modo diferente em populações e culturas diferentes.

Um dos modos de avaliar a equivalência das escalas é através da inspecção das propriedades psicométricas, garantindo que elas são semelhantes nas duas versões. Outro método é o de verificar, através da análise factorial, se as soluções factoriais são idênticas e se a variância de cada factor é semelhante.

- 4) *Equivalência métrica da escala* – refere-se a se a medida ordena os indivíduos num contínuo, de modo semelhante em ambas as culturas. Uma abordagem adequada seria escalar os itens relativamente às medidas de tendência central da cultura estudada, ou classificar os itens de acordo com uma escala externa que servisse de referencial, e compará-las nas duas culturas, para ver se a distribuição seria equivalente.

Outra proposta para implementar a equivalência entre instrumentos em culturas diferentes é proposto por Herdman, Fox-Rushby e Badia (1998), proposta que intitulam de Abordagem Universalista. Esta, como já se referiu, salienta que há a possibilidade de haver variações na natureza de construtos multidimensionais em diferentes culturas, pelo que é importante investigar alguns aspectos,

nomeadamente: (1) que domínios são importantes para o construto na cultura em jogo, e quais são as relações entre eles (equivalência conceptual), 2) examinar de modo crítico os itens utilizados para avaliar esses domínios, e se a relevância desses itens é idêntica nas duas culturas (equivalência do item), (3) assegurar que a tradução respeita a equivalência semântica dos itens (equivalência semântica), (4) assegurar que os métodos de medição utilizados são adequados para a cultura em questão (equivalência operacional), (5) inspeccionar as propriedades psicométricas do instrumento (equivalência de medida) e, (6) examinar o resultado do processo em termos de comportamento do instrumento (equivalência funcional).

- 1) Equivalência conceptual – esta equivalência é alcançada quando o questionário tem a mesma relação com o construto subjacente em ambas as culturas, quer em termos dos domínios incluídos, quer da ênfase colocada nos diferentes domínios. Este estágio do processo fornece uma base que permite julgar a legitimidade para adaptar o questionário e, simultaneamente, fornece as bases para a interpretação dos resultados. A definição de equivalência conceptual implica que não é possível alcançar ou examinar a equivalência conceptual com base numa simples tradução.

Maneiras de investigar a equivalência conceptual envolvem, por exemplo a pesquisa da literatura local, registos etnográficos, investigação. Pode recorrer-se, também, à consulta ampla de especialistas da cultura alvo, tais como antropólogos, sociólogos, linguistas, etc. Uma terceira abordagem implica o envolvimento da população que se pronuncia acerca do que pensa sobre o construto, ou seja uma perspectiva mais etnográfica.

- 2) Equivalência do item – Do mesmo modo que a relevância dos domínios do construto podem variar através de diferentes culturas, assim também pode variar a validade dos itens enquanto medidas de dado domínio. Existe equivalência do item quando estima o mesmo parâmetro do domínio em avaliação e quando é igualmente relevante e aceitável em ambas as culturas.
- 3) Equivalência semântica – Esta equivalência diz respeito à variação de significado existente entre as línguas, e expressa-se pela consecução de efeitos semelhantes nos respondentes nas diversas línguas. Estes autores referem uma detalhada classificação dos significados, nomeadamente: significado de referência, conotativo, de estilo, afectivo, de reflexo, coloquial, e temático. Não têm todos a mesma relevância e uns são mais importantes do que outros consoante os construtos, ou o tipo de avaliação. Para facilitar a equivalência semântica é importante que os termos principais utilizados na versão original tenham sido cuidadosamente descritas como é o caso, por exemplo, do questionário de avaliação da qualidade de vida da OMS referido acima. A adaptação de um questionário não é, só, um problema de técnica linguista. Os

tradutores deverão estar atentos à população a que se dirige e recorrer a um registo ou dialecto apropriado. A tradução deve ser verificada e depois testada na população-alvo para parafrasear o item traduzido para se poder entender se o significado atribuído é o pretendido.

- 4) Equivalência operacional – Esta equivalência existe quando o formato, instruções, modo de administração e métodos de medição não afectam os resultados. Aspectos como a forma de resposta, que é exigida pelos questionários, pode não ser igualmente acessível em todas as culturas. O preenchimento pelo próprio, por exemplo, é susceptível de ser influenciado pelo nível de escolaridade. O tipo de escala de resposta é, igualmente, menos conhecido numas culturas que noutras, e em níveis de literacia diferentes. O tipo de resposta em “sempre” ou “nunca”, “concordo totalmente” ou “discordo totalmente”, “totalmente de acordo” ou “totalmente em desacordo”, são expressões que não têm de ser entendidas do mesmo modo em diferentes culturas. O hábito de responder a questionários deste tipo também difere em diferentes culturas. Em Portugal, o nível de literacia baixo influencia a compreensão das perguntas. Por outro lado a falta de hábito, quer de responder a questionários auto-preenchidos, quer de responder ao formato de resposta naquele tipo de escalas é susceptível de influenciar as respostas.
- 5) Equivalência de medida – Refere-se ao grau ou extensão em que as propriedades psicométricas das versões do mesmo instrumento em diferentes línguas são semelhantes. Nomeadamente a consistência interna deve ser idêntica. Através da análise factorial deve-se verificar se a distribuição de itens pelos factores é a mesma na versão traduzida, e se a variância explicada pela solução factorial é idêntica à original.
- 6) Equivalência funcional – Pode definir-se como a extensão em que o instrumento funcione de modo equivalente nas duas culturas. Se as formas de equivalência descritas anteriormente forem conseguidas então é grande a probabilidade desta equivalência ser, também, elevada. Demonstra-se esta equivalência, primeiro analisando como é que a dimensão, o traço, em avaliação, é definido ou conceptualizado na cultura alvo, em segundo lugar quanto bem o instrumento reflecte essa característica ou traço e, finalmente, as semelhanças dos resultados obtidos nas diferentes culturas.

Os cuidados exigidos por uma adaptação cuidadosa que garanta que as versões são equivalentes são, como se vê, dispendiosos. Soluções menos exigentes têm sido tentadas. Por exemplo, Mathias, Fifer e Patrick (1994) apresentam um questionário traduzido de forma “minimalista” do inglês dos Estados Unidos da América para oito línguas diferentes (uma delas o inglês do Reino Unido), em que uma das preocupações foi simplificar o processo técnico referido acima,

mas garantindo níveis de qualidade mínimos. Os resultados apresentados pelos autores são, segundo eles próprios, satisfatórios, embora se deduza pelo implícito mais do que pelo explícito, que não completamente satisfatório. Inspeccionando os resultados psicométricos das diversas versões que os autores apresentam encontram-se variações substanciais, mostrando haver diferenças sensíveis entre grupos no emparelhamento escalas-culturas.

Como se constrói uma escala

Há inúmeros procedimentos que podem ser adoptados para desenvolver um teste. De entre eles propomos os seguintes passos:

1. O primeiro passo consiste na *definição da variável*. A escolha deve ser feita numa *variável natural* que, por definição é uma variável que (a) todas as pessoas têm e, (b) ninguém tem mais do que um nível dessa variável num dado momento. A altura, a inteligência, são variáveis naturais na medida em que todas as pessoas têm, por exemplo, um nível de inteligência, e só têm um nível nunca dois níveis (Reckase, 1990).

Conceber uma variável inclui, diz Dawis (1987), clarificar as distinções – o que é que a variável é e o que não é –, as dependências – como é que a variável é função de conceitos básicos ou de regras previamente definidos –, e relações – como é que a variável se relaciona com outras variáveis –. Como a variável será medida dependerá de considerações várias, nomeadamente como é que ela será melhor expressa e como é que será o desenho de investigação, quem serão os respondentes, em que contextos ela será aplicada.

2. O passo seguinte consiste em *relacionar a variável com um construto* que lhe dá origem. Este construto deve ter uma paternidade teórica e esta clarificação terá influência decisiva na escolha dos itens e no modo de resposta. Por exemplo, a variável “auto conceito” existe em quase todas as teorias psicológicas e cada uma a explica de modo substancialmente diferente. Do construto emergem aspectos, dimensões, componentes, características, que o esclarecem e iluminam a variável. Desta forma o construto começa a ser clarificado permitindo o desenvolvimento de indicadores observáveis desse construto/variável.

Deve ainda identificar-se quais as medidas ou realizações existentes, testes ou outros critérios com os quais o actual construto se relaciona e,

simultaneamente clarificar com quais medidas ou critérios ela se contrasta. Esta identificação clarifica o construto e facilita, nomeadamente, a exploração psicométrica da validade estrutural como sejam a validade convergente e discriminante ou divergente.

3. Antes de se iniciar a construção de uma escala deve-se pesquisar a literatura para saber se já existe uma escala para avaliar a variável em jogo. Como regra deve-se recorrer a uma escala já existente se ela possuir as características apropriadas, nomeadamente, validade, fidelidade, utilidade e apropriabilidade. Em Portugal é recente o costume de publicar estudos que se debrucem sobre as próprias escalas de modo a que outros investigadores possam recorrer a elas. Se a escala não existir em português ela existirá, provavelmente, em inglês. No entanto a tradução simples, como foi referido acima, dá origem a escalas com propriedades diferentes da original pelo que a adaptação da escala à língua portuguesa constituirá um processo muito semelhante ao de construção. Quando se recorre à tradução, o elemento essencial consiste na *tradução e clarificação do construto* subjacente à escala (e não na dos itens). Esse será o elemento de referência da escala ao qual todos os itens devem estar indexados. Para avaliar o construto pode, nomeadamente, ter de se alterar o item tal como ele fica depois de uma tradução correcta do ponto de vista lexical.
4. Numa fase seguinte o investigador terá de desenvolver itens (indicadores) que meçam o construto. Tal pode ser feito de vários modos sendo o mais fácil pesquisar de entre itens existentes noutras escalas desde que satisfaçam dois critérios (Comrey, 1988), (a) terem sido desenvolvidos para avaliar o mesmo construto, e, (b) relacionarem-se estatisticamente uns com os outros com uma magnitude suficiente de modo a poderem agregar-se na escala pretendida.

Ercikan (1998) sugere que a geração de itens (ou tradução) se respeitem as seguintes regras: (a) use frases simples e curtas (menos de 16 palavras); (b) Utilize a voz activa em vez da voz passiva; (c) repita os substantivos em vez de utilizar pronomes; (d) evite metáforas; evite formas verbais complexas; (e) evite advérbios e preposições; (f) evite forma possessiva se possível; (g) utilize termos específicos em vez de termos genéricos; (h) evite palavras vagas; (i) evite frases com dois verbos que indiquem acções diferentes.

Deve decidir-se se a escala é unidimensional ou se ela é composta por várias sub-escalas. Depois dos itens escolhidos pode ser utilizado um processo de retrotradução (*back translation*), como lhe chama Dawis (1987) e que consiste em pedir a especialistas que não estiveram envolvidos na geração dos itens que identifiquem a que construtos pertencem tais itens.

5. Deve decidir-se, no seguimento do tipo de agrupamento dos itens por sub-escalas, qual a magnitude da relação entre os itens. Esta expressa-se pela consistência interna. Como explica Rotter (1990) “podemos ter alfas muito elevados se fizermos a mesma pergunta inúmeras vezes, mas as limitações preditivas de tal teste são óbvias (...) e um teste com um alfa de 0,90 é suspeito na medida em que pode estar a medir o estilo de resposta em vez de fornecer uma amostra representativa do comportamento e de situações que se desejam prever, ou pode simplesmente consistir em itens redundantes” (p. 491). Relativamente à Escala I-E de que é autor, Rotter assume que, pelo contrário, o que lhe interessa nessa escala é listar uma amostra de itens representativa de muitas situações e não um grupo de itens que representam a mesma situação e, por isso, assume que a correlação entre tais itens deve ser positiva mas baixa. Convém não esquecer que a escala I-E se inscreve na teoria da aprendizagem social em que a assunção básica é que a unidade de investigação no estudo da personalidade é a interacção do indivíduo com o seu meio ambiente significativo, assumindo-se que o comportamento em diferentes situações varia. Ora, esta questão está dependente da teoria que sustenta o construto, e que deve, de algum modo ser clarificada.

Como se referiu quando se falou da consistência interna do item, esta pode significar a) uma elevada correlação entre o item e o critério externo, ou, b) uma elevada correlação com a nota da escala a que pertence. Se a primeira hipótese for a definida à priori, a correlação do item com a nota da escala a que pertence pode ser (ou não) baixa, e então, a homogeneidade da escala fica diminuída o que não significa que a qualidade do item ou a qualidade da escala seja igualmente diminuída. Tudo depende do critério definido à priori.

Por outro lado deve-se clarificar a estabilidade esperada da escala. Por exemplo, se se tratar de um traço (estável, perene), é esperado que o teste reteste com um mês de intervalo mostre uma correlação mais elevada do que se tratar de um atributo menos estável, nomeadamente um estado.

6. Num primeiro momento a escolha e inclusão dos itens é feita somente com base na teoria. A selecção final dos itens deve, no entanto ser feita com base em critérios empíricos. Loevinger (1957) chama ao primeiro critério (com base na teoria) a *componente substantiva da validade* enquanto ao segundo chama *componente estrutural da validade* que, segundo Jackson (1970) se expressa pela homogeneidade dos itens.
7. O formato da escala também é objecto de decisão consoante a população a que se dirige (crianças, analfabetos, universitários, etc.) e consoante o contexto em que se aplica (laboratório, acamados, estudantes, etc.). Uma escala verbal é formada por duas partes: uma parte que constitui o estímulo,

normalmente uma afirmação, questão, uma frase, ou mesmo uma palavra, que é o estímulo que desencadeia a resposta, e uma parte que constitui a resposta. A parte estímulo deve ser simples e incluir um único componente. A parte resposta deve ser adaptada aos respondentes: pode constituir uma escolha forçada tipo “sim” ou “não”, “concordo” ou “discordo” pode ser mais ampla variando entre “totalmente verdade”, “em parte verdade”, “indeciso”, “em parte falso”, ou “totalmente falso”, com a amplitude de escolhas variando entre duas e cinco ou mais. A escala de resposta pode ser numérica (de 1 ou zero a cinco ou mais), podem os números estar expressos de mais três a menos três passando por uma posição intermédia zero, etc. Pode ainda exigir uma resposta que está expressa em figuras tais como faces da mais satisfeita ou sorridente à mais triste, ou pode ter uma forma de escada. Pode ainda, tal como nas escalas análogas visuais ser representadas numa linha que varia entre dois extremos em que um expressa o máximo e o oposto o mínimo ou nada. A escolha do formato de resposta dependerá, nomeadamente, da população a quem o questionário se dirige, e o critério é escolher o formato de resposta mais simples ou mais adequado à população a que se destina.

8. O número de itens que uma escala deve ter é variável. Um número recomendado é de 20 itens por escala. 10 itens por escala, no entanto, podem chegar. Um teste com várias dimensões, idealmente deverá incluir 20 itens por dimensão: para que depois de seleccionados os itens que têm as propriedades psicométricas apropriadas chegue ao número de 20, devem existir à partida cerca do dobro deste número (Reckase, 1990).

Por vezes em contexto de saúde os testes têm de ser reduzidos e conter poucos itens. Boyle e Harrison (1981) propõem um critério, a saber: a solução factorial só é útil do ponto de vista teórico se inclui, para cada factor, pelo menos três itens que possuam uma carga que respeite a relação entre o valor da carga factorial mais elevada com a menor, segundo a seguinte fórmula: $C_i \geq H_i/2$, em que H é o valor do item com a carga factorial mais elevada e C o valor mínimo de carga factorial considerado aceitável para um item que define o factor em jogo.

9. Escrever o protocolo de passagem do teste. Como começar, que esclarecimentos devem ser dados ou não, como fazer para passar e recolher o questionário, e outras explicações que forneçam informação mínima aos especialistas que vão passar o questionário de modo a que todos utilizem procedimentos idênticos.
10. Formatar o questionário que deve incluir uma folha de rosto ou um texto introdutório em que se explica o que se pretende com o questionário. Esta

folha de apresentação é tanto mais importante se o questionário for de auto-preenchimento. Em contextos de cuidados de saúde, as comissões de ética solicitam por vezes que nesta folha de rosto exista um espaço em que o respondente assina um compromisso de que aceita participar na investigação e que está a par do que lhe está a ser pedido e das implicações que tem tal participação (consentimento informado).

11. O passo seguinte consiste na verificação se a população alvo compreende os itens do modo que os especialistas esperam. Procede-se para isso a uma entrevista em que o respondente, um membro (ou vários) da população alvo, responde ao questionário de forma aberta, conversando com o entrevistador de modo a identificar se está a entender as questões do modo suposto, se as compreende, e se é acessível o modo de responder.
12. Com base nos procedimentos anteriores produzir a versão de estudo definitiva. Dá-se início, então, ao estudo piloto em que a versão definitiva do questionário deve ser passada a uma amostra da população alvo. O estudo piloto deve registar minuciosamente todos os acontecimentos que ocorram desde o primeiro contacto com a população alvo, passando por eventuais incidentes de preenchimento, até a questões levantadas com a análise dos dados de modo a que se procedam às rectificações finais.
13. Com base nos resultados passa-se à exploração das propriedades métricas do questionário. O primeiro passo consiste em inspeccionar a distribuição das respostas pelos itens, nomeadamente (a) se todas as alternativas de resposta são utilizadas pelos respondentes, (b) se a distribuição das respostas segue um padrão estatístico normal (por exemplo, utilizar o teste K-S do SPSS). Esta inspecção pode permitir, eliminar (no caso de a distribuição não ser normal) a possibilidade de se adoptar pelo uso de uma estatística paramétrica na análise dos resultados do questionário, (c) verificar através da análise factorial (atenção ao padrão de distribuição das respostas dado que a análise factorial é um procedimento paramétrico) se os itens se distribuem pelos factores (dimensões) que foram definidos à priori, e que são os supostos, ou se é unidimensional, (d) Inspeccionar se a consistência interna da escala ou das sub-escalas é satisfatória, (e) inspeccionar a validade convergente – discriminante dos itens.
14. Finalmente produzir o relatório do teste com os procedimentos e valores psicométricos, clinimétricos, normativos e outros, que foram encontrados, e publicar num jornal da especialidade os resultados do estudo. Deste modo, todos os interessados podem conhecer em sumário o teste e solicitar ao autor o relatório do teste e autorização para o utilizar.

Ética na avaliação

Os códigos de ética recomendam que relativamente às técnicas de avaliação deve haver o cuidado de, ao apresentar os resultados, descrever os objectivos, as normas, a validade e fidelidade, e as aplicações do teste. Quando se apresentam resultados de um teste, quem os apresenta deve estar capacitado para assumir a responsabilidade pela aplicação, interpretação, e uso desses instrumentos.

Numa equipa de saúde, provavelmente quem está mais apto a utilizar técnicas deste tipo é o psicólogo. Estes profissionais estão subordinados a uma ética que define exigências sobre a escolha, administração, classificação, interpretação e uso de técnicas de avaliação deste tipo. Do mesmo modo a ética da avaliação psicológica impõe que não se utilizem técnicas de avaliação inapropriadas nomeadamente, porque são obsoletas, ou porque não se sabem dados sobre as propriedades psicométricas ou clinimétricas.

Direitos de autor dos testes

Quanto a autoria há dois grandes grupos de instrumentos de avaliação: (a) os que são produzidos por organizações empresariais com objectivos comerciais que desenvolvem testes, e (b) os que são desenvolvidos por organizações científicas com objectivos de investigação.

O desenvolvimento de um teste implica um estudo cuidadoso, nomeadamente, o estudo das propriedades psicométricas, e o desenvolvimento de valores normativos. Para tal, e no caso de alguns testes clássicos como os testes de Wechsler, para avaliação da inteligência o desenvolvimento dos testes demora alguns anos. Por essa razão a compra do teste e do direito de utilização é dispendiosa.

Muitos testes são vendidos por organizações que os comercializam a pessoas que estão legalmente capacitadas para os utilizar. No acto da compra assume-se o compromisso de utilizar o teste de um determinado modo. Tal implica, nomeadamente, que o teste seja utilizado com respeito da ética que rege os profissionais que o utilizam, e no estrito respeito das normas que acompanham o teste. A utilização de testes, que são comercializados, está protegida por direitos de autor cuja utilização se rege pela lei geral que define as regras para fazer cópias.

As leis que protegem os direitos de autor visam, fundamentalmente a protecção de aspectos económicos. Elas tendem a ser universais e os investigadores portugueses devem submeter-se a essas leis. Por outro lado a lei que protege os direitos de autor tende a proteger os usos que são feitos sem objectivos de lucro e com fins de investigação.

Como explicam Berzon, Patrick, Guyatt e Conley (1994), as leis que protegem os direitos de autor têm limitações: primeiro elas protegem somente o trabalho expresso e não as ideias que lhe dão origem; a lei protege somente a reprodução do trabalho expresso. Nada impede que se copiem as ideias.

No caso da utilização por investigadores portugueses acrescenta-se que os testes são normalmente desenvolvidos em língua inglesa tal como é falada nos Estados Unidos da América. Como já se referiu, a utilização destas técnicas, deve ser estudada e adaptada à realidade portuguesa. Ora, a adaptação das técnicas, como já foi referido deve começar pelos construtos e não pelos conteúdos, o que, não violaria a lei do ponto de vista dos direitos de autor. A violação seria do ponto de vista científico, e enquadrava-se no âmbito do plágio violando o código de ética da investigação científica.

Na área da saúde utilizam-se inúmeras técnicas de avaliação que são desenvolvidas por investigadores no âmbito de projectos de investigação, que são publicadas em revistas científicas. Muitas vezes estes instrumentos foram desenvolvidos com o apoio de subsídios do estado visando o desenvolvimento da ciência. O objectivo destes investigadores não é comercial e, se o seu estudo foi apoiado por fundos públicos, têm o dever de facilitar o uso das técnicas que desenvolveram.

Se um investigador pretende utilizar uma técnica de avaliação que tenha sido desenvolvida por uma organização com fins comerciais, o que tem a fazer é comprar o direito de utilização. Se a organização científica tem fins científicos e não comerciais, ou se se trata de um investigador individual, deve-se contactar esse investigador e pedir autorização para utilizar o instrumento.

De um modo geral, os investigadores facilitam, ou mesmo estimulam, que as suas técnicas sejam utilizadas por outros. Tal valoriza, ou acrescenta credibilidade aos instrumentos que desenvolveram, por um lado e, por outro, fornece dados para confirmar ou para aperfeiçoar esses instrumentos. A nossa experiência mostra que os investigadores respondem com rapidez no sentido de disponibilizar o uso dos instrumentos. Nalguns casos pedem um pagamento simbólico para despesas de correio. Noutros casos exigem a assinatura de um compromisso por parte do investigador que pede para utilizar o instrumento. Tal

visa garantir o uso apropriado da técnica e que ela não seja utilizada indiscriminadamente (é dada autorização para aquele estudo que é nomeado no contrato que assinam e não para outros).

Alguns grupos recomendam os procedimentos para adaptação dos instrumentos de avaliação a outras culturas e, quem decide utilizar essas técnicas deve fazê-lo do modo que a organização recomenda. Outros não fazem recomendações sobre o modo de fazer a adaptação e, mesmo, recomendam flexibilidade para a adaptação do instrumento. Ou seja, há muitas maneiras de proceder por parte dos investigadores que desenvolvem instrumentos de avaliação. O investigador português interessado em utilizar um instrumento que tenha sido publicado, deve contactar o autor para pedir autorização. Se não conseguir contactar o autor a utilização do instrumento deve ser feita com prudência.

Em todos os casos deve haver o cuidado para utilizar a técnica de modo adequado, com rigoroso respeito pelas procedimentos e normas definidas pelo autor.

Muitos investigadores recorrem à utilização de partes de instrumentos em vez do todo. Neste caso não se está a utilizar a técnica original. A transformação dos itens, da constituição das sub-escalas, e da forma de passar o teste altera o teste. Neste caso, e como já se referiu, trata-se de outro instrumento, de outro teste, e não do original. Neste caso tem de se fazer de novo o estudo das propriedades psicométricas ou clinimétricas.

Há limites à quantidade de material que pode ser copiado. Alterar um item de um teste com n itens e dizer que tal teste é de autoria própria é plágio. Quando a construção de um teste se baseia noutro, tal deve ser referido, e deve ser apresentada e defendida a razão porque se procedeu à alteração em vez de manter a versão original.

REFERÊNCIAS

- Aday, L. A. (1989). *Designing and conducting health surveys: A comprehensive guide*. San Francisco: Jossey-Bass Publishers.
- Ajzen, I. (1988). *Attitudes, personality, and behavior*. Milton Keynes: Open University Press.
- Anastasi, A. (1990). *Psychological testing* (6th ed.). New York: Macmillan Publishing Company.
- APA. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47(12), 1597-1611.
- Azeredo, Z. (1997). A autoria de um artigo científico. *Acta Médica Portuguesa*, 11, 813-813.
- Baker, T., & Brandon, T. (1990). Validity of self-reports in basic research. *Behavioral Assessment*, 12, 33-51.
- Baltes, P., Reese, H., & Nesselroade, J. (1977). *Life-span development psychology: Introduction to research methods*. California: Brooks/Cole publishing company.
- Beck, A. (1972). *Depression: Causes and treatment*. Philadelphia: University of Philadelphia Press.
- Belar, C., & Deardorff, W. (1995). *Clinical health psychology in medical settings: A practitioner's guidebook*. Washington: American Psychological Association.
- Bem, S., & de Jong, H. (1998). *Theoretical issues in psychology: An introduction*. London: Sage Publications.
- Bernal, H., Wooley, S., & Schensul, J. (1997). The challenge of using Likert-type scales with low-literate ethnic populations. *Nursing Research*, 46(3), 179-181.

- Berzon, R., Patrick, D., Guyatt, G., & Conley, J. (1994). Intellectual property considerations in the development and use of HRQL measures for clinical trial research. *Quality of Life Research*, 3, 273-277.
- Binet, A. (1890). Perceptions d'enfants. *Revue Philosophique*, 30, 582-611.
- Bishop, G. D. (1994). *Health psychology: Integrating mind and body*. Boston: Allyn and Bacon.
- Blanchard, E., & Schwarz, S. (1988). Clinically significant changes in behavioral medicine. *Behavioral Assessment*, 10, 171-188.
- Bowling, A. (1991). *Measuring health: A review of quality of life measurement scales*. Milton Keynes: Open University Press.
- Bowling, A. (1998). *Research methods in health*. Buckingham: Open University Press.
- Boyle, E., & Harrison, B. (1981). Factor structure of the health locus of control scale. *Journal of Clinical Psychology*, 37(4), 819-824.
- Bradley, C. (1994). Translation of questionnaires for use in different languages and cultures. In C Bradley (Ed.), *Handbook of psychology and diabetes* (pp. 43-55) Switzerland: Hardwood Academic Press.
- Brannon, L., & Feist, J. (1992). *Health psychology: An introduction to behavior and health*. Pacific Grove: Brooks/Cole Publishing.
- Breakwell, G. (1995) Research: Theory and method. In G. Breakwell, S. Hammond, & C. Fife-Schaw (Eds.), *Research methods in psychology* (pp. 5-15). London: SAGE Publications.
- Brislin, R., Lonner, W., & Thorndike, R. (1973). *Cross cultural research methods*. New York: Wiley and Sons.
- Campbell, D., & Stanley, J. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.
- Carroll, J. (1982). The measurement of intelligence. In R. Sternberg (Ed.), *Handbook of human intelligence* (pp. 29-120). Cambridge: Cambridge University Press.

- Carroll, J. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.
- Cattel, R. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1-22.
- Comrey, A. L. (1973). *A first course in factor analysis*. New York: Academic Press.
- Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, 56(5), 754-761.
- Cone, J., & Foster S. (1993). *Dissertation and theses from start to finish: Psychology and related fields*. Washington, DC: American Psychological Association.
- Cronbach, L. (1970). *Essentials of psychological testing*. New York: Harper & Row Publishers.
- Davis, A. (1995). The experimental method in psychology. In G. Beakwell, S. Hammond, & C. Fife-Schaw (Eds.), *Research methods in psychology* (pp. 50-68). London: SAGE Publications.
- Dawis, R. (1987). Scale construction. *Journal of Counseling Psychology*, 34(4), 481-489.
- De Raad, B. (1998). Five big, big five issues: Rationale, content, structure, status, and crosscultural assessment. *European Psychologist*, 3(2), 113-124.
- Digman, J. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.
- Digman, J. (1994). Historical antecedents of the five-factor model. In P. Costa & T. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (pp. 13-18). Washington: American Psychological Association.
- Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: John Wiley & Sons.
- Dommermuth, W. (1975). *The use of sampling in marketing research*. Chicago: American Marketing Association.

- Edwards, A. (1957). *The social desirability variable in peersonality assessment and research*. New York: Dryden.
- Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, 196, 129-136.
- Erikan, K. (1998). Translation effects in international assessments. *International Journal of Educational Research*, 29, 543-553.
- Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health*, 1(3), 70-72.
- Feinstein, A. (1994). Clinical judgement revisited: The distraction of quantitative models. *Annals of Internal Medicine*, 120, 799-805.
- Fernandez-Ballesteros. R. (1994). Evolución histórica de evaluación conductual. In R. Fernández-Ballesteros (Ed.), *Evaluación conductal hoy: Un enfoque para el cambio en psicología clínica y de la salud* (pp. 25-59). Madrid: Ediciones Pirámide, S.A.
- Fife-Schaw, C. (1995). Levels of measurement. In G. Beakwell, S. Hammond, & C. Fife-Schaw (Eds.), *Research methods in psychology* (pp. 38-49). London: SAGE Publications.
- Finsterbusch, K., Liewellyn, L., & Wolf, C.P. (1983). *Social impact assessment methods*. Beverly Hills: Sage Publications.
- Fitzpatrick, R., Fletcher, A., Gore, S., Jones, D., Spiegelhalter, D., & Cox, D. (1992). Quality of life measures in health care. I: Applications and issues in assessment. *British Medical Journal*, 305, 1074-1077.
- Fletcher, A., Gore, S., Jones, D., Fitzpatrick, R., Spiegelhalter, D., & Cox, D. (1992). Quality of life measures in health care. II: Design, analysis, and interpretation. *British Medical Journal*, 305, 1145-1148.
- Gardner, D., Cummings, L., Dunham, R., & Pierce, J. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement*, 58(6), 898-915.
- Goldfried, M., & Kent, R. (1972). Traditional versus behavioral personality assessment: A comparison of methodological and theoretical assumptions. *Psychological Bulletin*, 77(6), 409-420.

- Green, C. J. (1985). The use of psychodiagnostic questionnaires in predicting risk factors and health outcomes. In P. Karoly (Ed.), *Measurement strategies in health psychology* (pp. 301-333). New York: John Wiley & Sons.
- Guilford, J. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Guyatt, G. (1993). The philosophy of health-related quality of life translation. *Quality of Life Research*, 2, 461-465.
- Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, 6, 278-296.
- Hays, R., Anderson, R., & Revicki, D. (1993). Psychometric considerations in evaluating health-related quality of life measures. *Quality of Life Research*, 2, 441-449.
- Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: The universalist approach. *Quality of Life Research*, 7, 323-335.
- Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58, 295-300.
- Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, 11, 213-218.
- Huber, W. (1977). *Introduction à la psychologie de la personnalité*. Bruxelles: Dessart et Mardaga, Editeurs.
- Jackson, D. (1970). A sequential system for personality scale development. In C. Spielberger (Ed.), *Current topics in clinical and community psychology* (vol. 2, pp. 61-96). New York: Academic Press.
- JAMA. (1996). JAMA instructions for international authors. *The Journal of the American Medical Association*, 275(1), 5-9.
- Jerison, H. J. (1982). The evolution of biological intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 723-791). Cambridge: Cambridge University Press.
- Johnston, M. (1997). How health psychology makes a difference. *The Irish Journal of Psychology*, 18(1), 4-12.

- Kaplan, R., & Saccuzzo, D. (1993). *Psychological testing: Principles, applications, and issues* (3rd ed.). Pacific Grove: Brooks/Cole Publishing Company.
- Knape, K. (1994). initial management of acute pain. In S. Ramamurthy & J. Rogers (Eds.), *Decision making in Pain management* (pp. 2-3). New York: Mosby-Year Book, Inc.
- Kokmen, E., Smith, G., Petersen, R., Tangalos, E., & Ivnik, R. (1991). The Short Test of Mental Status: Correlations with standardized psychometric testing. *Archives of Neurology*, 48, 725-728.
- Lafon, R. (1973). *Vocabulaire de psychopédagogie et de psychiatrie de l'enfant*. Paris: Presses Universitaires de France.
- Lersch, P. (1954). *Aufbau der person*. Munique: Barth.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3(supl), 635-694.
- Magnusson, D., & Torestad, B. (1993). A holistic view of personality: A model revisited. *Annual Review of Psychology*, 44, 427-452.
- Mason, E., & Bramble, W. (1978). *Understanding and conducting research/ Applications in education and the behavioral sciences*. New York: McGraw-Hill Book Company.
- Matarazzo, J. D. (1987). Relationship of health psychology to other segments of psychology. In G. Stone, S. Weiss, J. Matarazzo, N. Miller, J. Rodin, C. Belar, M. Follick, & J. Singer (Eds.), *Health psychology: A discipline and a profession* (pp. 41-59). Chicago: The University of Chicago Press.
- Mathias, S., Fifer, S., & Patrick, D. (1994). Rapid translation of quality of life measures for international clinical trials: Avoiding errors in the minimalist approach. *Quality of Life Research*, 3, 403-412.
- McDowell, I., & Newell, C. (1987). *Measuring health: A guide to rating scales and questionnaires*. New York: Oxford University Press.
- Miaoulis, G., & Michener, R. (1976). *An introduction to sampling*. Iowa: Kendal/Hunt Publishing Company.
- Miller, S. (1984). *Experimental design and statistics*. London: Methuen.

- Miller, D. (1970) *Handbook of research design and social measurement* (2nd ed.). New York: David McKay Company, Inc.
- Murphy, K., & Davidshofer, C. (1998). *Psychological testing: principles and applications* (4th ed.). New Jersey: Prentice-Hall International, Inc.
- Norusis, M. (1991). *The SPSS guide to data analysis*. Chicago: SPSS Inc.
- Orley, J. (1994). The World Health Organization (WHO) quality of life project. In M. Trimble & W. Dpdsen (Eds.), *Epilepsy and quality of life* (pp. 99-107). New York: Raven Press.
- Ozer, D., & Reise, S. (1994). Personality assessment. *Annual Review of Psychology*, 45, 357-388.
- Pais-Ribeiro, J. L. (1994a). Adaptação do *Self-perception profile for college students* à população portuguesa: Sua utilização no contexto da psicologia da saúde. In L. Almeida & I. Ribeiro (Eds.), *Avaliação Psicológica: Formas e contextos* (pp. 129-138) Braga: APPORT.
- Pais-Ribeiro, J. L. (1994b). Reconstrução de uma escala de locus-de-controlo-de-saúde. *Psiquiatria Clínica*, 15(4), 207-214.
- Pais-Ribeiro, J. L. (1995). Adaptação de uma escala de avaliação da auto-eficácia geral. *Actas da III Conferência Internacional: Formas e Contextos* (pp. 163-176). Braga: APPORT.
- Pais-Ribeiro, J. L. (1995b). O padrão de respostas a questionários de avaliação psicológica: Questionário preenchido na presença do investigador ou quando os indivíduos estão sós? In L. Almeida & I. Ribeiro (Orgs.), *Avaliação Psicológica Formas e Contextos* (vol. III, pp.189-196). Braga: APPORT.
- Pais-Ribeiro, J. L., Mendonça, D., & Martins-da-Silva, A. (1997). Construção de uma escala de avaliação da qualidade de vida para doentes epilépticos. *Boletim da Epilepsia*, 2(2), 32-42.
- Pais-Ribeiro, J. L., Mendonça, D., & Martins-da-Silva, A. (1998) Impact of epilepsy on QOL in a portuguese population: exploratory study. *Acta Neurologica Scandinavica*, 97, 287-294.
- Pais-Ribeiro, J. L. (1998). *Psicologia e Saúde*. Lisboa: ISPA.

- Pais-Ribeiro, J. L. (2001). Mental health inventory: um estudo de adaptação. *Psicologia, Saúde & Doenças*, 2(1), 77-99.
- Pais-Ribeiro, J., & Leal, I. (1996). Psicologia clínica da saúde. *Análise Psicológica*, XIV(4), 589-599.
- Pais-Ribeiro, J., & Leal, I. (1997). Avaliação e psicodiagnóstico em psicologia da saúde. *Psicologia: Teoria, Investigação e Prática*, 2, 357-370.
- Perron-Borelli, M., & Perron, R. (1970). *L'examen psychologique de l'enfant*. Paris: Presses Universitaires de France.
- Pinto, A. (1990). *Metodologia da investigação científica*. Porto: Edições do Jornal de Psicologia.
- Polgar, S., & Thomas, S. (1988). *Introduction to research in the health sciences*. Melbourne: Churchill Livingstone.
- Poch, J., & Talarn, A. (1991). Entrevista psicológica e história clínica. In J.V. Ruiloba (Ed.), *Introducción a psicopatología y la psiquiatria* (pp. 64-75). Barcelona: Salvat.
- Popper, K. (1959). *The logic of Scientific Discovery*. London: Hutchinson.
- Reckase, M. (1990). Scaling techniques. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (2nd ed., pp.41-56). New York: Pergamon Press.
- Rotter, J. (1966). General expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80 (whole no. 609).
- Rotter, J. (1990). Internal versus external control of reinforcement: A case history of a variable. *American Psychologist*, 45(4), 489-493.
- SAC. (1995). Instrument review criteria. *Bulletin da Medical Outcomes Trust*, 3(4), I-IV.
- Sampaio, C., & Ferreira, J. (1997). Avaliação crítica de artigos relativos a ensaios clínicos controlados. I: Planeamento. *Acta Médica Portuguesa*, 11, 749-754
- Schoenfeld, L., & Costello, R. (1994). Psychological evaluation. In S. Ramamurthy & J. Rogers (Eds.), *Decision making in Pain management* (p. 6). New York: Mosby- Year Book, Inc.

- Siegler, H., & Richards, D. (1982). The development of intelligence. In R. Sternberg (Ed.), *Handbook of human intelligence* (pp. 897-971). Cambridge: Cambridge University Press.
- Smith, H. (1975). *Strategies of social research: the methodological imagination*. New Jersey: Prentice-Hall, Inc.
- Spearman, C. (1927). *The habilities of man*. New York: Macmillan.
- Spielberger, C., Gorsuch, R., Lushene, R., Vagg P., & Jacobs G. (1983). *Manual for the State Trait Anxiety Inventory (Form Y)*. Palo Alto: Consulting Psychologists Press, Inc.
- Sternberg, R., & Salter, W. (1982). Conceptions of intelligence. In R. Sternberg (Ed.), *Handbook of human intelligence* (pp. 3-28). Cambridge: Cambridge University Press.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.
- Strahan, R., & Gerbasi, K. (1972). Short homogeneous versions of the Marlowe-Crowne Social Desirability Scale. *Jornal of Clinical Psychology*, 28, 191-193.
- Sundberg, N. (1977). *Assessment of persons*. New Jersey: Prentice-Hall, Inc.
- Swencionis, C., & Hall, J. (1987). Ethical concerns in health psychology. In G. Stone, S. Weiss, J. Matarazzo, N. Miller, J. Rodin, C. Belar, M. Follick, & J. Singer (Eds.), *Health psychology: A discipline and a profession* (pp. 203-215). Chicago: The University of Chicago Press.
- Thorndike, R., & Hagen, E. (1977). *Measurement and evaluation in psychology and education* (4th ed.). New York: John Wiley & Sons.
- Thurstone, L. (1935). *Vectors of the mind: Multiple-factor analysis for the isolation of primary traits*. Chicago: University of Chicago Press.
- Tinsley, H., & Tinsley, D. (1987). Use of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, 34(4), 414-424.
- Todd, C., & Bradley, C. (1994) Evaluating the design and development of psychological scales. In C. Bradley (Ed.), *Handbook of psychology and diabetes* (pp. 15-42) Switzerland: Hardwood Academic Press.

- Truog, R., Robinson, W., Randolph A., & Morris, A. (1999). Is informed consent always necessary for randomized controlled trials? *The New England Journal of Medicine*, 340(10), 804-807.
- Turk, D. C., & Kerns, R. D. (1985). Assessment in health psychology: A cognitive behavioral perspective. In P. Karoly (Ed.), *Measurement strategies in health psychology* (pp. 335-372). New York: John Wiley & Sons.
- Vernon, P. (1960). *The structure of human abilities*. London: Methuen.
- West, S. G. (1985). Beyond the laboratory experiment: Experimental and quasi-experimental designs for interventions in naturalistic settings. In P. Karoly (Ed.), *Measurement strategies in health psychology* (pp. 183-233). New York: John Wiley & Sons.
- Wijsman, D., Hekster, Y., Keyser, A., Renier, W., & Meinardi, H. (1991). Clinimetrics and epilepsy care. *Pharmacy World & Science*, 13(4), 182-188.
- Willard, R. (1995). Assessment of pain and pain behavior. In L. Cushman & M. Scherer (Eds.), *Psychological assessment in medical rehabilitation* (pp. 237-273). Washington: American Psychological Association.