

## Aspectos Transculturais na Adaptação de Instrumentos de Avaliação Psicológica

---

Ilda Cardoso

### PRINCÍPIOS DE MEDIDA

A pesquisa em avaliação psicológica tem vindo a assumir grande relevo, sobretudo na prática da psicologia e psiquiatria, nos diversos contextos destas profissões. O interesse por esta área do conhecimento, da qual a psicometria é ‘responsável’, tem as suas origens na tentativa de resolução de vários problemas, em particular, saber avaliar objectivamente as aptidões humanas (Pasquali 2003: 14). Esta avaliação só se tornava possível, na condição de se poder quantificar essas mesmas aptidões, já que, segundo Pasquali e a título de exemplo, embora o teste de inteligência de Binet tenha sido fundamental para o progresso da psicologia, não foi Binet quem deu origem à psicometria, tendo contribuído para fomentar o seu desenvolvimento, mas não lhe conferindo o relevo primordial da quantificação. De acordo ainda com Pasquali, ‘a Psicometria poderia ter tido as suas origens em duas situações distintas: (1) a psicologia de orientação empirista ou (2) a psicologia mais mentalista de Binet, na França’ (Pasquali 2003: 14).

Deste modo, quando abordamos o tema da avaliação psicológica, não devemos negligenciar a complexidade no quadro de um instrumento de avaliação. A sua importância é de tal forma determinante

que permite tornar mensurável o que pensamos não o ser, como é o caso das aptidões humanas. Estas não possuem medidas directas, necessitando, por isso, de recorrer a indicadores.

Para que possamos quantificar alguma coisa, isto é, medir, é necessário recorrer à instrumentação que é um dos componentes maiores de medida. A instrumentação consiste, assim, em aplicar normas precisas na elaboração de métodos e instrumentos para medir características e atributos (Waltz, Strikland e Lenz 1991).

Uma vez que se torna necessário aplicar normas precisas para a elaboração dos instrumentos, coloca-se uma questão. Como será possível clarificar os objectos a medir, os números a utilizar e as regras de atribuição dos valores aos objectos ou aos sujeitos? Com efeito, actualmente, com o método dos testes é possível medir as emoções, os comportamentos, a inteligência, as características de personalidade, o desempenho, enfim, variadas situações particulares. No entanto, as interrogações incidem na definição do 'conceito' a medir. Tal como é defendido por Kerlinger (1973, citado por Fortin 2000: 36), um conceito é uma abstracção formada pela generalização de situações particulares, de enunciados, de observações ou de comportamentos, o que torna a maior parte dos conceitos não directamente observáveis ou mensuráveis como tais. Para serem mensuráveis, os conceitos devem ser convertidos em indicadores.

Definem-se como indicadores todas as medidas indirectas, escolhidas para quantificar os conceitos. As medidas directas são, de facto, bastante raras, porque se referem, na maior parte das vezes, a um instrumento construído a partir de uma medida padrão. Como exemplos mais comuns, temos a régua, o termómetro, o cronómetro, a balança.

As medidas indirectas, traduzidas por indicadores indirectos, referem-se, particularmente, aos conceitos mais abstractos, como o stresse, crenças, ansiedade, depressão, etc. Por esse motivo, é necessária a definição prévia conceptual, uma vez que é essa definição que fornece a significação teórica do construto e serve de guia à elaboração do indicador apropriado para medir o conceito. Não é, portanto, estranho que a decisão na escolha dos indicadores, ou 'itens' (como se diz em linguagem psicométrica), para a construção de um instrumento de medida, seja intensamente estudada e cuidada. Os itens são conceitos operacionalizáveis, isto é, passíveis de serem medidos. O processo que permite tornar um 'conceito operacional faz ressaltar as ligações entre o pensamento e a experiência, fixando, assim, como um conceito ou uma ideia abstracta serão identificados e medi-

dos em termos concretos' (Fortin 2000: 38). Desta forma, é fundamental que os investigadores, ao construírem um instrumento, atendem às bases conceptual e metodológicas dos conceitos que pretendem medir.

A construção integral de instrumentos de avaliação não é uma actividade ligeira. Trata-se de um processo bastante complexo que consome recursos, capacidades e conhecimentos, não sendo, por esses motivos, acessível a todos. Nesse sentido, é comum encontrarmos estudos realizados com instrumentos de avaliação originalmente de outros países que não o nosso, interrogando-nos por que não construímos um semelhante. Esta questão terá, com certeza, uma resposta, necessitando, no entanto, de um espaço próprio de reflexão. Irei debruçar-me, neste artigo, sobre os aspectos transculturais na adaptação de instrumentos de avaliação psicológica.

### **ASPECTOS TRANSCULTURAIS DOS INSTRUMENTOS DE AVALIAÇÃO**

Sendo a cultura um determinante crítico do comportamento humano, a definição conceptual e linguística dos instrumentos de avaliação deve ser apreciada de forma a contemplar as características específicas de cada população a ser investigada.

A linguagem é a principal forma pela qual conseguimos transmitir sintomas, preocupações, desejos, afectos, aliada à comunicação não verbal, igualmente importante na relação com os outros. Os instrumentos de avaliação, de que se ocupa a avaliação psicológica, são fundamentalmente dependentes da escrita, pelo que as preocupações se voltam para a necessidade de equivalências conceptuais e linguísticas, de forma a manter uma equivalência cultural do conceito a ser medido.

Para SAC (1995, citado por Ribeiro 1999: 131) e Ferreira (1998) e de acordo ainda com o European Group on Health Outcomes (ERGHO), a adaptação inter-cultural de um instrumento envolve duas fases fundamentais, designadamente: a avaliação das equivalências conceptuais e linguísticas e a avaliação das propriedades psicométricas.

Para Hayes, Anderson e Revicki (1993, citado por Ribeiro 1999) as quatro equivalências fundamentais a considerar, na tradução de um instrumento de avaliação, são: a equivalência da tradução do item; a equivalência operacional do item; a equivalência da escala; a equivalência métrica da escala. Podemos traduzi-las em equivalência de

conteúdo, de critério e técnica. Estas últimas são já contempladas, ou, pelo menos, deverão sê-lo, aquando das validações psicométricas, através da validade de conteúdo, validade de critério e validade de construto.

A equivalência conceptual refere-se 'à relevância e ao significado, noutra cultura, de um conceito ou dimensão tidos por relevantes na cultura original' (Ferreira 1998: 20). Por seu lado, segundo o mesmo autor, a equivalência linguística ou semântica prevê a verificação da construção das perguntas ou itens do instrumento, de modo a que estas mantenham o seu significado que tinham na sua língua original.

A simples tradução de um teste de uma língua para a outra, mesmo assegurando as regras lexicais, não mantém a originalidade de quem consumiu bastantes recursos a criar a versão original. Desta forma, para De Raad (1998, citado por Ribeiro 1999: 128), 'as línguas diferem muito umas das outras: traços descritos numa língua não têm necessariamente tradução correspondente noutra língua'. Por esta razão, a simples tradução não é suficiente para a correcta utilização de instrumentos, necessitando de ser feita a adaptação cultural e linguística.

De acordo com Hutchinson et al. (1997, citado por Ferreira et al. 1998: 19), este processo inicia-se pela tradução do instrumento, tendo sempre em atenção a existência de conceitos potencialmente ambíguos. Por essa razão, o tradutor/investigador deverá, sempre que possível, estar em sintonia com o autor do instrumento original para rever assuntos relativos à semântica dos conceitos utilizados. Posteriormente, a metodologia mais correcta compreende a retroversão para a língua original. A tradução deverá 'ser realizada por dois ou mais tradutores independentes, bilingues e multiprofissionais ou tradutores oficiais cuja língua mãe é a da versão a adaptar'.

Para Fletcher (1992, citado por Ribeiro 1999: 128) os investigadores devem ser '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'.

Note-se que os vieses de resposta podem ocorrer independentemente da qualidade dos itens ser boa ou da tradução estar convenientemente bem elaborada, uma vez que as respostas dadas podem ser desvirtuadas por factores culturais relacionados com os sujeitos que respondem, falseando os dados. O viés é, assim, causado pelas diferenças culturais e não pela tradução dos itens propriamente dita. De

acordo com Pasquali (2003: 146), a cultura pode ser ‘uma causa de erros de respostas que se relaciona com os problemas de transferência de instrumentos psicológicos para outras populações para as quais eles não foram especificamente construídos e validados’.

O debate sobre o uso de testes por culturas diferentes daquelas que foram inicialmente previstas, aquando da sua construção, tem sido frequente, nomeadamente, nos Estados Unidos, onde existe um vasto mundo de sub-culturas. São vários os métodos estatísticos que têm sido utilizados de forma a permitir reduzir o problema dos vieses das respostas, em função da cultura dos respondentes, salientando-se o método delta (Angoff e Ford 1972; Angoff 1982, citado por Pasquali 2003: 146) e o método dif da psicometria moderna (Ironson 1982; Hambleton 1991; Ellis 1991; Jackson 1991, citado por Pasquali 2003: 146).

Neste sentido, a versão traduzida deve ser retrovertida para a língua original por tradutores, cuja língua mãe é a da versão original. As duas versões na língua original deverão ser comparadas e, se houver necessidade de modificações no instrumento, deverão ser realizadas as retroversões necessárias até se conseguir estabelecer a referida equivalência. Somente através deste processo, afirma Hutchinson (1997), podemos concluir que o instrumento se encontra cultural e linguisticamente equivalente.

É essencial enfatizar que, no processo de adaptação de instrumentos originais de uma língua para outra, no caso de estar protegido por direitos de autor, é necessária a solicitação do pedido de autorização à editora responsável pela sua comercialização, por parte do investigador interessado na utilização do teste ou, em alguns casos, directamente ao autor da versão original, ou ainda a quem detenha a sua patente.

De seguida, passo a apresentar as considerações teóricas sobre os conceitos mais relevantes no processo de equivalência das escalas, a validade e fidedignidade dos instrumentos de avaliação psicológica.

## **VALIDADE E FIDEDIGNIDADE DOS INSTRUMENTOS DE AVALIAÇÃO PSICOLÓGICA**

Nenhum instrumento de avaliação psicológica é válido como instrumento de medida se não passar pelos crivos da validade e fidedignidade. Este procedimento aplica-se tanto para a construção de um novo instrumento como para a adaptação transcultural de um já existente.

A validade de um teste constitui um parâmetro essencial e indispensável na teoria da medida. A validade relaciona-se com a capaci-

dade de um instrumento medir o construto para o qual foi inicialmente construído. Desta forma, qualquer instrumento de avaliação psicológica que pretendamos adaptar para outra população tem, necessariamente, de passar pelos estudos de validade. De acordo com Pasquali (2003: 158):

A validade diz respeito ao aspecto da medida ser congruente com a propriedade medida dos objectos e não com a exactidão com que a mensuração, que descreve esta propriedade do objecto, é feita. Em Física, o instrumento é um objecto físico que mede propriedades físicas; então parece fácil ver que a propriedade do objecto mensurante é ou não congruente com a propriedade do objecto medido.

Para ilustrar estas definições, o autor refere que, se quisermos avaliar o ‘comprimento’ de um determinado objecto, utilizamos uma régua ou um metro para o fazer. A régua usa a sua propriedade de ‘comprimento’ para medir o comprimento de outro objecto. Neste caso, não necessitamos de provar a congruência das propriedades, uma vez que os termos são unívocos, isto é, conceptualmente equivalentes.

A história do parâmetro da validade pode traduzir-se em três períodos, que deram origem às principais validades actualmente estudadas em psicometria. O primeiro período, que corresponde ao Predomínio da Validade de Conteúdo (1900-1950), é caracterizado pelas teorias da personalidade (psicanálise, fenomenologia, gestaltismo, etc.) onde a fundamentação empírica não tinha lugar, prevalecendo a teoria. O segundo período ou Predomínio da Validade de Critério (1950-1970) dava grande relevo ao behaviorismo skinneriano, que veio influenciar a psicometria, sendo a estatística um recurso fundamental na análise dos itens. Por fim, o último período correspondente ao Predomínio da Validade de Construto (desde 1970), tendo sido Cronbach e Meehl os responsáveis pelo seu aparecimento com um artigo de 1955 sobre o *modelo trinitário da validade* (conteúdo, critério e construto) (Pasquali 2003: 161).

Como foi referido anteriormente, a validade de um instrumento é definida como a propriedade de medir aquilo que se pretende ser medido por esse instrumento (Edenborough 1998). Assim, a validade diz respeito à exactidão com que um determinado conceito é medido. Como os conceitos em estudo são representados pelos enunciados específicos do instrumento de medida, a validade depende dessa mesma representação.

Assim, com os testes de validação pretende-se saber se os indicadores medem, de facto, os atributos que lhes estão subjacentes. Não sendo apenas os únicos, os critérios, porém, mais frequentemente utilizados no processo de validação são: a validade de conteúdo, a validade de critério e a validade de construto.

Apresento, seguidamente, breves definições de alguns tipos de validade considerados relevantes na adaptação de instrumentos de avaliação psicológica.

#### *Validade facial*

A validade facial não é considerada, por alguns autores, como sendo validade em sentido técnico, uma vez que se relaciona apenas com a aparência do instrumento (teste), sendo um conceito limitado, que não pode reflectir a intenção do conteúdo. Para Nunnally (1995), este conceito de validade é, por definição, bastante vago e pode, inclusivamente, confundir-se validade facial com validade de conteúdo. Por essa razão, o autor defende que os aspectos relacionados com a validade facial se referem à análise dos itens, *depois* do instrumento ser construído.

Por outro lado, a validade de conteúdo fica garantida pela definição do conteúdo e escolha dos itens *antes* de o instrumento estar construído. Assim, a validade facial pode ser considerada como um aspecto limitativo da validade de conteúdo (Nunnally 1995). De qualquer modo, desempenha um importante papel na construção e adaptação de um teste.

A aparência do teste pode, reconhecidamente, interferir nas respostas do sujeito e, consequentemente, no resultado final da avaliação. Segundo Edenborough (1998), se quem responde aos itens os considera *bizarros*, provavelmente, todo o processo ficará seriamente comprometido. Este autor sugere, por isso, que é importante para a validade facial de um teste a forma como todo o processo é conduzido, particularmente, na administração do instrumento.

É através da validade facial que os sujeitos obtêm a sensação daquilo que o teste pretende medir. Se um teste aparentar ser demasiado fácil ou, pelo contrário, apresentar perguntas muito elaboradas ou difíceis, pode suscitar nos sujeitos efeitos comprometedores para a verificação da validade.

#### *Validade de conteúdo*

Dizemos que um determinado teste tem validade de conteúdo quando existe adequação do instrumento ao conteúdo a mensurar.

Este tipo de validade fica determinado pela concordância entre os vários peritos ou juízes especialistas no conteúdo do domínio em avaliação, através de processos estatísticos adequados a um refinamento da selecção original dos itens. Note-se que os conceitos teóricos utilizados nesta área científica (ciências sociais) são muito abstractos, o que nem sempre promove o consenso quanto à sua aceitação.

Cada item do teste é avaliado individualmente e em conjunto, de modo a que cada um seja uma amostra de conhecimento específico do que o teste se propõe medir. Além disso, cada item deve cair em, pelo menos, uma das áreas de conteúdo. Se isso não acontecer é porque não é relevante para os objectivos da escala ou estes não são suficientemente exaustivos (Ferreira 1998).

A validade de conteúdo de um instrumento baseia-se num critério; não existem métodos objectivos para assegurar que um instrumento apresenta de forma adequada o conteúdo, uma vez que a validade pode estar ameaçada, não só pela estrutura do teste, mas pelas próprias condições em que o teste está a ser aplicado. Assim, a validade de conteúdo fica, nomeadamente, ameaçada pelo tempo de duração do teste ou condições de aplicação, como é o caso das diferenças nas instruções e procedimentos de pontuação.

Através da validade de conteúdo pretende-se demonstrar que o domínio do conteúdo de um instrumento de medição é apropriado relativamente aos objectivos esperados. Deste modo, o teste deste tipo de validade pretende dar resposta às seguintes questões (Ferreira 1998: 12): 'Será que as componentes da escala ou item cobrem todos os aspectos do atributo a ser medido? Será que o conteúdo da variável corresponde à designação que se lhe atribuiu?'

Para a verificação deste tipo de validade, Ferreira (1998) sugere a necessidade de observação dos seguintes passos interrelacionados:

1. Especificar o domínio completo do conteúdo.
2. Especificar os procedimentos de amostragem usada na construção do instrumento de medição.
3. Proceder à análise da apresentação do instrumento de medição.

#### *Validade concorrente*

Este tipo de validade permite avaliar os resultados do novo teste, em comparação com outros critérios que medem dimensões semelhantes.

É possível apurar a validade concorrente, através da administração de outros instrumentos já validados para a medição do critério que se pretende medir. A validade concorrente envolve a predição de um

método alternativo de mensuração de algumas características que nos interessam. De acordo com Edenborough (1998: 47), 'se um grupo de indivíduos estabeleceu altas performances no seu trabalho, então eles obtêm altos scores no teste; às baixas performances de produção correspondem scores também baixos, de forma efectiva a validade concorrente é estabelecida'.

#### *Validade preditiva*

A validade preditiva refere-se à capacidade de um teste poder prever um comportamento futuro. Para estabelecer a validade preditiva, é necessário ter em atenção o período de tempo em que a investigação ocorre (Edenborough 1998).

No caso da utilização de testes de inteligência ou mesmo de aptidão que são administrados a jovens, estes, para além de poderem determinar as suas aptidões no aqui e no agora, podem, igualmente, prever comportamentos futuros, isto é, quais são os mais competentes para uma ou outra área. Neste sentido, é admitido que o teste tem validade preditiva.

#### *Validade de construto*

Um dos problemas mais difíceis e árduos que um investigador enfrenta na validação de um instrumento é relacionado com a validade de construto (Polit e Hungler 1991). O que se pretende saber relativamente à validade é se o instrumento realmente mede o construto que se pretende ser medido. Uma dificuldade da medição, como foi referido, é que a natureza abstracta de muitos construtos torna difícil a obtenção dos resultados, por não ser possível observá-los directamente.

A diferença entre validade de critério e validade de construto prende-se com o facto deste último se orientar mais para o atributo subjacente (também designados como construtos hipotéticos). A validade de construto baseia-se em implicações teóricas associadas às construções, isto é, ao paradigma teórico. Para Ferreira (1998: 13) 'este é o tipo de validação que deve ser investigada sempre que não exista critério ou universo de conteúdo aceite como sendo inteiramente adequado à definição do conceito que se pretende medir'.

O método mais comumente utilizado para o cálculo deste tipo de validade é a análise factorial, uma vez que permite avaliar o número de dimensões que estão subjacentes a um conjunto de variáveis. Pode também recorrer-se ao cálculo da validade convergente (correlacionando as variáveis relacionadas) e à validade divergente (ausência de correlação com variáveis não relacionadas).

## ANÁLISE FACTORIAL

O procedimento estatístico mais antigo e mais bem conhecido para investigar a relação entre um conjunto de variáveis observadas e variáveis latentes é a Análise Factorial. A utilização desta técnica permite ao investigador examinar a covariação entre um conjunto de variáveis observadas (por ex., itens de um teste) informações sobre os construtos latentes que lhe são subjacentes (ex.: ansiedade).

Podemos falar de dois tipos de análise factorial, a Análise Factorial Exploratória (*Exploratory Factor Analysis* – EFA) e a Análise Factorial Confirmatória (AFC).

A Análise Factorial Exploratória (EFA) é utilizada nas situações em que as ligações entre as variáveis observadas (itens) e as latentes são desconhecidas ou incertas. A análise efectua-se de uma forma exploratória, a fim de determinar como as variáveis observadas se ligam com os seus factores subjacentes. De uma forma típica, o investigador pretende identificar o número mínimo de factores subjacentes (isto é a covariação comum) relacionados com as variáveis observadas. Por exemplo, suponhamos que o investigador desenvolve ou adapta um novo instrumento destinado a avaliar as múltiplas facetas da ansiedade (por ex., ansiedade estado-cognitiva e somática). Após a elaboração dos itens do questionário destinado a medir os dois construtos subjacentes, o investigador conduzirá, então, uma EFA para determinar de que forma os itens utilizados (variáveis observadas) se relacionam com os construtos subjacentes. Em análise factorial, essas relações denominam-se saturações factoriais.

O investigador, ao realizar uma EFA, desejará que os itens utilizados ou concebidos para medir determinado construto, por exemplo, ansiedade cognitiva, apresentem saturações factoriais elevadas nesse factor e saturações baixas ou negligenciáveis no outro factor. Considera-se este processo uma abordagem analítica factorial exploratória, uma vez que o investigador não tem conhecimento prévio de como os itens se comportam na medição dos factores previamente definidos.

Por outro lado, na Análise Factorial Confirmatória (AFC), pretende-se testar a hipótese da existência de uma ligação particular entre as variáveis observadas (itens) e os factores que lhe são subjacentes, isto é, se existe uma ligação de facto ou não. Detendo o conhecimento teórico sobre determinado fenómeno, o investigador, a priori, estará em condições de poder postular um padrão de ligação existente de forma a testar a sua hipótese estatisticamente.

A especificação do modelo da AFC, de acordo com os exemplos que tenho vindo a apresentar, permite que todos os itens da ansiedade cognitiva sejam 'livres' de saturar nesse factor, mas com a restrição de terem saturações 'Zero' nos outros factores. Este modelo é, então, avaliado, estatisticamente, de forma a poder determinar-se a adequação do ajustamento dos dados ao modelo previamente elaborado. Dito por outras palavras, este modelo diz respeito à amplitude e à intensidade com que as variáveis observadas são geradas pelos construtos que lhe são subjacentes e, como consequência, os padrões de regressão dos factores com as variáveis observadas, que são de um interesse fundamental.

## FIDEDIGNIDADE

A fidedignidade de um teste refere-se à consistência dos resultados obtidos pelas mesmas pessoas em ocasiões diferentes, ou com conjuntos de itens equivalentes (Anastasi 1990). O que se pretende com a fidedignidade de um teste é perceber se as diferenças individuais na realização de um teste podem ser atribuídas a diferenças reais das características avaliadas. O resultado/estimativa dos dois conjuntos de avaliações independentes são traduzidas por uma correlação.

Os procedimentos mais comuns no cálculo da fidedignidade de um teste são: teste-reteste (estabilidade), formas alternativas (equivalência), consistência interna (metade-metade, Alpha de Cronbach, correlação item-restante, Kuder-Richardson) e inter avaliadores.

A fidedignidade teste-reteste fornece os resultados obtidos pelos mesmos sujeitos em momentos diferentes. Assim, permite avaliar os resultados em dois momentos diferentes. No entanto, é necessário garantir, tanto quanto possível, que estes sujeitos não foram expostos a condições que possam alterar os domínios que se pretendem medir, uma vez que podem ocorrer situações que comprometem os resultados entre as duas medições. Em alguns casos, 'devido à aprendizagem, como acontece com os testes de inteligência, os respondentes podem aprender a resposta' (Ribeiro 1999: 112). Noutras situações, é possível que haja alterações na variável, como é o caso 'das medições da ansiedade estado, onde as correlações deverão, caso o tempo entre as duas passagens seja muito grande, ser mais moderadas' (ibid.).

O processo de Formas Equivalentes (também designadas Formas Paralelas) não difere muito do teste-reteste. Trata-se de um procedi-

mento diferente, na medida em que são feitas duas passagens de testes de duas formas equivalentes (dois testes semelhantes que avaliam o mesmo construto). Esta administração pode ser realizada no mesmo momento ou em momentos diferentes e apresenta vantagens relativamente ao anterior, na medida em que reduz o efeito de memória, mas acarreta, porém, a dificuldade da construção de formas equivalentes.

A consistência interna é avaliada, através de vários procedimentos, dos quais se destaca a metade-metade que pressupõe a divisão do teste em duas metades. Pode recorrer-se à divisão pelos itens pares e ímpares ou, simplesmente, dividir os itens da escala, de modo a que corresponda 50% dos itens na primeira parte e 50% na segunda parte. No cálculo da fidedignidade metade-metade, o coeficiente de correlação entre ambas as metades é efectuado, utilizando-se quando necessário, através da correcção de Spearman Brown ou consistência interna pela metade-metade, de acordo com os procedimentos de Guttman.

Outro procedimento igualmente importante no cálculo da consistência interna é o coeficiente alfa de Cronbach e está 'matematicamente ligado ao método das duas metades' (Murphy e Davidshofer 1998, citado por Ribeiro 1999: 113). O cálculo do coeficiente alfa permite estimar até que ponto cada item da escala mede de forma equivalente o mesmo conceito. O valor considerado bom para o alfa deverá situar-se entre .70 e .80 (Nunnally e Bernstein 1995). No entanto, são aceitáveis valores acima dos .60. Se a escala for constituída por um reduzido número de itens, o valor da correlação poderá ser influenciado (Ribeiro 1999). O coeficiente tende a ser mais elevado, se o instrumento possuir um maior número de itens.

Através do coeficiente de correlação Kuder-Richardson, estima-se a consistência interna dos itens da escala entre si. A particularidade de utilização desta técnica encontra-se no tipo de teste, constituído por itens dicotómicos (sim ou não; verdadeiro ou falso).

A correlação item-restante estabelece o grau de correlação entre cada item e o *score* total da escala, permitindo identificar os itens com correlações mais baixas e eliminá-los, se for caso disso, uma vez que podem ameaçar a consistência interna do instrumento.

Por fim, a fidedignidade interavaliadores e intercodificadores (também designado por harmonia entre juízes ou fidelidade do cotador) refere a concordância entre os vários avaliadores e cotadores, avaliando a concordância dos resultados obtido nos testes, quando estes são recolhidos por dois ou mais avaliadores. Para Fortin (2000: 228),

‘a harmonia entre juízes refere-se ao grau de concordância entre os resultados de dois ou mais observadores independentes que utilizaram os mesmos instrumentos de medida junto dos sujeitos, ou observaram e registaram o mesmo acontecimento’. Para esta autora, o cálculo é efectuado tendo por base dois processos distintos:

1. O cálculo da proporção dos julgamentos atribuídos às mesmas categorias pelos mesmos avaliadores e permite obter uma percentagem de concordância normalmente entre os 80 e 100%.
2. A correlação dos resultados obtidos pelos observadores em cada enunciado (item) com a ajuda da estatística K (Kappa).

$$K = \frac{Po - Pc}{1 - Pc}$$

**Po** é a proporção de concordância observada entre os scores e **Pc** é a proporção de concordância esperada pelo acaso.

## CONCLUSÃO

O propósito crítico deste artigo é chamar a atenção para a necessidade, junto daqueles que se dedicam à investigação, da correcta utilização de instrumentos de avaliação psicológica. A crescente proliferação de instrumentos de medida e, conseqüentemente, o aumento dos seus utilizadores, faz com que as preocupações se acentuem, quando nos damos conta, empiricamente, que vários desses instrumentos, que consumiram muito tempo de trabalho e recursos aos seus autores, são utilizados de forma pouco metódico e, por vezes, duvidosa. As razões desta situação oscilam entre o desconhecimento e a pura falta de rigor científico.

## REFERÊNCIAS

- Anastasi, A.  
1990 *Psychological Testing*. Nova Iorque: Macmillan Publishing Company.
- Edenborough Robert  
1998 *Using Psychometrics: A Practical Guide to Testing and Assessment*. Londres: Edenborough Robert.
- Ferreira, P. Marques, F.  
1998 *Avaliação Psicométrica e Adaptação Cultural e Linguística de Instrumentos de Medição em Saúde: Princípios Metodológicos Gerais*. Faculdade de Economia. Coimbra: Centro de Estudos e Investigação em Saúde.
- Ferreira, P.  
2000 'Criação da Versão Portuguesa do MOS SF – 36 Parte I – Adaptação cultural e linguística'. *Acta Médica Portuguesa* 13. pp.55-63.
- Hutchinson, A.; Bentzen, N.; Koning-Zahn, C. (eds.)  
1997 *Cross-cultural Health Outcome Assessment: A User's Guide*. Ruiner, NL: ERGHO.
- Fortin, M.F.  
2000 *O Processo de Investigação: Da Concepção à Realização*. Loures: Lusociência.
- Nunnally, J.C.  
1995 *Psychometric Testing*. Nova Iorque: McGraw-Hill Book Company.
- Nunnally, J.C.; Bernstein, I.  
1995 *Psychometric Theory*. Nova Iorque: McGraw-Hill Book Company.
- Pasquali, L.  
2003 *Psicometria: Teoria dos Testes na Psicologia e na Educação*. Petrópolis: Vozes.
- Polit, D.F.; Hungler, B.P.  
1991 *Nursing Research: Principles and Methods*. Filadélfia: J.B. Lippincott.
- Ribeiro, J.L.P.  
1999 *Investigação e Avaliação em Psicologia da Saúde*. Lisboa: Climepsi Editores.
- Waltz, C.F.; Strickland, O.; Lenz, E.R.  
1991 *Measurement in Nursing Research*. Filadélfia: F. A. Davis Company.

**Aspectos Transculturais na Adaptação  
de Instrumentos de Avaliação Psicológica**

**Trans-cultural Aspects in the Adaptation of  
Instruments of Psychological Assessment**

***Sumário***

***Summary***

---

Os testes de inteligência contribuíram, significativamente, para o desenvolvimento na avaliação e compreensão das competências humanas. Este artigo destaca a dimensão transcultural como um dos aspectos mais importantes na adaptação de instrumentos de medição em psicologia. O objectivo e a responsabilidade crítica do investigador é garantir a correcta aplicação desses instrumentos na diversidade de contextos socio-culturais, promovendo a validade transcultural do conhecimento científico.

Intelligence tests have contributed, significantly, to the development of assessment and comprehension of human competences. This article enhances the trans-cultural dimension as one of the most important aspects in the adaptation of measurement instruments in psychology. The purpose and critical responsibility of the researcher is to ensure the proper application of those instruments in the diversity of socio-cultural contexts, promoting the trans-cultural validity of scientific knowledge.