

**RECOMENDAÇÕES
PARA A UTILIZAÇÃO DE TESTES PSICOLÓGICOS EM INVESTIGAÇÃO
E PARA A
TRADUÇÃO E ADAPTAÇÃO DE TESTES PARA FINS DE INVESTIGAÇÃO**

INTRODUÇÃO

Este documento destina-se a apoiar a submissão de projectos de investigação à COMISSÃO DE ÉTICA DA FACULDADE DE MEDICINA DA UNIVERSIDADE DE COIMBRA, nomeadamente que incluam estudos que impliquem a tradução e adaptação de testes para fins de investigação ou simplesmente a utilização de testes em contextos de investigação.

A preparação deste documento baseou-se essencialmente, mas não exclusivamente, em dois documentos publicados pela *International Test Commission* (ITC) (<https://www.intestcom.org>), que inclui a *European Association of Psychological Assessment* (EAPA) (<https://www.eapa.science>), cujas referências são as seguintes:

- *International Test Commission. (2017). The ITC Guidelines for Translating and Adapting Tests (Second edition). Document reference: ITC-G-TU-20131008*
- *International Test Commission. (2014). The ITC Statement On the Use of Tests and Other Assessment Instruments for Research Purposes. Document reference: ITC-S-TU-20140410*

Disponíveis aqui: www.InTestCom.org

Recomenda-se a leitura prévia do documento “Diretrizes Internacionais para a Utilização de Testes - Versão portuguesa”, cuja referência é a seguinte:

- *International Test Commission. (2003). Diretrizes internacionais para a utilização de testes—versão portuguesa. Cegoc-Tea: Lisboa.*

Disponível aqui: www.InTestCom.org

Aconselha-se também que os investigadores se familiarizem com conceitos básicos da estratégia psicométrica. Existe muita literatura científica sobre este tópico. Como exemplo de referência sintética e prática, escrita em língua portuguesa, por investigadores portugueses, com experiência no desenvolvimento, adaptação e validação de testes para fins de investigação, nomeadamente em áreas da saúde, sugere-se a consulta da seguinte publicação:

- Pereira A, Marques C (2018). Fundamentos da psicometria. In, A Macedo, AT Pereira & Madeira N. (Coords.) *Psicologia na Medicina*. pp. 151-178, Lidel.

PRIMEIRA PARTE: RECOMENDAÇÕES PARA A UTILIZAÇÃO DE TESTES PSICOLÓGICOS EM INVESTIGAÇÃO

Testes e outros instrumentos de avaliação (por exemplo, testes de aptidão, guiões de entrevistas) geralmente são desenvolvidos para contextos de investigação e também para diversos contextos da prática profissional (por exemplo, na clínica ou na educação).

Em atividades relacionadas com a investigação, os dados dos testes são usados especificamente para avaliar/medir/quantificar fenómenos (por exemplo, as características dos testes ou as relações entre os resultados dos testes, ou seja, entre diversas variáveis) com o objetivo de avançar o conhecimento científico.

Em contraste, em práticas institucionais e clínicas, os dados dos testes são usados especificamente para avaliação e tomada de decisão que podem impactar diretamente os indivíduos a quem os testes são aplicados.

Este documento pretende promover a convicção de que o uso de testes em investigação deve respeitar princípios legais e éticos, bem como padrões de boas práticas. Os testes devem ser utilizados por profissionais competentes que foquem a sua atenção nos direitos dos indivíduos nos quais os testes são aplicados e nos direitos das outras partes envolvidas no processo de testagem.

Antes de avançarmos, apresentamos uma **definição de teste psicológico**:

Podemos definir testes (psicológicos) como procedimentos sistemáticos e estandardizados de observação e registo de amostras de comportamentos ou de funções cognitivas de indivíduos com o objetivo de descrever e/ou mensurar de forma válida e fidedigna traços, estados e processos psicológicos, em áreas como emoção, cognição, motivação, personalidade, psicomotricidade, atenção, memória, perceção, entre outras. Para isso, os testes compreendem um conjunto de itens (por exemplo, questões, afirmações, estímulos ou tarefas) que são corrigidos de uma maneira padronizada e usados para quantificar diferenças individuais (Anastasi, 2011).

Esta definição inclui testes psicológicos e educacionais em todas as formas de apresentação (por exemplo, “cadernos” para resposta com “papel e lápis”, testes para realizar num programa de computador ou numa aplicação para *smartphone*, jogos interativos...).

Devemos começar por distinguir dois tipos de testes: a) os que possuem direitos de autor, e que são, portanto, do domínio privado; b) os que não possuem direitos de autor, e que são, portanto, do domínio público. Não obstante esta diferença, muitos dos princípios a discutir de seguida aplicam-se a ambos os tipos. Atualmente, a maioria dos testes são do domínio público, ou seja, de acesso livre, mas mesmo assim, o seu uso requer um conjunto de boas práticas.

De seguida, iremos abordar sete tópicos relativos ao uso dos testes em contextos de investigação:

1. Permissão para o uso de testes em investigação
2. Permissão para impressão e reimpressão

3. Modificação do teste ou dos seus componentes
4. Uso ético dos testes
5. Documentação
6. Conflitos de interesse
7. Divulgação

1. Permissão para o uso de testes em investigação

1.1. Quando os testes são protegidos por direitos de autor apresentam restrições impostas por aqueles que detêm os direitos ou que podem usar os materiais ou instrumentos.

Normalmente, os testes são desenvolvidos por um ou mais investigadores/autores, por agências públicas ou privadas ou ainda por empresas especializadas no desenvolvimento de testes. Uma vez desenvolvidos, o autor automaticamente detém a propriedade dos direitos de autor, ou seja, a sua proteção legal. O direito de autor protege o título do teste, as instruções, os itens, a estrutura, os manuais, as tabelas de pontuação e, por vezes, outros equipamentos e componentes. Informações sobre quem detém os direitos de autor geralmente encontram-se na primeira página do manual do teste, nos formulários ou nas páginas de internet (websites) do teste.

O autor ou autores do teste detém os direitos autorais, a menos que os transfira por meio de um acordo escrito para outra pessoa ou entidade (por exemplo, editoras que publicam os testes).

Os autores de testes, cujas publicações em revistas incluem a totalidade do instrumento (uma ocorrência rara) ou itens do teste, renunciam aos direitos autorais destes materiais à revista, a não ser que um acordo por escrito estipule que o autor mantém a sua propriedade. Editoras de testes e empresas de desenvolvimento de testes retêm os direitos autorais dos testes que desenvolvem.

1.2. Profissionais responsáveis por conduzir ou supervisionar projetos de investigação que utilizam testes devem obter permissão dos detentores dos direitos de autor antes de fazerem uso daqueles. Esta prerrogativa é válida tanto para os testes originais quanto para qualquer trabalho que derive deles (por exemplo, a versão traduzida de um teste). Alguns proprietários de direitos de autor (por exemplo, especialistas) podem conceder permissão sem custos. Outros proprietários de direitos de autor (por exemplo, editores de testes) normalmente estabelecem um acordo contratual com o investigador que define os termos (por exemplo, o uso do teste por um período limitado e/ou por um número limitado de aplicações) e condições (por exemplo, fornecer ao proprietário do direito de autor uma cópia de todas as publicações nas quais dados referentes às suas produções são relatados). O investigador deve garantir que o acordo contratual não comprometa as regras de privacidade dos dados, outras leis, bem como as normas impostas pelas agências. O investigador deve respeitar o acordo contratual.

1.3. O detentor dos direitos de autor tem o direito de recusar a permissão para o uso livre de um teste para investigação. Este direito aplica-se tanto para as versões originais como para as solicitações de traduções. O detentor dos direitos de autor não deve estar em posição de determinar a natureza da investigação feita com suas medidas.

1.4. Quando os testes estão no domínio público, ou seja, são de acesso livre, podem ser utilizados por qualquer pessoa; são considerados propriedades comunitárias e podem ser

utilizados sem a permissão do autor. Neste caso, os materiais dos testes podem ser reproduzidos, distribuídos, exibidos publicamente ou transformados em atividades derivadas. Os investigadores não precisam de obter permissão para o uso de testes que não possuem direito de autor. No entanto, a boa prática dita que os investigadores informem os autores dos testes, quando possível, acerca do seu uso, bem como que referenciem as fontes bibliográficas adequadas.

1.5. Estudos recentes derivados de instrumentos são frequentemente publicadas em revistas *online* ou em *websites* dos autores. Quando publicados em fontes de acesso aberto, os testes podem dar a impressão de estar no âmbito do domínio público. No entanto, regra geral, o autor é o detentor do direito de autor, a menos que o atribua a outra entidade, ou dê permissão explícita para acesso livre e aberto. Daí que seja altamente recomendável que os investigadores consultem os autores quando não estão totalmente seguros do *status* de direito autoral de um teste. Por outras palavras, o facto de um teste estar facilmente acessível numa publicação ou num *website*, não significa que este seja de livre acesso.

2. Permissão para impressão e reimpressão

2.1. Trabalhos com direito de autor não devem ser reproduzidos, distribuídos, exibidos publicamente ou transformados em outros trabalhos decorrentes destes sem a permissão do detentor do direito de autor. A segurança do teste deve ser garantida. Os itens dos testes não devem ser impressos em publicações, uma vez que isso os tornaria disponíveis para utilizadores não autorizados e colocaria em risco a sua segurança. Por exemplo, a impressão do teste original, bem como de qualquer versão adaptada ou traduzida dela ou dos seus itens numa publicação (por exemplo, artigo de revista, dissertação ou site da *web*) é proibida.

2.2. Ao relatarem os seus resultados, pode ser solicitado aos investigadores que forneçam itens do teste que exemplifiquem as qualidades avaliadas. Quando os autores fizerem uso de testes de domínio público, podem fornecer itens para exemplificação e citar as suas fontes. Tratando-se de testes com direito de autor protegido, deve ser obtida permissão do proprietário dos direitos de autor para reproduzir um número limitado de itens (geralmente um ou dois por escala).

3. Modificação do teste ou dos seus componentes

3.1. Quando os investigadores usam testes que são protegidos por direitos de autor, a menos que estejam autorizados, não deve modificar nenhum dos seus componentes, pois isso colocaria em risco a integridade do instrumento, violaria os direitos de autor e seria ilegal.

3.2. Existem diferentes normas para a utilização de testes de domínio público. Ao utilizá-los, os investigadores podem modificá-los adicionando ou excluindo itens, bem como alterando a sua escrita ou ordem, instruções de aplicação ou tabelas de correção, a fim de torná-los mais adequados aos propósitos do estudo, aos grupos populacionais que responderão ou às condições contextuais. Essas versões modificadas dos testes são consideradas atividades derivadas e permitidas para testes de domínio público. Por exemplo, uma adaptação de um teste de domínio público para uso noutra cultura,

diferente daquela para a qual foi originalmente desenvolvido, pode exigir uma tradução, exclusão, reescrita ou acréscimo de itens e/ou outras mudanças nas instruções e no dimensionamento dos itens.

3.3. Em testes protegidos por direitos de autor, alterações deste tipo só podem ser feitas com a permissão escrita e expressa dos proprietários dos direitos de autor. Qualquer modificação da forma original do teste deve ser documentada, como destacado no ponto (5) abaixo.

4. Questões éticas a considerar no uso de testes

4.1. A ética para os usos de testes em investigação é normalmente coerente com o seu uso em contextos profissionais. Por isso, espera-se que os profissionais envolvidos em projetos de investigação ajam num escrupuloso cumprimento das boas práticas profissionais e normas éticas aplicáveis à investigação, demonstrem competência no uso de testes, sejam responsáveis no momento da sua utilização, garantam a segurança dos materiais e assegurem que os resultados sejam devidamente protegidos no seu tratamento e armazenamento.

4.2. Os utilizadores de testes devem cumprir com as suas atribuições éticas. Tais obrigações incluem a necessidade de obter consentimento informado, defender a segurança dos instrumentos utilizados e garantir a proteção dos dados pessoais. Um documento fornecido aos participantes da investigação deve informá-los sobre o(s) investigador(es), o objetivo e plano de estudo e os direitos do participante. A obrigação de oferecer um *feedback* aos participantes da pesquisa sobre seus resultados pode ser impraticável ou irrealizável e, por isso, pode não ser garantida.

Os investigadores devem também seguir os princípios éticos e legais da região ou país em que o seu projeto de investigação está a ser conduzido e demonstrar sensibilidade para com os aspetos culturais da população alvo do seu estudo.

.

A pessoa a quem é dada a permissão para o uso do teste em investigação deve ser qualificada para o usar, ou, pelo menos, deve ser supervisionada por um profissional qualificado (por exemplo, um colega, um professor ou um orientador). A responsabilidade do profissional qualificado estende-se ao processo de testagem mais amplo, incluindo a recolha de dados, a codificação, a análise e o relatório.

4. Documentação

A documentação permite à comunidade científica avaliar a qualidade e os resultados da investigação e facilita a sua replicação. Assim, o uso de materiais e métodos de investigação, incluindo o uso de testes, deve ser documentado na seção de métodos de todos os relatórios de investigação. A documentação deve fazer menção ao nome do teste, ao autor ou autores e à data da publicação do instrumento original e/ou adaptado.

Adicionalmente, se for apropriado, o investigador deve documentar se a permissão para o uso do teste foi fornecida pelo detentor do direito de autor.

Investigadores que traduzem e/ou adaptam um teste devem declarar a maneira como estas modificações foram feitas e fornecer evidências de validade, fidelidade e equivalência das formas traduzidas ou adaptadas (ITC, 2017).

Investigadores que desenvolvem uma nova medida devem fornecer, no mínimo, as seguintes informações sobre a mesma: a base teórica e o propósito do teste, a forma como os itens iniciais foram selecionados e depois analisados, o número de itens por domínio e pelo menos um exemplo de cada domínio, os métodos de escalonamento usados, as evidências de validade e precisão. Todas as publicações que utilizam dados de testes recentemente desenvolvidos devem indicar se o investigador oferece a possibilidade do envio de cópias das medidas a outros investigadores.

Exemplo: Almeida e Freire (2008) indicam quais os parâmetros e análises que devem estar concluídos antes da utilização de instrumento para avaliação psicológica, tanto em contextos de investigação como de prática profissional. A tabela 1 lista esses parâmetros e exemplifica cada um deles, recorrendo a um questionário traduzido e validado para a população portuguesa, o Questionário de Pensamento Perseverativo (QPP; Chaves et al. 2013; Azevedo et al. 2017).

Tabela 1: Parâmetros e análises que devem estar concluídos antes da utilização de um teste para avaliação psicológica [O caso do Questionário de Pensamento Perseverativo (Chaves et al. 2013; Azevedo et al. 2017)] – Adaptado de Pereira et al. 2018.

O que é que a escala mede Pensamento repetitivo negativo enquanto processo transdiagnóstico.
Enquadramento teórico subjacente O pensamento repetitivo negativo é um estilo de pensamento sobre os próprios problemas (atuais, passados ou futuros) ou experiências negativas (passadas ou antecipadas), que ocorre em diversas perturbações afetivas, como as depressivas e da ansiedade – processo transdiagnóstico. Nestas apresenta as mesmas características processuais (repetitivo; intrusivo; difícil desligar dele; improdutivo e captura a capacidade mental), sendo que apenas o conteúdo é específico das perturbações.
Justificação da escala e a sua utilidade prática Não havia nenhuma escala cujos itens não fizessem referência aos conteúdos específicos de cada perturbação, pelo que os que existiam apenas podiam ser utilizados com amostra de pessoas afetadas por essa perturbação.
Como é que o primeiro conjunto de itens foi definido Os autores da versão original desenvolveram três itens relativos a cada uma das características processuais, partindo dos itens de outros questionários de ruminação e de preocupação, transformando-os, nomeadamente pela exclusão de conteúdos que remetiam diretamente para as perturbações.
Tipo de itens e escala de resposta Os itens são afirmações às quais se responde numa escala de <i>Likert</i> clássica, com cinco pontos, desde “nunca” (0) a “sempre” (4).
Tradução e retroversão

<p>Primeiro, psicólogos e psiquiatras (sete ao todo) com bom domínio da língua inglesa e com experiência clínica com doentes com perturbações depressivas e da ansiedade, traduziram independentemente as instruções e os itens da escala; depois discutiram em conjunto as respetivas traduções para chegarem a uma tradução de consenso, que fosse fiel à versão original e facilmente compreensível (português coloquial); por fim, uma psicóloga bilingue voltou a traduzir esta versão quase final para inglês, verificando-se total sobreposição com a versão original.</p>
<p>Descrição das amostras usadas para testar a escala</p> <p>Amostra 1 1º estudo psicométrico (Fidelidade, AFE, validade concorrente): 344 estudantes (68.4% sexo feminino) da FMUC, com idade média de 20.69 anos (DP=1.586; variação: 17-24), não se distinguindo significativamente por género. Um grupo de 129 estudantes (65.1% do sexo feminino; idade média=21.49±.821 anos) voltaram a preencher alguns questionários após quatro a seis semanas, de modo a estudar a estabilidade temporal dos mesmos.</p> <p>Amostras 2 e 3 2º estudo psicométrico (AFC): 256 estudantes de diversas licenciaturas (78.1% sexo feminino), com idade média de 20.58 ±1.870 anos; 480 adultos da comunidade (pais dos estudantes); 53.1% do sexo feminino; com a idade média de 50.84 ±5.310.</p>
<p>Estatísticas descritivas dos resultados nos itens</p> <p>Todos os itens podem ser considerados “bons” itens, pois todos se correlacionam acima de .20 com o total (quando este não contém o item) (e.g. Pasquali, 2003), com coeficientes a variarem de .50 (item 14 “Os meus pensamentos não me ajudam muito”) a .73 (item 11 “Estou continuamente a pensar na mesma coisa”). Além disso, todos os itens contribuem para a consistência interna, ou seja, se retirados, fariam diminuir o α global.</p>
<p>Fidelidade da escala</p> <p>Consistência interna: O coeficiente α de foi de .93, o qual pode ser considerado “muito bom” (DeVellis, 1998), apontando para a uniformidade e coerência entre as respostas dos sujeitos a cada um dos itens.</p> <p>Estabilidade temporal: O coeficiente de correlação entre as pontuações totais no teste e no reteste, de .79 ($p < .001$), foi abonatório deste parâmetro.</p>
<p>Validade da escala</p> <p>Validade de constructo: Para a extração dos factores guiámo-nos pelos Critério de Kaiser (1958) e do <i>scree test</i> de Cattell (Cattell, 1966). Seguindo a sugestão de Kline (2000), consideramos que os “pesos” (<i>loadings</i>) com valores $\geq .30$ estão associados com o fator. Antes de prosseguirmos com a análise fatorial, realizámos o teste KMO (<i>Kaiser-Meyer-Olkin</i>) e o teste de esfericidade de <i>Bartlett</i>, que são dois procedimentos estatísticos que permitem aferir a qualidade das correlações entre as variáveis, de forma a sabermos se é plausível realizar a análise fatorial (Pestana & Gageiro, 2003). Para o podermos fazer com segurança, o primeiro deve aproximar-se da unidade, sendo “bom” se for $> .80$, o que se verificou com a nossa amostra, em que KMO foi igual a .93; o segundo deve levar a rejeitar a hipótese nula, o que também aconteceu com os nossos dados ($p < .001$). Perante a observação do <i>scree plot</i> e a interpretabilidade dos factores, considerámos que a estrutura de dois factores seria efetivamente a mais compreensível. Os factores 1 e 2 explicam respetivamente 49.48% e 10.48% desta. Considerando o conteúdo dos itens que apresentam o seu peso máximo em cada um dos factores, estes foram denominados F1-Pensamento repetitivo (7 itens; $\alpha = .90$) e F2-Interferência e improdutividade (8 itens; $\alpha = .87$). As correlações de <i>Pearson</i> entre as pontuações fatoriais e a pontuação total do QPP-15 foram elevadas e significativas ($p < .001$): entre o total e o F1 o coeficiente foi de .91 e entre o total e o F2 foi de .92; entre os dois factores foi de .67.</p> <p>Validade de critério (concorrente): Os coeficientes de correlação de <i>Pearson</i> entre as pontuações (total e dimensionais/fatoriais) no QPP e no Questionário de Pensamento Repetitivo-10 (Macedo et al. 2014) foram significativos, positivos e elevados, todos $> .60$ ($p < .001$).</p>
<p>Procedimentos de cálculo e interpretação de resultados obtidos</p> <p>A média e o desvio padrão foram os seguintes:</p>

Amostra 1: 22.93±80890

Amostra 2: 23.01±9.012

Amostra 3: 23.71±8.882

O teste estatístico One Way ANOVA revelou que estes valores não se distinguem significativamente entre si.

6. Conflitos de interesse

Os investigadores devem identificar as fontes de financiamentos, se houver, que permitiram o desenvolvimento do teste ou da investigação na qual o teste foi usado. Investigações financiadas por agências externas podem ter um interesse investido nos seus resultados (por exemplo, uma empresa de testes que tem a propriedade dos direitos de autor do teste que financia ou patrocina o projeto de investigação). Os investigadores devem estar cientes de tais situações e manter a sua dedicação em prol da objetividade e imparcialidade. Os investigadores devem também declarar relações que possam sugerir conflitos de interesse (por exemplo, quando são pagos por um editor de testes).

7. Divulgação na prática profissional

Na prática profissional onde os testes são utilizados para avaliar uma característica ou fornecer uma classificação, os testes precisam de ser fidedignos e válidos ao nível individual, de forma a garantir uma estimativa correta da característica alvo ou uma classificação correta da mesma. Em contextos de investigação, os testes nem sempre precisam de satisfazer este requisito.

Há diversas razões para a flexibilização em relação aos critérios técnicos para o uso de testes em investigação. Primeiro, os testes podem ser desenvolvidos e usados para testar hipóteses inovadoras e construtos que ainda não tenham sido suficientemente estudados ao ponto de garantir a obtenção de modelos de medidas robustos. Segundo, os testes usados em contextos de investigação podem não ter (ainda) as suas propriedades psicométricas e estatísticas populacionais (normas) conhecidas antes da investigação. Terceiro, os dados do teste podem ser usados apenas a um nível grupal e não para definir conclusões sobre indivíduos a quem os testes foram aplicados.

No entanto, versões de investigação de testes são por vezes usadas na prática profissional sem um reconhecimento das suas limitações, o que pode tornar-se problemático. O autor do teste tem a responsabilidade de não permitir o uso de instrumentos desenvolvidos para fins de investigação para usos profissionais, antes de dispor de evidências suficientes acerca das suas propriedades psicométricas.

SEGUNDA PARTE:

RECOMENDAÇÕES PARA A TRADUÇÃO E ADAPTAÇÃO DE TESTES

A segunda edição das Diretrizes do ITC para a Tradução e Adaptação de Testes, publicada em 2017, compreende 18 pontos, organizados em seis categorias:

1. Pré-condições
2. Desenvolvimento do teste
3. Confirmação
4. Administração
5. Cotação/pontuação e interpretação
6. Documentação

Para cada diretriz, é fornecida uma explicação e algumas sugestões para prática.

Em termos gerais, a primeira seção, “Pré-condições”, destaca as que decisões devem ser tomadas antes do início do processo de tradução/adaptação. A segunda seção “Desenvolvimento do teste” foca o processo de adaptação de um teste. A terceira seção “Confirmação” inclui as diretrizes associadas à compilação de evidências empíricas para abordar a equivalência, fidelidade e validade de um teste em vários idiomas e culturas. As três seções finais - “Administração”, “Cotação/pontuação e interpretação” e “Documentação” tendem a ser mais negligenciadas nas iniciativas de adaptação de testes psicológicos; recomenda-se maior exigência da parte das editoras de publicações científicas e das agências de financiamento quando se trata de documentação relativa a trabalhos de adaptação de testes.

De seguida apresentamos um resumo para cada uma destas categorias, contemplando, quando necessária uma breve explicação e/ou algumas breves sugestões para implementar as recomendações na prática.

1. Pré-condições:

- 1.1. Obter autorização do titular dos direitos de propriedade intelectual relativos ao teste (ou seja, o autor ou o editor) antes de realizar qualquer adaptação. Antes de iniciar a adaptação de um teste, deve ser estabelecido um acordo, que deve especificar as modificações no teste adaptado, que serão aceitáveis em relação às características do teste original; neste caso, o acordo deve clarificar quem deterá os direitos de propriedade intelectual da versão adaptada.
- 1.2. Avaliar se a quantidade de sobreposição no conteúdo do construto medido pelo teste nos grupos populacionais de interesse é suficiente para o uso pretendido.

Sugestões para a prática: Indivíduos especialistas no construto medido e familiarizados com os grupos culturais a testar devem ser recrutados para avaliar a legitimidade do construto medido em cada um dos grupos culturais/linguísticos. Estes devem tentar responder à seguinte questão: o construto faz

sentido nas culturas de ambos os grupos? Métodos como grupos focais e entrevistas podem ser usados para obter informações estruturadas sobre o grau de sobreposição entre construtos.

1.3. Minimizar a influência de quaisquer diferenças culturais e linguísticas que sejam irrelevantes para os usos pretendidos do teste nas populações de interesse.

Explicação: As características culturais e linguísticas irrelevantes para as variáveis que o teste pretende medir devem ser identificadas na fase inicial do projeto. Estas podem estar relacionados com o formato dos itens, o material (por exemplo, uso de computador, imagens ou ideogramas...), os limites de tempo, etc..

Sugestões para a prática: Podem ser recolhidas algumas evidências qualitativas, nomeadamente através de entrevistas a realizar através da estratégia de reflexão falada e grupos focais de potenciais destinatários, para tentar determinar níveis motivacionais dos participantes, compreensão das instruções, experiência prévia com testes psicológicos, rapidez associada à administração do teste, familiaridade com as escalas de resposta e diferenças culturais.

2. Desenvolvimento do teste

2.1. Assegurar que os processos de tradução e adaptação consideram as diferenças linguísticas, psicológicas e culturais nas populações pretendidas através da seleção de especialistas com experiência relevante.

Explicação: Além do bom conhecimento dos dois idiomas envolvidos na adaptação do teste (ver, por exemplo, Grisay, 2003), importa que os tradutores possuam bom conhecimento das culturas, e pelo menos, conhecimento geral do assunto do teste. Importa também que estejam envolvidos pelo menos dois tradutores em cada projeto de tradução. A prática de confiar em apenas um tradutor, por mais qualificado que seja, não é aceite atualmente.

A garantia de conhecimento/experiência na cultura-alvo resulta do recrutamento de tradutores nativos da língua-alvo e residentes na região/ cultura-alvo, sendo o primeiro requisito essencial e o segundo altamente desejável. Um falante nativo do idioma de destino produzirá uma tradução precisa e atualizada.

A definição de "especialista" requer, assim, que as pessoas da equipa de tradução tenham conhecimentos suficientes: dos (1) idiomas envolvidos, das (2) culturas, dos (3) conteúdos e princípios gerais do teste (por exemplo, que não existem respostas certas nem erradas; que a rapidez de resposta aumenta a pontuação, etc.), para produzir uma tradução/ adaptação de qualidade profissional.

2.2. Usar procedimentos de tradução apropriados para maximizar a adequação da adaptação do teste nas populações pretendidas.

Sugestões para a prática: Um procedimento de dupla tradução e reconciliação permite diminuir as deficiências ou incongruências e os riscos de confiar nas idiossincrasias das traduções simples. Nesta abordagem, um terceiro tradutor independente ou um painel de especialistas identifica e resolve quaisquer discrepâncias entre traduções alternativas e reconcilia-as numa única versão.

2.3. Fornecer evidências de que as instruções do teste e o conteúdo do item têm significado semelhante e são compreendidas de igual modo para todas as populações pretendidas.

Sugestões para a prática: Conduzir um pequeno ensaio da versão adaptada do teste é uma boa estratégia. Neste ensaio, pode-se administrar o teste e solicitar que tanto os examinadores como os examinados possam partilhar as suas impressões (dificuldades, críticas, sugestões). Quando se trata de testes com resposta certas e erradas e, portanto, com determinados índices de dificuldade, pode ser muito útil solicitar o apoio de especialistas bilingues, capazes de avaliar a similaridade da dificuldade dos formatos dos itens e do conteúdo dos mesmos (original e o traduzido). Note-se que tais análises qualitativas não dispensam as análises quantitativas adequadas para esse efeito e a efectuar em fases subsequentes.

2.4. Fornecer evidências de que os formatos dos itens, escalas de resposta, categorias de pontuação, convenções, modos de administração e outros procedimentos são adequados para todas as populações pretendidas.

Explicação: Alguns formatos de itens podem ser confusos para pessoas não familiarizadas com este tipo de tarefas, por exemplo, escalas de cinco pontos ou novos formatos de itens, como "arrastar e soltar" ou "responder a todas as respostas corretas" ou até mesmo "responder uma e apenas uma opção de resposta". Se os respondentes não forem treinados nestes novos formatos de itens, o avaliador estará a introduzir uma fonte de viés que pode distorcer qualquer resultado do teste.

Um problema emergente remete para a administração de testes por computador. Se os participantes não estiverem familiarizados com a plataforma de teste, torna-se necessário preparar um tutorial que garanta que obtenham a familiaridade necessária para que um teste administrado por computador forneça pontuações válidas e fidedignas.

2.5. Recolher dados do teste adaptado, no âmbito de um projecto-piloto, para permitir análises de itens, de confiabilidade e de validade em pequena escala, para que quaisquer revisões necessárias possam ser feitas, antes da realização de estudos psicométricos em grande escala.

Explicação: Antes de iniciar um estudo psicométrico de larga escala, que exige mais recursos de tempo e dinheiro, de fidelidade e validade das pontuações de teste e/ou de aferição/normalização, é importante ter evidências preliminares acerca da a qualidade psicométrica do teste adaptado. Na fase de

desenvolvimento do teste, proceder a análises preliminares, com um tamanho de amostra modesto (por exemplo, 100) pode fornecer informação útil sobre o funcionamento dos itens de teste específicos. Por exemplo, itens que apresentam poder discriminativo baixo ou mesmo negativo, podem ser revistos ou retirados e pode-se analisar o modo como afetam o coeficiente alfa de consistência interna.

3. Confirmação

As seguintes recomendações dizem respeito às análises empíricas decorrentes de estudos de fidelidade e validade a realizar com amostras representativas do grupo populacional ao qual se pretende aplicar o teste.

3.1. Selecionar amostras representativas do grupo populacional ao qual se pretende aplicar o teste e de tamanho e relevância suficientes para as análises empíricas tenham poder estatístico.

Explicação: Embora este requisito seja válido para qualquer tipo de investigação científica, é particularmente relevante no contexto de um estudo de validação de instrumento psicométrico, porque as técnicas estatísticas necessárias para estabelecer a equivalência entre duas medidas (por exemplo, análise fatorial confirmatória) podem ser aplicadas de forma mais significativa se as amostras forem suficientemente grandes para estimar de forma confiável os parâmetros dos modelos de medida (o tamanho da amostra recomendado dependerá da complexidade do modelo e da qualidade dos dados).

Além disso, só com amostras representativas será válida a comparação dos resultados de pessoas ou grupos com as normas obtidas nos estudos de validação e aferição.

Sugestões para a prática: Recolher uma amostra tão grande quanto possível ou razoável; considera-se que estudos para identificar itens de teste potencialmente tendenciosos exigem um mínimo de 200 respondentes por versão do teste (Subok, 2017); estudos para investigar a estrutura fatorial de um teste requerem tamanhos de amostra de 300 ou mais respondentes (Wolf, Harrington, Clark & Miller, 2013); uma regra consensualmente aceita para este tipo de análise é de um tamanho de amostra que seja equivalente a 10 respondentes por cada variável/ item do teste (Marôco 2010); por exemplo, se um questionário tem 21 itens, será necessário recolher dados de uma amostra de 210 respondentes para analisar a sua estrutura factorial. Análises com amostras menores também são aceitáveis – mas a primeira regra é gerar grandes amostras sempre que possível.

3.2. Fornecer evidências estatísticas relevantes sobre a equivalência do construto e dos itens para todas as populações pretendidas.

De acordo com van de Vijver e Poortinga (1991), a análise fatorial (tanto AFE quanto AFC) é a técnica estatística mais utilizada e recomendada para avaliar se um construto encontrado na mesma forma e frequência em culturas diferentes. Embora as abordagens de modelagem estatística tenham avançado consideravelmente na última década, esta afirmação de 1991 permanece

verdadeira nos dias de hoje (ver, por exemplo, Hambleton & Lee, 2013). Uma vez que, com AFE, é difícil comparar estruturas fatoriais separadas, e não há regras comumente acordadas para decidir quando as estruturas podem ser consideradas equivalentes, abordagens estatísticas como AFC (ver, por exemplo, Byrne 2008) são mais desejáveis.

Sugestões para a prática: Procurar evidências adicionais de validade concorrente, preditiva, convergente-divergente e discriminante, para além de evidências de validade de construto, é uma boa prática que pode fornecer conhecimento adicional acerca da validade do teste na população à qual o mesmo se pretende aplicar.

3.3. Fornecer evidências que sustentem as normas, a fidelidade e a validade da versão adaptada do teste nas populações pretendidas.

Explicação: As normas e as evidências de fidelidade e de validade e evidência de um teste na sua versão original e no seu idioma de origem não se aplicam automaticamente a outras possíveis adaptações do teste em diferentes culturas e idiomas. Portanto, evidências de validade empírica de quaisquer novas versões desenvolvidas também devem ser apresentadas.

4. Administração

4.1. Preparar materiais e instruções de administração que minimizem quaisquer problemas relacionados com a cultura e o idioma que possam afetar a validade das inferências extraídas das pontuações.

Sugestões para a prática: É importante antecipar fatores que possam criar problemas na administração do teste, tais como: falta de clareza nas instruções do teste, uso da folha de respostas, tempo permitido, grau de conhecimento sobre o propósito do teste e como será pontuado.

4.2. Especificar as condições de teste que devem ser seguidas em todas as populações de interesse, ou seja, garantir a standardização do teste.

Explicação: Quaisquer variações das condições padronizadas de teste devem ser observadas, para que, posteriormente, essas variações e o seu impacto nas generalizações e interpretações possam ser consideradas.

5. Cotação/pontuação e interpretação

5.1. Interpretar quaisquer pontuações somente com referência a normas ou outras informações relevantes disponíveis.

Explicação: Mesmo que um teste tenha sido adaptado por meio de procedimentos tecnicamente sólidos e a validade das pontuações do teste já tenha sido estabelecida até certo ponto, deve-se ter em mente que o significado das pontuações individuais ou grupais pode ser interpretado de muitas maneiras,

devido a fatores individuais e culturais ou outras diferenças entre os países e/ou culturas participantes.

5.2. Comparar pontuações entre populações somente quando o nível de invariância tiver sido estabelecido na escala na qual as pontuações são relatadas.

Sugestões para a prática: o ponto-chave aqui é que as pontuações dos testes não devem ser sobre-interpretadas:

6. Documentação

6.1. Fornecer documentação técnica de quaisquer alterações, incluindo um relato das evidências obtidas para apoiar a equivalência, quando um teste é adaptado para uso em outra população.

A documentação técnica também deve conter detalhes suficientes da metodologia, de modo a que possa ser replicada em investigações futuras, na mesma ou em outras populações. Deve conter informações suficientes da evidência de equivalência de construto para apoiar o uso do instrumento na nova população.

Sugestões para a prática: Os testes adaptados devem dispor de um manual técnico que documente todas as evidências qualitativas e quantitativas associadas ao processo de tradução, adaptação e validação. É especialmente útil documentar quaisquer alterações feitas para adaptar o teste a outro idioma e cultura. Técnicos e editores devem exigir informação acerca do processo que foi seguido para desenvolver e validar a versão do teste no idioma e cultura de destino, bem como acerca dos resultados de todas as análises. As perguntas que precisam de ser respondidas no manual ou no trabalho científico de divulgação do trabalho de tradução, adaptação e validação do teste são as seguintes: que evidências estão disponíveis para apoiar a utilidade, validade e fidelidade do teste adaptado na nova população? que dados de itens foram recolhidos e de que amostras? que outros dados foram obtidos para avaliar a validade de conteúdo, de critério e de construto? como é que os vários conjuntos de dados foram analisados? quais foram os resultados? que conclusões podem ser retiradas?

6.2. Fornecer documentação para utilizadores do teste que sustente boas práticas no uso do teste adaptado no contexto da nova população.

Explicação: A documentação deve ser escrita para pessoas que usarão o teste em ambientes de avaliação prática. Deve ser consistente com a boa prática definida pelas Diretrizes da Comissão Internacional de Testes sobre Uso de Testes (consulte www.InTestCom.org).

Sugestões para a prática: O manual do usuário deve: descrever o(s) constructo(s) medido(s) pelo teste; descrever o processo de adaptação; resumir as evidências que apoiam a adaptação, incluindo evidências de adequação cultural do conteúdo dos itens, instruções do teste, formatos de resposta, etc; informar acerca da adequação de usar o teste com vários subgrupos dentro da população e acerca de quaisquer outras restrições de uso; explicar quaisquer que

devam ser consideradas em relação às boas práticas na administração do teste; explicar se e como as comparações inter-populacionais podem ser feitas; fornecer informações sobre as implicações da validade e fidelidade dos dados nas inferências que podem ser extraídas dos resultados dos testes; fornecer informações necessárias para obter e interpretar pontuações e cotações (por exemplo, tabelas de consulta de normas por grupos etários) ou descrever como os utilizadores podem aceder aos procedimentos de pontuação (por exemplo, num sítio da internet).

RESUMO

Sobre o uso dos testes em contextos de investigação:

- Mesmo quando um teste está no domínio público, ou seja, é de acesso livre, os investigadores devem sempre informar os autores dos testes acerca do seu uso, bem como que referenciar as fontes bibliográficas adequadas;
- Qualquer modificação da forma original do teste deve ser documentada;
- Quando os investigadores usam testes que são protegidos por direitos de autor, a menos que estejam autorizados, não deve modificar nenhum dos seus componentes;
- Os profissionais envolvidos em projetos de investigação devem agir num escrupuloso cumprimento das boas práticas profissionais e normas éticas aplicáveis à investigação, nomeadamente: competência no uso de testes e responsabilidade na sua utilização; obtenção de consentimento informado e garantia de protecção no seu tratamento e armazenamento.
- O uso de testes deve ser documentado na seção de métodos de todos os projectos e relatórios de investigação (nome do teste, autor(es), data da publicação do instrumento original e/ou adaptado), bem como que a permissão para o seu uso foi fornecida pelo detentor do direito de autor.
- Investigadores que traduzem e/ou adaptam um teste devem declarar a maneira como estas modificações foram feitas e fornecer evidências de validade, fidelidade e equivalência das formas traduzidas ou adaptadas.
- O autor do teste tem a responsabilidade de não permitir o uso de instrumentos desenvolvidos para fins de investigação para usos profissionais, antes de dispor de evidências suficientes acerca das suas propriedades psicométricas.

Sobre a tradução e adaptação de testes:

- Obter autorização do titular dos direitos de propriedade intelectual relativos ao teste (ou seja, o autor ou o editor) antes de realizar qualquer tradução e/ou adaptação;
- Avaliar se a sobreposição no conteúdo do teste é suficiente para o uso pretendido (estudos qualitativos);
- Minimizar a influência de quaisquer diferenças culturais e linguísticas que sejam irrelevantes para os usos pretendidos do teste nas populações de interesse;
- Assegurar que os processos de tradução e adaptação consideram as diferenças linguísticas, psicológicas e culturais nas populações pretendidas através da seleção de especialistas com experiência relevante.
- Usar procedimentos de tradução apropriados para maximizar a adequação da adaptação do teste nas populações pretendidas.

- Fornecer evidências de que as instruções do teste e o conteúdo dos itens têm significado semelhante e são compreendidas de igual modo para todas as populações pretendidas e de que os formatos dos itens, escalas de resposta, categorias de pontuação, convenções, modos de administração e outros procedimentos são adequados para todas as populações pretendidas.
 - Selecionar amostras representativas do grupo populacional ao qual se pretende aplicar o teste e de tamanho e relevância suficientes para as análises empíricas tenham poder estatístico e fornecer evidências que sustentem as normas, a fidelidade e a validade da versão adaptada do teste nas populações pretendidas.
 - Preparar materiais e instruções de administração e cotação que minimizem quaisquer problemas relacionados com a cultura e o idioma que possam afetar a validade das inferências extraídas das pontuações;
- Interpretar quaisquer pontuações somente com referência a normas ou outras informações relevantes disponíveis.