

Michael Eduardo Reichenheim

Claudia Leite Moraes

Operacionalização de adaptação transcultural de instrumentos de aferição usados em epidemiologia

Operationalizing the cross-cultural adaptation of epidemiological measurement instruments

RESUMO

O objetivo do artigo foi propor uma sistemática operativa para avaliar o processo de adaptação transcultural de instrumentos desenvolvidos em outros contextos lingüístico sócio-culturais. São feitas considerações sobre a necessidade do uso de instrumentos de aferição robustos; a importância de instrumentos "universais" que permitam sintonias transculturais; e a necessidade de partir para adaptação em contraposição ao desenvolvimento de um instrumento novo. Aborda-se a existência de controvérsias e diferentes propostas processuais na literatura e apresenta-se um modelo de adaptação. Essa síntese envolve os passos operativos dos processos de avaliação de equivalência conceitual, de itens, semântica e operacional, e apresenta um roteiro de análise psicométrica que alicerça a avaliação de equivalência de mensuração. Finalmente, reforça-se a necessidade de cuidar da qualidade da informação em estudos epidemiológicos, incluindo meticolosos processos de adaptação transcultural nas agendas de pesquisas.

DESCRITORES: Comparação transcultural. Diferencial semântico. Tradução (Processo). Medidas em epidemiologia. Técnicas de diagnóstico e procedimentos. Estudos de validação.

ABSTRACT

The objective of the article was to offer an operational framework to assess cross cultural adaptation processes of instruments developed in other linguistic, social and cultural contexts. It covers the need for using robust measurement tools; the importance of 'universal' instruments that permit cross cultural fine-tuning; and stresses the need for adapting existent instruments rather than developing new ones. Existing controversies and proposals for different procedures in current literature are reviewed and a model for adapting instruments is presented. This synthesis covers the operational steps involved in evaluating concepts, semantic and operational items, and presents psychometric analysis guidelines that underlay an evaluation of measurement equivalence. Finally, the need for adequately controlling the quality of information presented in epidemiological studies, including a meticulous cross-cultural adaptation of research agendas, is reinforced.

KEY WORDS: Cross-cultural comparison. Semantic differential. Translating. Epidemiologic measurements. Diagnostic techniques and procedures. Validation studies.

Departamento de Epidemiologia. Instituto de Medicina Social. Universidade do Estado do Rio de Janeiro. Rio de Janeiro, RJ, Brasil

Correspondência | Correspondence:
Michael Eduardo Reichenheim
R. São Francisco Xavier, 524, 7º andar
20559-900 Rio de Janeiro, RJ, Brasil
E-mail: michael@ims.uerj.br

Recebido: 8/1/2007
Aprovado: 23/4/2007

INTRODUÇÃO

Estudos epidemiológicos com pretensões explicativas (determinantes, fatores de risco ou proteção, fatores etiológicos, e outros), a rigor, tendem a utilizar questionários multitemáticos. Comumente, estes questionários são compostos por diferentes módulos abarcando um ou mais construtos (dimensões*) de um modelo teórico a ser testado. Nesse sentido, cada construto implica em um instrumento epidemiológico que necessita ser incorporado ao questionário principal.

A primeira tarefa à montagem de um questionário modular consiste em uma detalhada revisão bibliográfica envolvendo o escrutínio dos programas de investigação referentes aos instrumentos disponíveis sobre cada um dos construtos de interesse. A compilação do histórico de cada instrumento candidato deve conter uma apreciação sobre o grau de utilização prévia e, principalmente, uma avaliação do estágio de desenvolvimento do programa de investigação. Para isto, é crucial examinar as evidências de adequação e suficiência da trajetória psicométrica existente até então. Esse escrutínio indica ao pesquisador se realmente há ou não instrumentos satisfatórios para captar o objeto em pauta. Para o caso de serem identificados instrumentos desenvolvidos e consolidados fora da cultura em questão, é também importante investigar se já passaram por um processo formal de adaptação transcultural (ATC).

O pesquisador pode assim decidir, para um determinado construto, se vale a pena admitir incondicionalmente um instrumento; ou, à inexistência ou insuficiência do que pôde ser identificado, se há necessidade de partir para a construção de um novo instrumento;⁵³ ou ainda, se é preciso iniciar um programa de investigação ancilar de ATC de um ou mais instrumentos julgados meritórios. Diferenças entre definições, crenças e comportamentos relacionados a um construto a ser usado em uma pesquisa epidemiológica impõem que a utilização de um instrumento elaborado em outros contextos culturais seja precedida de uma avaliação metódica da equivalência entre o original e sua versão. A necessidade de adaptação de instrumentos de aferição não se restringe às situações que envolvem países e/ou idiomas distintos. Ajustes locais e regionais também requerem atenção. É difícil caracterizar se foi ou não alcançada uma sintonia com a população na qual a versão será usada. Uma decisão deve ponderar o quanto se ganha com a aproximação cultural e o quanto se perde em termos de generalização e comparabilidade. Em países com raízes culturais heterogêneas como o Brasil, a proposição de termos coloquiais típicos, facilmente aceitos e compreendidos em uma região ou estado do país poderia não ser pertinente em outro. Também merece menção que adaptações transculturais

não se restringem ao espaço. Mudanças linguísticas acontecem em uma mesma população ao longo de anos e, logo, adaptações temporais são possíveis e, por vezes, necessárias.

O pouco rigor quanto ao uso de instrumentos de aferição desenvolvidos em outras localidades é um problema a ser enfrentado. Não é incomum um pesquisador traduzir informalmente um instrumento ou mesmo alterar o número e conteúdo dos itens constituintes. Ainda que possivelmente bem intencionado, não sintonizar as escolhas terminológicas à população-alvo, incluir itens novos e/ou excluir outros sem subseqüentemente implementar testes rigorosos pode comprometer sobremaneira a qualidade da informação.⁴¹ No limite, pode até impedir a comparação de casuísticas e estudos sobre o mesmo tema.

PROCESSO DE ADAPTAÇÃO TRANSCULTURAL

Historicamente, a adaptação de instrumentos elaborados em uma outra cultura e/ou idioma se detinha à simples tradução do original ou, excepcionalmente, à comparação literal desta com uma retradução. Há algum tempo, pesquisadores de diferentes áreas temáticas vêm sugerindo que a avaliação semântica constitua apenas um dos passos necessários ao processo de ATC.^{2,5,7,8,24,27,43} Eles recomendam que o processo seja uma combinação entre um componente de tradução literal de palavras e frases de um idioma ao outro, e um processo metódico de sintonização que contemple o contexto cultural e estilo de vida da população-alvo da versão.^{4,24,28}

Existem vários artigos na literatura contendo excelentes apreciações de abordagens teóricas e propostas processuais que, *grosso modo*, têm essa visão ampliada.^{3,4,9,19,24,28,35,46,59} Todavia, não há consenso quanto às estratégias de execução, o que faz da síntese operacional um mosaico de procedimentos oriundos de diversas fontes. Ainda assim, dirigida pela própria prática dos autores, faz-se uma escolha aqui usando como norte um dos modelos possíveis.^{27,28}

Lapidada na área de desenvolvimento de instrumentos de aferição sobre qualidade de vida, a proposta de Herdman et al²⁸ se alicerça em uma interessante revisão sobre o tema, na qual os autores identificam a pleora terminológica encontrada na literatura e a confusão que a conseqüente superposição gera entre os pesquisadores da área.²⁷ No primeiro de dois importantes artigos também apontam quatro perspectivas que tendem a reger os programas de investigação de ATC.²⁷ A primeira, denominada "ingênua" (*naive*), se baseia apenas no

* Distingue-se *construto de dimensão*, entendendo que um construto pode ser composto por várias dimensões. Por extensão, entende-se que uma dimensão tem na *escala* o seu representante empírico que, por sua vez, tem no *escore* a ordenação numérica subjacente.

Tabela. Principais etapas para avaliação da equivalência transcultural de instrumentos de aferição.

Aspecto avaliado	Estratégia para avaliação
Equivalência conceitual	<ul style="list-style-type: none"> • Revisão bibliográfica envolvendo publicações da cultura do instrumento original e da população-alvo • Discussão com especialistas • Discussão com população-alvo
Equivalência de itens	<ul style="list-style-type: none"> • Discussão com especialistas • Discussão com população-alvo
Equivalência semântica	<ul style="list-style-type: none"> • Traduções • Retraduções • Avaliação da equivalência semântica entre as retraduições e o original • Discussão com população-alvo • Discussão com especialistas para ajustes finais • Pré-teste da versão
Equivalência operacional	<ul style="list-style-type: none"> • Avaliação pelo grupo de pesquisa quanto à pertinência e adequação do: <ul style="list-style-type: none"> - Veículo e formato das questões/instruções. - Cenário de administração. - Modo de aplicação. - Modo de categorização.
Equivalência mensuração	<ul style="list-style-type: none"> • Estudos psicométricos <ul style="list-style-type: none"> - Enfoque 1: Avaliação de validade dimensional e adequação de itens componentes. - Enfoque 2: Avaliação de confiabilidade. - Enfoque 3: Avaliação de validade de construto de validade de critério
Equivalência funcional	<ul style="list-style-type: none"> • Dada pelas equivalências identificadas nas demais etapas de avaliação

processo de tradução simples e informal do instrumento original. A segunda, denominada “relativista” afirma a impossibilidade do uso de instrumentos padronizados em diferentes culturas e propõe que somente aqueles concebidos localmente devam ser utilizados. Nesse caso, a noção de equivalência não é pertinente e, por extensão, a possibilidade de interlocução. A terceira perspectiva, cunhada de “absolutista”, assume que a cultura tem um impacto mínimo nos construtos a serem mensurados e que estes são invariantes em diferentes contextos. Em termos metodológicos, a ênfase é toda no processo de tradução e retradução do instrumento. A última perspectiva, denominada “universalista”, não assume *a priori* que os construtos são os mesmos em diferentes culturas. Nesse sentido, é necessário primeiro investigar se um conceito efetivamente existe ou se é interpretado similarmente na nova cultura, para depois se estabelecer sua equivalência transcultural por meio de metodologia própria.

No artigo de seqüência publicado em 1998, Herdman et al²⁸ propõem um roteiro básico. Assumindo a postura “universalista”, apresentam um modelo de avaliação do processo de ATC que abrange a apreciação de equivalência entre o instrumento original e aquele a ser adaptado. Definições e detalhes são oferecidos a respeito de seis tipos, a saber, equivalência conceitual, de item, semântica, operacional, de mensuração e funcional.

A seguir, sugere-se uma sistemática operacional para utilizar instrumentos desenvolvidos em outros contextos lingüístico-sócio-culturais. Motiva o artigo a perspectiva de que há interesse comparar perfis epidemiológicos e achados de pesquisas realizadas em diferentes localidades e culturas. Uma outra razão é a relativa carência de textos estruturados em português sobre o “que e como fazer”, uma real lacuna diante da recente, mas crescente presença de estudos desse tipo na literatura brasileira em saúde coletiva. Filtrando-se os títulos que envolvem ATC de instrumentos de aplicação direta (verbal), uma busca na Rede SciELO usando os descritores [*questionario or instrumento*] and [*adaptacao or traducaao or confiabilidade or validade*] identificou 121 artigos nacionais. Destes, 36 eram em periódicos específicos da área de saúde coletiva* e, excetuando-se um de 1999, todos publicados na presente década. Algumas dessas experiências brasileiras têm se guiado pelo roteiro de Herdman et al,²⁸ quer usado *in totum*,^{39,40,47,51,52} quer parcialmente.^{1,11,16,21-23,26,45,54,64}

PROPOSTA DE OPERACIONALIZAÇÃO

Uma síntese do processo de avaliação de ATC está resumida na Tabela. Cada uma das etapas necessárias à apreciação dos diferentes aspectos da equivalência é aprofundada a seguir. A equivalência funcional não é elaborada, pois representa uma síntese das pregressas,

* Cadernos de Saúde Pública (20), Revista de Saúde Pública (15) e Revista Brasileira de Epidemiologia (1).

como os próprios proponentes do modelo definem. Esta síntese capta se a eficiência de um instrumento é similarmemente satisfatória em duas ou mais culturas.²⁸

Equivalência conceitual e de itens

A avaliação da equivalência conceitual consiste na exploração do construto de interesse e dos pesos dados aos seus diferentes domínios constituintes no local (país, região, cidade) de origem e na população-alvo onde o instrumento será utilizado. Conforme apresentado na Tabela, em geral, esta etapa envolve uma discussão com um grupo de especialistas. Isto tem como finalidade explorar se os diferentes domínios abarcados pelo instrumento original na definição dos conceitos de interesse seriam relevantes e pertinentes ao novo contexto ao qual está sendo adaptado. No processo, avalia-se a pertinência dos itens para a captação de cada um desses domínios. As discussões se dão à luz de uma revisão bibliográfica que prioriza as publicações sobre os processos envolvidos na construção do instrumento-fonte e do material bibliográfico disponível no contexto local. Membros seletos e indivíduos representativos⁶ da população-alvo devem ser envolvidos, seja por meio de entrevistas abertas individualizadas com seletos e representativos indivíduos da comunidade ou atividades coletivas como, por exemplo, os grupos focais.^{14,32}

Equivalência semântica

A avaliação da equivalência semântica envolve a capacidade de transferência de sentido dos conceitos contidos no instrumento original para a versão, propiciando um efeito nos respondentes semelhante nas duas culturas. O roteiro de avaliação desse aspecto de equivalência deve envolver vários passos.²⁸ Voltando à Tabela, o processo começa pela tradução do instrumento original para o idioma da cultura-alvo. Sugere-se que duas ou mais versões sejam obtidas de forma independente para que, oportunamente, se tenha mais opção para definir os termos a serem utilizados na versão a ser posta a teste. A seguir, as versões são retraduzidas para o original por outros tradutores, também de forma independente. O perfil dos tradutores também importa. Alguns autores recomendam que o processo de tradução seja realizado por profissionais cuja língua-mãe e cultura sejam aquelas para as quais se está realizando a tradução.^{24,28,46} Por exemplo, no contexto de um instrumento desenvolvido na Inglaterra a ser adaptado para uso no Brasil, as traduções do original para o português devem ser realizadas por brasileiros com bom domínio do inglês e as retraduições executadas por ingleses com bom domínio do português.

Na seqüência, um novo tradutor bilíngüe avalia formalmente a equivalência entre as retraduições e o instrumento original. Além de independente, essa avaliação deve ser cega em relação aos tradutores e retradutores.

Preferencialmente, os formulários que são apresentados ao profissional não devem indicar qual "vinheta" se refere à re-tradução e qual ao original. Uma forma de alcançar isso é aleatorizar a ordem de aparecimento. Para o caso de avaliação simultânea de mais de uma retradução, além de um formulário para cada par contendo o original e uma retradução, formulários com pares de retraduições também devem ser submetidos, para que o avaliador não tenha como identificar a "vinheta" original no conjunto. Evidentemente, esses formulários não são efetivamente analisados, só servindo como "despistadores".

Ainda que Herdman et al²⁸ aduzam vários tipos de significados lingüísticos a apreciar, dois merecem menção. O primeiro concerne uma avaliação de equivalência entre o original e cada uma das retraduições sob a perspectiva do significado referencial (denotativo) dos termos/palavras constituintes, i.e., as idéias ou objetos do mundo a que uma ou várias palavras se referem. Se há o mesmo significado referencial de uma palavra no original e na respectiva tradução, presume-se que existe uma correspondência literal entre elas. O segundo aspecto diz respeito ao significado geral (conotativo) de cada item do instrumento, contrastando-se o original com o que foi captado na tradução para o idioma-alvo. Essa correspondência transcende a literalidade das palavras, encampando também aspectos mais sutis, como o impacto que um termo tem no contexto cultural da população-alvo. A apreciação é necessária porque a correspondência literal de um termo não implica que a mesma reação emocional ou afetiva seja evocada em diferentes culturas. É indispensável a sintonia fina que alcance também uma correspondência de percepção e impacto no respondente. Essa questão é particularmente relevante em instrumentos para a captação empírica de conceitos culturalmente construídos, pois uma palavra ou assertiva usada com uma determinada intenção no contexto de origem pode não produzir o mesmo efeito na população-alvo da nova versão. A substituição por outro termo permitiria resgatar plenamente a equivalência desejada. Nesse ponto, é valioso retornar à população-alvo para que sutilezas suscitadas pelas várias propostas de versão sejam manejadas e debatidas. Esse quarto passo pode ser alcançado, por exemplo, retornando-se aos grupos focais.^{14,32}

O quinto passo da avaliação semântica envolve o mesmo grupo de especialistas que participou da etapa de avaliação da equivalência conceitual e itens, e tem como objetivo identificar e encaminhar os problemas de cada uma das atividades pregressas. Se possível, a equipe deve ser complementada por pelo menos um dos tradutores envolvidos anteriormente, de preferência aquele encarregado pela comparação formal entre as re-traduições e o instrumento original descrita acima. Dirimidos os entraves, propõe-se uma versão-síntese, ora incorporando itens oriundos de uma das versões

trabalhadas, ora optando-se por certas modificações para melhor atender os critérios expostos acima.

O sexto e último passo da etapa de avaliação de equivalência semântica envolve um pré-teste. A versão-síntese do instrumento é aplicada a grupos de indivíduos da população-alvo com vista a uma intensa avaliação de aceitabilidade, compreensão e impacto emocional. Uma técnica a ser utilizada no pré-teste é solicitar aos respondentes que parafraseiem cada item, devendo o entrevistador anotar em uma questão adicional se houve ou não compreensão do referido item. Tantas "séries" de n entrevistas (e.g., 30 a 40) são realizadas até que um percentual pré-estabelecido de ajustamento (entendimento) em todos os itens seja alcançado (e.g., $\geq 90\%$). Essas avaliações interinas podem ser realizadas pela própria equipe de pesquisa ou, melhor ainda, por um grupo de especialistas convocados para tal. A partir das evidências encontradas nesse pré-teste, fazem-se os ajustes semânticos finais da versão-síntese a ser posta a prova subseqüentemente.

Equivalência operacional

A equivalência operacional se refere a uma comparação entre os aspectos de utilização de um instrumento nas populações-alvo e fonte, de modo que eficácia seja semelhante mesmo que os *modus operandi* não sejam os mesmos. Importa escrutinar as possíveis influências de certas características do instrumento, tais como: veículo e formato das questões/instruções (e.g., se em papel impresso ou em forma eletrônica); cenário de administração (e.g., se intra-hospitalar ou domiciliar); e o modo de aplicação (e.g., se via entrevista face-a-face ou por autopreenchimento). Também está em pauta aqui a equivalência de especificação do "espaço do desfecho",⁶⁶ i.e., da escalonabilidade de cada item. Importa, assim, observar o modo de categorização e as possíveis repercussões de se optar por alguma modificação. Como exemplo, uma modificação perfeitamente cabível em situações em que um instrumento é aplicado com outros em um questionário multidimensional, mas em que o tempo de aplicação projetado é forçosamente curto, seria transformar itens originalmente propostos em cinco níveis (escala Likert⁶²) em itens dicotomizados (0/1).

Mudanças operacionais são muitas vezes decorrentes de circunstâncias em que o instrumento deve ou pode ser usado, e não do arbítrio do pesquisador. Assim, na perspectiva da ação, avaliações de equivalência operacional entre situações de aplicação encontradas à concepção do instrumento na cultura-fonte e as preponderantes na cultura-alvo requerem inicialmente uma apreciação eminentemente qualitativa sobre a possibilidade de êxito. Os grupos de expertos acionados nas etapas anteriores também poderiam ser consultados nessas discussões.

Uma vez estabelecido um consenso sobre a viabilidade e adequação de uma ou mais estratégias de ação, essas são incorporadas ao estudo que subjazerá as análises psicométricas implementadas na etapa de equivalência de mensuração. Nesse sentido, são as evidências "duras" exploradas na psicometria e as possíveis discrepâncias entre propostas operacionais competidoras que corroborarão ou não as premissas de adequação propostas inicialmente pelos especialistas. Claramente, uma evidência de equivalência psicométrica entre o original e a versão sob escrutínio também atesta positivamente sobre a adequação da operacionalização do instrumental e, por contigüidade, afirma a equivalência operacional.

Equivalência de mensuração

Como aludido anteriormente, a equivalência de mensuração se baseia na investigação das propriedades psicométricas do instrumento vertido. Em princípio, a execução de um estudo psicométrico em nada difere de qualquer estudo epidemiológico clássico, requerendo os mesmos rigores processuais.^{49,50} Em particular, é de grande interesse identificar o domínio da população captada no estudo, vislumbrando-se a extensibilidade dos achados para uma população em que efetivamente o instrumento vertido será utilizado.

Planejada e executada a etapa de campo, procede-se à análise dos dados. Nos mesmos moldes do que se poderia propor em uma avaliação do desenvolvimento de um instrumento novo, três enfoques psicométricos também podem ser sugeridos: avaliações de estrutura dimensional, incluindo a adequação dos itens componentes; de confiabilidade das informações a partir do processo que utiliza as escalas em teste; e de validade das mesmas em seus diversos matizes.⁵³ A perspectiva do presente artigo, no entanto, difere um pouco daquela ensejada no processo de geração de um instrumento. Lembrando que de fato se está à procura do estabelecimento de equivalência (de mensuração), o foco central não é tanto na magnitude dos valores das estimativas psicométricas em si, mas na comparação sistemática destes com os obtidos nos estudos progressos sobre o instrumento em sua língua/cultura original. Por exemplo, ao se apreciar algum aspecto da confiabilidade como um eixo de adequação da equivalência de mensuração, não interessa tanto apreciar os valores absolutos de um coeficiente de correlação intraclasses,⁵⁶ mas se esse converge ao encontrado nos estudos que embasam o instrumento original. Obviamente que um valor relativamente alto já seria esperado, pois, conforme indicado na introdução do texto, a escolha do próprio instrumento a ser adaptado pressuporia um histórico psicométrico positivo.

Conforme mostrado na Tabela, a primeira tarefa é explorar a estrutura dimensional do instrumento e a

adequação dos itens componentes. Métodos multivariados estão no âmago do processo. A dimensionalidade constatada em estudos anteriores pode ser acessada inicialmente por uma Análise de Fatores Exploratória (AFE),^{12,20,30,34,48,55,58} ainda que de certa forma já se tenha uma estrutura postulada a priori quanto à dimensionalidade e itens participantes. Lembrando que o termo “exploratória” descreve mais a técnica em si do que a perspectiva substantiva nesse caso. A AFE é um bom começo para que se possa implementar subsequentemente uma Análise de Fatores Confirmatória (AFC)^{31,34,36,58} com bases firmes. A AFE seguida de AFC não só ajuda a explorar se efetivamente existe a estrutura multidimensional conjecturada, mas também permite explorar o comportamento dos itens *vis-à-vis* às escalas previstas.

Para o caso de escalas formadas por itens dicotômicos e ordinais, as propriedades psicométricas dos itens e das escalas que estes formam são mais bem acomodadas nos modelos de Teoria de Resposta ao Item (TRI),^{10,15,18,25,37,57,62,63,65} uma forma de AFC não linear na perspectiva dos modelos de variáveis latentes generalizados.⁵⁸ Além do foco habitual nas cargas (*loadings*) dos itens, algumas outras questões centrais devem ser observadas em cada uma das escalas dimensionais constituintes do instrumento. Interessa ratificar a presença de escalonabilidade conjunta dos itens e a capacidade discriminante de cada um. O posicionamento absoluto e relativo ao longo do contínuo da variável latente (fator/dimensão) que a escala do instrumento aspira captar requer escrutínio para que se identifique o aparecimento indevido da presença (indesejável) ou não (desejável) de lacunas de informação ao longo do espectro. Da mesma forma, avaliar o grau de informatividade do conjunto de itens ao longo do espectro da escala e a precisão da informação ao longo do contínuo da variável latente também são aspectos a comparar com o instrumento original.

Uma alternativa mais simples quando os itens são dicotômicos ou ordinais é usar, respectivamente, matrizes de correlações tetracóricas e policóricas, obtidas por transformações das matrizes gaussianas antes da submissão à análise de fatores, seja esta exploratória (AFE) ou confirmatória (AFC).^{17,*} O uso inadvertido de matrizes de correlações gaussianas nessas situações tende a acarretar em má especificação de modelo, o que pode levar a resultados espúrios^{20,29,55} e um falso julgamento de ausência ou presença de equivalência.

O segundo enfoque psicométrico envolve avaliações formais de confiabilidade das escalas (consistência interna, estabilidade e reprodutibilidade intra ou interobservador^{13,33,41,42,62}). O objetivo é avaliar em que medida os escores de um instrumento (i.e., das escalas componentes) estão livres de erro aleatório,⁴⁴ o que

serve não apenas para robustecer a qualidade do estudo relacionado ao processo de ATC em si, mas como mais uma instância de adequação processual. Em longo prazo, uma série de estudos usando o instrumento e revelando consistentemente uma boa confiabilidade da mensuração (informação), acaba também atestando sobre a sua imanente qualidade.

Mesmo que tenha sido possível sancionar a estrutura dimensional, a adequação de itens (via TRI, por exemplo) e a confiabilidade do processo usando o instrumento adaptado, efetivamente não há nenhuma garantia de que a ATC tenha sido exitosa sem uma avaliação explícita dos aspectos de validade. Nesse sentido, enfatiza-se a perspectiva dada por Streiner & Norman,⁶² na qual estabelecer a validade de um instrumento, em última instância, é estabelecer a adequação da teoria que a suporta.

Várias estratégias têm sido usadas. Estudos de validade de construto são frequentes quando não há instrumento de referência (padrão-ouro) para o contraste. Avaliam-se as relações entre as dimensões supostamente captadas pelas diferentes escalas do instrumento, bem como as relações com outros conceitos, atributos e características ligadas à teoria geral na qual se insere o construto sob escrutínio. O encontro de associações previstas ou afinadas com evidências pregressas corrobora e reforça a validade do instrumento e no presente artigo, a adequação da ATC. O inverso também pode ser relevante ao se constatar a inexistência de relações entre os conceitos teóricos manifestos pelas escalas adaptadas e outras reconhecidamente fora do escopo da teoria geral envolvendo o fenômeno de interesse.

Ainda que não seja impeditivo buscar-se a validade de construto quando existe um instrumento, exame ou teste de referência para contrastar o instrumento sob escrutínio, é próprio avaliar-se a validade critério (concorrente ou preditiva) nesse caso. Visto sob a ótica precípua do contexto da ATC de um instrumento, avaliações da capacidade discriminante das escalas de um instrumento podem ser extremamente esclarecedoras. Saber que um instrumento de aplicação em estudos epidemiológicos capta não somente o contínuo da variável latente subjacente, mas também está substantivamente “colado” ao que um exame ou instrumento de referência encontraria, é claramente profícuo e atraente.

Ainda que o foco primordial seja na equivalência de mensuração, problemas e incongruências que transcendem a ATC também podem aparecer no processo, remetendo as discussões para um plano maior sobre a qualidade do instrumento em si. Discernir as duas questões nem sempre é possível. Por exemplo, na ATC da *Conflict Tactics Scales: Parent-Child*,⁵² o item

* Uebersax JS. The tetrachoric and polychoric correlation coefficients [monografia na internet]. [S.1.]; 2006. Disponível em: <http://ourworld.compuserve.com/homepages/jsuebersax/tetra.htm>.

abordando o ato de “dar um tapa/bofetada no rosto, na cabeça ou nas orelhas da criança” mostrou-se muito mais afim aos atos de violência física do que com os de punição corporal, esta última sendo a dimensão na qual os autores do instrumento localizam o item.^{60,61} Resta perguntar se efetivamente houve um comprometimento de equivalência de mensuração devido a um problema de significado conotativo discrepante entre as duas culturas ou se a inadequação está no próprio instrumento proposto originalmente. Nesse caso, o item em questão pertenceria à dimensão de violência física, independentemente da cultura em questão.

Ao encontrar inconsistências, várias possibilidades devem ser consideradas, sob perspectivas diferenciadas. Em princípio, deve-se questionar a qualidade da adaptação e procurar as falhas em uma ou mais etapas do processo. No entanto, é necessário observar algumas argúcias de interpretação. Focalizando a confiabilidade, por exemplo, estimativas inferiores às encontradas no original não necessariamente indicam problemas. A confiabilidade é um indicador conjuntural, refletindo tanto a qualidade da mensuração (presença ou não de erro de aferição), quanto à variabilidade do evento sob estudo na base populacional.³⁸ Questões relativas aos domínios dos estudos como objeto de comparação também devem ser debatidas. Diferenças populacionais específicas entre os estudos, tais como nível de escolaridade, gênero e faixa etária dos respondentes, podem interferir na performance dos instrumentos. Discrepâncias psicométricas não significam necessariamente alguma falha importante no processo de adaptação em si e os resultados merecem ser debatidos caso a caso.

CONSIDERAÇÕES FINAIS

Ao fechar o texto sobre os pilares que sustentam a validade de estudos epidemiológicos, Reichenheim & Moraes⁴⁹ sugerem que os muitos percalços metodológicos que dominam esses estudos desbancam por completo a idéia de “verdade objetiva”. Os autores propõem, assim, a noção de “veracidade construída” como uma base para afirmar o conhecimento. Apontam como o “construir” do conhecimento fica evidente quando se percebe o quanto a validade de especificação de um modelo estatístico necessita de um quadro teórico para sua instalação. Por sua vez, esse modelo teórico tende a crescer e se consolidar lentamente ao longo de um processo interativo de teoria e experimentação. No cerne dessa argumentação está o rigor nos métodos ao se planejar e executar um estudo epidemiológico.

No mesmo movimento, talvez não seja exagerado propor a qualidade da informação como o elo central entre a teoria e o empírico, e que para tal os rigores dispensados aos processos de aferição ocupem um *locus* privilegiado na prática da investigação epidemiológica. E não se trata de somente cuidar da coleta de dados, ainda que esta etapa sele todo um preparativo. Um exame atento ao instrumental é igualmente prudente, mantendo-se sempre a perspectiva de que haja investimento em adaptações formais de instrumentos elaborados em outros contextos. Como já mencionado, essa questão se torna indispensável se é do interesse comparar resultados de pesquisas epidemiológicas realizadas em diferentes localidades e culturas.

REFERÊNCIAS

1. Avanci JQ, Assis SG, Santos NC, Oliveira RVC. Escala de violência psicológica contra adolescentes. *Rev Saude Publica*. 2005;39(5):702-8.
2. Badia X, Alonso J. Re-scaling the Spanish version of the Sickness Impact Profile: an opportunity for the assessment of cross-cultural equivalence. *J Clin Epidemiol*. 1995;48(7):949-57.
3. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;25(24):3186-91.
4. Behling O, Law KS. Translating questionnaires and other research instruments: problems and solutions. Thousand Oaks: Sage; 2000.
5. Berkanovich E. The effect of inadequate language translation of Hispanics' responses to health surveys. *Am J Public Health*. 1980;70(12):1273-6.
6. Bowling A. Research Methods in Health. Investigating Health and Health Services. Buckingham: Open University Press; 1997.
7. Bravo M, Canino GJ, Rubio-Stipec M, Woodbury-Farina M. A cross-cultural adaptation of a psychiatric epidemiologic instrument: the diagnostic interview schedule's adaptation in Puerto Rico. *Cult Med Psychiatry*. 1991;15(1):1-18.
8. Bucquet D, Condon S, Ritchie K. The French version of the Nottingham Health Profile. A comparison of items weights with those of the source version. *Soc Sci Med*. 1990;30(7):829-35.
9. Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual Life Res*. 1993;2(6):451-9.
10. Cella D, Chang CH. A discussion of item response theory and its application in health status assessment. *Med Care*. 2000;38 (9 Supl):1166-72.
11. Chor D, Griep RH, Lopes CS, Faerstein E. Medidas de rede e apoio social no Estudo Pró-Saúde: pré-testes e estudo piloto. *Cad Saude Publica*. 2001;17(4):887-96.
12. Comrey AL, Lee HB. A first course in factor analysis. Hillsdale: Lawrence Erlbaum; 1992.
13. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297-334.
14. Dawson S, Manderson L, Tallo VL. The Focus Group Manual. Methods for Social Research in Tropical Diseases. Geneva: World Health Organization; 1992.
15. De Boeck P, Wilson M. Explanatory item response models: a generalized linear and nonlinear approach. New York: Springer-Verlag; 2004.
16. Alves MGM, Chor D, Faerstein E, Lopes CS, Werneck GL. Versão resumida da job stress scale: adaptação para o português. *Rev Saude Publica*. 2004;38(2):164-71.
17. Divgi DR. Calculation of the tetrachoric correlation coefficient. *Psychometrika*. 1979;44(2):169-72.
18. Embretson SE, Reise SP. Item response theory for psychologists. New Jersey: Lawrence Erlbaum Associates; 2000.
19. Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof*. 2005;28(2):212-32.
20. Gorsuch RL. Factor analysis. 2. ed. Hillsdale: Lawrence Erlbaum; 1983.
21. Grassi-Oliveira R, Stein LM, Pezzi JC. Tradução e validação de conteúdo da versão em português do Childhood Trauma Questionnaire. *Rev Saude Publica*. 2006;40(2):249-55.
22. Griep RH, Chor D, Faerstein E, Lopes CS. Apoio social: confiabilidade teste-reteste de escala no Estudo Pró-Saúde. *Cad Saude Publica*. 2003;19(2):625-34.
23. Griep RH, Chor D, Faerstein E, Werneck GL, Lopes CS. Validade de constructo de escala de apoio social do Medical Outcomes Study adaptada para o português no Estudo Pró-Saúde. *Cad Saude Publica*. 2005;21(3):703-14.
24. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46(12):1417-32.
25. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of item response theory. Newbury park: Sage; 1991.
26. Hasselmann MH, Reichenheim ME. Adaptação transcultural da versão em português das "Conflict Tactics Scales Form R" (CTS-1) usada para aferir violência no casal: equivalências semântica e de mensuração. *Cad Saude Publica*. 2003;19(4):1083-93.
27. Herdman M, Fox-Rushby J, Badia X. "Equivalence" and the translation and adaptation of health-related quality of life questionnaires. *Qual Life Res*. 1997;6(3):237-47.
28. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res*. 1998;7(4):323-35.
29. Jöreskog KG, Sörbom D. LISREL 8 User's Reference Guide. Chicago: Scientific Software International; 1996.
30. Kline P. An easy guide to factor analysis. New York: Routledge; 1994.
31. Kline RB. Principles and practice of structural equation modeling. 2. ed. New York: The Guilford Press; 2005.
32. Krueger R. Focus Groups: a Practical Guide for Applied Research. 2. ed. London: SAGE; 1994.
33. Kuder GF, Richardson MW. The theory of estimation of test reliability. *Psychometrika*. 1937;2(3):151-60.
34. Loehlin JC. Latent variable models. An introduction to factor, path and structural equation analysis. 4. ed. Mahwah: Lawrence Erlbaum Associates; 2004.
35. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. *J Adv Nurs*. 2004;48(2):175-86.
36. Maruyama GM. Basics of structural equation modeling. Thousand Oaks: Sage; 1998.

37. Mellenbergh GJ. Generalized linear item response theory. *Psychol Bull.* 1994;115(22):300-7.
38. Miettinen O. Design options in epidemiologic research. An update. *Scand J Work Environ Health.* 1982;8(Supl1):7-14.
39. Moraes CL, Hasselmann MH, Reichenheim ME. Adaptação transcultural para o português do instrumento "Revised Conflict Tactics Scales (CTS2)" utilizado para identificar a violência entre casais. *Cad Saude Publica.* 2002;18(1):163-75.
40. Moraes CL, Reichenheim ME. Cross-cultural measurement equivalence of the Revised Conflict Tactics Scales (CTS2) Portuguese version used to identify violence within couples. *Cad Saude Publica.* 2002;18(3):783-96.
41. Nunnally JC, Bernstein I. Psychometric theory. 2. ed. New York: McGraw-Hill; 1995.
42. Osburn HG. Coefficient alpha and related internal consistency reliability coefficients. *Psychol Methods.* 2000;5(3):343-55.
43. Patrick DL, Sittampalam Y, Somerville SM, Carter WB, Bergner M. A cross-cultural comparison of health status values. *Am J Public Health.* 1985;75(12):1402-7.
44. Pedhazur EJ, Schmelkin LP. Measurement, design, and analysis: an integrated approach. Hillsdale: Lawrence Erlbaum Associates; 1991.
45. Pereira LSM, Marra TA, Faria CD, Pereira DS, Martins MAA, Dias JMD, et al. Adaptação transcultural e análise da confiabilidade do Southampton Assessment of Mobility para avaliar a mobilidade de idosos brasileiros com demência. *Cad Saude Publica.* 2006;22(10):2085-95.
46. Perneger TV, Leplège A, Etter JF. Cross-cultural adaptation of a psychometric instrument: two methods compared. *J Clin Epidemiol.* 1999;52(11):1037-46.
47. Pesce RP, Assis SG, Avanci JQ, Santos NC, Malaquias JV, Carvalhaes R. Adaptação transcultural, confiabilidade e validade da Escala de Resiliência. *Cad Saude Publica.* 2005;21(2):436-48.
48. Pett MA, Lackey NR, Sullivan JJ. Making sense of factor analysis: the use of factor analysis for instrument development in health care research. Thousand Oaks: Sage; 2003.
49. Reichenheim ME, Moraes CL. Alguns pilares para a apreciação da validade de estudos epidemiológicos. *Rev Bras Epidemiol.* 1998;1(2):131-48.
50. Reichenheim ME, Moraes CL, Hasselmann MH. Equivalência semântica da versão em português do instrumento *Abuse Assessment Screen* para rastrear a violência contra a mulher grávida. *Rev Saude Publica.* 2000;34(6):610-6.
51. Reichenheim ME, Moraes CL. Adaptação transcultural do instrumento "Parent-Child Conflict Tactics Scales (CTS2)" utilizado para identificar a violência contra a criança. *Cad Saude Publica.* 2003;19(6):1701-12.
52. Reichenheim ME, Moraes CL. Psychometric properties of the Portuguese version of the Conflict Tactics Scales: Parent-child Version (CTS2) used to identify child abuse. *Cad Saude Publica.* 2006;22(3):503-15.
53. Reichenheim ME, Moraes CL. Desenvolvimento de instrumentos de aferição epidemiológicos. In: Kac G, Schieri R, Gigante D, organizadores. *Epidemiologia Nutricional.* Rio de Janeiro: Fiocruz; 2007. No prelo.
54. Rothman KJ, Greenland S. Modern Epidemiology. 2. ed. Philadelphia: Lippincott-Raven; 1998.
55. Rummel RJ. Applied Factor Analysis. 4. ed. Evanston: Northwest University Press; 1988.
56. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86(2):420-8.
57. Sijtsma K, Molenaar IW. Introduction to nonparametric item response theory. Thousand Oaks: Sage; 2002.
58. Skrondal A, Rabe-Hesketh S. Generalized latent variable modeling: multilevel, longitudinal, and structural equation models. Boca Raton: Chapman & Hall/CRC; 2004.
59. Sperber AD. Translation and validation of study instruments for cross-cultural research. *Gastroenterology.* 2004;126(1Supl 1):S124-8.
60. Straus MA, Hamby SH, Finkelhor D, Moore DW, Runyan D. Identification of child maltreatment with Parent-Child Conflict Tactics Scales: development and psychometric data for a national sample of American parents. *Child Abuse Negl.* 1998;22(4):249-70. Errata em: *Child Abuse Negl.* 1998;22(11):1177.
61. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 3. ed. Oxford: Oxford University Press; 2003.
63. Van der Linden WJ, Hambleton RK. Handbook of Modern Item Response Theory. New York: Springer; 1996.
64. Vilete LMP, Coutinho ESF, Figueira ILV. Confiabilidade da versão em Português do Inventário de Fobia Social (SPIN) entre adolescentes estudantes do Município do Rio de Janeiro. *Cad Saude Publica.* 2004;20(1):89-99.
65. Wilson M. Constructing measures. An item response modeling approach. Mahwah: Lawrence Erlbaum Associates; 2005.