

em que um instrumento é aplicado em conjunto com outros em um questionário multidimensional, mas para o qual é projetado um tempo de aplicação relativamente curto, é transformar itens originalmente propostos em cinco níveis [escala *Likert* (Streiner & Norman, 2008)] em itens dicotomizados (0/1).

Fica claro que mudanças operacionais são muitas vezes dependentes de circunstâncias em que o instrumento deve ou pode ser usado, e não do arbítrio do pesquisador. Assim sendo, na perspectiva da ação, avaliações de equivalência operacional entre situações de aplicação encontradas na concepção do instrumento na cultura fonte e as preponderantes na cultura alvo requerem inicialmente uma apreciação eminentemente qualitativa sobre a possibilidade de êxito. O grupo de especialistas acionado nas etapas anteriores também poderia ser utilizado como palco dessas discussões.

Uma vez estabelecido um consenso sobre a viabilidade e a adequação de uma ou mais estratégias de ação, essas são incorporadas ao estudo que subjazera às análises psicométricas implementadas na etapa de equivalência de mensuração (ver seção “Equivalência de mensuração” adiante). Nesse sentido, são as evidências “duras” exploradas na psicomетria 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 contiguidade, afirma a equivalência operacional.

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. Um roteiro de avaliação desse aspecto de equivalência deve envolver vários passos, como mostrado no Quadro 13.6. Seguem alguns detalhes importantes.

Tradução e retrotradução

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 tenham mais opções para decidir sobre os termos a serem utilizados na versão a ser testada mais à frente. A seguir, as versões são retrotraduzidas para o original por outros tradutores, também de forma independente.

O perfil dos tradutores também importa, alguns autores recomendando que o processo de tradução seja realizado por profissionais cuja língua-mãe e cultura sejam aquelas para as quais se está realizando a tradução (Guillemin *et al.*, 1993; Herdman *et al.*, 1998; Perneger *et al.*, 1999). Por exemplo, no contexto de um instrumento originalmente 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 retrotraduções executadas por ingleses com bom comando do português.

Avaliação entre retrotradução e original

Na sequência, um novo tradutor bilíngue avalia formalmente a equivalência entre as retrotraduções e o instrumento original. Além de independente, essa avaliação deve ser cega em relação aos tradutores e retrotradutores. Preferencialmente, os formulários que são apresentados ao profissional não devem indicar qual “vinheta” se refere à retrotradução e qual ao original. Uma forma de alcançar isso é aleatorizar a ordem de aparecimento. Para o caso de mais retrotraduções estarem sendo avaliadas simultaneamente, além de um formulário para cada par contendo o original e uma retrotradução, formulários com pares de retrotraduções também necessitam ser utilizados, de tal sorte 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 al.* (1998) aduzam vários tipos de significados linguísticos a apreciar, dois merecem menção. O pri-

■ **Quadro 13.6** Principais etapas para avaliação da equivalência transcultural de instrumentos de aferição

Aspectos da equivalência avaliada	Estratégias 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 indicadores (itens)	<ul style="list-style-type: none"> • Discussão com especialistas • Discussão com população alvo
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 semântica	<ul style="list-style-type: none"> • Traduções • Retrotraduções • Avaliação da equivalência semântica entre as retrotraduções e o original • Discussão com população-alvo • Discussão com especialistas para ajustes finais • Pré-teste da versão
Equivalência de 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 constructo de validade de critério
Equivalência funcional	<ul style="list-style-type: none"> • Dada pelas equivalências identificadas nas demais etapas de avaliação

os mesmos moldes, envolvendo as avaliações "internas" (dimensão de confiabilidade e de confiabilidade das escalas) e "externas" (validade mediante comparações com outros componentes da teoria geral e quando há um instrumento de referência). A perspectiva aqui, no entanto, difere daquela ensinada no processo de geração de um instrumento. Enfim, quando que, de fato, se está a procurar 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 prévios sobre o instrumento em sua língua/cultura original. Por exemplo, ao se apreciar algum aspecto dimensional como um eixo de adequação da equivalência de mensuração, não interessa tanto apreciar os valores absolutos de cargas fatoriais ou magnitude dos erros (*inaccuracy*) de indicadores (Brown, 2006), mas se estes convergem aos encontrados nos estudos que embasam o instrumento original. Obviamente que um valor relativamente alto já seria esperado, pois, conforme indicado na introdução do texto (e na Seção 3 do capítulo conexo), a escolha do próprio instrumento a ser adaptado já pressupõe um histórico psicométrico positivo.

► Síntese e decisões

Uma questão fundamental que um/a pesquisador/a deve se colocar no final do processo é se, em face das evidências em mãos, o instrumento está efetivamente apto para já ser usado na população-alvo. No ato de se usar imediatamente o instrumento verificado, não é incomum optar-se por um "sim" frente a ainda poucas evidências, como, por exemplo, um achado de "satisfação" confiabilidade teste-reteste (intraobservador), consistência interna ou, quiçá, algumas correlações "estatisticamente significantes" entre (a) escala(s) em estudo e variáveis pertencentes ao quadro teórico geral.⁹ Como pode ser aprendido neste capítulo, o limite entre o que é ou não aceitável não tem como ser claramente demarcado, o "quanto é suficiente" necessita ser acordado entre pares. De toda sorte, há uma clara sinalização de que se trata de um processo longo, envolvendo vários estudos (análises/avaliações) sequenciais e que, frequentemente, acaba gerando um quadro bastante intrincado. Reconhecidamente simplificado, o Quadro 13.7 ilustra um conjunto de hipotéticos resultados obtidos em um estudo de ATC envolvendo um instrumento de duas dimensões (escalas) com três indicadores cada. O quadro esboça uma síntese que poderia perfeitamente ser usada como substrato para um processo de decisão em um painel de especialistas. A gama de possíveis combinações de problemas (e soluções) encontradas nas diversas etapas é notória; da mesma forma, o leque de decisões que daí decorre. No caso deste exemplo, fica claro que nem sempre um instrumento está incondicionalmente pronto para uso após um único estudo. Sob perspectivas diversas, várias possibilidades devem ser consideradas ao encontro de inconsistências. Por 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,

⁹ Trata-se, por exemplo, de estimativas obtidas via coeficiente *kappa*, *alpha* Cronbach e *tau-b* de Kendall, respectivamente. Para obter uma perspectiva melhor sobre a argumentação em pauta, o leitor é convidado a ler a seção sobre "Especificação do modelo de mensuração" na Parte I e, em particular, os Quadros 13.2 a 13.5 que a acompanham.

meio concerne a uma avaliação de equivalência entre o original e cada uma das retratadas sob a perspectiva do significado referencial (denotativo) dos termos/palavras constituintes. Este conceito se refere a que uma ou várias palavras referencialmente significadas referem a 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, encarnando também aspectos mais sutis, tais 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 uma sintonia fina que alcance também uma correspondência de percepção e impacto no respondente. Essa questão é particularmente relevante quando se está diante de instrumentos a serem utilizados para a captação em prática 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. Uma substituição por outro termo permitiria resgatar plenamente a equivalência desejada.

Preparando os protótipos de versões para aplicação na população-alvo

Nesse ponto, é de valia retornar à população-alvo para que as sutilezas suscitadas pelas várias propostas de versão (traduções) sejam manejadas e debatidas. Esse passo pode ser alcançado, por exemplo, retornando-se aos grupos focais (Dawson et al., 1992; Krueger, 1994). Adicionalmente, o processo ganharia se o mesmo grupo de especialistas que participou das etapas descritas nas seções sobre equivalência conceitual/ítems e operacional estivesse envolvido aqui também. Tendo como objetivo debater e encaminhar os problemas já identificados, a equipe certamente se beneficiaria se complementada por, pelo menos, um dos tradutores envolvidos nas etapas pregressas, de preferência aquele encarregado pela comparação formal entre as retratadas e o instrumento original. Dirimidos os entraves, é proposta uma versão-síntese, ora incorporando ítems oriundos de uma das versões trabalhadas, ora optando-se por certas modificações para melhor atender aos critérios semânticos expostos acima.

Equivalência de mensuração

Pre-teste O último passo da etapa de avaliação de equivalência semântica envolve uma "ida ao campo", da mesma forma como se procede no caso do desenvolvimento de um instrumento novo. 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. A operação em si deve seguir a proposta exposta na seção "Especificação do espaço de desfecho" apresentada na Parte I.

Como aludido acima, a equivalência de mensuração se baseia na investigação das propriedades psicométricas do instrumento verificado, como as outras equivalências, esta nitidamente conectada à estrutura proposta para o desenvolvimento de novos instrumentos. Do ponto de vista processual e operacional, principalmente, as tarefas em nada diferem do que é apresentado na seção "Especificação do modelo de mensuração" da Parte I. Planejada e executada a etapa de campo, a análise segue

■ **Quadro 13.7** Um exemplo de quadro sinóptico de resultados sobre indicadores (itens) de um instrumento vertido para ser usado como substrato de um processo de decisão

Dimensão (escala)	Problemas de equivalência?						Localização do problema	Tipo de problema	Decisão sobre item
	Identificados antes do estudo psicométrico (incluindo pré-teste)?	"Internos"? (validade dimensional)			"Externos"? (comparações com outras variáveis)				
		Análise fatorial	Confiabilidade	Item-resto	Escala				
A	#1	—	—	—	—	—	Nenhum	Admissível	
	#2	Especialistas: inadequado para crianças? Entrevistadoras: mães parecem não entender a pergunta.	—	Correlação diminui discretamente quando item é retirado.	—	—	Nenhum	Admissível	
	#3	Especialistas: teoricamente, o item não qualifica a fazer parte do constructo e, logo, de qualquer uma das escalas.	Carga no fator A: $\lambda = 0,201$, e se transferido ao fator B: $\lambda = 0,019$ Resíduo $\epsilon = 0,899$	Correlação da escala (sem o item) com o item aumenta muito (>10%) quando este é retirado.	Consistência interna melhora quando item é removido	—	Maior Conjetura corroborada	Retirar item? Procurar outro item que possa substituí-lo?	
B	#4	Grupos focais: item pertenceria ao fator A na nova cultura?	Carga no fator B: $\lambda = 0,184$, mas se transferido ao fator A: $\lambda = 0,671$	Correlação da escala (sem o item) com o item aumenta muito (>10%) quando este é retirado.	Consistência interna e estabilidade temporal melhoram quando item é removido.	Maior Relação entre fator B item #4 mal especificada	Parece haver "desajuste" cultural	Retirar item? Transferir item para fator A? Substituir item?	
	#5	4ª reunião interina: redação em português necessita aprimoramento	Carga no fator antes B: $\lambda = 0,488$ Resíduo $\epsilon = 0,687$	—	—	Menor	Permanece algum "desajuste" semântico (ATC)	Admissível, mas necessitando de acertos.	
	#6	Entrevistadoras: mães se retraem com a pergunta	—	—	—	—	Nenhum ou menor	Admissível, com ligeira reserva.	

- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297-334, 1951.
- Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. The dependability of Behavioral Measurement: Theory of Generalizability for Scores and Profiles. New York: Wiley & Sons, 1972.
- Dans R. *Handbook of Cross-cultural and Multicultural Personality Assessment*. (1st ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2000.
- Dawson S, Manderson L, Tallo V. *Social and Economic Research (SER). The Focus Group Manual*. World Health Organization, 1992.
- De Boeck P, Wilson M. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer, 2004.
- Denzin NK, Lincoln YS. *Handbook of Qualitative Research*. London: SAGE Publications, Inc., 1994.
- Dyvig DR. Calculation of the tetrachoric correlation coefficient. *Psychometrika* 44:169-172, 1979.
- Donner A, Eliasziw M. A goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance-testing and sample size estimation [see comments]. *Statistics in Medicine* 11:1511-1519, 1992.
- Elmore JG, Feinstein AR. A bibliography of publications on observer variability (final installment). *Journal of Clinical Epidemiology* 45:567-580, 1992.
- Emberson SE, Reize SP. *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers, 2000.
- Ermencio SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation & the Health Professions*, 28:212-232, 2005.
- Finney SJ, Distefano C. Non-normal and categorical data in structural equation modeling. In: Hancock GR, Mueller RO (eds.). *Structural Equation Modeling: A Second Course*. Greenwich, CT: Information Age Publishing, 2006. pp. 269-314.
- Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: John Wiley & Sons, 1981.
- Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18:39-50, 1981.
- Fox J. 2006, 19 of October. *Polychoric and Polyserial Correlations, Function for R* [0.7-2]. CRAN (R-project). Available: <http://cran.r-project.org/src/contrib/Descriptions/polychor.html> [2006, November 19].
- Gorsuch RL. *Factor Analysis*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1983.
- Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology* 46:1417-1432, 1993.
- Hair JF, Anderson RE, Tatham RL, Black WC (Ed). *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1998.
- Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage, 1991.
- Hancock GR, Mueller RO. *Structural Equation Modeling: A Second Course*. 1st ed. Greenwich, CT: Information Age Publishing, 2006.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29-36, 1982.
- Herdman M, Fox-Rushby J, Badia X. "Equivalence" and the translation and adaptation of health-related quality of life questionnaires. *Quality of Life Research*, 6:237-247, 1997.
- Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Quality of Life Research* 7:323-335, 1998.
- Jöreskog KG, Sörbom D. *LISREL 8 User's Reference Guide*. Chicago: Scientific Software International, 1996.
- Jöreskog KG, Sörbom D. *Interactive LISREL*. 8.8. Chicago: Scientific Software International, 2006.
- Jöreskog KG, 2005. *Structural Equation Modeling with Ordinal Variables using LISREL (Revised version)*: Manuscript available at <http://www.ssi.com/lisrel/ordinal.htm>.
- Kline R. *An easy guide to factor analysis*. 2nd ed. New York: Routledge, 1994.
- Kline RB. *Principles and practice of structural equation modeling*. 2nd ed. London: The Guilford Press, 2005.
- Kolmikov S. 2006. *Polychoric - The polychoric correlation package for Stata*. Stata [2006, November 19].
- Krueger R. *Focus Groups: A Practical Guide for Applied Research*. 2nd ed. London: SAGE Publications, 1994.
- Krug EG, Dahlberg LL, Mercy JA, Zwi AB, Lozano R. *World report on violence and health*. Geneva: World Health Organization, 2002.
- Kuder GF, Richardson MW. The theory of estimation of test reliability. *Psychometrika* 2:151-160, 1937.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*, 33:159-174, 1977.
- Landy FJ. Stamp collection versus science. *American Psychologist* 35:1012-1027, 1986.
- Arbuckle JL. *Amos 7.0 User's Guide*. Chicago: SPSS, 2006.
- Armstrong R, Berry G. *Statistical Methods in Medical Research*. 3rd ed. London: Blackwell Scientific Publications, 1994.
- Badia X, Alonso J. Re-scaling the Spanish version of the sickness impact profile: an opportunity for the assessment of cross-cultural equivalence. *Journal of Clinical Epidemiology* 48:949-957, 1995.
- Barcko JJ. On various intraclass correlation reliability coefficients. *Psychological Bulletin* 83:762-765, 1976.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25:3186-3191, 2000.
- Behling O, Law KS. *Translating Questionnaires and Other Research Instruments: Problems and Solutions*. (Vol. 133). Thousand Oaks, CA: Sage Publications, 2000.
- Benler F. EQS 6.1: *Structural Equations Program Manual*. Encino, CA: Multi-variate Software Inc., 2004.
- Berkmanovich E. The effect of inadequate language translation of Hispanics' responses to health surveys. *American Journal of Public Health* 70:1273-1276, 1980.
- Blackock JF. *HLM: Social Statistics*. 2nd ed. London: McGraw-Hill Book Company, 1985.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two views of clinical measurement. *The Lancet* 8:307-310, Feb 1986.
- Bollen KA. *Structural Equations with Latent Variables*. New York: John Wiley & Sons, 1989.
- Bowling A. *Research Methods in Health, Investigating Health and Health Services*. Buckingham: Open University Press, 1997.
- Bravo M, Canino GJ, Rubio-Stepec M, Woodbury-Farria M. A cross-cultural adaptation of a psychiatric epidemiologic instrument: the diagnostic interview Schedule adaptation in Puerto Rico. *Culture, Medicine and Psychiatry*, 15:1-18, 1991.
- Brom TA. *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press, 2006.
- Bucquet S, Gondon S, Ritchie K. The French version of the Nottingham Health Profile: a comparison of items weights with those of the source version. *Social Science and Medicine* 30:829-835, 1990.
- Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Quality of Life Research* 2:451-459, 1993.
- Byrne BM. *Structural Equation Modeling with AMOS - Basic Concepts, Applications and Programming*. Mahwah, NJ: Lawrence Erlbaum Associates, 2001.
- Bryt T, Bishop J, Carlin JF. Bias, prevalence and kappa. *Journal of Clinical Epidemiology* 46:423-429, 1993.
- Cella D, Chang CH. A discussion of item response theory and its application in paradoxes. *Journal of Clinical Epidemiology* 43:551-558, 1990.
- Conroy AL, Lee HB. *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum, 1992.
- Converse JM, Presser S. *Survey Questions: Handcrafting the Standardized Questionnaire*. London: Sage Publications, 1986.
- Coughlin SS, Pickle LW. Sensitivity and specificity-like measures of the validity of a diagnostic test that are corrected for chance agreement. *Epidemiology* 3:178-181, 1992.
- Coughlin SS, Pickle LW. Sensitivity and specificity of a diagnostic test. *Journal of Clinical Epidemiology* 45:1-7, 1992.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297-334, 1951.
- Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. The dependability of Behavioral Measurement: Theory of Generalizability for Scores and Profiles. New York: Wiley & Sons, 1972.
- Dans R. *Handbook of Cross-cultural and Multicultural Personality Assessment*. (1st ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2000.
- Dawson S, Manderson L, Tallo V. *Social and Economic Research (SER). The Focus Group Manual*. World Health Organization, 1992.
- De Boeck P, Wilson M. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer, 2004.
- Denzin NK, Lincoln YS. *Handbook of Qualitative Research*. London: SAGE Publications, Inc., 1994.
- Dyvig DR. Calculation of the tetrachoric correlation coefficient. *Psychometrika* 44:169-172, 1979.
- Donner A, Eliasziw M. A goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance-testing and sample size estimation [see comments]. *Statistics in Medicine* 11:1511-1519, 1992.
- Elmore JG, Feinstein AR. A bibliography of publications on observer variability (final installment). *Journal of Clinical Epidemiology* 45:567-580, 1992.
- Emberson SE, Reize SP. *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers, 2000.
- Ermencio SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation & the Health Professions*, 28:212-232, 2005.
- Finney SJ, Distefano C. Non-normal and categorical data in structural equation modeling. In: Hancock GR, Mueller RO (eds.). *Structural Equation Modeling: A Second Course*. Greenwich, CT: Information Age Publishing, 2006. pp. 269-314.
- Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: John Wiley & Sons, 1981.
- Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18:39-50, 1981.
- Fox J. 2006, 19 of October. *Polychoric and Polyserial Correlations, Function for R* [0.7-2]. CRAN (R-project). Available: <http://cran.r-project.org/src/contrib/Descriptions/polychor.html> [2006, November 19].
- Gorsuch RL. *Factor Analysis*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1983.
- Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology* 46:1417-1432, 1993.
- Hair JF, Anderson RE, Tatham RL, Black WC (Ed). *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1998.
- Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage, 1991.
- Hancock GR, Mueller RO. *Structural Equation Modeling: A Second Course*. 1st ed. Greenwich, CT: Information Age Publishing, 2006.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29-36, 1982.
- Herdman M, Fox-Rushby J, Badia X. "Equivalence" and the translation and adaptation of health-related quality of life questionnaires. *Quality of Life Research*, 6:237-247, 1997.
- Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Quality of Life Research* 7:323-335, 1998.
- Jöreskog KG, Sörbom D. *LISREL 8 User's Reference Guide*. Chicago: Scientific Software International, 1996.
- Jöreskog KG, Sörbom D. *Interactive LISREL*. 8.8. Chicago: Scientific Software International, 2006.
- Jöreskog KG, 2005. *Structural Equation Modeling with Ordinal Variables using LISREL (Revised version)*: Manuscript available at <http://www.ssi.com/lisrel/ordinal.htm>.
- Kline R. *An easy guide to factor analysis*. 2nd ed. New York: Routledge, 1994.
- Kline RB. *Principles and practice of structural equation modeling*. 2nd ed. London: The Guilford Press, 2005.
- Kolmikov S. 2006. *Polychoric - The polychoric correlation package for Stata*. Stata [2006, November 19].
- Krueger R. *Focus Groups: A Practical Guide for Applied Research*. 2nd ed. London: SAGE Publications, 1994.
- Krug EG, Dahlberg LL, Mercy JA, Zwi AB, Lozano R. *World report on violence and health*. Geneva: World Health Organization, 2002.
- Kuder GF, Richardson MW. The theory of estimation of test reliability. *Psychometrika* 2:151-160, 1937.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*, 33:159-174, 1977.
- Landy FJ. Stamp collection versus science. *American Psychologist* 35:1012-1027, 1986.
- Arbuckle JL. *Amos 7.0 User's Guide*. Chicago: SPSS, 2006.
- Armstrong R, Berry G. *Statistical Methods in Medical Research*. 3rd ed. London: Blackwell Scientific Publications, 1994.
- Badia X, Alonso J. Re-scaling the Spanish version of the sickness impact profile: an opportunity for the assessment of cross-cultural equivalence. *Journal of Clinical Epidemiology* 48:949-957, 1995.
- Barcko JJ. On various intraclass correlation reliability coefficients. *Psychological Bulletin* 83:762-765, 1976.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25:3186-3191, 2000.
- Behling O, Law KS. *Translating Questionnaires and Other Research Instruments: Problems and Solutions*. (Vol. 133). Thousand Oaks, CA: Sage Publications, 2000.
- Benler F. EQS 6.1: *Structural Equations Program Manual*. Encino, CA: Multi-variate Software Inc., 2004.
- Berkmanovich E. The effect of inadequate language translation of Hispanics' responses to health surveys. *American Journal of Public Health* 70:1273-1276, 1980.
- Blackock JF. *HLM: Social Statistics*. 2nd ed. London: McGraw-Hill Book Company, 1985.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two views of clinical measurement. *The Lancet* 8:307-310, Feb 1986.
- Bollen KA. *Structural Equations with Latent Variables*. New York: John Wiley & Sons, 1989.
- Bowling A. *Research Methods in Health, Investigating Health and Health Services*. Buckingham: Open University Press, 1997.
- Bravo M, Canino GJ, Rubio-Stepec M, Woodbury-Farria M. A cross-cultural adaptation of a psychiatric epidemiologic instrument: the diagnostic interview Schedule adaptation in Puerto Rico. *Culture, Medicine and Psychiatry*, 15:1-18, 1991.
- Brom TA. *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press, 2006.
- Bucquet S, Gondon S, Ritchie K. The French version of the Nottingham Health Profile: a comparison of items weights with those of the source version. *Social Science and Medicine* 30:829-835, 1990.
- Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Quality of Life Research* 2:451-459, 1993.
- Byrne BM. *Structural Equation Modeling with AMOS - Basic Concepts, Applications and Programming*. Mahwah, NJ: Lawrence Erlbaum Associates, 2001.
- Bryt T, Bishop J, Carlin JF. Bias, prevalence and kappa. *Journal of Clinical Epidemiology* 46:423-429, 1993.
- Cella D, Chang CH. A discussion of item response theory and its application in paradoxes. *Journal of Clinical Epidemiology* 43:551-558, 1990.
- Conroy AL, Lee HB. *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum, 1992.
- Converse JM, Presser S. *Survey Questions: Handcrafting the Standardized Questionnaire*. London: Sage Publications, 1986.
- Coughlin SS, Pickle LW. Sensitivity and specificity-like measures of the validity of a diagnostic test that are corrected for chance agreement. *Epidemiology* 3:178-181, 1992.
- Coughlin SS, Pickle LW. Sensitivity and specificity of a diagnostic test. *Journal of Clinical Epidemiology* 45:1-7, 1992.

Referências bibliográficas

por exemplo, estimações inferiores às encontradas no original não indicam problemas necessariamente. Vale lembrar que a confiabilidade é um indicador conjuntural, refletindo tanto a qualidade da mensuração (presença ou não de erro de aferição) quanto a variabilidade do evento sob estudo na base populacional (Miettinen, 1982). Também devem ser debatidas questões relativas aos domínios dos estudos que estão sendo objeto de comparação. Diferenças populacionais específicas entre os estudos – nível de escolaridade, gênero e faixa etária dos respondentes, por exemplo – podem interferir no desempenho de instrumentos. Discrepâncias psicométricas não significam necessariamente alguma falha importante no processo de adaptação em si, os resultados merecendo ser debatidos caso a caso.

- Leovinger J. The technique of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. *Psychological Bulletin* 45:507-530, 1948.
- Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45:255-268, 1989.
- Loehlin JC. *Latent variable models. An introduction to factor, path and structural equation analysis*. 4th ed. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers, 2004.
- Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. *Journal of Advanced Nursing* 48:175-186, 2004.
- Maruyama GM. *Basics of Structural Equation Modeling*. Thousand Oaks: SAGE Publications, 1998.
- Medronho R, Block KV, Luiz RR, Werneck GL. *Epidemiologia*. 2nd ed. Rio de Janeiro: Atheneu, 2009.
- Mellenbergh GJ. Generalized linear item response theory. *Psychological Bulletin*, 115:300-307, 1994.
- Merenda PF. An overview of adapting educational and psychological assessment instruments: past and present. *Psychological Reports* 99:307-314, 2006.
- Miettinen O. Design options in epidemiologic research. *Scandinavian Journal of Work, Environment and Health* 8 (suppl 1):7-14, 1982.
- Moraes CL, Reichenheim ME. Cross-cultural measurement equivalence of the Revised Conflict Tactics Scales (CTS2) Portuguese version used to identify violence within couples. *Cadernos de Saúde Pública* 18:783-796, 2002.
- Moser CA, Kalton G. *Survey Methods in Social Investigation*. 2nd ed. London: Heinemann, 1984.
- Muthén LK, Muthén BO. *Mplus User's Guide. Fifth Edition*. Los Angeles, CA: Muthén & Muthén, 1998-2007.
- Nunnally JC, Bernstein I. *Psychometric theory*. 2nd New York: McGraw-Hill, 1995.
- Osburn HG. Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods* 5:343-355, 2000.
- Patrick DL, Sittampalam Y, Somerville SM, Carter WB, Bergner M. A cross-cultural comparison of health status values. *American Journal of Public Health* 75:1402-1407, 1985.
- Pedhazur EJ, Schmelkin LP. *Measurement, Design, and Analysis: An Integrated Approach*. Illsdale NJ: Lawrence Erlbaum, 1991.
- Perneger TV, Leplège A, Etter J-F. Cross-cultural adaptation of a psychometric instrument: two methods compared. *Journal of Clinical Epidemiology* 52:1037-1046, 1999.
- Pesce RP, Assis SG, Avanci JQ, Santos NC, Malaquias JV, Caravallhaes R. Adaptação transcultural, confiabilidade e validade da Escala de Resiliência. *Cadernos de Saúde Pública* 21:436-448, 2005.
- 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, Ca: Sage Publication Ltd., 2003.
- Prince M. Measurement validity in cross-cultural comparative research. *Epidemiologia e Psiquiatria Sociale* 17:211-220, 2008.
- Rabe-Hesketh S, Skrondal A. *Multilevel and longitudinal modeling using Stata*. College Station, TX: Stata Press Publication, 2005.
- Raykov T, Shrout P. Reliability of scales with general structure: point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling* 9:195-212, 2002.
- Raykov T. Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research* 32:329-353, 1997a.
- Raykov T. Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement* 21:173, 1997b.
- Raykov T. Analytic estimation of standard error and confidence interval for scale reliability. *Multivariate Behavioral Research* 37:89-103, 2002.
- Reichenheim ME, Moraes CL. Alguns pilares para a apreciação da validade de estudos epidemiológicos. *Revista Brasileira de Epidemiologia* 1:131-148, 1998.
- Reichenheim ME, Moraes CL. Buscando a qualidade das informações em pesquisas epidemiológicas. In: Minayo, MCS, Deslandes SF. *Caminhos do Pensamento: Epistemologia e Método* (ed.). Rio de Janeiro: Editora Fiocruz, 2002. pp. 227-254.
- Reichenheim ME, Moraes CL. Psychometric properties of the Portuguese version of the Conflict Tactics Scales: Parent-child Version (CTSPC) used to identify child abuse. *Cadernos de Saúde Pública* 22:503-515, 2006.
- Reichenheim ME, Moraes CL. Adaptação transcultural de instrumentos de aferição epidemiológicos: uma proposta de operacionalização. *Revista de Saúde Pública* 41:665-673, 2007a.
- Reichenheim ME, Moraes CL. Desenvolvimento de instrumentos de aferição epidemiológicos. In: Kac G, Schieri R, Gigante D. *Epidemiologia Nutricional*. (Ed.). Rio de Janeiro: Editora Fiocruz, 2007b. pp. 227-243.
- Reichenheim ME, Klein R, Moraes CL. Assessing the physical violence component of the Revised Conflict Tactics Scales when used in heterosexual couples: An item response theory analysis. *Cadernos de Saúde Pública* 23:53-62, 2007.
- Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2008.
- Rummel RJ. *Applied Factor Analysis*. 4th ed. Evanston: Northwest University Press, 1988.
- Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 2nd ed. Boston: Little, Brown & Co., 1991.
- Schmitt DP, Allik J. Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology*, 89:623-42, 2005.
- Shavelson RJ, Webb NM. *Generalizability theory. A primer*. Newbury Park: SAGE Publications, 1991.
- Sherbourne CD, Stewart AL. The MOS social support survey. *Social Science and Medicine* 32:705-714, 1991.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 86:420-428, 1979.
- Shrout PE. Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research* 7:301-317, 1998.
- Sijtsma K, Molenaar IW. *Introduction to Nonparametric Item Response Theory*. Thousand Oaks: Sage Publications, Inc., 2002.
- Skrondal A, Rabe-Hesketh S. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton: Chapman & Hall/CRC, 2004.
- Sperber AD. Translation and validation of study instruments for cross-cultural research. *Gastroenterology* 126:S124-128, 2004.
- StataCorp. *Tetrachoric - Tetrachoric correlations for binary variables. Program in Stata Statistical Software, Release 9*. College Station (TX): Stata Corporation, 2005.
- Straus MA, Hamby SL, Boney-McCoy S, Sugarman DB. The revised conflict tactics scales (CTS2). Development and preliminary psychometric data. *Journal of Family Issues* 17:283-316, 1996.
- Streiner DL, Norman GR. *Health Measurement Scales. A Practical Guide to Their Development and Use*. 4th ed. Oxford: Oxford University Press, 2008.
- Tanner WPJ, Swets JA. A decision making theory of visual detection. *Psychological Review* 61:401-409, 1954.
- Teixeira-Salmela LF, Magalhaes LC, Souza AC, Lima MC, Lima RC, Goulart F. Adaptação do Perfil de Saúde de Nottingham: um instrumento simples de avaliação da qualidade de vida. *Cadernos de Saúde Pública* 20:905-914, 2004.
- Uebersax JS. *The tetrachoric and polychoric correlation coefficients*. Statistical Methods for Rater Agreement web site. Available: <http://ourworld.computer.com/homepages/jsuebersax/tetra.htm>. Acesso em nov. 2006.
- Van der Linden WJ, Hambleton RK. *Handbook of Modern Item Response Theory*. New York: Springer, 1996.
- Wagnild GM, Young HM. Development and psychometric evaluation of the Resilience Scale. *Journal of Nursing Measurement* 1:165-78, 1993.
- Wilson M. *Constructing Measures. An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers, 2005.
- Zimowski MF, Muraki E, Mislevy R, Bock RD. *BILOG-MG for Windows 3.0*. Lincolnwood (IL): Scientific Software International, 2002.
- Zumbo BD. *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense, 1999.