

## Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol: *Measurement in medicine: a practical guide*

Cambridge University Press, 2011, 348 pp, ISBN: 978-0-521-13385-2 Paperback (\$45.00), ISBN: 978-0-521-11820-0 Hardback (\$99.00), ISBN: 978-1-139-14244-1 eBook (\$36.00)

Brian D. Stucky · Claudia C. A. Pereira

Published online: 1 February 2012  
© Springer Science+Business Media B.V. 2012

### Developing and assessing medical and health measures

Medical and health research presents many unique challenges for developing and assessing measurement instruments. While the fields of psychometrics and clinimetrics have long and for the most part distinct traditions, there has been surprisingly little work aimed at integrating the two disciplines. In *Measurement in Medicine*, de Vet et al. [1] nicely fill that gap by bringing measurement concepts traditionally restricted to psychological and educational testing, to the medical and health research community. To that extent, it provides researchers with practical guidance and advice to aid in the development and review of health-specific measurement instruments. The book is not intended to present state-of-the-art measurement techniques but rather focuses on the techniques that are well established and widely used.

This work may be viewed as an outgrowth of the COSMIN initiative (Conesus-based Standards for the selection of health Measurement Instruments; [4]). The COSMIN checklist provides a standardized tool for evaluating the methodological properties of health-related patient-reported outcomes (HR-PRO) measurement instruments (Mokkink et al. [5]). By adopting the COSMIN terminology and taxonomy, this book provides a uniform set of guidelines and standards that researchers and health professionals can use to develop and assess measurement instruments (Mokkink et al. [6]).

### An overview of the book

The chapters may be broadly categorized into three sections. Section one (Chapters 1–3) introduces measurement theories [classical test theory (CTT) and item response theory (IRT)] and provides an overview of measurement instrument development. Section two (Chapters 4–7) describes in detail various aspects of measurement, including the analysis of inter-item correlations (internal consistency and factor analysis), reliability (Chapter 5), and validity (Chapter 6). Finally, the third section synthesizes the previous chapters, focuses on responsiveness (Chapter 7), interpreting scores and change scores (Chapter 8), and concludes with a chapter devoted to systematic reviews of measurement properties (Chapter 9).

These chapters are arranged to provide comprehensive coverage of the methods needed to develop and assess measurement instruments. The reader is guided from the initial stages of literature review and item writing, through field testing and psychometric analysis, and ultimately to score interpretation and reporting. Each topic is first discussed theoretically before examples are used to provide real-world explanations. The examples and illustrations provided are one of the book's great strengths. Each example seems carefully selected to better describe the measurement topic. It is our experience that measurement issues lend themselves well to graphical illustrations and real-data applications, and this book offers a wealth of both. Finally, an "assignment" section follows each chapter that could prove useful in classroom settings. Intended for students, the questions focus on key aspects from each chapter and require a solid understanding of the material covered.

While master and dissertation-level students will find this work useful, the book is primarily written for

---

B. D. Stucky (✉)  
RAND Corporation, Santa Monica, CA 90407, USA  
e-mail: bstucky@RAND.org

C. C. A. Pereira  
Oswaldo Cruz Foundation, ENSP/Fiocruz, Rio de Janeiro, Brazil

researchers and clinicians interested in developing new instruments, or assessing the measurement properties of existing instruments. The text is not overly technical and does not require an extensive statistical background (a first course in statistics is likely sufficient). The authors' use of uniform and consistent language will also likely appeal to those with minimal background in statistics or test theory.

### Chapter reviews

Following an introductory chapter, Chapters 2 and 3 review key concepts used in scale development and provide a framework for later chapters on CTT and IRT. Appropriate for those with no prior measurement background, the general statistical models that underlie item responses are introduced and discussed with examples from Wilson and Clearly's [10] health-related quality of life model as a conceptual framework. The authors begin by underscoring that new instruments should only be developed when no other instruments are available for the researcher's needs. Essentially, these chapters function as a review of the prerequisite material needed to better understand the material presented in Chapters 4 and 5.

Chapters 4 and 6 guide the reader through all phases of scale development, starting with field testing, data collection, and item reduction through practical examples of exploratory factor analysis (EFA). While interested readers will want more details provided by primary sources (e.g., [3], the chapter does include a useful introduction to EFA that the novice user will no doubt find helpful. Reliability is the focus of Chapter 5 and begins with a reasonably thorough discussion of CTT conceptions of internal consistency and ends with a more brief description modern IRT-based conceptions of score reliability as a function of the participant's location on the latent continuum (for more descriptions of IRT models, see [2, 9]. Along the way, an extensive section is included on methods for assessing the reliability of categorical variables (e.g., Kappa statistics) and continuous variables obtained from data collection plans that utilize single and multiple raters (e.g., intraclass correlation coefficients). These sections are strong and likely to be useful when used along side primary sources (see [8]. While the reliability chapter is the most analytically heavy, it is not expected to be taxing for most readers given the ample number of examples the authors provide. Following this, Chapter 6 presents validity from the perspective of content, criterion, and construct validity using regression and CFA examples when useful.

The remaining chapters largely cover post-production topics. Chapter 7 discusses the concept of the scale's "responsiveness" to patient-level changes in the construct or criterion over time (including an effective presentation

of minimally important change (MIC) and smallest detectable change (SDC). Chapter 8 provides an overview of response shift bias, or the response variation over time can be attributed to an internal change in the patient's interpretation of a scale or item (see Schwartz and Spangers [7]). To our knowledge, this is the first health-based measurement book to cover both the theoretical models that describe response shift and the qualitative and quantitative methods used to assess its magnitude. Finally, for researchers considering developing a scale, Chapter 9 provides a practical description of how to systematically review instruments in the research literature.

### Conclusions

This book does not provide a scale development "recipe" but rather tries to address briefly many of the issues that commonly occur during scale development. This trade-off is both a strength and a weakness of the book. To the extent that the reader is interested in learning latent variable techniques in order to analyze large data sets with many items, the book is likely only a helpful introduction; however, because the book covers such a broad range of topics, researchers at any stage of the scale development or assessment process can be expected to find some utility in the advice and excellent examples the book provides. As research on health measurement expands, we expect this book to serve as a helpful companion to the medical and health researcher.

### References

1. de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide*. Cambridge, UK: Cambridge University Press.
2. Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
3. Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.
4. Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., et al. (2006). Protocol of the COSMIN study: consensus-based standards for the selection of health measurement instruments. *BMC Medical Research Methodology*, 6, 2.
5. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: Results of the COSMIN study. *Journal of Clinical Epidemiology*, 63, 737–745.
6. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, 19, 539–549.

7. Schwarz, C. E., & Spangers, M. A. G. (1999). Methodological approaches to assessing response shift in longitudinal health-related quality-of-life research. *Social Science and Medicine*, *48*, 1531–1548.
8. Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.
9. Thissen, D., & Wainer, H. (Eds.). (2001). *Test scoring*. Hillsdale, NJ: Lawrence Erlbaum Associates.
10. Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health related quality of life. *Journal of the American Medical Association*, *273*, 59–65.