

This article was downloaded by: [National Institute of Education]

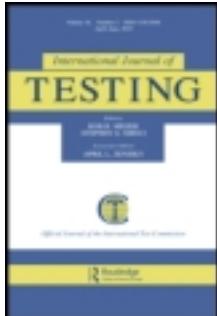
On: 03 May 2014, At: 01:19

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954

Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH,

UK



International Journal of Testing

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/hijt20>

Book Review

Antony John Kunnan

Published online: 13 Nov 2009.

To cite this article: Antony John Kunnan (2005) Book Review, International Journal of Testing, 5:1, 97-101, DOI: [10.1207/s15327574ijt0501_8](https://doi.org/10.1207/s15327574ijt0501_8)

To link to this article: http://dx.doi.org/10.1207/s15327574ijt0501_8

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

BOOK REVIEW

Scaling Procedures: Issues and Applications. Richard Netemeyer, William Beardon, and Subhash Sharma, Thousand Oaks, CA: Sage, 2003, 224 pages, \$32.95 (paperback).

Reviewed by Antony John Kunnan
*Division of Applied & Advanced Studies in Education
California State University, Los Angeles*

Measurement of latent social-psychological constructs such as self-esteem, job satisfaction, and consumer self-confidence through scales is a critical component in social science research. However, not much time is devoted to developing expertise in this area in research programs because it is unfortunately assumed that trainee researchers might pick up the expertise along the way either through research methods or through statistics courses. Devellis (1991) offered a first volume on this topic with an overview of measurement and the rationale and methods of scale development in detail. Using the Devellis outline to a large extent, Netemeyer, Beardon, and Sharma update and expand the Devellis work in *Scaling Procedures: Issues and Applications* by discussing issues in developing and validating multi-item scales.

The book begins with the importance of theory in scale development; brief overviews of the concepts of dimensionality, reliability, and validity; and recommended procedures and steps in scale development. At the outset, the authors clarify that the focus of the book is on theory-based scale development of constructs and not on empirical or atheoretical scale development constructs such as opinion polls assessing political liberalism or conservatism and the question of gun control or not (that could lead to a theory that politically liberals favor gun control). This focus is reiterated with the example of the construct of job satisfaction that is at the center of theories of employee turnover and the content domain of satisfaction with pay, satisfaction with supervisors, and satisfaction with coworkers. However, in the field of social sciences, there are examples of empirical scale development that correlate with external criteria. The Minnesota Multiphasic Personality Inventory

Requests for reprints should be sent to Antony John Kunnan, Division of Applied & Advanced Studies in Education, California State University, Los Angeles, 5151 State University Dr., Los Angeles, CA 9003228143. E-mail: akunnan@calstatela.edu

(MMPI) is an example of this type of scale development, a personality inventory based on criterion-referenced items (Holtzman, 2002). Therefore, a discussion of the two approaches—before the focus on theory-based scale development—would have been useful in the introductory chapter because advocates of both approaches share some common beliefs and scale development procedures.

The authors also emphasize the value of beginning scale development activities with a thorough review of literature for two reasons: one, to identify previous attempts to conceptualize the construct of interest and also to provide “a more precise conceptualization of the construct, its boundaries and content domain, and potential antecedents and consequences” (p. 8); and two, to answer the question: “Is a scale need at all?” because if good measures of a scale already exist then a new measure should offer some advantages over the existing measures of the same scale. However, the authors do not discuss how a construct is defined and selected and the related question of how a domain is defined, organized, and structured in scale development.

Chapters on dimensionality, reliability, and validity follow the introduction. In the chapter on dimensionality of constructs, the authors clearly demonstrate how to use exploratory and confirmatory factor analyses in assessing dimensionality of a construct, items, and a set of items. The discussion centers on unidimensionality of items, although it is well known that the measurement of high generality constructs requires heterogenous measurement scale items (Gustafsson, 2002). The authors do not offer much in terms of multidimensional scales. In addition, the authors focus on analytical measures such as exploratory and confirmatory linear factor analysis that assume continuous ratings (e.g., in a Likert scale) and normal distributions, two assumptions that are often violated in typical research contexts. Research has shown that violations of these assumptions have led to underestimates of factor loadings and overestimates of latent dimensions (Bock, Gibbons, & Muraki, 1988). Therefore, linear factor analysis is not appropriate for item-level analysis of scale items that violate these assumptions because they have dichotomous or polytomous response formats. The appropriate technique to assess dimensionality then would be either nonlinear factor analysis (Waller, Tellgren, McDonald, & Lykken, 1996) or multidimensional item response theory (IRT) available in TESTFACT (Wilson, Wood, & Gibbons, 1991). An explication of IRT for assessment of dimensionality with a worked example using TESTFACT or any other software program would have offered very useful insights.

In the chapter on reliability, the authors go over the well-known types of reliability (test-retest, alternative form, and internal consistency) and a brief overview of generalizability theory with worked examples. The authors make an often misunderstood point about coefficient alpha clearly: “a high coefficient alpha is considered evidence of the unidimensionality of the construct; unfortunately this is not true ... coefficient alpha is meaningful only for a unidimensional set of items” (p. 27). It is well known that a scale or test that is homogenous could be internally con-

sistent, but a scale or test that is internally consistent is often not homogenous. Again, this chapter would have benefited from a discussion of test information statistics from the IRT perspective.

The chapter on validity is least satisfying. Although the authors begin with the general concept of construct validity and the various ways of collecting evidence of construct validity, they group the types of validity using classification similar to those proposed by other authors. The three types of validity are grouped as follows: translation validity (content and face validity), criterion-related validity (predictive and postdictive validity and concurrent, convergent, discriminant, and known-group validity), and nomological validity. First, the term *translation validity* is confusing although the authors explain it: “content and face validity reflect the extent to which a construct is translated into the operationalization of the construct” (p. 72). In addition, the idea that face validity could provide valuable evidence of construct validity does not seem defensible. An examination of the scale in terms of content representativeness and relevance could do this job with much more accuracy. Second, *known-group validity* defined as the “extent to which a measure differs as predicted between groups who should score low and high on a trait” (p. 86) does not seem to be a valuable source of evidence of construct validity either. Third, nomological validity (resurrected from Cronbach & Meehl, 1955), generally considered an outdated concept, is problematic because a scale measure does not need to fit lawfully into a network of relations (nomological network) to be capable of providing valid inferences. Fourth, the impression is left that it is the scale that is valid rather than the inferences or actions based on the responses to a scale that are considered valid, as in Messick’s (1989) articulation. Finally, there is no discussion of how a scale developer might bring all these pieces of evidence from various analyses together to make a case for the validity of the inferences of a scale. A discussion along Kane’s (1992) interpretive or argument-based validity would have been useful.

The remaining chapters offer step-by-step practical guidance in scale development (developing construct definition and generating and judging measurement items, designing and conducting studies to develop a scale, and finalizing a scale). Three scale developments are used for illustration and discussion: the Consumer Self-Confidence Scales, the Consumer Frugality Scale, and the Work-Family Conflict and Family-Work Conflict Scales. Empirical examples illustrating exploratory and confirmatory factor and G-theory are presented in a step-by-step fashion. Unfortunately, the authors do not include the use of structural equation modeling (SEM) in scale development (e.g., Noar, 2003; Rubio, Berg-Weger, & Tebb, 2001), although latent construct modeling is the focus of SEM and this fits well with the authors’ focus of theory-based scale development of constructs.

In summary, the book is clearly satisfactory and useful for novice researchers who are beginning research based on scale development with analytical proce-

dures from classical test theory (CTT); a few shortcomings that can easily be corrected are briefly noted. However, a major problem with the book arises when it is looked at through the prism of “the new rules of measurement” that promote the use of IRT (Embretson & Hershberger, 1999). This major deficiency characterizes the book as one that is not forward-looking in preparing such researchers. In this case, the authors missed the opportunity to integrate CTT with IRT approaches to measurement issues, although there has been recent compelling research in cognitive and personality measurement with IRT (see e.g., Embretson & Reise, 2000; Reise, 1999). Furthermore, looking beyond, topics such as the measurement of intraindividual differences and interindividual differences in behavior, attitudes and feelings on different occasions of measurement (Anastasi, 1983; Eid & Langeheine, 2003), and modeling the changes of such differences across time (Duncan, Duncan, Stryker, Li, & Alpert, 1999) have become important. Researchers working in this area will not find any discussion in terms of scale development and psychometric modeling in this volume.

REFERENCES

- Anastasi, A. (1983). Traits, states, and situations: A comprehensive view. In H. Wainer & S. Messick (Eds.), *Principles of modern psychological measurement* (pp. 345–356). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Bock, R., Gibbons, R., & Muraki, E. (1988). Full information item factor analysis. *Applied Psychological Measurement*, 12, 261–280.
- Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–303.
- Devellis, R. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.
- Duncan, T., Duncan, S., Stryker, L., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Eid, M., & Langeheine, R. (2003). Separating stable from variable individuals in longitudinal studies by mixture distribution models. *Measurement*, 1, 179–206.
- Embretson, S. E., & Hershberger, S. (Eds.). (1999). *The new rules of measurement*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Gustafsson, J.-E. (2002). Measurement from a hierarchical point of view. In H. Braun, D. Jackson, & D. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 73–95). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Holtzman, W. (2002). Personality theory and assessment: Current and timeless issues. In H. Braun, D. Jackson, & D. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 37–48). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527–535.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education.
- Noar, S. (2003). The role of structural equation modeling in scale development. *Structural Equation Modeling*, 10, 622–647.

- Reise, S. (1999). Personality measurement issues viewed through the eyes of IRT. In S. E. Embretson & S. Hershberger (Eds.), *The new rules of measurement* (pp. 219–241). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Rubio, D., Berg-Weger, M., & Tebb, S. (2001). Using structural equation modeling to test for multidimensionality. *Structural Equation Modeling*, 8, 613–626.
- Waller, N., Tellgren, A., McDonald R., & Lykken, D. (1996). Exploring nonlinear models in personality assessment: Development and preliminary validation of a negative emotionality scale. *Journal of Personality*, 64, 545–576.
- Wilson, D., Wood, R., & Gibbons, R. (1991). *TESTFACT: Test scoring, item statistics, and item factor analysis*. Mooresville, IN: Scientific Software.