

Instrument/Scale Development

OVERVIEW

Instrument/scale development is the systematic process of creating and testing questionnaire, survey, or rating scale items and response options. Scales are sets of questions that collectively are expected to measure theoretical constructs. The term instrument may refer to a question set as well, or it can refer to a collection of question sets (e.g., four scales that measure aspects of classroom engagement). Scale development may start with qualitative exploration of an unstudied construct or with the combining of items in an existing dataset that were not originally designed to measure a particular construct of interest. Often the goal of scale development is a quantitative measure of a theoretical construct that can be used in research and/or practice. Qualitative procedures that are important components of scale development include literature review of a construct, expert review of preliminary items, and cognitive testing of items with members of the target population of the scale. Quantitative steps include descriptive univariate and bivariate statistical analyses, exploratory factor analyses (EFA), confirmatory factor analyses (CFA), and invariance tests. Item response theory (IRT) and Rasch methods are also commonly used quantitative scale development methods.

ARTICLES AND CHAPTERS

Description of Method

Bowen, N. K., Bowen, G. L., & Woolley, M. E. (2004). Constructing and validating assessment tools for school-based practitioners: The Elementary School Success Profile. In A. R. Roberts & K. Yeager (Eds.), *Evidence-based practice manual: Research and outcome measures in health and human services* (pp. 509-517). New York: Oxford University Press.

This book chapter describes a mixed-methods scale development process that includes cognitive interviewing of children.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education*, 15(3), 309-333.

Provides a thorough review of guidelines for writing multiple-choice items.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research a content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806-838.
Recommends best practices for developing an instrument in counseling psychology.

Applications

McCoach, D. B., & Siegle, D. (2003). The school attitude assessment survey-revised: A new instrument to identify academically able students who underachieve. *Educational and Psychological Measurement*, 63(3), 414-429.

Describes the design of a psychometrically sound instrument to measure adolescents' attitudes and self-perceptions.

Rambo, K. E. & McCoach, D. B. (2012). Teacher attitudes toward subject-specific acceleration: Instrument development and validation. *Journal for the Education of the Gifted*, 35(2), 129-152.

Describes the creation of an instrument to measure teachers' attitudes and the methods they used to examine the validity of the instrument.

TEXTS

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

This book of standards is the product of an authoritative, interdisciplinary collaboration. Many previous editions exist; each edition represents the latest thinking and methods about validity and high-quality instrument development. The standards stress that validity refers to the interpretations and uses of assessment scores. In addition to the standards themselves, the book contains extensive information about the process of scale development.

DeVellis, R. F. (2011). *Scale development: Theory and applications* (3rd ed.). Los Angeles: Sage.

This book provides a user-friendly description of traditional (primarily quantitative) scale development procedures. It includes guidelines for item-writing, factor analysis, and construct validity and reliability tests. This edition includes a chapter on item response theory.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York: Routledge.

A thorough guide for developing and validating both selected and constructed response items. Includes a discussion of the role of validity in item development and current best practice regarding test development.

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain*. New York, NY: Springer.

This book provides step-by-step instructions for developing an instrument. Includes the defining, measuring, and scaling of affective constructs and provides a thorough treatment of validity and reliability. Includes details about EFA, CFA, latent class analysis, IRT, construct invariance and major steps in the development process.

- The e-book is available through the OSU library at [link](#).

Salkind, N. J. (2006). Tests & measurement for people who (think they) hate tests & measurement. Thousand Oaks: Sage.

A useful resource for educational researchers and students. The book provides detailed basic information on items, scales, and measurement characteristics. The book also includes discussion of the types of scores that can be calculated from raw scale scores and how they differ (e.g., stanines).