Book Review

Morgan Earp[1]

Francesco Bartolucci, Silvia Bacci, and Michela Gnaldi. *Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata.* CRC Press, 2016. ISBN 978-1-4665-68495, 305 pp, $76USD.

Measurement error is an important aspect of total survey error and is discussed at length in the *Journal of Official Statistics*, but there is very little discussion and/or application of psychometric tools such as Classical Test Theory (CTT) or Item Response Theory (IRT) being used to evaluate measurement error of latent traits (traits that are not directly observed, like respondent burden). CTT and IRT are specifically designed to evaluate the validity and reliability of measures dealing with latent traits, and are commonly used to assess the validity and reliability of psychological and educational measures; however this does not appear to be a common practice in the construction of the survey questionnaires discussed in the pages of *JOS*. In fact, while fifty-five articles come up in *JOS* searching "measurement," only twelve come up searching "reliability," nine searching "validity," six searching "latent," four searching "item response theory," and zero articles come up when using the search terms "Classical Test Theory" or "psychometric." While psychometric assessment is commonly taught in education and psychological research methods programs, it is not generally understood outside of those fields. The authors of the book *Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata,* Bartoulucci, Bacci, and Gnaldi provide a rich and easy to follow overview of psychometric theory that can easily be understood and applied by survey methodologists outside of the fields of education and psychology; they discuss the theoretical framework behind these types of models as well as provide a practical guide for assessing questionnaire latent constructs using psychometric evaluation methods.

Many of the surveys used outside of education and psychology also contain latent traits, traits that we cannot directly observe, but can indirectly measure using a series of related questions (i.e., respondent burden). CTT and IRT can be used to compare and contrast the relationship between items and/or factors to assess and compare the reliability of items or latent construct measures as a whole. CTT assumes that the standard error is the same for all scores in a given population, where IRT assumes the standard errors vary. The authors discuss psychometrics from both a CTT and an IRT perspective. The book is designed specifically for graduate psychometric and statistics courses with an emphasis on measurement via questionnaires, but can be used by any survey methodologist and or practitioner interested in evaluating the reliability of latent construct measurement in their survey. The book is not only rich in theory and provides a thorough background, but it also

[1] Bureau of Labor Statistics – Office of Survey Methods Research, PSB Suite 1950, 2 Massachusetts Avenue, NE Washington District of Columbia 20212, U.S.A. Email: earp.morgan@bls.gov

provides an easy user-guide to building and assessing CTT and IRT models using Stata and R.

The book is organized into six chapters. The first three chapters provide an easy to follow and solid introduction to the psychometric evaluation of questionnaires, including both CTT and IRT, and the subsequent chapters provide both the theory and the programming for working with various types of items and constructs using R and Stata. The first chapter provides an overview of measurement as it relates to item development and questionnaires in the social sciences. The second chapter discusses construct reliability and validity using CTT. Chapters three through six discuss the overall assumptions behind IRT and the varying types of IRT models. Chapter three focuses on dichotomously scored items and discusses the three main types of IRT models including Rasch, two-parameter, and three-parameter models. Chapter four discusses the use of IRT to examine categorical responses including binary, polytomous, and ordinal data. Chapter five compares the different IRT estimation methods as well as goodness-of-fit measures. Lastly, Chapter six discusses other extensions of the IRT model including multi-level/longitudinal data, multidimensional constructs, and ends by comparing IRT to the use of structural equation modeling.

While intended for the psychometrician developing psychological or educational measures, survey practitioners will benefit from using the methods laid out in this text to evaluate the potential for measurement error within surveys and questionnaires designed to measure any latent constructs. CTT and IRT can be used to determine which items have large amounts of measurement error and therefore may not be reliable, and IRT can be used to determine for which persons/households/establishments items may not be reliable, or in some cases invalid given how easy or difficult they are to answer or agree with. Both CTT and IRT can be used to determine if the measurement error of latent constructs is higher for certain groups/types than others using invariance/differential item functioning testing. IRT can also be used to assess the utility of an item scale, to determine which response options distinguish between trait levels (as well as those that do not). Both CTT and IRT can serve as a nice compliment when evaluating the quality of survey items.

This book provides an easy to follow overview of psychometric evaluation methods that can be used to assess questionnaires and surveys designed to measure latent constructs, as well as easy to follow documentation in both R and Stata for the various types of models. This book is of value to both the novice and the experienced psychometrician as well as those who are completely new to the field. With plenty of introduction as well as detail for those who wish to delve deeper into the theory, math, and programming behind the different types of psychometric models, the text serves as an excellent manual for those looking to evaluate latent constructs within questionnaires (whether in the pre-testing phase or generally) or to better understand and review psychometric evaluation research. It provides easy to follow syntaxes and annotated output for both Stata and R, making it a simple and yet rich enough for graduate level students and practitioners alike. This book brings together a vast amount of psychometric knowledge as well as software package programming into a single cohesive book, which is something that has been missing for a long time. Most psychometric books focus on specific types of models or specific software packages, but none to date have provided the type of overview and straight forward explanations, not to mention the number of examples (including syntaxes) that this text does.