

BOOK REVIEW

- G. ENGELHARD, JR. (2013) *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*. New York: Routledge. 312 pages. US\$44.95. ISBN: 978-0-415-87125-9

Invariant Measurement fills a much needed void in the library of Rasch measurement books. Presently, there are a number of books on Rasch models but most are edited volumes that combine discussion of Rasch models with various item response theory (IRT) models. Very few books focus specifically on Rasch models, and most that do are also an edited compendium of chapters by multiple authors. While edited volumes often possess a number of strengths, one significant weakness is that a larger text can sometimes appear disjointed. One of the most popular books on Rasch models to date has been Trevor Bond and Christine Fox's (2007) *Applying the Rasch Model*. This book is a favorite among Rasch measurement enthusiasts for many reasons, not the least of which is the authors' use of practical language coupled with a consistent tone. *Invariant Measurement*, however, is sure to be another favorite in the Rasch measurement anthology as its crisp, clear writing and practical approach will be easily digestible for both measurement experts and curious students alike.

Invariant Measurement offers something much different than previously published books about Rasch measurement. As the title suggests, the book focuses on the quest for invariant measurement, the very cornerstone of item response theory. The beginning of the book focuses on conceptual and theoretical issues of invariant measurement. Readers are presented with an overview of important psychometric theories and developments that led to modern day "objective measurement."¹ A strong case is presented for what measurement in the social, behavioral, and health sciences ought to be. Engelhard appropriately tempers the utility of Rasch models by carefully pointing out that Rasch models are simply ideal-type measurement models that possess the requirements for invariant measurement. He says "the challenge is to create operational assessments and variable maps that meet these requirements" (p. 58).

Chapter 3 provides a conceptual and technical overview of various Rasch models and discusses the context in which each model would be appropriate for use. While a number of resources are available to understand the mathematical underpinnings of Rasch models, there is somewhat a shortage of literature illustrating exactly how one might construct quality measures. In Chapter 4, Engelhard presents Wilson's (2005) building block framework, discusses the process for constructing measures, and provides two straightforward examples. Chapter 5 focuses on historical perspectives of measurement and summarizes the primary viewpoints from the test score (e.g., classical test theory, generalizability theory, factor analysis, and structural equation modeling) and scaling (e.g., psychophysics, absolute scaling, IRT, and non-parametric IRT) traditions. Because the scaling tradition has become the predominant paradigm in psychometrics in recent years, Engelhard provides further discussion on this topic in Chapter 6.

Chapter 7 largely picks up where Chapter 3 left off with regard to technical issues. In this chapter various estimation procedures are presented and discussed. Chapter 8 follows with discussion of model-to-data fit. Much of the current Rasch measurement literature seems to suggest

¹Note: objective measurement is defined as "the repetition of a unit amount that maintains its size, within an allowable range of error, no matter which instrument, intended to measure the variable of interest, is used and no matter who or what relevant person or thing is measured" (Institute for Objective Measurement, 2000).

a limited number of options for discerning fit, but Engelhard presents a variety of alternatives. One notable suggestion was to consider the use of Power-Divergence Statistics described by Cressie and Read (1988) as adapted to Rasch models (see Engelhard, 2008). Engelhard discusses the advantages and disadvantages of various fit indicators and suggests that “all. . . indicators of model-data fit should be interpreted based on ‘rules of thumb’ that develop within a particular context” (p. 177).

Chapters 9–11 focus on assessments involving raters. The push for “authentic assessment” (Wiggins, 1989) has created a groundswell of rater-mediated assessments in virtually every educational arena today. Authentic assessments attempt to replicate authentic intellectual and performance challenges and typically involve expert raters making judgments about an examinee’s performance. Although such assessments benefit from increased realism, they are often burdened by additional measurement problems. Further, very few assessment practitioners are even aware of the many measurement problems that may result from assessments involving raters, and even fewer are equipped to handle these issues. Engelhard capitalizes on the opportunity to discuss measurement problems in this context and demonstrates how many of the problems can be overcome by Linacre’s (1994) Many-Faceted Rasch Model (MFRM). Chapters 10 and 11 go into further detail and discuss issues of rater agreement, errors, biases and accuracy. Several examples are provided to help readers understand the steps involved in constructing quality measures involving raters. Engelhard closes the text with a summary and discussion of invariant measurement. He offers additional insights into particularly fertile grounds for researchers and practitioners looking to advance objective measurement in the social, behavioral, and health sciences, and discusses ways in which advances in invariant measurement applications can improve the quality of research and practice in these fields.

Collectively, *Invariant Measurement* has a significant number of strengths. Among the greatest strengths is that the text offers a different perspective than most previously published books and the language is much accessible to both graduate students and seasoned researchers. The practical tone throughout helps deconstruct complex ideas into easily digestible form for readers. For example, a major practical problem in the larger assessment arena is a considerable number of assessment practitioners that have very little, if any, formal training in psychometrics. The average practitioner likely uses classical test theory and assumes a total score is all one needs to make quality inferences about what a person knows or can do. Engelhard makes one of the strongest cases I have seen, not only for the need to investigate data structures, but also to understand what data structures when coupled with the requirements for objective measurement can do as they relate to producing meaningful information about a latent trait. I was also pleased to see Engelhard repeatedly mention item maps (also called Wright maps, construct maps, etc.) throughout the text. Item maps are visual illustrations of a psychometric ruler containing measures of person ability and item difficulty on the same linear continuum. These maps are easily comprehensible for persons with little to no understanding of psychometrics because almost everyone can relate to the concepts that underscore a ruler. The continued focus on establishing psychometric rulers is certain to help readers understand the central concepts of Rasch models. As an added bonus, Engelhard has created a website (http://www.GeorgeEngelhard.com/invariant_measurement) that currently includes four data files complete with sample syntax to help readers become familiar with data analysis using the Facets software package.

In sum, I believe Engelhard has produced an incredibly valuable text that will be warmly embraced by measurement students, practitioners, and scholars for many years to come.

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

Kenneth D. Royal

References

- Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences* (2nd ed.). Mahwah: Erlbaum.

- Cressie, N., & Read, T.R.C. (1988). Cressie–Read statistic. In S. Kotz & N.L. Johnson (Eds.), *Encyclopedia of statistical sciences, supplementary volume* (pp. 37–39). New York: Wiley.
- Engelhard, G. (2008). Tuneable goodness-of-fit statistics. *Rasch Measurement Transactions*, 22(1), 1158–1159.
- Institute for Objective Measurement (2000). Definition of objective measurement. Available at <http://www.rasch.org/define.htm>.
- Linacre, J.M. (1994). *Many-facet Rasch measurement* (2nd ed.). Chicago: MESA Press.
- Wiggins, G. (1989). A true test: toward more authentic and equitable assessment. *Phi Delta Kappan*, 79, 703–713.
- Wilson, M. (2005). *Constructing measures: an item response modeling approach* (2nd ed.). Mahwah: Erlbaum.

Published Online Date: 18 DEC 2013