Chapter 1

INTRODUCTION

In any statistical modeling scenario, whether the model represents atoms or galaxies or the human brain, it is essential that all variables are measured with optimal precision. Without exact and meticulous measurement, the model may not be an accurate representation of the real-world phenomena under investigation. In the field of psychometrics, the variables in the model are psychological in nature—academic proficiency, personality traits, severity of psychiatric symptoms, and so on. Psychological constructs such as these are inherently complicated and multifaceted, and relatively simple models that only measure a single construct are often insufficient approximations of complex data. As Zhang (2007) noted, "the unidimensionality of a set of items usually cannot be met and most tests are actually multidimensional to some extent" (p. 69). Accordingly, several decades of psychometric research have led to the development of sophisticated models for multidimensional test data, and in recent years, multidimensional item response theory (MIRT) has become a burgeoning topic in psychological and educational measurement. With regard to theoretical development, MIRT is the focus of ongoing research by many leading quantitative methodologists, who are continually supplying the psychometric community with novel and innovative statistical techniques. In terms of application, MIRT has been successfully implemented not only in psychology and education but also in economics, biostatistics, psychiatry, and a number of other scientific disciplines that demand precise measurement of multidimensional psychological constructs.

MIRT is rightly considered to be a cutting-edge statistical technique; indeed, the methodology underlying MIRT can become exceedingly complex, and many leading psychometricians and researchers are actively building upon the foundations of MIRT in increasingly sophisticated ways. As a result, this topic may not receive much attention in an introductory item response theory (IRT) course. In this author's opinion, however, it is a major misperception to regard MIRT as too advanced or intimidating for inexpert audiences. While MIRT offers many technical challenges, it can certainly be understood and applied by readers who have a firm grounding in unidimensional IRT modeling. As with other titles in the Quantitative Applications in the Social Sciences (QASS) series, the purpose of the book is to present the foundations of an advanced quantitative topic in a palatable and concise format to students, instructors, and researchers. It includes

1

many practical examples and illustrations, along with numerous intuitive and informative figures and diagrams. In addition, many high-quality applied MIRT research articles are cited and discussed throughout the text to demonstrate how the various models and methods are being used in the real world. A particularly useful accompaniment to this volume is the freely available irtDemo package (Bulus & Bonifay, 2016) for the R statistical software environment (R Core Team, 2018). This package was specifically designed to provide students and other learners with a hands-on approach to IRT modeling via a suite of interactive applets. By using these applets, readers can easily manipulate and inspect the complex output produced by several common MIRT models and gain a greater understanding of potentially difficult topics. Furthermore, the irtDemo package can be used to create MIRT figures for use in academic publications; in fact, many of the figures presented in this book were created using irtDemo. This package can be downloaded from the Comprehensive R Archive Network at **https://cran.r-project.org/package=irtDemo**.

In addition, brief snippets of R code are interspersed throughout the text (with the complete R code included on the Companion Student Study Site at **study.sagepub.com/researchmethods/qass/bonifay-multidimensional-item-response-theory-1e**) to guide readers in exploring MIRT models, estimating the model parameters, generating plots, and implementing the various procedures and applications discussed throughout this book. The R code is primarily based on the mirt package (Chalmers, 2012), which provides a powerful, flexible toolkit for advanced psychometric modeling. With the hands-on interactive irtDemo applet and the implementable R code, readers will be well equipped to conduct MIRT analyses, interpret results, communicate findings, and even provide instruction in this advanced statistical topic.[1]

Finally, it is important to note that this book, like many titles in the QASS series, is more of an overview of an advanced statistical method than a technical reference source. In an effort to present MIRT methods and models in a palatable and usable format, the rigorous mathematical underpinnings of MIRT are not discussed herein. If you require derivations and proofs, then you may find a volume such as Baker and Kim (2004) more beneficial.

This book is structured as follows: Chapter 2 offers a brief review of unidimensional IRT (UIRT), covering data assumptions, dichotomous and polytomous UIRT models, descriptive statistics, and UIRT parameter

---

[1] A thorough and user-friendly guide to IRT in R is offered by Desjardins and Bulut (2018). Also, readers who are less familiar with R should note that many of the models and methods discussed in this volume can be implemented in general statistical software like Mplus (L. Muthén & Muthén, 2017), SAS (SAS Institute Inc., 2015), and SPSS (via the SPIRIT macro; DiTrapani, Rockwood, & Jeon, 2018).

estimation. Each of the sections in Chapter 2 includes only the essential details with limited exposition regarding the finer points of IRT. The goal is to refresh your memory of unidimensional IRT models and methods in preparation for the subsequent chapters on MIRT.

Chapters 3 and 4 expand upon the common UIRT measurement models to include multiple latent traits. Chapter 3 presents MIRT models for dichotomous response data, whereas Chapter 4 focuses on polytomous MIRT models. After describing the testing scenarios in which these models are appropriate, each of these chapters then presents the relevant equations along with numerous visualizations and interpretations of these models and their parameters. Readers should note that this book offers limited information on multidimensional Rasch models. Interested readers may consult Briggs and Wilson (2003) for an introduction to multidimensional Rasch modeling.

Chapter 5 covers several ways of describing the results of a MIRT analysis. One of the many challenges in understanding MIRT models is how to make sense of the parameter estimates and other statistical properties of multidimensional items/tests. This chapter presents both item- and test-level descriptives, including multidimensional item response surfaces, information functions, and other important components of standard MIRT output.

Chapter 6 explores several common MIRT factor structures. These structures describe the overall arrangement of the latent variables, their connections to one another, and their relationships with the item responses. This chapter focuses primarily on a recent development in MIRT modeling: the flexible two-tier item factor analysis model, which encompasses a number of simpler item factor structures, including the popular bifactor model.

Chapter 7 presents several methods of estimating the parameters in the MIRT models discussed in Chapters 3 through 5. This chapter introduces the estimation complications that arise due to the presence of multiple latent traits and then reviews three contemporary techniques of estimating the MIRT model parameters.

Chapter 8 focuses on an important component of any statistical modeling endeavor: the diagnosis and evaluation of the model. This chapter details several MIRT model diagnostics, including dimensionality assessment, test-level goodness of fit, and the evaluation of item-level fit. This chapter will provide you with the tools to carefully appraise the quality of a MIRT model and thereby uncover its strengths and/or shortcomings.

Finally, Chapter 9 presents several informative and cutting-edge applications of MIRT. This chapter presents a handful of the many exciting MIRT applications that have been developed in recent years, including large-scale assessment analysis, longitudinal modeling, linking and equating, differential item functioning, and computerized adaptive testing.