Digital Module 19: Foundations of IRT Estimation

Zhuoran Wang, Prometric
Nathan Thompson, Assessment Systems Corporation (ASC)
Available in the ITEMS Portal at https://ncme.elevate.commpartners.com

Module Overview

In this digital ITEMS module, Dr. Zhuoran Wang and Dr. Nathan Thompson introduce the basic item response theory (IRT) item calibration and examinee scoring procedures as well as strategies to improve estimation accuracy. They begin the module with a conceptual review of IRT that includes core advantages of the IRT framework, commonly used IRT models, and essential components such as information and likelihood functions. In the second part of the module, they illustrate the structure and inner workings of calibration and scoring algorithms such as the MMLE/EM algorithm for item parameter calibration and the MLE, EAP, and MAP algorithms for examinee scoring. In part three, they demonstrate the influence of multiple factors on estimation accuracy and provide strategies for maximizing accuracy. In addition to audio-narrated slides, the digital module contains sample *R* code, quiz questions with diagnostic feedback, curated resources, and a glossary.

Keywords: calibration, EM algorithm, estimation accuracy, item response theory (IRT), maximum likelihood estimation (MLE), maximum a posteriori (MAP), expected a posteriori (EAP), marginal maximum likelihood estimation (MMLE), scoring

Prerequisite Knowledge

This ITEMS module assumes no prior knowledge of IRT estimation algorithms. However, to get the most of this module, it might be beneficial to have a basic understanding of:

- key components of IRT such as item response functions for 1PL, 2PL, and 3PL models
- the goals and logic of item calibration and examinee scoring (i.e., to recover item and examinee parameters using the item response matrix and possible input parameters)
- foundations of probability theory, including Bayesian variants (e.g., prior and posterior distributions)

Experience with *R* statistical software would help enhance learners' experience with the module, in particular the "factors influencing estimation accuracy" section.

Learning Objectives

Upon completion of this module, learners should be able to:

- demonstrate a preliminary understanding of item calibration and examinee scoring algorithms
- conduct item calibration and examinee scoring using *R*
- read and interpret the calibration and scoring results
- evaluate estimation accuracy

- identify potential need to increase estimation accuracy
- use multiple strategies to facilitate estimation

Module Structure

The digital module is divided into the following sections, which can be reviewed sequentially or independently [approximate completion times in parentheses].

- Module Introduction [*5 Minutes*]
- Section 1: IRT Basics [*15 Minutes*]
- Section 2: Scoring and Calibration [*20 Minutes*]
- Section 3: Influencers on Estimation Accuracy [*20 Minutes*]
- Section 4: Quizzes [*10 minutes*]
- Section 5: Data Activity [*30 minutes*]

In the portal site, you can also find a video version of the core content as well as a handout with all core slides along with other materials.

Module Components

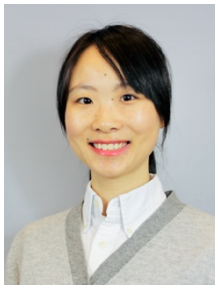
This ITEMS module includes the following components, which are delivered within a web-delivered unified design shell that is compatible across platforms (i.e., laptops, desktops, tablets, cell phones) and was created with modern course development software (*Articulate 360*):

- integrated content slides that provide a structured walk-through of the content with suitable voice-over
- embedded didactic videos to demonstrate software implementations
- interactive quiz questions
- data activity sample *R* code and annotated solutions
- glossary of key terms
- supplementary digital resources

Additional materials may be added over time so check back periodically!

Instructors

Zhuoran Wang, *Prometric*



Zhuoran Wang is a psychometrician at Prometric. She provides psychometric expertise for a number of clients, mostly professional associations and state-based licensure exams. Tasks mainly includes classical test theory analysis, item response theory analysis, computerized adaptive testing development, equating, scaling, job analysis, and standard setting. Zhuoran graduated from University of Minnesota with a Ph.D. in Psychometrics/ Quantitative Methods. Her research interests span a wide range of topics in psychometrics including multidimensional item response theory, cognitive diagnostic modeling, differential item functioning, computerized adaptive testing, and multistage testing. When in graduate school, she taught multiple undergraduate level and graduate level statistic and data analysis exercise courses. She also provided R tutorial in two workshops on multidimensional IRT models in IACAT conference 2017 and 2019.

Nathan Thompson, *Assessment Systems Corporation (ASC)*

Nate currently serve as VP for ASC (www.assess.com). His focus is on bringing quality measurement to more organizations, either through software that automates the work, makes it easier to implement sophisticated psychometrics like item response theory, or directly consulting with organizations to ensure that they align with best practices. Nate earned his PhD in Psychometrics from the University of Minnesota, with a focus on computerized adaptive testing. His undergraduate degree was from Luther College with a triple major of Mathematics, Psychology, and Latin. He is primarily interested in the use of artificial intelligence and software automation to augment and replace the work done by psychometricians, which has provided extensive experience in software design and programming. He has published over 100 journal articles and conference presentations. In addition to research, Nate has abundant teaching experiences. He served as an Instructor for the Department of Psychology for two undergraduate statistic courses while in graduate school. When working as adjunct faculty in University of Cincinnati, he taught an online class on measurement and assessment as part of a Master's in Medical Education program.

Instructional Design Team

André A. Rupp, *Mindful Measurement*

André A. Rupp is the co-author and co-editor of two award-winning interdisciplinary books entitled *Diagnostic Measurement: Theory, Methods, and Applications* (2010) and *The Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications* (2016) and has just published the *Handbook of Automated Scoring: Theory into Practice* (2020). His research synthesis- and framework-oriented work has appeared in a wide variety of prestigious peer-reviewed journals. Among other things, he is passionate about improving processes for interdisciplinary collaborations during the development and implementation of scoring solutions for digitally-delivered assessments. Consequently, he is very excited to serve as the associate editor / lead instructional designer of the ITEMS portal for NCME whose mission is to provide free digital resources to support self-directed learning and professional development.

Jonathan Lehrfeld, *Educational Testing Service*

Jonathan Lehrfeld graduated from Fordham University in 2016 with a Ph.D. in psychometrics and quantitative psychology, where his dissertation focused on integrating propensity score methods with structural equation modeling. After graduating, he worked at the Council for Aid to Education (CAE) for three years, serving as their psychometrician and Associate Director of Measurement Science. While at CAE, his operational and research work focused on practical problems in low-stakes testing. He most recently joined ETS where he currently works as a psychometrician on a large-scale state assessment team.