
Digital Module 08: Foundations of Operational Item Analysis

Hanwook [Henry] Yoo, Educational Testing Service
Ronald K. Hambleton, University of Massachusetts Amherst

Module Overview

Item analysis is an integral part of operational test development and is typically conducted within two popular statistical frameworks: classical test theory (CTT) and item response theory (IRT). In this digital ITEMS module, Hanwook Yoo and Ronald K. Hambleton provide an accessible overview of operational item analysis approaches within these frameworks. They review the different stages of test development and associated item analyses to identify poorly performing items and effective item selection. Moreover, they walk through the computational and interpretational steps for CTT- and IRT-based evaluation statistics using simulated data examples and review various graphical displays such as distractor response curves, item characteristic curves, and item information curves. The digital module contains sample data, Excel sheets with various templates and examples, diagnostic quiz questions, data-based activities, curated resources, and a glossary.

Key words: Classical test theory, corrections, difficulty, discrimination, distractors, item analysis, item response theory, R Shiny, TAP, test development

Prerequisite Knowledge

This ITEMS module assumes no prior knowledge of item analysis and test development. However, to get the most out of this module, it might be beneficial to have a basic understanding of:

- Descriptive statistics such as means, standard deviations, and correlations
- Types of test items such as selected-response and constructed-response
- Relationships between observed item responses and unobserved learner abilities

The following NCME ITEMS modules may serve as a useful introduction to the prerequisite knowledge:

- Module 7: Comparison of 1-, 2-, and 3-Parameter IRT Models
- Module 17: Item Bank Development
- Module 19: Differential Item Functioning

These modules and others are available for free in the ITEMS portal.

Learning Objectives

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

- Describe the key phases in test development and the role of item analysis within it
- Discuss issues associated with item analysis such as sample size, missing data, and item bias
- Interpret and use CTT-based item statistics in item review and test development
- Interpret and use IRT-based item statistics in item review and test development
- Identify the pros and cons of CTT- and IRT-based item statistics
- Conduct item analysis using software packages such as *Test Analysis Program* (TAP) and *R Shiny*

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: Test Development [20 Minutes]
- Section 2: Item Analysis with CTT [30 Minutes]
- Section 3: Item Analysis with IRT [30 Minutes]
- Section 4: Data Activity using Shiny Item Analysis [45 Minutes]
- Section 5: Data Activity using Test Analysis Program [45 Minutes]
- Section 6: Quizzes [15 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

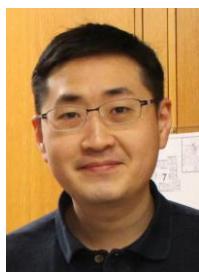
This ITEMS module includes the following components, which are delivered within a web-delivered unified design shell that is compatible across platforms (i.e., laptops, desktops, tablets, cell phones) and was created with modern course development software (*Articulate 360*):

- integrated content slides that provide a structured walk-through of the content
- embedded didactic videos to demonstrate software implementations
- interactive quiz questions with diagnostic feedback
- sample data, sample excel computation files, and annotated solutions for data activities
- glossary of key terms
- supplementary digital resources

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

Instructors

Hanwook [Henry] Yoo, Managing Senior Psychometrician at Educational Testing Service (ETS)



Henry is a managing senior psychometrician in the Psychometric Analysis and Research division at Educational Testing Service (ETS). At ETS, he manages operational psychometric work for graduate admissions programs. He received his Ed.D. in Research and Evaluation Methods Program from the University of Massachusetts, Amherst. His research interests include measurement invariance across subgroups, innovative score reporting, construct validity of English language proficiency assessment, and applications of IRT to computer-based testing. He is a co-author of a bibliography of research on test score reporting, which is available at the NCME website (<https://ncme.connectedcommunity.org/ncmedev/viewdocument/score-reporting-bibliography>).

Ronald K. Hambleton, Professor Emeritus at the University of Massachusetts Amherst



Ronald holds the titles of Distinguished University Professor and Executive Director of the Center for Educational Assessment at the University of Massachusetts, Amherst. He earned his Ph.D. in 1969 from the University of Toronto with specialties in psychometric methods and statistics. He is the co-author or co-editor of eight measurement books as well as author or co-author of many research papers, reports, and reviews spanning 50 years on topics such as standard-setting, score reporting, test adaptation, and applications of IRT. He is currently conducting research on a number of topics: computer-based testing, methods and guidelines for adapting tests from one language and culture to another, and design and field-testing of new approaches for reporting test scores.

Instructional Design Team

André A. Rupp, Research Director at Educational Testing Service (ETS)



André is a research director in the psychometrics, statistics, and data sciences area at ETS. He 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); he is currently working on the *Handbook of Automated Scoring: Theory into Practice*. His synthesis- and framework-oriented research 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 developer of the ITEMS portal for NCME whose mission is to provide free digital resources in educational measurement to support self-directed learning and professional development.

Xi Lu, Doctoral Candidate at the Florida State University (FSU)



Xi is a doctoral candidate in the Instructional Systems and Learning Technologies program at the Florida State University. Her current research interest focuses on designing and developing optimal learning supports to facilitate STEM learning in digital interactive environments. She also works as a research assistant with Dr. Val Shute on an NSF project targeted at designing various learning supports for a 2D physics game called *Physics Playground* to help middle school kids learn physics. Before coming to FSU, Xi taught Chinese for six years in Monterey Bay, California.

This is the pre-peer reviewed version of the following article: Yoo, H., & Hambleton, R. K. (2019). Digital Module 08: Foundations of operational item analysis. Educational Measurement: Issues and Practice, 38(3), 116-117. It has been published in final form at <https://doi.org/10.1111/emip.12289>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.
