PSYCHOMETRICS

IGOR HIMELFARB, PH.D., MBA

DEPARTMENT OF PSYCHOMETRICS AND RESEARCH

NBCE

NATIONAL BOARD OF CHIROPRACTIC EXAMINERS

PSYCHOMETRICS

- Psychometrics = Educational Measurement
- Measurement is a process of assignment of numbers to a phenomenon (Stevens, 1946)
- What gets measured gets managed (P.Drucker)

PSYCHOMETRICS

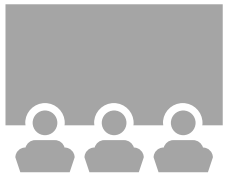
VALID TESTS

RELIABLE TESTS

FAIR TESTS

DEPARTMENT OF PSYCHOMETRICS AND RESEARCH

Operational (80%)



Research (20%)

Operational research (90%)

Non-operational research (10%)

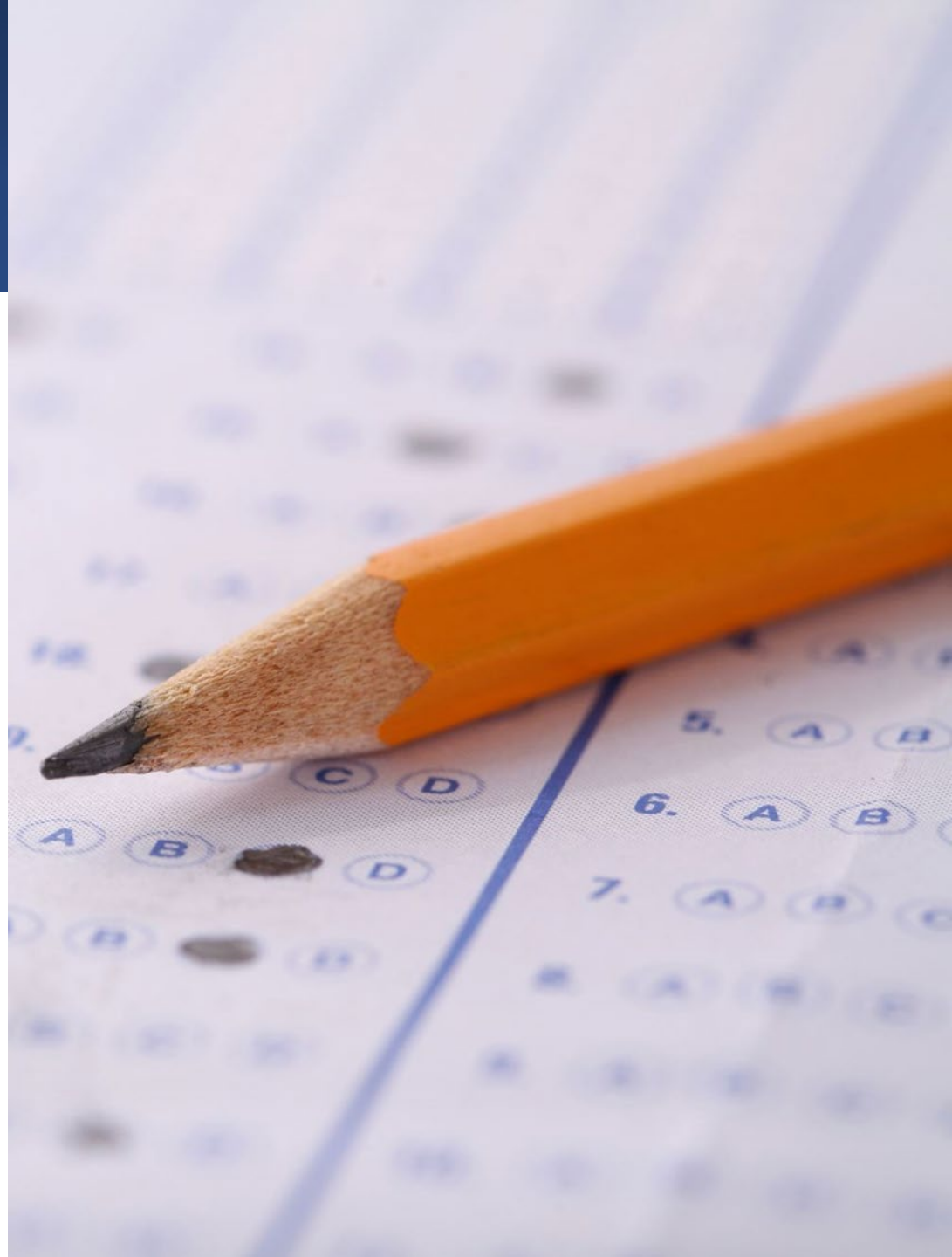
The model is:

Director/Psychometrician

- Associate Psychometrician
- Senior/Data Analyst

OPERATIONAL WORK

- Producing Scores (each administration of CBT and Part IV)
  - Data validity
  - Scoring
  - DIA
  - Post-Exam Review
  - Calibration
  - Equating
  - Scaling
  - Reporting



RESEARCH WORK

EDUCATION RESEARCH SERVICE

REVIEW OF THE LITERATURE

A primer on standardized testing:
History, measurement, classical test theory, item response theory, and equating

Igor Himelfarb, PhD

Objective: This article presents health science educators and researchers with an educational measurement. The history, theoretical frameworks of classical test theo the most common IRT models used in modern testing are presented.
Methods: A narrative overview of the history, theoretical concepts, test theory, an reader with these concepts of modern testing. Examples of data analyses using simulated data sets. One set consisted of a sample of 2000 item responses to 40 m items. This set was used to fit 1-parameter logistic (PL) model, 2PL, and 3PL IRT m of 1500 item responses to 10 polytomously scored items. The second data set was **Results:** Model-based item parameter estimates for 1PL, 2PL, 3PL, and graded re explained.
Conclusion: This study provides health science educators and education re educational measurement. The history of standardized testing, the frameworks of c logic of scaling and equating are presented. This introductory article will aid read

Key Indexing Terms: Chiropractic; Education; Educational Measurement

J Chiropr Educ 2019;33(2):151–163 DOI 10.7899/JCE-18-22

INTRODUCTION

In the 20th century, the concept of public protection dictated implementation of licensing laws to those professions having a direct relationship to public health and safety.[1] A plethora of discipline-specific prelicensure standardized assessment instruments (tests) exists to ensure compliance with the disciplinary standards. In the chiropractic profession, every year thousands of students take the prelicensure Part I, II, III, and IV examinations of the National Board of Chiropractic Examiners. As with any examination, some students feel that these standardized tests are unfair and have little relevance to clinical practice. Even faculty members often understand little about the boards. This article aims to provide an introduction to the world of standardized assessment not only for chiropractic educators but also for any health sciences educator or educational researcher.

OVERVIEW AND SIMULATED ANALYSES

China, to qualify for
examined for their
manship, calligraphy
edge. Later, the exa
law, military affairs,
and poetry.[2,3] The
qualified to serve t
The exams were a
solemnity and atten
to be scrutinized for
of the exams were fr
he often examined t
the competition.

In the late 1880s
work of his cousin,
of species and becan
intelligence and the
developed the theore
of a series of identica
and the statistical

EDUCATION RESEARCH SERVICE

ORIGINAL ARTICLE

Score production and quantitative methods used by the National Board of Chiropractic Examiners for postexam analyses

Igor Himelfarb, PhD, Bruce L. Shotts, DC, Nai-En Tang, PhD, and Margaret Smith

Objective: The National Board of Chiropractic Examiners (NBCE) uses a robust system for data analysis. The this work is to introduce the reader to the process of score production and the quantitative methods used psychometrician and data analysts of the NBCE.
Methods: The NBCE employs data validation, diagnostic analyses, and item response theory–based model responses to estimate test takers' abilities and item-related parameters. For this article, the authors generate synthetic item responses to 20 multiple-choice items with 4 response options to each item. These data were u illustrate and explain the processes of data validation, diagnostic item analysis, and item calibration based o response theory.
Results: The diagnostic item analysis is presented for items 1 and 5 of the data set. The 3-parameter logisti response theory model was used for calibration. Numerical and graphical results are presented and discussed.
Conclusion: Demands for data-driven decision making and evidence-based effectiveness create a need for ob measures to be used in educational program reviews and evaluations. Standardized test scores are often included array of objective measures. With this article, we offer transparency of score production used for NBCE testin

Key Indexing Terms: Chiropractic; Education; Psychometrics; Scoring Methods; Statistical Data Analysis

J Chiropr Educ 2020;34(1):35–42 DOI 10.7899/JCE-18-27

INTRODUCTION

According to the Standards for Educational and Psychological Testing, assessment is among the most important contributions of cognitive and behavioral sciences to our society.[1] Decision processes in health care professional testing are often complex and ongoing and pose additional challenges due to numerous regulations and their enforcement by the agencies responsible for safety of the general public.[2]

The National Board of Chiropractic Examiners (NBCE) adheres to the Standards for Educational and Psychological Testing.[1] These principles dictate that the NBCE provide accurate, fair, valid, and reliable assessment results to the intended score recipients, follow specific guidance in assessment development, obey psychometric standards, and respect the rights and responsibilities of the test takers. The goal of the NBCE is to produce scores that

Therefore, our objective for this article is to chiropractic educators to the field of measurem demystify the complex and laborious process NBCE exams. This article provides an ov principle concepts and modern-day best practice in testing. This is followed by a description and i of the multiphase process of score development u NBCE using a generated data set that mimics structure of NBCE Part I and Part II exams.

OVERVIEW OF TESTING CONCEP

Data Validity

Tabachnick and Fidell[3] supplied a checklist fo data prior to statistical analysis: (1) inspect descriptive statistics for accuracy of input, (2) ev amount and distribution of missing data and de problem, (3) check pairwise plots for nonlin heteroscedasticity, (4) identify and deal with

Article

A Two-Level Alternating Direction Model for Polytomous Items With Local Dependence

Educational and Psychological Measurement
2020, Vol. 80(2) 293–311

DOI: 10.1177/0013164419871597
journals.sagepub.com/home/epm
SAGE

Igor Himelfarb[1], Katerina M. Marcoulides[2], Guoliang Fang[3] and Bruce L. Shotts[1]

Abstract

The chiropractic clinical competency examination uses groups of items that are integrated by a common case vignette. The nature of the vignette items violates the assumption of local independence for items nested within a vignette. This study examines via simulation a new algorithmic approach for addressing the local independence violation problem using a two-level alternating directions testlet model. Parameter values for item difficulty, discrimination, test-taker ability, and test-taker secondary abilities associated with a particular testlet are generated and parameter recovery through Markov Chain Monte Carlo Bayesian methods and generalized maximum likelihood estimation methods are compared. To aid with the complex computational efforts, the novel so-called TensorFlow platform is used. Both estimation methods provided satisfactory parameter recovery, although the Bayesian methods were found to be somewhat superior in recovering item discrimination parameters. The practical significance of the results are discussed in relation to obtaining accurate estimates of item, test, ability parameters, and measurement reliability information.

Keywords

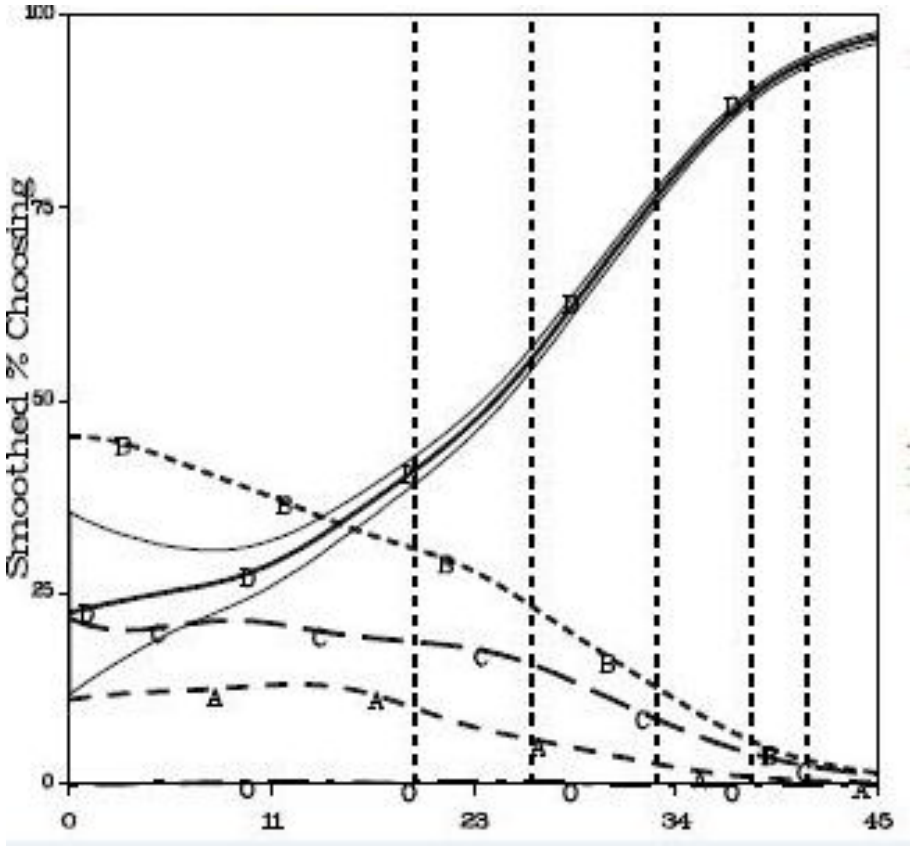
testlet response theory (TRT), violation of local independence, Bayesian methods, Markov Chain Monte Carlo (MCMC), generalized maximum likelihood estimation (GMLE)

[1]National Board of Chiropractic Examiners, Greeley, CO, USA
[2]University of Minnesota, Minneapolis, MN, USA
[3]The Pennsylvania State University, University Park, PA, USA

Corresponding Author:
Igor Himelfarb, National Board of Chiropractic Examiners, 901 54th avenue, Greeley, CO 80634, USA.
Email: ihimelfarb@nbce.org

ITEM ANALYSIS

Multiple Choice: 4 options					
Response	**N**	**%Total**	**Mean**	**SD**	**Top 10%**
A	371	4	23.3	7.1	0.1
B	1407	15.1	25.1	7.4	0.8
C	908	9.7	25.7	7.3	0.6
D*	6652	71.2	33.9	7.1	98.5
Omt	6	0.1	21.7	5.1	0
NR	0				



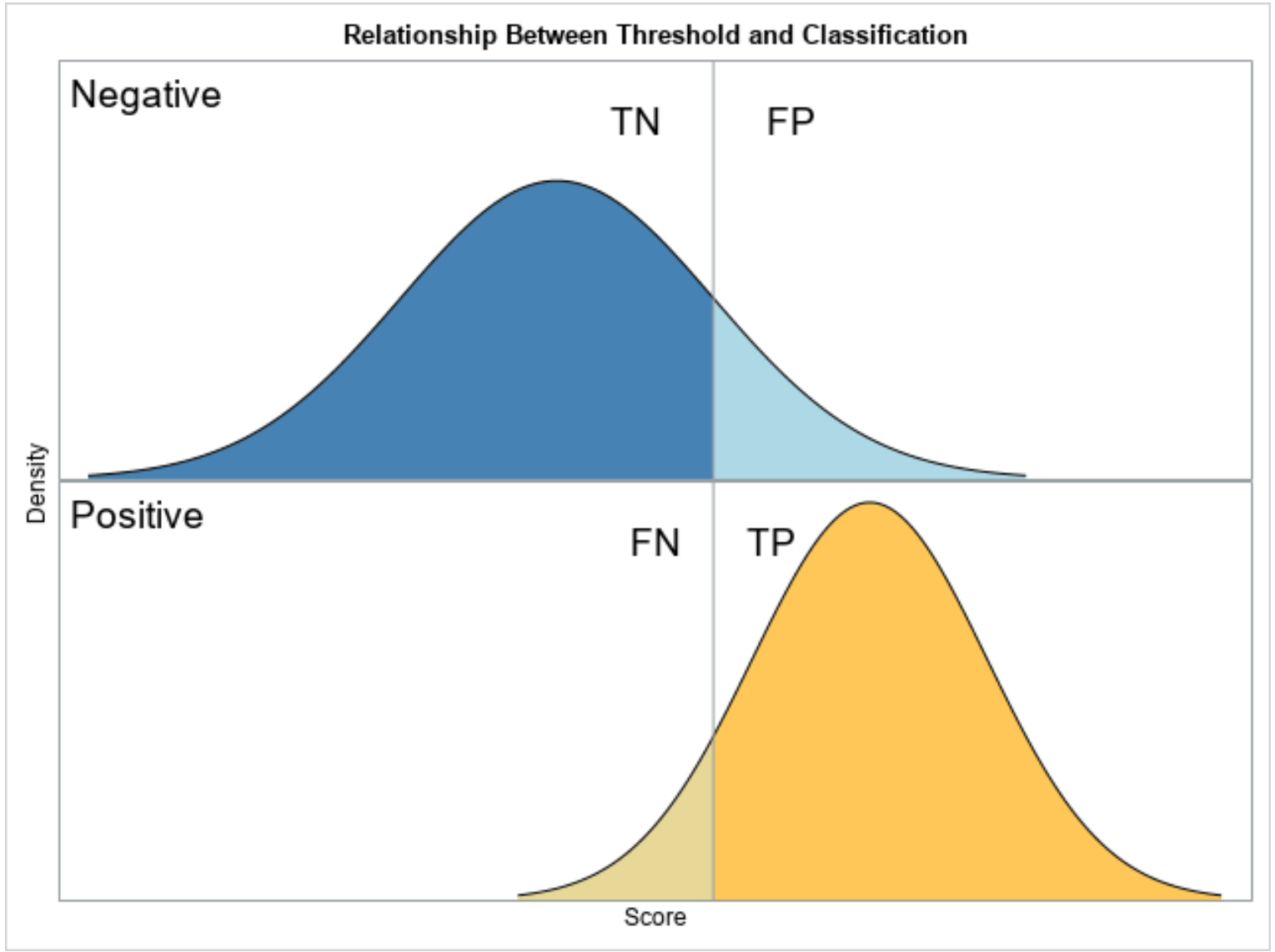
ITEM RESPONSE THEORY

$$f_i(u_i|\theta) = P_i(\theta)^{u_i} Q_i(\theta)^{1-u_i}$$

$$f_i(u_i|\theta) = P_i(\theta) \; if \; u_i = 1$$

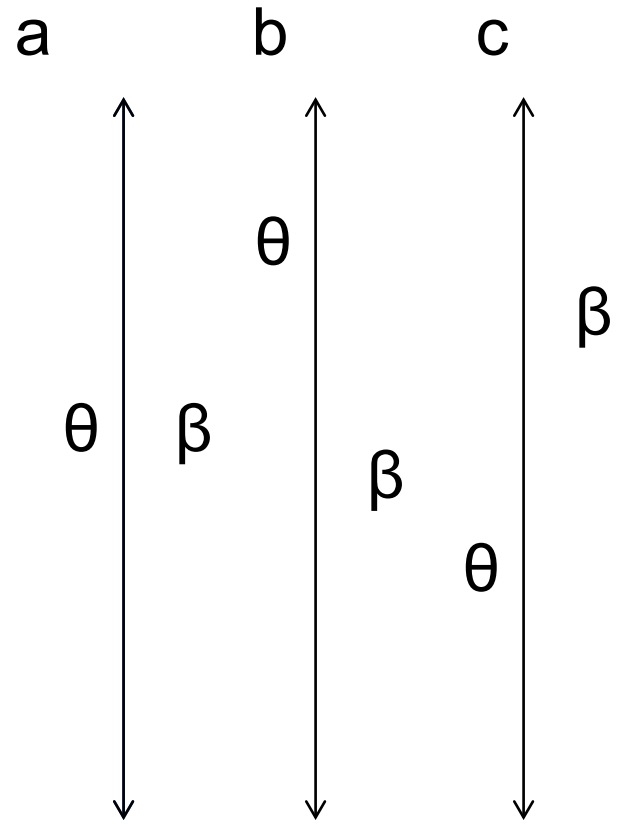
$$f_i(u_i|\theta) = Q_i(\theta) \; if \; u_i = 0$$

CLASSIFICATION THEORY



ITEM RESPONSE THEORY

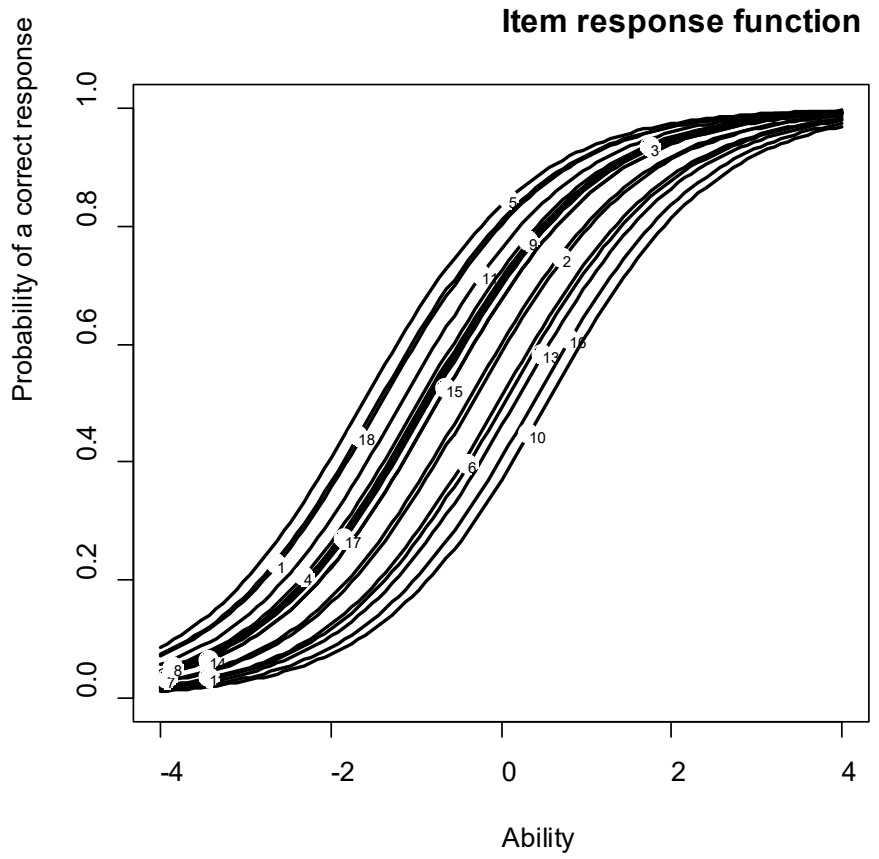
a. When the amounts are the same, $P(Y_i = 1|\theta, \beta) = 0.5$
b. When $\theta > \beta$, $P(Y_i = 1|\theta, \beta) > 0.5$
c. When $\theta < \beta$, $P(Y_i = 1|\theta, \beta) < 0.5$

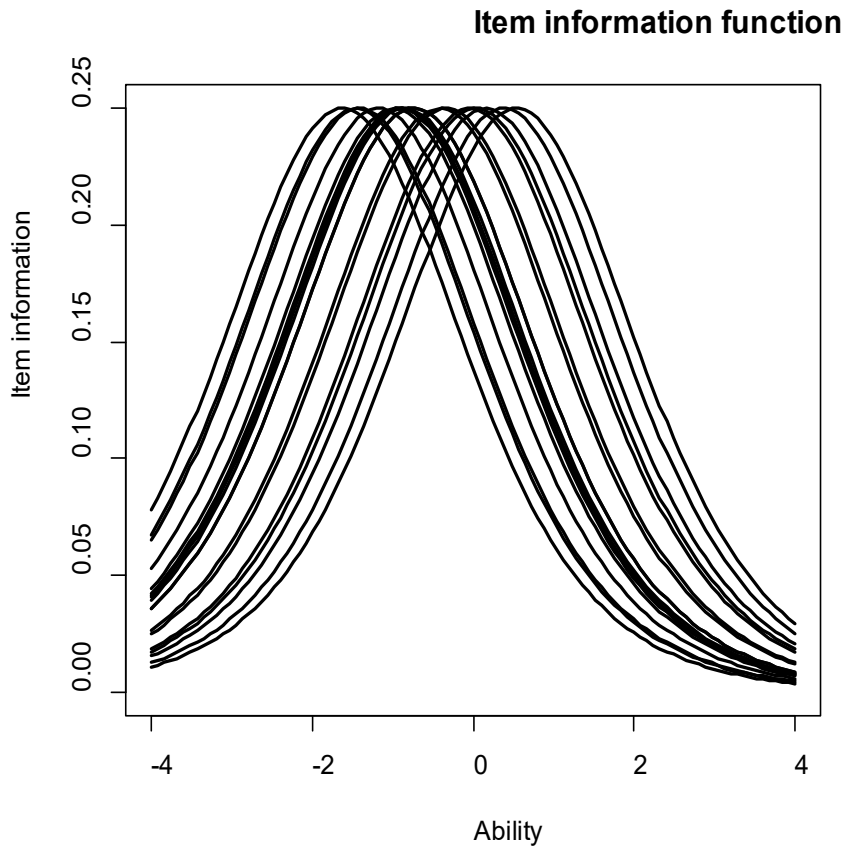


ITEM RESPONSE THEORY (1PL)

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

ITEMR ESPONSE THEORY (1PL)

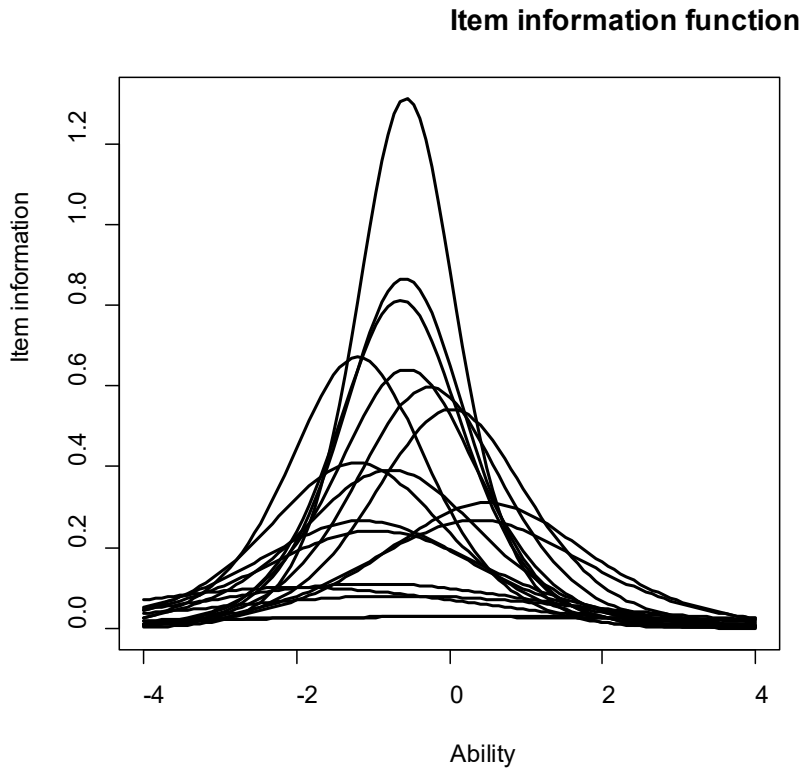


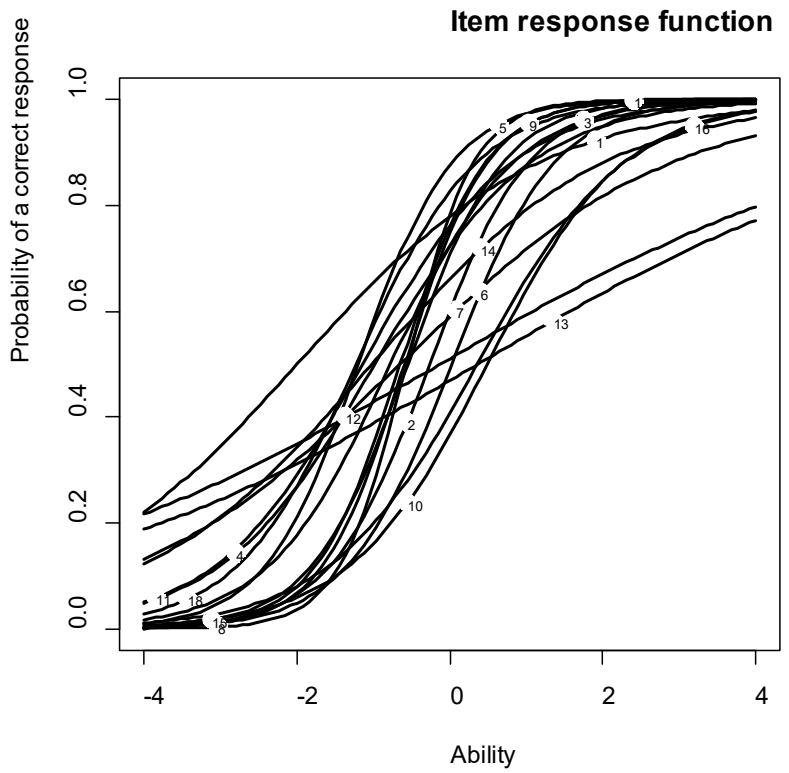


ITEM RESPONSE THEORY (2PL)

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

ITEM RESPONSE THEORY (2PL)

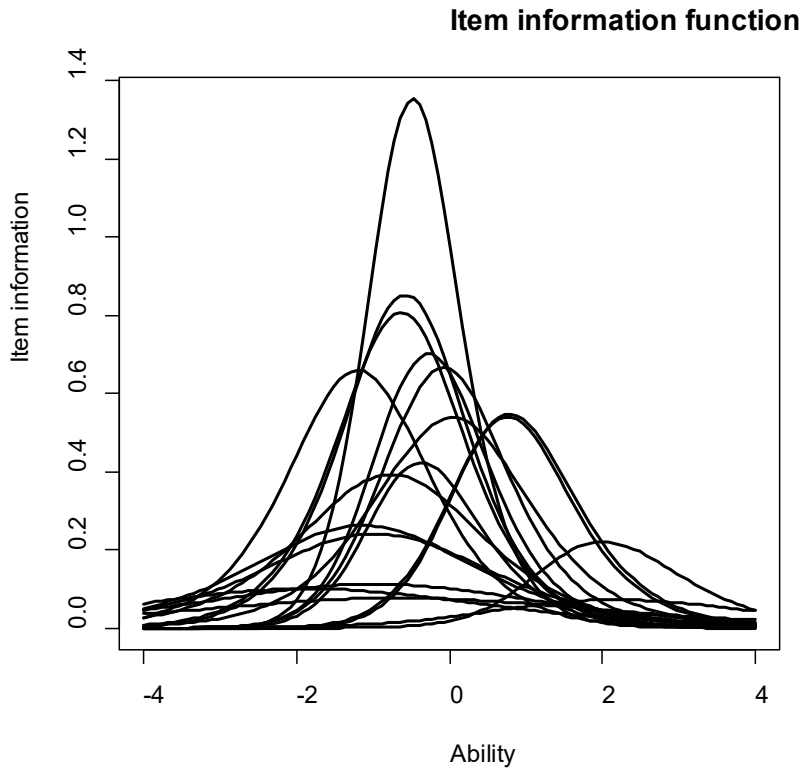


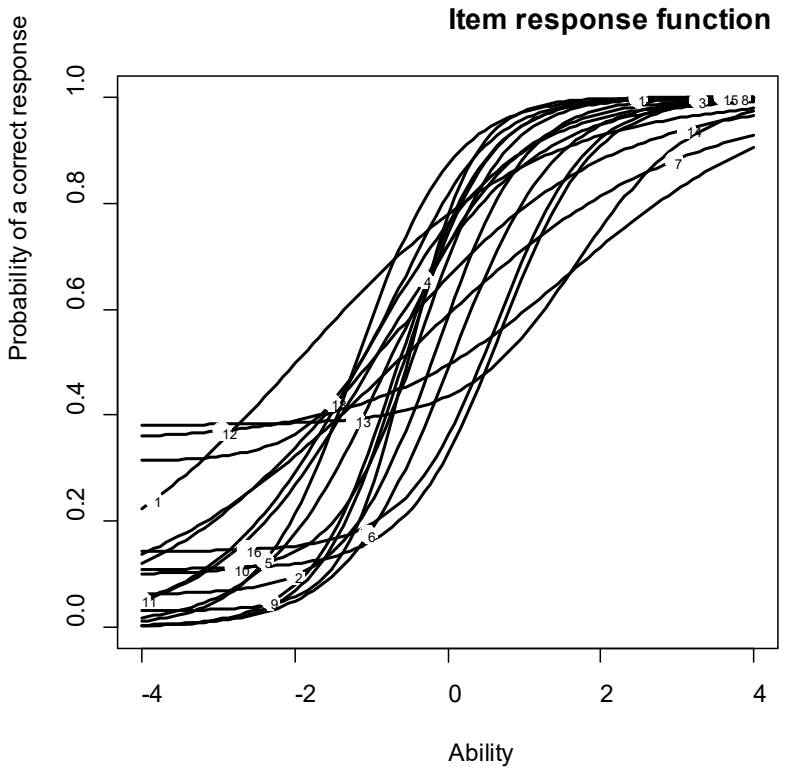


ITEM RESPONSE THEORY (3PL)

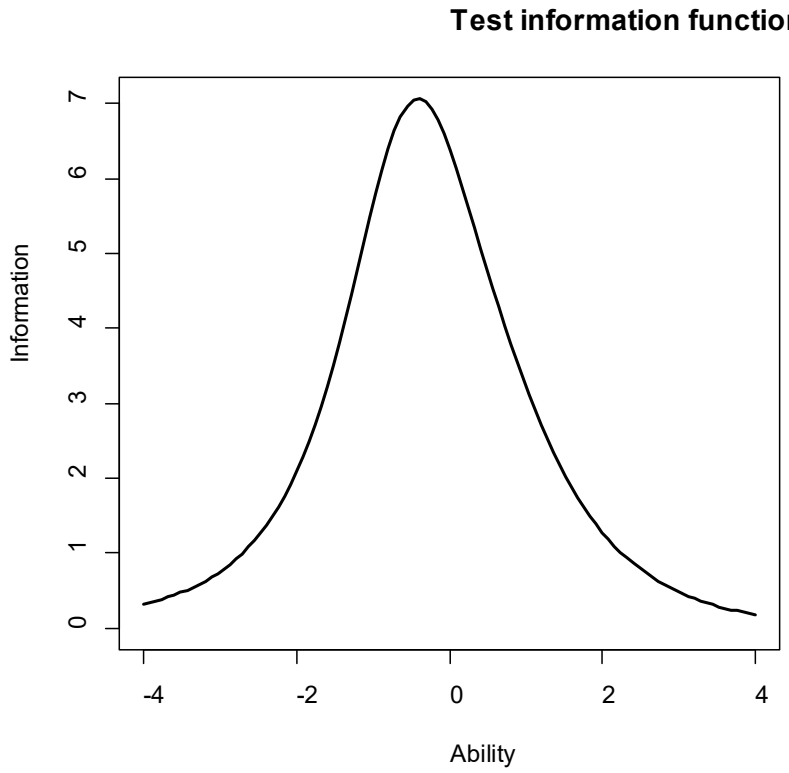
$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

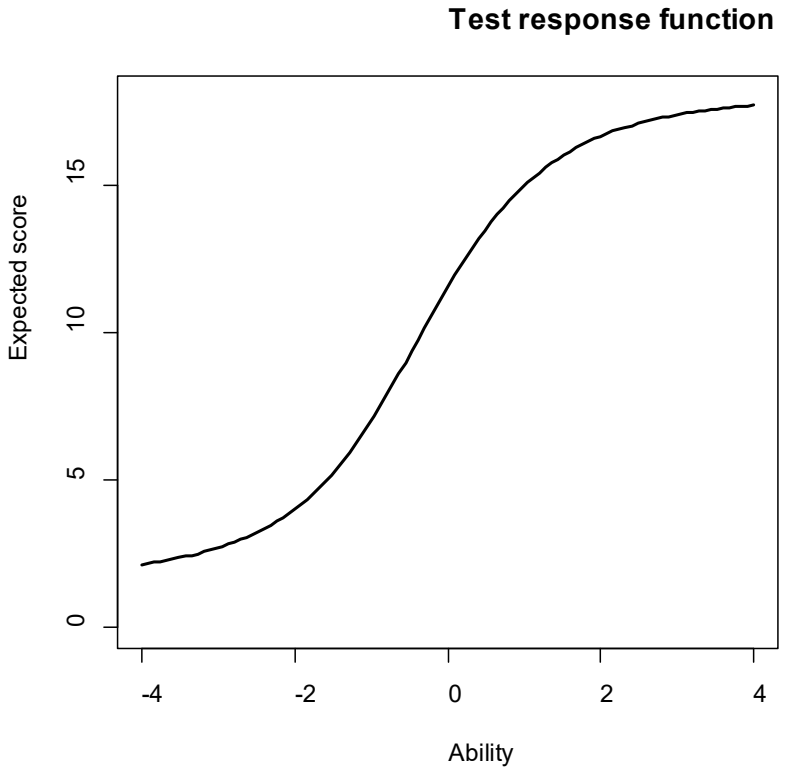
ITEM RESPONSE THEORY (3PL)

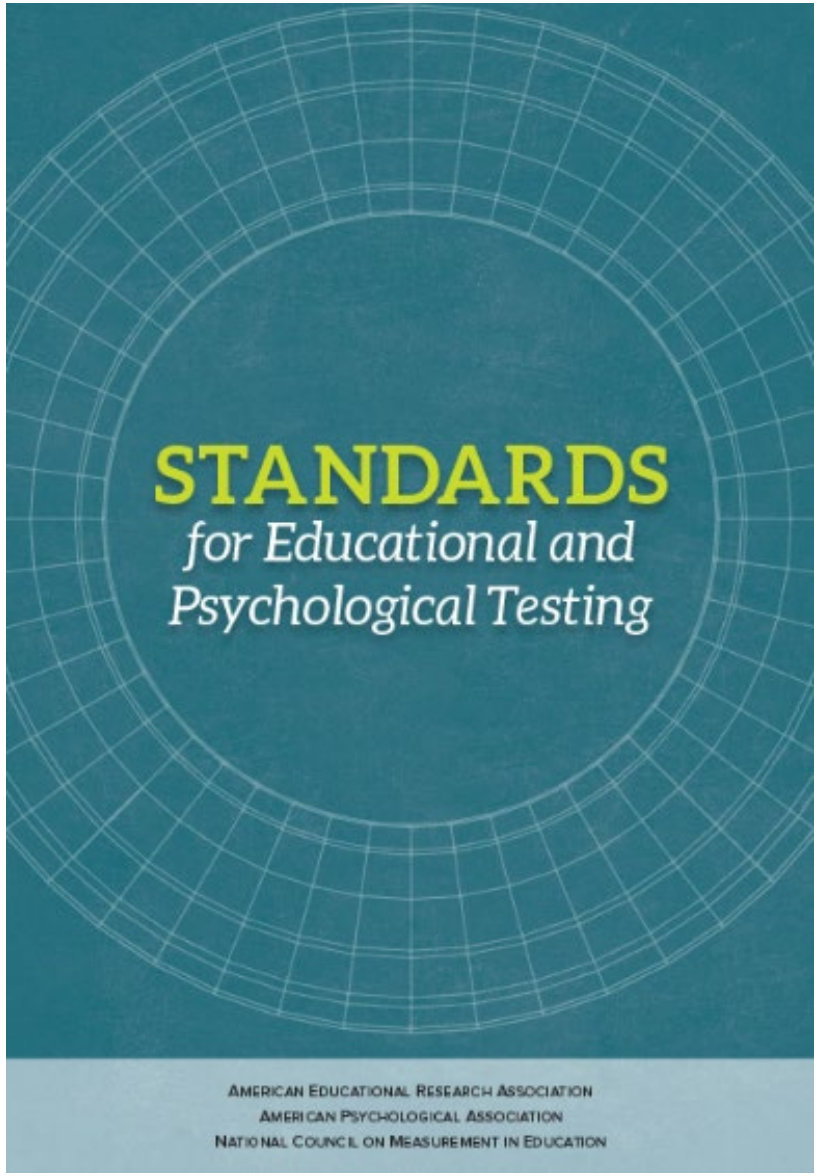




TEST-LEVEL CURVES






STANDARDS
for Educational and
Psychological Testing
American Educational Research Association
American Psychological Association
National Council on Measurement in Education