

ITEM RESPONSE THEORY

Ronald J. Nungester, National Board of Medical examiners

Marc E. Gessaroli, National Board of Medical Examiners

Zarrin Siddiqui, Faculty of Medicine, Dentistry and Health Sciences, University of Western Australia

Introduction

The widespread use of item response theory (IRT) over the past four decades attests to the importance of these models in the development and analysis of tests and items in medical education. IRT models have been used by both testing organizations and departments of medical education for a variety of purposes including item analysis, score equating, in the assessment of fairness issues as well as in the implementation of computer-based testing programs.

The properties of common IRT models, namely that “sample-free” item parameter estimates and “test-free” ability estimates can be obtained, make them particularly appealing to address a number of measurement issues. IRT models allow for the comparison of several groups of examinees who may have responded to different test forms (with some overlap). Additionally, IRT models enable the user to assess whether the test is well-targeted, in terms of its overall difficulty level, to the proficiency level of candidates.

Goals and Objectives

- (1) Provide participants with an overview of common IRT models, including the Rasch model, as well as two- and three-parameter logistic models.
- (2) Discuss practical applications of IRT in various aspects of medical education and assessment including item analysis, test development, comparability of performance across candidates and time, and issues relating to fairness.
- (3) Provide a comprehensive list of software packages that implement IRT models with advantages and limitations.
- (4) Integrate aspects covered in (1) through (3) by going through an actual data set, with emphasis on discussion of findings and interactive participation among attendees.