

Differential Item Functioning Analyses in Large-Scale Educational Surveys: Key Concepts and Modeling Approaches for Secondary Analysts

Xiao-Shu Zhu

Department of Measurement,
Statistics & Evaluation,
University of Maryland
Graduate Student

André A. Rupp

Department of Measurement,
Statistics & Evaluation,
University of Maryland
Associate Professor

Jing Gao

Office of Assessment & Evaluation,
University College
University of Maryland
Psychometrician

Abstract

Many educational surveys employ a multi-stage sampling design for students, which makes use of stratification and/or clustering of population units, as well as a complex booklet design for items from an item pool. In these surveys, the reliable detection of item bias or differential item functioning (DIF) across student groups is a key component for ensuring fair representations of different student groups. In this paper, we describe several modeling approaches that can be useful for detecting DIF in educational surveys. We illustrate the key ideas by investigating the performance of six hierarchical generalized linear models (HGLMs) using a small simulation study and by applying them to real data from the Trends in Mathematics and Science Study (TIMSS) study where we use them to investigate potential uniform gender DIF.

Keywords: complex booklet design, DIF, HGLMs, multi-stage sampling design