

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

教育科學研究期刊 第五十六卷第一期
2011 年，56 (1)，91-127

大型教育調查研究中的差別試題功能： 次級分析中的核心概念及建模方法

朱小妹

馬里蘭大學
測量統計評估系
研究生

安德魯·儒普

馬里蘭大學
測量統計評估系
副教授

高靜

馬里蘭大學
大學學院
測量評估辦公室
心理測評師

摘要

大型教育評量研究常採用多階段抽樣的設計 (multi-stage sampling design)，透過對母群體之抽樣單位進行分層以抽取受測者。此外，還會採用複雜題本設計 (complex booklet design) 的方式將題目組成多份測驗題本。在此情況下，欲確保公正測量出不同受測群體的能力，關鍵在於能夠有效偵測所採用的題目是否具差別試題功能 (differential item functioning, DIF)。本文旨在介紹探討在大型教育評量複雜設計之下能用以偵測差別試題功能的建模方法，並應用六種可用於偵測 DIF 的多階層廣義線性模式 (hierarchical generalized linear models, HGLMs)，再透過電腦模擬比較它們偵測 DIF 的效力。接著又將這些模式應用到國際數學與科學教育成就趨勢調查研究 (TIMSS) 的實證數據上，藉以探測是否存在一致性的性別 DIF (uniform gender DIF)。

關鍵字：複雜題本設計、差別試題功能、多階層廣義線性模式、多階段抽樣設計