

差異試題功能分析—以全民英檢為例

Shih-Sheng Huang & Chia-Lung Lee
The Language Training & Testing Center (LTTC)
stat@lttc.ntu.edu.tw

全民英檢測驗(General English Proficiency Test, 簡稱 GEPT)自民國 89 年開辦至今已邁入第 11 年, 每年約有 40 萬考生參加 GEPT 測驗, GEPT 成績已成為國內許多學校及公、私營單位在申請入學、畢業、甄選及升遷的參考依據, 測驗結果攸關考生的權益, 因此 GEPT 是否具有「公平性(test fairness)」是一項不容忽視的重要議題。測驗公平性自 1960 年代即普遍被歐美國家所重視, 當相同能力卻隸屬不同群組的考生(例如男、女)在作答相同試題卻有不同的得分情形或答對機率, 則試題具有試題差異功能 (Differential Item Functioning, 簡稱 DIF), 而可能違反測驗公平性。

本研究分析 97 至 99 年的初、中及中高級聽力測驗在性別與居住區域之試題差異功能。DIF 試題的檢測方法採用 Mantel-Haenszel 法(M-H 法)與 Logistic regression 法(LR 法), 並使用 DIF 嚴重度分類系統(Dorans & Holland, 1993)判斷 DIF 試題的嚴重程度。研究結果顯示: 97 至 99 年的初、中及中高級聽力測驗在性別與居住區域的 DIF 試題比例相當低, 表示全民英檢聽力測驗具有測驗「公平性」。