

Yes/No Angoff 法在華語文聽力測驗與閱讀測驗的應用

藍珮君^{1*}、陳柏熹²、林玲英¹、李嘉真¹、孫雪芹¹、曾文璇¹

¹ 國家華語測驗推動工作委員會

² 國立台灣師範大學教育心理與輔導學系

摘要

標準設定為一透過合理的、規範的程序和步驟建立測驗通過分數的方法，因而逐漸廣為許多大型語言測驗或學習成就評量資料庫所使用，藉以將學習者的學習成果或成就做適當的分類。

由國家華語測驗推動工作委員會(簡稱華測會)研發之華語文能力測驗(Test of Chinese as a Foreign Language, 簡稱 TOCFL), 自 2008 年起進行改版, 調整題型與命題取向, 使測驗內容更符合「歐洲語言共同架構 (The Common European Framework of Reference for Languages, 簡稱 CEFR)」(Council of Europe, 2001)溝通任務之精神, 並與 CEFR 等級接軌, 讓考生清楚瞭解測驗成績所代表的語言能力。也進一步調整測驗架構, 於 2013 年推出新版華語文能力測驗, 將語言能力分成三等六級, 三等分別為入門基礎級、進階高階級及流利精通級; 每一等可再依據測驗成績細分為兩級, 分別為入門級、基礎級、進階級、高階級、流利級、精通級, 共六級。

由於測驗架構已做了改變, 訂定新的通過門檻乃有其必要性。本研究目的在採用 Yes/No Angoff 法(Impara & Plake, 1997)對華語文聽力與閱讀測驗進行通過門檻的標準設定研究, 使其符合 CEFR A1 至 C2 各級之能力描述, 並對結果進行效度檢核。

本研究邀請華語文領域教授及資深華語教師參加, 一共舉辦了六場標準設定會議, 每次參加人數為 9 至 13 人次不等。每等試卷標準設定程序大致相同, 以入門基礎級為例, 研究者首先說明標準設定流程, 接著簡介試題反應理論, 目的為讓成員理解試題難度參數的意義, 之後提供的試題難度才能有助於成員做判斷。然後介紹 CEFR 架構與內容, 講述入門級與基礎級學習者能力的最低表現描述, 確認與會成員皆理解後, 發下入門基礎級 50 道試題測驗卷與填答表, 請成員實際進行作答並逐題判斷。

若成員認為入門級最低學習者可以答對此題, 則於入門級相對應欄位打勾標記, 判斷下一題; 若認為入門級最低學習者無法答對, 則再判斷基礎級最低學習者能否答對, 若能答對, 同樣在基礎級相對應欄位打勾標記, 判斷下一題; 若認為基礎級最低學習者亦無法答對, 則不做任何標記, 進入下一題。

¹ 藍珮君, 國家華語測驗推動工作委員會研究助理

電子郵件: martinalan@sc-top.org.tw

² 陳柏熹, 國立台灣師範大學教育心理與輔導學系副教授

電子郵件: chenph@ntnu.edu.tw

研究者在第一與第二回合結束後，根據成員判斷結果與經由預試獲得之試題難度參數，計算出入門級與基礎級通過門檻的答對題數與考生能力值；此外，也將 50 道試題之難度參數表，與以此暫時門檻(考生能力值)計算過去正式考試考生通過入門級與基礎級之比例，一併提供給成員參考，以促進討論的進行、凝聚判斷的共識。最後，研究者於第三回合結束後進行問卷調查，了解與會成員在會議過程中的想法，並作為後續標準設定會議流程調整的參考。

三等卷的標準設定結果，無論是聽力或閱讀測驗，各等級通過門檻平均答對題數，皆隨著每回合討論進行而逐漸降低。考生能力值的部分，考量各個等級之間能力值的差距不宜過於接近等因素，最終在部分等級進行微調，完成了聽力與閱讀測驗六個等級的通過門檻。

標準設定結果的效度檢核分為程序性效度、內部效度，以及外部效度三部分進行說明。首先，程序性效度方面，每次標準設定會議皆按照固定流程進行，且會議後的問卷調查結果顯示，成員大多同意會議帶領者對於標準設定方法的操作流程說明得很清楚、每回合後團體中的討論和分享，有助於進行下一回合的判斷、對於自己所設定的切截分數有信心……等等，可做為本研究之程序性效度來源。

內部效度則由每一回合估計考生能力值之標準差，及成員判斷試題難度與實徵難度參數之積差相關作為依據。標準差部分，無論是聽力或閱讀測驗，六個等級第二回合的標準差均低於第一回合；第三回合的標準差結果，除了閱讀測驗基礎級、進階級與流利級略高於第二回合以外，其餘皆低於第二回合。表示經過討論，判斷結果的一致性逐漸提高。從相關分析結果可發現，三回合的討論中，成員判斷試題難度與實徵難度間的相關越來越高，第三回合聽力測驗六個等級的相關係數介於.557 至.763 之間($p < 0.05$)，閱讀測驗則介於.492 至.816 之間($p < 0.05$)，皆有中等至高等的正相關存在。上述二項分析結果顯示標準設定結果具有內部效度。

至於外部效度，過去研究指出語言學習者的聽力與閱讀能力皆屬於語言的接收能力(receptive skill)，兩者間的表現具有正相關(Bae & Bachman, 1998)，故以 2013 年五月華測會舉辦之正式考試中，報考同等級聽力與閱讀測驗考生之通過等級進行交叉分析。結果發現，入門基礎級、進階高階級與流利精通級考生聽力與閱讀測驗結果被歸類為相同等級的比率分別為 76.5%、59.0% 以及 63.0%；Kappa 分析的結果依序為.45($p < .01$)、.30($p < .01$)以及.39($p < .01$)，依照 Landis 和 Koch(1977)提出的判斷標準，數值介於.21~.40 為普通一致(fair)、.41~.61 為中度一致(moderate)，說明本研究之標準設定結果的分類一致性在進階高階級與流利精通級已達到普通一致，入門基礎級更達到了中度一致的程度。

華語文聽力與閱讀測驗入門基礎級、進階高階級與流利精通級三等試卷的標準設定結果，在程序性效度、內部效度以及外部效度三項效度證據均獲得支持，表示新版華語文聽力與閱讀測驗，將華語學習者依其報考等級與測驗表現，區分為 CEFR A1 至 C2 六個等級的有效性得到驗證。

關鍵詞：華語文、聽力測驗、閱讀測驗、Yes/No Angoff 法、標準設定