

壹、前言

隨著國際的趨勢與潮流，臺灣教育決策者漸漸意識到，建置長期學習成就評量資料庫的需求，以作為研訂課程與教學政策之重要參考依據；臺灣學生學習成就評量資料庫（Taiwan Assessment of Student Achievement, TASA）便在此背景下應運而生。其建置目的，在於追蹤、分析臺灣國小、國中，以及高中職學生之學習成就表現，藉以探討學生學習成就上之表現差異與學習上發展之趨勢，進而檢視當前國家教育體制與政策實施之成效；據此，其設置最重要的目的之一，乃在於描述臺灣學生於各學科的學習成就表現概況。學習成就表現的描述即牽涉到標準設定議題，而世界各大型學習成就資料庫，對於標準設定之方式不盡相同。國際數學與科學教育成就趨勢調查（Trends in International Mathematics and Science Study, TIMSS）、國際學生成就評比計畫（Programme for International Student Assessment, PISA），對於學生表現概況的描述，乃建立在相對標準上面。TIMSS 以固定間距（即百分位數）進行切截分數訂立（Mullis, Erberber, & Preuschoff, 2008），例如：美國國小四年級學生在 TIMSS 2007 自然科，有 6% 低於初級國際基準點（low benchmark）以下，而臺灣只有 3% 的學生低於初級國際基準點以下。而 PISA 對切截分數設立的想法與作法，某程度亦雷同於 TIMSS（Organization for Economic Cooperation and Development [OECD], 2009），決策者對於學生表現的詮釋方式，乃透過相對比較來呈現；接續，再以實徵數據篩選適當試題，進行各國際基準點的質性能力描述，而此詮釋模式的產生亦多為呼應其國際評比的評量目標。

另一方面，就本研究探討之 TASA 自然科學習成就評量而言，其 2009 年的評量架構係參酌國民中小學九年一貫課程綱要中，自然與生活科技學習領域的分段能力指標（教育部，2003），以效標參照評量（criterion-based assessment）的精神進行測驗的設計；亦即對於測驗表現結果的詮釋建立在絕對標準上，探討各年段的學生經過學習後，對應課程所規劃的內容標準（content standards）時，位於不同水平（level）之學生，所具備之不同能力或知識。TASA 不採用國人較熟悉的常模參照來詮釋測驗結果，而是藉由標準設定（standard setting）的方式來闡釋，期望盡可能降低學生同儕競爭之壓力。在此理念下，吳宜芳、鄒慧英及林娟如（2010）曾針對 TASA 數學科建立各水平的標準，而吳毓瑩等（2009）則就 TASA 之英語建立通過／不通過的標準，並同時輔以適切的標準設定效度證據，以支持其標準設定結果。

時至今日，TASA 自然科標準設定之後繼研究較少，且隨著現行評量架構之調整，致使沿用的表現標準已多不符合實務運用，而興起重新修訂之舉。據此，本研究目的為評估 2009 年臺灣 TASA 國小四年級自然科標準設定結果，輔以過程、內部及外部等多元效度證據，最後提出數項建議，供未來標準設定研究者參考。

貳、文獻探討

本節首先說明 2009 年 TASA 自然科所引用標準設定過程、技術之文獻基礎，以確立其理論依據，接續，再陳述評估標準設定結果可能的效度證據來源及本研究採行之準則，茲逐一陳述如下：

一、標準設定流程之理論基礎

在 2000 年以後，對於執行標準設定的流程或技術，各學者所提出的概念十分豐富及多元。本研究主要擷取 Cizek 與 Bunch (2007) 所著專書 *Standard setting: A guide to establishing and evaluating performance standards on tests* 所提及之標準設定共同核心元素為架構，再輔以美國國家教育進展評估 (National Assessment of Educational Progress, NAEP) 執行標準設定時，所提供的實務經驗為標準設定藍圖。此舉乃因本研究標的之建置理念和目的，十分貼近 NAEP 評量的取向，因而適合作為標竿。

Cizek 與 Bunch (2007) 認為執行一項標準設定，其流程大致包含幾個重大面向，分別是：澄清測驗目的、訂定表現標準個數並命名、建立表現標準描述、選擇及訓練標準設定成員、執行標準設定、技術選擇、訊息回饋、過程監控、形成切截分數、以及評估標準設定過程與結果。其概念的組成大致可形成如圖 1 所示，茲說明如下。

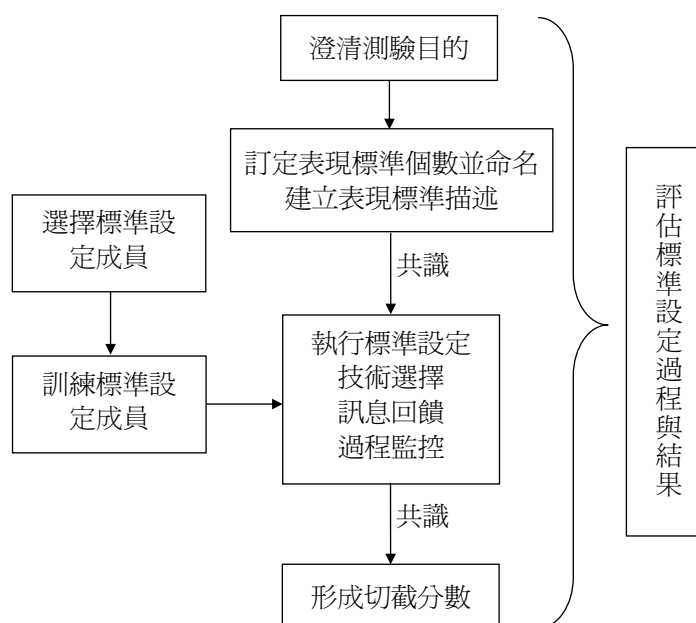


圖1 標準設定流程