

壹、緒論

隨著國際的趨勢與潮流，國內教育決策者漸漸意識到建置長期學習成就評量資料庫的重要性，臺灣學生學習成就評量資料庫（Taiwan Assessment of Student Achievement, TASA）在此背景下應運而生，其建置目的在追蹤與分析學生學習成就表現，藉以探討學生學習成就之表現差異與學習變遷趨勢，進而檢視當前國家教育體制與政策實施之成效。因此，評定或描述學生目前的學力表現，是否達到政策決策者所預期的程度或標準則是一項重要的工作，爲了能達成此任務，必須進行學習成就之標準設定（standard setting）。標準設定能夠幫助受試者了解本身成就表現水準，許多大型測驗通常會建立測驗本身的決斷分數（cut scores）。例如：國家教育進展評量（The National Assessment of Educational Progress, NAEP）、國際學生評量（The Programme for International Student Assessment, PISA）、國際數理趨勢研究（The Trends in International Mathematics and Science Study, TIMSS）等。此外，在教育測驗情境中，不僅是受試者的表現水準，甚至學校、學區或縣市也常需要進行評估分類。因此，標準設定不論是在教育或證照發放之用途，都扮演著相當重要的角色。近幾年，伴隨著試題反

應理論（item response theory, IRT）與相關統計、評量、電腦科技的發展，更使得標準設定的相關議題愈受到重視（謝進昌，2006）。標準設定原本是指在測驗中建構一個或多個切割分數的步驟，至於是一個或多個切割分數，則需根據這個測驗的性質而定。本研究鑑於過去TASA學習成就之標準設定缺乏系統性，且各學科評量架構已有所修訂，是故，若沿用舊有設定之標準較不符合實務運用，實有重新修訂之需求。

由於在學習的表現或是一些政策的擬訂，這些標準所該具備的條件，決不是僅靠專家或學科領域的專家各自表述，或是一些約定俗成的門檻來做判斷的依據。必須要考量學生的能力，以及其他的相關因素，例如：特定政策、整體的社會環境發展或是一些學生的實徵資料。因此，從1980年開始，就不斷的有學者提出一些標準設定的方式，企圖利用一些科學的程序來建立一個最合適的標準。大致而言，在教育測驗的相關文獻中已有許多設定標準的方法被提出來，主要以Nedelsky（1954）、Angoff（1971）與Ebel（1972）等人所提出的三種方法較常被使用。其中，NAEP自1994年起，使用多種延伸的Angoff方法進行成就水準的設定，並將其使用於二元計分試題（dichotomously scored items）與多點計分試題（polytomously scored items）中。Bourque（2009）指

出NAEP透過模式將標準設定成員評判之結果對應到NAEP能力量尺中，以確認在測量量尺上對應成就水準的決斷點，並依據決斷點合併內容線索再重新量尺化至NAEP的分數量尺，以確認每個年級基本、精熟、進階成就水準的決斷分數。Bourque與Hambleton（1993）亦指出Angoff提出的方法能適用於多種試題形式，且該方法能結合檢視試題內容與真實受試者反應的技術，相當符合NAEP評量試題研發之取向。此外，Lewis、Mitzel與Green（1996）提出之書籤標定法（bookmark method）融合IRT技術與Angoff方法之概念，因此，Lewis、Green、Mitzel、Baum與Patz（1998）將其稱為修正的IRT-Angoff方法（IRT-Modified Angoff Procedure）。Cizek與Bunch（2007）指出書籤標定法包括以下幾個優點，使其逐漸讓學者所採用。

一、能應用於較複雜的測驗題型，例如：混合式的測驗題型（mixed-format），包括選擇題（selected-response）與建構式反應題（constructed-response）。

二、相對於其他標準設定方法而言較為簡單，使得參與的成員能夠簡化標準設定程序，且在進行標準設定會議時，此方法能減少潛在的誤差與節省時間。

三、由於以IRT模式為基礎，標準

設定分數與測驗結果解釋以同樣方式進行，因此，在沒有其他IRT模式為基礎的標準設定程序下，此方法具有一定優勢。

是故，本研究以TASA國小六年級社會科實徵資料為例，建立一套適合TASA的標準設定流程，於建立後提出具體建議供學習成就評量實務參考。並監控標準設定過程，包含：團體成員間評定差異、不同背景團體成員間評定差異、過程的評估問卷、標準的一致性、成員對於極端判定之評估性反應、各輪間判定結果的改變、提供影響訊息後成員的判定改變等指標。以評估設定過程的適切性，同時將評估的結果做為過程效度證據。

貳、文獻探討

一、標準設定實施程序

自2000年起，對於執行標準設定的流程或技術，許多學者所提出的概念十分豐富及多元。由於TASA建置之理念與目的與NAEP相當貼近，適合做為本研究設計之標竿。因此，本研究擷取Cizek與Bunch（2007）提出標準設定的共同核心架構，並以NAEP執行標準設定之實務經驗為藍圖，規劃適合TASA的標準設定流程。Cizek與Bunch指出在執行一項標準設定時，主要流程包含幾個面向，茲說明如下。

（一）澄清測驗目標