

第一章 緒論

數學是科學的基礎，國民小學學生數學成就水準的評估及學習困難的診斷，一直深受各先進國家之重視。如美國及澳洲等均有「基本能力測驗」，此種測驗在性質上比較接近標準參照測驗。實施基本能力測驗除了可了解一般學生之數學成就水準之外，也可作為國民小學教學成效之評鑑和改進教材教法或實施個別輔導之參考，甚至也可以作為品質管制（如，發給畢業證書）之重要依據。所有這些措施，均是為了提昇未來主人翁的數學及科學水準，以強化個體社會適應和問題解決的能力。

教育部委託國立台南師範學院測驗發展中心編擬「國民小學數學標準參照測驗」，主要目的是在發展一套適用於國小學生的數學標準參照測驗，以作為評估我國國小學生數學成就水準、協助診斷學習困難、及特殊學生篩選之參考工具。研究小組考慮經濟的因素，希望能以最低的成本建構出品質優良，而且生機充沛的測驗工具，以利不同目的的應用及後續的經營發展。因此，本研究以年級為單位，建構出二～五式一般學童適用的成就水準鑑別工具，同年級各式測驗間有共同試題，以題目反應理論（item response theory, IRT）模式，進行同時估計，以達成題目參數量尺化的目的。題目參數量尺化後，再依課程結構及題目難度重新編輯出基礎版和進階版兩套測驗，以利特殊兒童篩選的應用。此外為了考慮成長的測量，相鄰兩年級也安排共同試題，建立跨年級題庫，一方面利於年級間數學成長的探討，另一方面更可活化年級間題目共享的可能。換句話說本研究共包含 20 式測驗常模對照資料及分年級的四個年級題庫，同時也量尺化三和四、四和五、五和六及三至六年級等另外四個跨年級題庫。

除了經濟的考慮外，研究小組人員也儘力兼顧所研發工具的前瞻壽命。由於目前數學課程已準備進入另一個全面改革的階段，部份學校已開始採用新課程的教科書進行試用，其他非官方的教科書也即將進入教育市場，因此目前數學教育改革的重點絕對是必需納入考慮，以避免工具研發完竣立即失效的命運。研究小組在新的趨勢，舊的素材和教學文化中力求均衡。在尋求作業新穎、認知導向、內容複雜似真的努力中，我們發現可能源於小組人員經驗不足，也可能離現行教學和測驗模式太遠，考生挫折很大，預試結果多半不甚理想。基於對教學現況的尊重，小組人員不得不將嘗試的幅度縮小，大部份題目以問題解決為焦點，降低計算的比重，題目型式多採選擇式，只保留部份四、五年級數學推理溝通的半建構式題型。

另外一項主要的決定在於測量焦點和範圍的大小，大部份標準參照測驗多在非常同質的內涵中作精熟與否的推論，但是這樣一來測驗時間、題數就會非常龐大，研究小組在人力物力和數學教育前瞻的考慮下，選擇以較廣大範圍、較複雜作業的取向來搜集考生學習成就水準的資訊。但是單一個分數的摘要是不具太多描述性資訊，因此本研究採認知層次鑑別的方式，希望能提供考生目前精熟水準的判別資訊，同時能描述出該考生精熟水準的作業內涵及臨屆精熟水準的可強化方向。我們深信，每一位學生都需要受到肯定，也都需要在支持鼓舞的環境中繼續朝向明確的目標前進。測驗的回饋也許是粗糙的，但研究小組深信這樣的企圖是必要的，而且這些訊息可以隨著努力逐漸趨向豐富與精緻。

本研究擬以問題解決為主要焦點，兼顧課程內容及數學核心概念的認知層次來發展國小三至六年級標準參照測驗題庫，擷取課程核心概念，編製複雜而接近生活經驗的作業來檢視重要學習成果。我們在建立題庫時，同時兼顧各個能力層次考生能力估計的準確程度，尤其以篩選數學學習困難學生和數學資優學生為主要應用之考慮。研究中並透過層次的分析，就標準參照所提供的描述訊息，估計目前學生通過各認知層次的比率，同時討論認知層次結構適切性及各年級數學能力的成長差異，以及對資源班學童學習困難篩選與診斷的效益評估。最後，呈現為因應新課程標準的新評量方的努力—發展四、五年級數學溝通能力測驗的初步成果。