

## 壹、緒 論

性向測驗 (aptitude tests) 是設計來預測某些活動未來的表現。目前學校所使用的性向測驗，除了傳統的學術性向測驗 (scholastic aptitude tests) 及較綜合性的區分性向測驗 (defferential aptitude tests) 外，也包括創造力、藝術、音樂等特殊性向測驗。這些性向測驗中，使用最廣的是學術性向測驗。

由於學校的一項主要目標是協助學生充分發揮其潛能，因此，標準化性向測驗顯得相當重要。在決定學生的學習準備度 (learning readiness)、個別化教學、編班或分組、鑑定低成就者、診斷學習問題、及協助學生擬訂其教育和職業計劃時，均需估計學生的能力。雖然成就測驗也能提供上述重要決定的有用資訊，但是，性向測驗有其特殊的貢獻。

目前國內各級學校普設資賦優異及智能不足等特殊班級，顯示國內的教育越來越重視學生的個別差異，尤其對特殊兒童的鑑定與提供適性的教學更加重視。特殊兒童的鑑定屬於分類決定，錯誤的分類除了浪費人力、時間及金錢外，可能對學生的身心發展也有不利影響，故如何減少分類誤差至為重要。

分類誤差與測驗效度具有函數關係。良好的性向測驗可以減少分類誤差，國內過去已修訂不少學術性向測驗，但大多數久未重新修訂，常模已過時，而且使用頻繁，已失去保密性，故測量所得結果之信效度都值得懷疑。為了提昇學術性向測驗之品質，宜有系統建立學術性向測驗題庫。

一個良好的題庫必須相當大，且能不斷加以擴充，並具有相當程度的保密性。題庫內的題目可依測驗對象及目的隨時挑選適當的題目來組成測驗。施測的方式除了可印成題本以傳統紙筆式測驗施測外，最好也能實施電腦化適性測驗 (computerized adaptive testing, CAT)。為了要達成此項目標，題庫內的每個題目均需有題目反應理論 (item response theory, IRT) 的題目參數，並採用題目參數連結 (linking) 方法來擴充題庫，這樣就可有系統改進題目品質，提高測驗的信效度，而且只要題庫夠大，並有適切的管理方法，那就不太需要顧慮試題外洩的問題。

大型題庫的建立不是一蹴可幾的，故題庫的建立雖然很理想，但需分

成數個階段來進行，第一個階段是先發展一份具有全台灣區常模的測驗，以後再以較小樣本，採共同試題設計的方法，逐漸將新試題加入原有題庫。本研究即為整個學術性向測驗電腦化題庫建立的第一階段工作。

## 貳、學術性向測驗編製的理論基礎

### 一、學術性向測驗之意義及與成就測驗之異同

成就測驗與性向測驗有某些相似的地方，但也有一些相異之處。一般認為二者主要不同在於「成就測驗是測量學生已學得的知能，而性向測驗是測量學生學習新知能的能力」。此種觀點似乎能明晰區分二者之不同，但是卻過分簡化這個問題，而且未能指出二者某些重要的相似點和差異處。事實上，這兩類測驗均在測量學生已學得的知能，而且二者對能否成功學習新工作的預測也都很有用。成就與性向測驗二者主要的不同是在二者所測量的學習類型及最有用的預測類型稍有不同（Gronlund，1985）。

(一)所測量的學習類型：能力譜（the ability spectrum）

Cronbach（1984）曾以圖 2.1 的能力譜來說明成就和性向測驗所測量的各類學習結果之同異。在這個能力測驗譜中，成就測驗含括層次 A 和 B，而性向測驗含括層次 C，D，和 E。此譜是依照測驗內容依賴特定學習經驗的程度來對各種測驗加以分類。最上端（層次 A）是測量特定課程內容知識的「內容導向成就測驗」。最下端（層次 E）是測量不易受直接訓練影響的那類學習的「文化導向非語文性向測驗」。當我們從層次 A 逐漸移至層次 E，測驗內容就越來越不受任何特定學習經驗的影響。能力測驗譜上越是相近的兩類測驗，其所測量的學習類型也就越相近。例如，設計來測量一般教育發展的成就測驗（層次 B）和以學校所學能力為依據的學術性向測驗（層次 C），二者所測量的能力很類似，因此，二者可預期有相當高的相關。同樣地，在能力測驗譜上距離較遠的兩個測驗，由於共同的部分較少，因此，其相關也就較低。此種資訊有助於性向測驗的選擇和使用。例如，我們可以預期層次 C 和 D 的性向測驗，對於學校成就的預測要優於層次 E 的性向測驗。另一方面，如果我們主要的興趣是在鑑定學生未開發的學習潛能（potential），那層次 E 的測驗就比較適合。圖 2.1