

壹、前言

日新月異的電腦科技與認知科學進展已對傳統測驗造成極大的變革。電腦科技推動測驗和心理計量理論快速演進，尤其在電腦化適性測驗（Computerized Adaptive Testing, CAT）的應用上發揮到淋漓盡致。儘管CAT日趨成熟使得更富彈性和人性的測驗安排變成常態。然而，背後的題庫建立卻是十分費時、費力和高成本的工作。相關的命題、組卷、實地預試、建立信效度和常模、校準量尺（calibration）等繁瑣事務都是高品質測驗所不可或缺的。然而，連續性測驗（continuous test）往往衍生出試題大量曝光的問題，危及測驗的保密性，試題的耗損率相對提高，需要後續的題庫修訂，補充更豐富的試題，以維持測驗品質。為了解決大量試題的需求，最近幾年許多學者積極地朝自動化命題領域進行探究（Glas & van der Linden, 2003; Irvine, 2002）。

在智力研究的領域中，推理能力測驗一直是智力測驗的核心，是了解人類智力時非常重要的一個面向（Sternberg, 1986）。其中又以瑞文氏圖形推理測驗屬於最典型和最廣為使用的非語文形式推理能力測驗，十分單純、經濟且有效（Raven, Raven, & Court, 2000）。

運用自動化命題技術於圖形推理測驗具有相當高可行性與應用的潛力，特別是發展適合於國小階段的圖形推理測驗值得深入探究，以便清晰掌握試題難度變異來源，一旦自動命題機制可運轉，試題的產出就變得更加有效率，能滿足CAT對大量試題的需求，挾帶著資通科技多媒體和互動的優勢，而後有關學童圖形推理能力的評估，就可以採取更人性有效的方式進行。

貳、研究背景與目的

一、試題複製方法——試題模型與試題結構

為了解決大量試題的需求，Bejar（1996）提出以試題模型（item model）的形式進行試題複製（item cloning）的概念。試題模型法是一種有別於傳統測驗發展的一種產生測驗的方法，是以具有良好心理計量的試題為基礎（模板或原型），在內容和難度上具有廣泛代表性，且經過校準好的試題作為起始點。在實做時，透過替換與問題解決歷程無關的特徵，如物體名稱、具體數字等，複製新的試題，也就是產生多個同構異形題（isomorphic instance）。此一方法假定這些新的試題外表看起

來與原來的試題不同，但實質內容和心理計量特性上都被視為與試題模型相似，因而可直接使用試題模型參數估計受試者的能力，而不需要對產生的試題再逐一進行實徵的參數校準。Bejar等人（2003）認為，試題模型法有助於減低命題成本，並以GRE數量推理測驗證實該方法在CAT應用的可行性。

此外，Embretson（1998, 1999）強調構念效度是探討測驗品質時最核心的議題，主張測驗發展必須與認知心理學緊密地結合，線上出題系統應當掌握測驗構念效度。基於構念效度系統化的概念，Embretson提出認知設計系統法，以試題結構（item structure）作為出題的依據。試題結構不僅界定與解題歷程關聯的刺激特徵組合，也能預測其心理計量參數，某些刺激項屬性可直接與試題難度或其他心理計量特徵（如鑑別度）直接關聯。

無論是試題模型法或試題結構法，主要差異在測驗內容領域定義對認知理論的依賴程度。一般而言，試題結構法對於認知心理學的依賴較高，適用的領域比較狹窄，侷限於具有完整的心理學理論或穩固認知研究支持的試題類型。

二、試題複製的參數變異對能力估計準確的影響

無論是試題模型法或認知系統設計法，在對命題設計原則（如item model或item structure）進行參數預測或校準時，不可避免會產生同構異形度（isomorphism）（Bejar et al., 2003）或參數不確定性（parameter uncertainty）（Embretson, 1999）的問題。Bejar等人（2003）利用所謂的期望反應函數（Expected Response Function, ERF）來提供試題模型的校準參數。ERF是一種以不確定性函數來調整（attenuate）參數估計的程序，以獲得更精確的測驗分數。一般的作法是將試題模型的所有試題特徵曲線（Item Characteristic Curve, ICC）加以平均，也就是將某一 θ 值的反應概率加以平均。愈高的同構異形度的試題模型，其ERF愈貼近於產出試題（instances）的ICC，因而能力估計的精確度愈高。計算ERF需要對每個試題模型的平均參數向量（ β ，包含鑑別度、難度和猜測度），以及參數間的變異數——共變數矩陣（variance-covariance matrix, Σ ）進行估計。透過對相等的受試樣本施測多個產出試題，將校準後的試題參數加以計算平均數和變異數——共變數矩陣（ β 和 Σ ），最後可獲得這些估計值。另外，Embretson（1999）也提出類似的概念，將參數估計的誤差變異分解為兩個來源：參數本身估計誤差變異與試題參數不確定性（item parameter uncertainty）誤差變異，並採用Rubin（1987）多重插補（multiple imputation）的方法，透過模擬研究不同模式下參數變異對能力估計的影響。