

壹、緒論

個體之心理能力與學習成就的評量，大致可分為心理計量與認知心理學兩種理論模式（丁振豐，1997；林世華、葉嘉惠，1999；Embretson, 1983; Embretson & Gorin, 2001）。心理計量學者致力於發展各種測驗與統計分析方法，用以區別受試者之能力與成就的差異，無論是古典測驗理論（classical test theory, CTT）或是試題反應理論（item response theory, IRT），均能將受試者能力與試題相對難度簡約成數據，提供教育決策實務重要的參考依據（Sheehan & Mislevy, 1990）。然因心理計量模式在測驗的編製上，著重在試題內容是否能代表測驗所要測量的構念（construct），僅以單一分數推估受試者在該構念之能力高低，不易得知受試者已具備哪些知識結構和解題策略，或是具有哪些認知成分的試題比較難（Sheehan & Mislevy, 1990）；此外，心理計量模式在測驗效度研究上，將測驗得分進行各項統計分析，以統計顯著性做為判斷的依據，雖具客觀性，卻因效度研究是在試題編製與施測後才進行，藉由測驗內試題間的相關、測驗得分和其他測驗的相關，以及比較不同群體受試者之得分差異等方法來推估測驗的建構效度（construct validation），對試題設計的貢獻相當有限（Embretson,

1998; Embretson & Gorin, 2001）。

認知心理學者以訊息處理理論為基礎，關注個體在學習與問題解決歷程的認知運作（cognitive operation），在評量受試者的能力或學習成就之前，會先界定一組能正確完成某項任務所需的程序、解題策略和知識結構（knowledge structures），再設計情境明確的作業（task），用來診斷受試者在哪些部分出現困難（Dimitrov, 2007; Embretson, 1983; Embretson & Gorin, 2001; Sheehan & Mislevy, 1990），有助於學習問題之診斷，也可根據受試者的反應調整試題內容結構以改變試題難度（Dimitrov, 2007）。但由於此類評量以質性分析來描述受試者在幾個特定作業的表現，評量結果的穩定性與可推論性顯然較弱（林世華、葉佳惠，1999；Sheehan & Mislevy, 1990）。

由上述可看出心理計量與認知心理學各有其擅長與不足之處，因此，學者們分別提出各種結合心理計量與認知心理學的研究模式，例如：Fischer（1973）的線性 logistic 測驗模式（linear logistic test model, LLTM），Embretson（1984）的一般多元成分潛在特質模式（general multicomponent latent trait model, GLTM），以及 Tatsuoka（1983）的規則空間（rule-space）等，其中最早且使用最多的是 LLTM（Chen, MacDonald, & Leu, 2011;

Dimitrov, 2007)。《心理科學季刊》(Psychology Science Quarterly) 於2008年以LLTM為主題刊出9篇論文，研究範圍涵蓋試題構念、試題出現的前後位置效果 (item-position effects)、不同國家學生在大型測驗的認知成分比較……等議題，可看出LLTM在評量研究的應用相當廣泛。在LLTM中，每個試題的難度參數 δ_i 均可線性分解為完成該試題所需的認知成分，例如：知識結構、解題規則與解題步驟等 (Dimitrov, 2007; Fischer, 1973; Kubinger, 2008)。試題認知成分分析可用來檢驗測驗的建構效度，測驗發展者於施測前即可根據試題的認知成分預測試題難度，或藉由變化試題認知成分操弄試題難度 (洪碧霞、蕭嘉偉、林素微, 2009; Dimitrov & Raykov, 2003)。Dimitrov和Raykov進一步應用測驗內各試題認知成分矩陣，藉助結構方程模式 (structure equation model, SEM) 以受試者在基礎認知成分試題的表現，預測其在認知成分較為複雜試題 (需包含基礎認知成分) 之得分，並與認知成分分析結果進行交互驗證。

色塊測驗的英文名稱是“The Token Test”，自從 De Renzi 和 Vignolo (1962) 在*Brain*期刊發表之後，它就成為一個極受歡迎的神經心理測驗，廣泛用於評量語言障礙者的語言理解能力。這個測驗要求受試者依循主試者的

口語指令，操作擺在桌上的塑膠色塊 (tokens)，藉此來評估受試者之聽覺理解能力。由於日常生活對話中，聽者可以透過說者的動作表情或情境線索猜測說話者的意圖，當個體有輕微注意力缺損與記憶廣度不足的問題時，並不容易被發現。色塊測驗在設計上盡量排除語句外的情境線索，僅使用動作、大小、顏色、形狀和方位等簡單詞彙，組成結構簡潔的常用語句，受試者必須注意聽試題中提到的每個詞彙，並有足夠的聽覺記憶廣度方能正確作答。因De Renzi和Vignolo僅於期刊發表色塊測驗的構想，色塊的規格與施測程序均沒有標準化也未正式發行，因此有許多研究人員參考The Token Test的理念發展測驗與進行研究，除了原始的The Token Test外，還有許多不同語言與版本的色塊測驗 (林月仙, 2007)。中文色塊測驗是林月仙和曾進興 (2006) 參考色塊測驗相關文獻及中文語句特性編製而成。就已出版的文獻來看，色塊測驗的效度研究以心理計量為主，無論以古典測驗理論或試題反應理論為依據的分析結果，均顯示色塊測驗有不錯的建構效度 (林月仙、曾進興、吳裕益, 2011)。至於組成色塊測驗的認知成分對色塊測驗試題難度的解釋力如何？哪些成分對試題難度較具影響力？以及是否可依據組成試題的認知成分建構試題間難易度的從屬關係？Willmes