

## 第五章 結論與建議

本研究旨在以廣義的測驗建構角度檢視精熟標準的設定，企圖發展一套以最大測驗訊息量法為核心的精熟標準設定方法，探討其理論性，並尋求最佳測驗結果的解釋方法，以提供多元的效度證據，並聯結其與測驗長度、測驗異質性間的關聯，評估可能存在影響分類精熟/未精熟者的因素。

為達到上述目的，首先，藉由文獻探討方式以定位出有效詮釋精熟標準設定方法的概念；其次，尋求支持運用最大測驗訊息量法於精熟標準設定上的理論基礎，並研討其分類精熟/未精熟者的效果；接續，探討換算古典測驗分數法、測驗特徵曲線構圖法於轉換分數上的效益，及定錨點於測驗結果解釋上的功能；最後，統整測驗長度、測驗異質性因素，評估其與上述方法間的互動、影響效果。

綜合上述，以下茲將相關結論與未來實務與研究建議描述如下。

### 第一節 結論

#### 壹、最大測驗訊息量法的詮釋概念與理論性

精熟標準設定方法在過去研究中，多半專注於所謂「方法內過程」的概念，觀念過於集中，並未能考慮到與上(測驗編製目的、試題來源)及下(測驗解釋結果)的融合，致使運用方法多所限制。因而，透過精熟標準設定的歷史演進與發展中，可發現詮釋本研究最大測驗訊息量法的三個面向，分別為：元素的搭配組合與調整、廣義測驗建構流程、多元效度等，以此概念即能充分解釋運用最大測驗訊息量於精熟標準設定時的合理性與適切性，更增廣了可提供的效度證據。因而在確立最大測驗訊息量法的公式意涵面向、試題選擇面向、統計考驗力面向的合理性後，以建立其於精熟標準上的理論基礎，而後，再輔以精熟/未精熟者分類一致性信度值以佐證多元效度。最後，探討其測驗分數的轉換方法、差異能力描述，期能同時獲得量與質的測驗結果解釋。

#### 貳、最大測驗訊息量法設定結果

##### 一、分類精熟/未精熟者一致性與轉換分數效果

在 2002 年國中基本學力測驗自然科資料分析中，由最大測驗訊息量法求得精熟標準及藉由換算古典測驗分數法、測驗特徵曲線構圖法於轉換分數上，經交叉驗證後，對於分類精熟/未精熟者一致性表現上，大致可獲得滿意的結果，皆有高達九成以上的分類一致性水準。另，在精熟標準調整面向上，藉由區間的概念亦能充分顯現出，以最大測驗訊息量法求得之精熟標準，具有作為參照基準點的優勢。而在完全控制測驗與受試者作答表現，探討轉換方法間的轉換效果時，雖然，換算古典測驗分數法皆較測驗特徵曲線構圖法為佳，但其分類差異程度多屬微小，皆在良好控制範圍內，顯示運用最大測驗訊息量法於精熟標準設定，不論搭配換算古典測驗分數法或測驗特徵曲線構圖法皆是值得參照的方法。

## 二、國中基本學力測驗精熟/未精熟者間差異能力

在運用定錨點以解釋由最大測驗訊息量法於國中基本學力測驗求得之精熟標準時，可發現未精熟者乃僅需具備基礎學科知識與簡易圖示理解能力，而對於精熟者而言，則仍需進一步擁有對於廣泛學科知識的了解、複雜問題、資料與圖表詮釋、邏輯推理與分析實驗結果以獲得相關論點等能力，或者更高階之具備進階學科知識與綜合、評鑑資料、情境傳遞之訊息的能力。若將之對照於國民中學學生基本學力測驗推動工作委員會(2002c)所認為，學生除需具備各學科的基礎知識外，尚需應用這些知識於詮釋問題、資料和圖表，甚至更進一步地應用學科知識及邏輯推理來整理出問題的答案，二者乃具有類似的核心概念。

## 參、測驗長度因素

### 一、分類精熟/未精熟者一致性與轉換分數效果

探討測驗長度因素時，不論採行最大測驗訊息量法、換算古典測驗分數法或測驗特徵曲線構圖法，皆受此因素的影響，顯示測驗長度愈長，分類一致性愈高，此項結果乃與過去大多數的研究一致；另，由本資料分析結果顯示，若研究者對分類精確性的要求是很一般化(約 80%分類一致性)，且在考量金錢、時間成本效益下，希望測驗的題數能儘量少時，則測驗長度 20 題時，會是必備的基本題數要求值。另一方面，在比較轉換分數方法間分類的差異程度時，因不論在何種測驗長度類型下，整體的信度值皆極高，因此，使得隨測驗長度增加，提昇信度的幅度較為有限。

從細部精確錯誤分類人數角度分析時，對於影響轉換分數間不一致因素，可從轉換的差異分數大小與差異分數點人數多寡兩方面檢視。其中，就前者而言，在換算古典測驗分數法下，可能因本研究乃採用受試者實際表現資料進行分析，有些反應組型並沒有呈現出來，或者出現某些極端的反應組型，因而使得分數轉換間有些許落差；而對於測驗特徵曲線構圖法，可能因轉換後之 TCC-Y 軸值與古典測驗答對題數間的落差，而出現分類不一致現象。就此，實質影響原因較不明顯，而對於後者，對照實務用途時，可發現對於影響轉換時差異分數的因素，決策者並不容易掌握與控制，但卻可藉由增加測驗長度，分散分數點人數，以彌平錯誤分類的影響。

### 二、定錨點題目篩選結果

在探討測驗長度因素與定錨點題目篩選間的互動關係，若從直觀的角度檢視時，顯示測驗長度愈長，相對能擷取到較多的定錨點題目，但若加以深究時，可發現測驗長度，並非是絕對影響定錨點題目篩選的因素，更重要的在於最大試題訊息量所對應之最適能力值是否能與定錨點相搭配。就此觀點，對照於本研究採用之最大測驗訊息量法，精熟標準所在位置，即為整份測驗多數試題最適於施測之能力值，可發現於精熟標準附近之定錨點多半能獲取較多定錨點題目，而於較

極端之定錨點，即使測驗長度較長，亦無法具備充足的定錨點題目。

## 肆、測驗異質性因素

### 一、分類精熟/未精熟者一致性與轉換分數效果

在探討測驗異質性因素下，最大測驗訊息量法乃具有因試題參數而調整估計受試者能力的特性，使得在異質測驗下，分類一致性仍能維持在不錯的水準之上。反觀換算古典測驗分數法與測驗特徵曲線構圖法，在固定精熟標準下，則有明顯的錯誤分類比率，會因不同測驗難度類型，使得精熟者被分類成未精熟者或未精熟者被分類成精熟者的人數大增，此現象也反應出現行以固定 60 分作為及格(精熟)標準的缺失。

在探討不同測驗難度比較中，採用簡易測驗、困難測驗或常態測驗間於轉換分數上之效果時，由換算古典測驗分數法或測驗特徵曲線構圖法轉換來自最大測驗訊息量法之精熟標準時，資料分析結果顯示，在不同測驗難度類型下，彼此間差異皆十分微小。整體而言，兩者皆具備良好百分比一致性與  $\kappa$  係數，顯示不論測驗比較中，於何種測驗難度類型中，採用何種轉換方式，並不致於影響轉換分數間一致性分類的效果。

從細部精確錯誤分類人數角度分析時，在簡易的測驗中受試者的得分通常會較偏高，因而人數多集中於高分數點上，而困難的測驗則是相反，人數會多集中於低分數點上，就此，若通過分數之位置位於人數較多之分數點上，自然會影響轉換分數的效果，而對照於本研究所採用之最大測驗訊息量法，此法乃具備隨測驗難易程度來決定門檻的特性，於簡易測驗中求得之精熟標準較低，而於困難測驗中求得之精熟標準相對較高，使得由此法求得之精熟標準附近人數並不會太多，相對於轉換概念上，代表著即使有較大的差異分數，亦不會造成嚴重的錯誤分類人數。

### 二、定錨點題目篩選結果

在探討測驗異質因素與定錨點題目篩選間的關聯，就初步觀點檢視時，資料顯示，在簡易測驗中，對於高能力定錨點而言，較難獲取適當的定錨點題目；相反的，於困難測驗中，對於低能力定錨點而言，亦無法產生適當定錨點題目，在此情況下，若搭配由簡易或困難測驗求得之偏低或偏高定錨點，則情況更加明顯，但若加以深究時，可發現真正影響因素，在於最大試題訊息量所對應之最適能力值是否能與定錨點相搭配。

## 第二節 建議

茲依據前述結論提出未來實務運用與研究之建議，陳述如下：

### 壹、實務運用建議

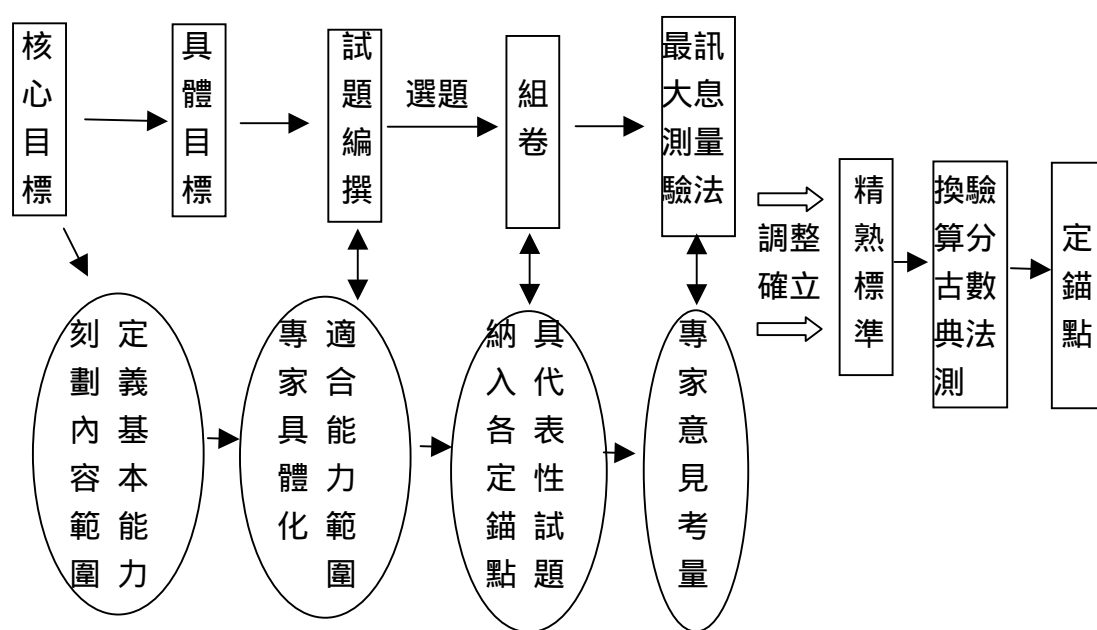
- 一、過去對於精熟標準設定的觀念，多半過於自限，因而，本研究乃建議研究者於運用精熟標準設定方法時，可將其視為一種「元素的搭配組合與調整」，自行根據測驗目的、實務考量等因素，建立一套適合的精熟標準，同時，以廣義測驗建構流程觀點，提供各面向的多元效度證據，以佐證該方法所得精熟標準是值得參照的。
- 二、本研究所採最大測驗訊息量法，經檢驗後，於分類一致性上有不錯的表現，且有相對強韌、嚴謹的理論支持，是最適合用於大型考試上使用。而對於現行考選部採行 60 分作為固定精熟標準作法的缺失，亦能提供有效的彌補。因此，建議未來政府單位或實務工作者在進行大型證照、資格檢定考試時，可考慮使用本策略。
- 三、最大測驗訊息量法本身雖然具有多項優勢，但於解釋測驗結果上，因其計算出的能力值與一般古典測驗分數的表達方式不同，使得無法同時與大多數建立在古典測驗理論上的精熟標準方法作有效對照，且對一般民眾之解釋上，亦十分不方便；若搭配換算古典測驗分數法或測驗特徵曲線構圖法則不失一項有效解決此問題的方法，足供相關測驗工作者參照運用。此外，若為更進階解釋精熟/未精熟者間的實質差異能力，定錨點則不失為一項有用的工具。
- 四、在過往諸多研究中，均傾向一致肯定，較長的測驗會擁有較高的分類一致性效果。由本研究中顯示，若在考量施測時間、成本、疲累等因素下，實務上無法提供受試者無限的應考試題時，乃建議長度在 20 題時，會是一不錯的基本值，可供往後編製測驗時的參考。

綜合上述，本研究乃提出如圖 5-1 所示之以最大測驗訊息量法為核心所建立之完整精熟標準設定模式，以茲後續實務應用，詳細描述如下：

在初次設定某資格檢定測驗之精熟標準時，於測驗前，往往未能確切知道所欲設定的通過分數為何？因此，多半需藉由首次測驗資料尋求出可供參照的標準，而本研究即希望在「元素的搭配組合與調整」、「廣義測驗建構流程」的精熟標準設定詮釋概念下，運用最大測驗訊息量具備定位客觀、穩定的精熟標準點方式，作為專家、學者建立特定通過分數之參照，以供第二次資格檢定之人才篩選標準。在如圖 5-1 流程中，精熟標準設定的開端從測驗編製初即已融入，在核心

目標主導下，決策者需針對人才甄選取向、最低能力要求等達成共識，以作為接續篩選人才的概念，進而，在確切的測驗內容範圍內，訂立具體的能力指標，同時具體化該測驗適合的能力範圍(如多少百分比的考生能正確作答)，並以此挑選具內容代表性試題，組卷施測(而題庫的建立可根據歷次的測驗逐步累積)。於施測後，藉由最大測驗訊息量法求得暫時的精熟標準，搭配著決策者相關因素考量以調整，而決定出最適切之精熟標準，接續，再運用換算古典測驗分數法或測驗特徵曲線構圖法轉換分數，搭配定錨點以進行測驗結果的解釋與報告撰寫。

在接續第二次資格檢定時，因應公平性的前提要求下，多半需採用首次建立之精熟標準以進行試題組卷，但另一方面，為有效解決利用試題訊息量作為挑選試題依據時，是否能選擇足以符合各定錨點內容代表性的定錨點题目的疑義，在此，可採行填滿目標訊息量函數的概念以編製平行測驗。首先，乃針對 4 個定錨點(各採 2 個以描述精熟/未精熟者能力，而此定錨點個數可由決策者自行決定欲解釋的面向，加以調整)，各選取 5 題最大試題訊息量所對應之能力值，為最適於該定錨點者之試題，以組卷形成至少 20 題之平行測驗卷(測驗總試題數可根據首次測驗自行調整訂定)，使得第二次欲施測之全部試題能填滿首次精熟標準能力值所對應之目標訊息量(即隱含第二次測驗最大測驗訊息量對應之能力值恰似第一次測驗精熟標準目標訊息量對應之能力值)。如此，以此測驗卷進行第二次(或往後)的資格檢定考試。



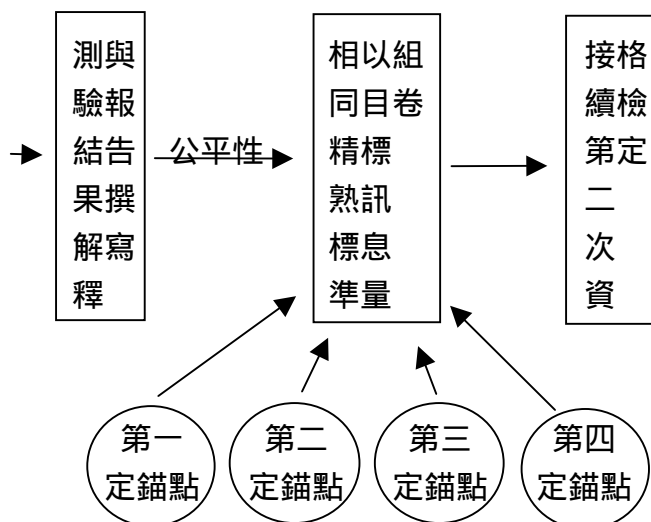


圖 5-1 以最大測驗訊息量法為核心建構之精熟標準設定模式

## 貳、未來研究建議

- 一、對於採用最大測驗訊息量法於精熟標準設定上，綜合本研究結果，可發現幾項疑義：1、利用試題訊息量作為挑選試題依據時，組成的測驗皆以準確測量精熟標準的能力值為目的，但其是否足以符合整個測驗內容代表性，則值得懷疑。2、最大測驗訊息量法求得之 IRT- 能力值，主要由測驗本身組成的試題決定，與傳統對於精熟標準乃屬檢測受試者是否適任工作似乎無關聯。對此，搭配著在探討測驗長度與測驗異質性議題時，曾提到如圖 2-11、2-12 以結合選題、定錨點與目標訊息量概念之解決策略，但有鑒於本研究非以驗證此模式為目的，因而在此並未加以證實。針對此，建議未來研究可運用相關資料以佐證其可行性。
- 二、在研究設計上，筆者雖企圖在各相關議題操弄時，儘量控制無關因素的影響，但實際進行時，發現研究間仍存有採用實徵資料不足之處，加上影響分類一致性信度的干擾因素十分複雜，在現有資料下，乃無法逐一控制。針對此，建議未來研究可採模擬資料，加以檢定相關因素的影響力。
- 三、本研究雖於探討最大測驗訊息量法於精熟/未精熟者的分類表現上，乃採以交叉驗證的方式，將某測驗求得精熟標準，由另一測驗視為效標，加以驗證，但嚴格而言，此仍屬內部一致性範疇，並未具外在推論的效果。因此，建議未來研究可蒐集真正具外在效標的資料，以佐證最大測驗訊息量法於精熟標準設定上之外在推論的可行性。
- 四、對於測驗工具的篩選上，本研究雖考量過許多因素，儘可能尋求符合期望與

假設之資料，但本研究所採用之國中基本學力測驗自然科兩次測驗，雖符合IRT單向度假設，但在模式適配度檢定上，對於某些試題而言，其卡方值仍偏高，顯示測驗試題並非完全適合該模式，因此，建議未來研究可尋求更適切之資格考試資料或採用其它模式進行檢定，以尋求更嚴謹的研究設計。

- 五、本研究在採用定錨點以描述精熟/未精熟者的實質差異能力，於初步篩選符合各能力點之範例試題時，乃採用 Huynh(1998)的建議準則，對此，乃僅考量到該試題“量”方面的適切性，並未能進一步考慮試題“質”方面的特點，因此，易使得某些具備良好“描述”某能力值之試題，卻因未符合“量”之準則，而被忽略。因此，建議未來的研究更深入探究此類雖未符合準則，但具備良好描述性之試題，考量是否亦納入描述差異能力的範例試題中。