

# 第一章 緒論

本研究乃以精熟標準設定作為探討核心，目的在於運用試題反應理論中最大測驗訊息量的概念，加以重新詮釋與定位，突顯其在精熟標準設定上的功能，同時探討其它相關議題，期能提供另一項解決現行實務應用上缺點的方式。

在本章緒論中，論述主題將含括研究動機、研究目的與待答問題、名詞釋義等，茲說明如下：

## 第一節 研究動機

讓我們將時光拉回當學生的年代，從國小、國中、高中，直到大學，是否還記得老師們所訂的及格標準是什麼呢？「60分」，一個永遠不變的數字，也象徵著國人根深蒂固的傳統精熟(或稱之為及格)概念。是否曾思索過，為何59分與60分間1分的差距，是如此的遙遠，可以判定一個人及格或不及格、通過或失敗，這個標準是如何製訂的，是否存在著合理性、公平性呢？

而在傳統學校評量外，讓我們將場景轉移至更具重要性的國家證照考試上，由於大學窄門的開放，大學文憑已有漸漸勢微的傾向，取而代之的，便是「證照」(license)的公信力，愈突顯其重要性，而國內目前實施的國家職業證照考試，其辦法主要可分為三種：第一、考試院主辦的國家考試；第二、由職訓局辦理的技術士技能檢定；第三、各行業主管機關辦理的考證(莊淑如，1997)。這些單位所舉辦的證照檢定，其中不乏是像醫師、律師、建築師等，悠關你我性命與權益的重量級考試，但其為我們所把守的通過門檻又是如何製訂的呢？同樣是60分或70分，一個長久以來約定成俗的習慣標準，吳裕益(1986)曾表示，這是一種比較屬於武斷、主觀決定精熟的方法(相關缺點於第二章文獻探討精熟標準設定相關議題探討一節中有詳細說明)。

此外，隨著科技訊息的進步與測驗理論的推行，亦能漸漸感受到這股測驗潮流的變化，諸如2001年起所舉辦之國民中學學生基本學力測驗(the basic competency test for junior high school students, 簡稱BCT)、2004年起社團法人台灣理財顧問認證協會所推行之理財規劃顧問(certified financial planner, 簡稱CFP)證照考試。為了確保其公信力、維護其專業權威形象，這些考試皆採用較嚴謹的試題反應理論(item response theory, 簡稱IRT)技術，但相較於國內多數重大考試，這僅是個開端，面對多年以來，對於傳統測驗固有的概念，的確是面臨到轉移的時刻。基於上述，本研究動機之一即是希望點出國人對於精熟標準上一些概念的謬誤，喚醒對此議題的重視，並提出建基在試題反應理論下，具合理、公平性的精熟標準設定方法與流程的重要性。

學術發展上，對於精熟標準設定方法的研究，於國外，至今仍是屬於熱門話題，所提出的方法十分多元與龐雜，尤其是伴隨著IRT與相關統計、評量、電腦科技的發展，更使得此議題愈趨活絡。Berk(1986)確認過有38種方式被發展出來；鄭明長、余民寧(1994)認為有超過40種方法，至今2005年，陸續也有許多

精進方式被研發，面對如此龐雜的方法，其相關討論議題更是複雜難解。基於此，本研究動機之二，是希望統整過去相關精熟標準設定方法之文獻，簡要的以歷史演進方式呈現，從中萃取重要概念，重新詮釋與定位本研究所採精熟標準設定方法，並釐清與各項議題間的關係。

反觀國內，相關研究主要是以吳裕益(1986)、林惠芬(1993)、鄭明長、余民寧(1994)等人為主，不僅年代較久遠，而此類研究主要還是以古典測驗理論(classical test theory, 簡稱 CTT)為探究核心，其理論背景、能力估計、靈活度皆較薄弱，雖然鄭明長、余民寧的研究中曾考量融入 IRT 的方法，但是對其運用於精熟標準設定時的理論基礎，並未能充分說明完整，且詮釋的概念亦無法有效化解實務應用上的質疑(於第二章文獻探討轉換通過分數與差異能力描述一節會有更詳盡探討)，同時並未針對可能影響分類一致性的主要因素(如：測驗長度、異質性等)同時納入考量，且計算出的能力值無法與古典測驗答對題數作對照，只能單純依分類一致性作對應，增加比較、解釋的不方便性。因而，本研究動機之三，乃企圖解決 IRT-能力值無法轉換為古典測驗答對題數的問題，同時，兼顧質與量關注於精熟/未精熟者能力描述方法，並檢視、討論其可能產生的議題。

## 第二節 研究目的與待答問題

基於上述研究動機，本研究期望達到下述之實際目的與描述其可能產生相對應之待答問題。茲介紹如下：

研究目的一：

簡要的以歷史演進方式呈現過去相關精熟標準設定方法之文獻，從中萃取相關訊息，以詮釋與定位本研究所採最大測驗訊息量函數概念於精熟標準設定上之運用。

待答問題：

- 一、幾十年間所研發之精熟標準設定方法，其演變的脈絡為何？
- 二、本研究用以詮釋最大測驗訊息量函數應用於精熟標準設定之概念為何？

研究目的二：

運用試題反應理論中最大測驗訊息量函數概念於精熟標準設定上，並探討與各項議題間之關係。

待答問題：

- 一、最大測驗訊息量函數概念應用於精熟標準設定上，其理論基礎為何？
- 二、最大測驗訊息量函數概念應用於國中基測自然科兩次測驗的精熟標準設定上，其分類一致性效果為何？
- 三、最大測驗訊息量函數概念應用於國中基測自然科兩次測驗的精熟標準設定上，與測驗長度互動關係為何？
- 四、最大測驗訊息量函數概念應用於國中基測自然科兩次測驗的精熟標準設定

上，受測驗異質性影響為何？

研究目的三：

運用換算古典測驗分數法與測驗特徵曲線構圖法於轉換 IRT- 能力值與古典測驗答對題數，並探討與各項議題間之關係。

待答問題：

- 一、運用換算古典測驗分數法於轉換 IRT- 能力值成古典測驗答對題數時，其分類一致性效果為何？
- 二、運用測驗特徵曲線構圖法於轉換 IRT- 能力值成古典測驗答對題數時，其分類一致性效果為何？
- 三、運用換算古典測驗分數法與測驗特徵曲線構圖法於轉換 IRT- 能力值成古典測驗答對題數時，何者具較佳的效用？
- 四、運用換算古典測驗分數法與測驗特徵曲線構圖法於轉換 IRT- 能力值成古典測驗答對題數時，與測驗長度互動關係為何？
- 五、運用換算古典測驗分數法與測驗特徵曲線構圖法於轉換 IRT- 能力值成古典測驗答對題數時，受測驗異質性影響為何？

研究目的四：

運用定錨點以增進解釋精熟/未精熟者間差異能力之描述，並探討與各項議題間之關係。

待答問題：

- 一、運用定錨點於解釋國中基測自然科兩次測驗精熟/未精熟者間差異能力結果為何？
- 二、運用定錨點於解釋精熟/未精熟者間差異能力時，與測驗長度互動關係為何？
- 三、運用定錨點於解釋精熟/未精熟者間差異能力時，受測驗異質性影響為何？

研究目的五：

點出國人對於精熟標準上一些概念的謬誤，喚醒對此議題的重視，並提出具理論、公平性的精熟標準設定流程。

簡言之，筆者是想統整過去有關於精熟標準設定的相關研究文獻，從其演進的脈絡中，找尋出支持運用最大測驗訊息量函數概念於精熟標準設定上具合理性之證據。同時，提出架構在此方法核心下之精熟標準設定流程，發展出能與古典測驗答對題數相互轉換及有效解釋精熟/未精熟者差異能力的方式，並探討與各項議題間的關聯。

### 第三節 名詞釋義

茲就本研究中所涉及相關重要名詞，分別界定如下：

#### 壹、效標參照測驗(criterion referenced test)

其相關同義詞彙係有領域參照測驗(domain-referenced test)、目標導向測驗(objective-based measurement)、精熟測驗(mastery test)、能力本位測驗(competency-based test)等，皆代表著在界定清楚的行為領域中，根據學習或行為目標訂定能力標準，以檢定個人是否能達到預設的標準，並用此作為表現程度解釋的測驗，於本研究中乃簡稱為 CRT。

#### 貳、精熟測驗(mastery test)

乃為效標參照測驗之同義詞，但有藉於本文研究基礎乃建立在試題反應理論上，因此，援用 Lord(1980)的說法，以「精熟測驗」一詞顯現於此方面的應用，以區別古典測驗理論所指稱「效標參照測驗」的不同。

#### 參、試題反應理論(item response theory)

試題反應理論乃是以機率的概念來解釋受試者能力和試題反應間關係之理論，亦即觀察其反應與表現結果，經由數學模式的運算，以推估受試者的能力(ability)或心理特質(latent trait)，於本研究中乃簡稱為 IRT。

#### 肆、最大測驗訊息量法(maximum test information approach)

運用試題反應理論中最大測驗訊息量的概念於精熟標準設定上的一種方法，其核心乃以整份測驗中最大測驗訊息量所對應之能力值，即視為該測驗最佳的精熟標準，可作為後續專家、學者考量此精熟標準對經濟、社會可能的影響，而加以調整之基礎。

#### 伍、內容標準(content standard)、能力標準(performance standards)、通過分數(passing score)

內容標準乃係指在某課程內容範圍內，受試者被期望應該知道與瞭解的知識，如懂得基本加、減、乘、除運算等；能力標準係將上述內容標準，進一步劃分，以區分出精熟/未精熟者間所具備的差異能力，如決策者認為精熟者至少需懂得二位數乘法運算，若受試者僅能正確作答一位數乘法運算，則僅能將之視為未精熟者；通過分數則將上述質性描述的能力標準轉換為測驗的實際通過分數，如 60、70 分等，以便解釋、運用，而其常見同義名詞係有及格分數(passing score)、切截分數、決斷分數(cutoff score)、精熟標準等，對照於本研究，則將以最大測驗訊息量法求得之最佳能力值，稱為精熟標準，而藉用換算古典測驗分數法與測驗特徵曲線構圖法轉換自精熟標準之古典測驗答對題數，稱為通過分數。

## 陸、精熟/未精熟(mastery/non-mastery)

在某個界定清楚的內容領域中，受試者的表現高於預先所設定的能力標準，即稱之為精熟；若低於預先所設定之能力標準，則稱之為未精熟。而本研究所指稱之精熟，乃係指學生於測驗上之得分，高於或等於由最大測驗訊息量法或其轉換分數方法求得之精熟標準或通過分數；反之，未精熟則係指學生得分低於該精熟標準或通過分數。

## 柒、第一類型錯誤、第二類型錯誤、分類不一致性

係指根據精熟標準及受試者實際表現作能力篩選時，可能犯下分類精熟/未精熟者之不一致現象。若受試者的實質能力未達精熟狀態，而進行分類時，卻將其歸類為精熟者，此時所犯錯誤，即稱之為第一類型錯誤；相反的，若受試者實質能力已達精熟狀態，而歸類時，卻將其視為未精熟者，此時，所犯錯誤，即稱之為第二類型錯誤。而此二類型錯誤，皆統稱為分類不一致性(或分類錯誤)。

## 捌、換算古典測驗分數法(transformed classical test scores approach)

係指運用余民寧、汪慧瑜(2005)所發展之「答對題數與量尺化能力分數」轉換概念，加以引用至精熟標準設定上，將計算出之 IRT 精熟標準 能力值與古典測驗理論下之答對題數作轉換的方法，於本研究中乃簡稱為 TCTS 法。

## 玖、測驗特徵曲線構圖法(test characteristics curve mapping method)

運用試題反應理論中測驗特徵曲線的概念，採用圖形對應方式進行 IRT 精熟標準 能力值與古典測驗答對機率(或題數)轉換的方法，其概念乃將 X 軸上求得之精熟能力值，畫一垂直於 X 軸的直線，交於 TCC 曲線上的一點，再經由此點，畫一垂直於 Y 軸的橫線，交於 Y 軸上的一點，該點即代表著轉換後之古典測驗答對機率值或題數，於本研究中乃簡稱為 TCCM 法。

## 拾、轉換分數效益

係指研究者所期望的 IRT 精熟標準 能力值與實際經轉換為古典測驗答對題數後，運用於測驗結果解釋時的助益效果，此乃同時包含分數解釋的便利性與分數轉換後分類的一致性效果。對照本研究，乃主要以評估換算古典測驗分數法與測驗特徵曲線構圖法於轉換 IRT- 能力值成古典測驗答對題數後之分類精熟/未精熟者的一致性效果。

## 拾壹、定錨點(anchor points)、定錨點題目(anchor items)

係指於精熟測驗中，藉以描述精熟/未精熟者間差異能力的方法，概念上乃需於量尺分數上任意選擇數個定錨分數點，以分別代表精熟/未精熟狀態，而後，根據此定錨點選擇一個或數個具內容代表性的範例試題(exemplary items)，對此稱之為定錨點題目(為區辨等化概念中之定錨試題，在本研究乃以定錨點題目稱之)，以此為基礎，作為解釋受試者於精熟/未精熟上實質的表現能力。

## 拾貳、測驗長度/測驗異質性議題

測驗長度議題係指探討在不同的測驗長度下，與最大測驗訊息量法、換算古典測驗分數法、測驗特徵曲線構圖法、定錨點間的互動效果；測驗異質性議題係指探討在兩次測驗試題組成難易度不一致時，可能影響上述方法於正確分類精熟/未精熟者的表現。