

第三節 國中基本學力測驗與聯考實務作法的差異

基本學力測驗(以下簡稱基本學測)與傳統聯招考試在作法上是有實質上的不同。這些差異主要是圍繞在四個相關聯的議題：(一)、基本學測的內容(基本學測到底是什麼?)；(二)、入闈組題；(三)、量尺分數；(四)、一年多次(目前是一年兩次)測驗等化的問題(林世華，民91)。

基本學測與傳統聯考最大的差異是：測驗內容與測驗工具發展方法的差異。基本學測主要是設計來評量國中學生在修完國中三年後，所學習到的基本知識與所培養的基本能力。所謂「基本」意指核心的、重要的、統整的。國中三年老師在學校所教的，學生所學習的，包羅萬象，其中主要的當然是基本的範疇，當然也不乏複雜專業方面的涉略。基本學測是排除複雜專業的部分。基本學測的發展方法是以標準化心理測驗發展的科學方法為依據。科學方法步驟中，包括：清楚界定基本學力；測驗試題編寫，審查與試驗研究；測驗分數如何使用等。基本學測研究團隊從八十七年起便著手建造題庫，其中每一試題均需經歷嚴密審查並預試，以便清楚掌握試題的品質。題庫中的試題是不停地在增加與淘汰。這與聯考在考前入闈後才開始命題形成測驗是截然不同的。

其次是入闈組題的議題。基本學測是以持續發展的題庫入闈，並在闈場以自行開發的組題電腦系統來選題，組成當次的基本學測。理論上，系統會組成

哪一種測驗試題，事先大家並不清楚。所能清楚的是組題的原則。由於基本學測的界定相當清楚，所以組題原則也隨之具體。幾個主要原則如：試題評量符合基本學測目的；試題取材均勻分佈（到國三下第二次段考前的範圍），生活化、統整化素材優先；試題難度以 50-75% 通過率為主。上述組題原則的目的是為了保證所組出來的測驗能吻合基本學測的原有目的。這也等同於是說，基本學測有一定的設計規格，基本上來說，基本學測是朝穩定性的方向來發展。傳統聯考中易出現「去年簡單今年會難」、「去年難今年會簡單」的現象，在基本學測中這樣的情況會比較少出現，但預測難度及真實難度若有差別的話則是在所難免。

至於量尺分數則是基本學測分數使用上的考量。測驗分數的使用有兩大重要考量；第一是，測驗相對機制的建立（專業術語稱建立常模），它是希望能建立一個類似於「IQ100 是中等」，「IQ100 以上是中等資質以上」的具體意義。這套作法，在理論上還可以將量尺分數作跨年的比較，然而由於教育部所公佈的命題範圍會因不同學年而異等因素，以測驗的角度來看量尺分數作跨年的比較，暫不可完全實行，但有其參考的價值。第二是測驗分數的採計與運算。基本學測是以答對題數作為原始分數。原始分數在測量性質上是屬於名次性質，它會有在不同區段中名次間的差距看似等距離，其實不是等距離（如第 1 與第 2 名；第 15 與 16 名，間距均為 1，但實質並不是相等的 1），因此將其轉換成一

個可以採計與運算的量尺分數（基本學測量尺分數是1~60）是一個測驗標準化方法中的重要步驟。

最後基本學測是一年多次的（目前是一年兩次），國人也不易接受。來自各方的質疑有兩方面，第一、「學生會因考到簡單題而佔便宜，也會因考到難題而吃虧」。其實基本學測採用量尺分數的目的之一，也是因應此一問題。不過處理這問題的基本作法有二：第一、不同次的基本學測內容不會一樣，但是組題分佈、難易程度均要控制成一樣（再一次保證了基本學測穩定性的必要）；第二、如果在控制上出現小瑕疵，則另一個專業機制：等化，便可以派上用場。簡單的說測驗等化設計可以讓學生在考到稍難試題時，只要答對較少題時，就可以得到考簡單題時要答對較多題時的量尺分數，舉例來說：第一次數學科測驗較簡單，第二次數學科測驗較難，那是有可能透過等化，第一次對10題與第二次對9題的量尺分數是一樣的。另外一個質疑則是兩次基本學測考生不同（有第二次考生資質較優，亦有較低之說），所建立量尺分數在相對機制上有公平性的問題。確實，基本學測如果每次測驗均獨立建造常模，那會是問題。但基本學測其實不是這樣做。以90年為例：只有第一次基本學力測驗真正建立量尺分數對照表，第二次其實是以第二次答對題數等化到第一次答對題數，再由第一次答對題數對照量尺分數。

基本學力測驗的分數等化過程(涂柏原，民92)：

- (1) 由於幾乎所有的國三學生都參加了第一次國中基本學力測驗，第一次測驗的結果非常適合用來建立量尺分數。因此，根據前面所提到的原始分數轉換成量尺分數的方法，我們先將原始分數與量尺分數對照表算出來。如此，我們可以得到每一位考生的原始分數與量尺分數等資料。因為要用 IRT 的方法來進行等化，所以我們得進一步用 IRT 的方法來估算每一位考生的能力參數，根據全部考生的資料，經過綜合整理之後，我們可以得到下表的第二、三及四欄的資料，也就是答對題數、能力值與量尺分數的對照表。
- (2) 根據參加第二次學力測驗的考生在第二次測驗上面的作答資料，用 IRT 的方法來估計考生的能力值。如此，每一位考生的答對題數原始分數與其能力估計值都可以被收集到，而將這些資料與從第一個步驟所得到的資料合併，我們可以得到如下的對照表範例。
- (3) 第二次國中基本學力測驗不再建立新的量尺分數，而是使用我們在上一個步驟所得到的表，來將考生在第二次學力測驗所得到的原始分數，對應到在第一次學力測驗時所建立的量尺分數。舉例而言，如果某一考生在第二次基本學力測驗某學科的原始分數為 58 分，從下表我們可以得知他所對應的量尺分數為 51，相當於在第一次基本學力測驗該科的原始分數為 62。換言之，對該科來說，在第一次基本學力測驗原始分數為 62 分的考生，與在第二次基本學力測驗原始分數得到 58 分的考生，有相同的量尺分數 51 分。

某科量尺分數對照表範例：

| 第二次答對題數 | 能力值 | 第一次答對題數 | 量尺分數 |
|---------|------|---------|------|
| 63→ | 4.93 | 66 | 60 |
| | 4.4 | | 59 |
| 62→ | 4.1 | | 58 |
| | 3.71 | 65 | 57 |
| 61→ | 3.35 | | 56 |
| | 2.99 | 64 | 55 |
| 60→ | 2.77 | | 54 |
| | 2.56 | 63 | 53 |
| 59→ | 2.4 | | 52 |
| 58→ | 2.25 | 62 | 51 |
| : | : | : | : |
| : | : | : | : |

IRT 最為有力的地方，是同一個考生若在同一個時段考兩次試，即使作答不同的試題，所估計出來的兩個能力參數的值，理論上是一樣大小的。如果考完第一次之後隔了一些時日再考第二次，以至於兩次考試之間有成長（或進步）的現象存在，那麼根據第二次測驗的結果所估計得到的能力參數值，將會大於

第一次測驗後所估計得到的。因為 IRT 具備這樣的特性，因此透過用 IRT 方法所進行的等化之後，我們有信心考生在第二次測驗所得到的量尺分數絕對是合理、公平、公正的；而且從兩次測驗所得到的量尺分數之間也是可以相互比較的。

影響整個基本學力測驗等化程序的關鍵點，其實是在於各個題目的試題參數 (item parameter) 是否被精確地估計出來。因為這裡所說明的等化方法在某個程度之下，其原理與目前常見的電腦化適性測驗 (computerized adaptive testing, CAT, 如 TOEFL、GRE 以及美國的護士執照考試等) 原理是一樣的，是在試題參數已知的情形下估計考生的能力參數 (ability parameter) (註：在 IRT 中，每一個試題會有幾個試題參數來描述該題的特性，而也有一個能力參數來描述考生的能力)。目前國中基本學力測驗每一道試題皆經過至少 240 至 320 位不同地區的國三學生「預試」過，截至目前為止，參與過預試的國三學生已超過數萬人，涵蓋全國各縣市的國民中學。以此大規模的預試工作所得的答題反應資料，可以用 IRT 來估計出每一個題目的試題參數 (難度、鑑別度) 等。經過實際驗證，九十年預試所得的題目難度與該年度第一次測驗三十萬人的資料，算出來的題目難度相當接近。即使如此，學力小組仍然依據從三十萬人的資料所得到的試題參數，來將題庫中每一個題目的試題參數加以校正。