

第五章 結論與建議

本章將本研究作一整體性的總結，並提出建議。本章共分為兩節，第一節結論，第二節建議。

第一節 結論

本研究主要探討民國九十年第一次學測與第二次學測，兩次學測間數學科量尺分數與等化，在 Rasch、1PL 和 3PL 等不同模式下，所估計的差異。具體的研究項目包括下列六項：一、探討民國 90 年第一次學測，數學科利用 1-PL 模式的校準後，所呈現的量尺分數與學力測驗的量尺分數改變情形。二、探討民國 90 年第二次學測，數學科利用 1-PL 模式的校準後，所呈現的量尺分數與學力測驗的量尺分數改變情形。三、探討民國 90 年第一次學測，數學科利用 3-PL 模式的校準後，所呈現的量尺分數與學力測驗的量尺分數改變情形。四、探討民國 90 年第二次學測，數學科利用 3-PL 模式的校準後，所呈現的量尺分數與學力測驗的量尺分數改變情形。五、探討民國 90 年第一次學測，數學科利用 1-PL 與 3PL 模式校準後之題目難度估計值，應用 IRT 真分數法所呈現真分數分佈情形。六、探討民國 90 年第二次學測，數學科利用 1-PL 與 3PL 模式校準後之題目難度估計值，應用 IRT 真分數法所呈現真分數分佈情形。

針對此等研究項目，本研究以國民中學學生基本學力測驗推動工作委員會所提供之資料，以 1PL 和 3PL 等不同模式進行分析，研究工具包括：一、民國

九十年國中基本學力測驗第一次和第二次數學科試題，二、電腦軟體：BILOG-MG、MATLAB 和 SPSS。經統計分析後，獲得下列發現：

1. 第一次學測樣本資料經過 1PL 模式校準後，結果發現答對題數在 19 題以下，除了答對 13 題高於學測所公布的量尺外，餘皆等於或小於學測所公布的量尺，答對題數高於 20 題的情形者剛好相反，因此量尺分數若是以此種情形呈現，將有利於前半段能力較好的考生，若以變化率而言，此種估計與學測只有 36.36%(12/33)相同，真分數範圍為 0.98~31.46。
2. 第一次學測以 3PL 模式校準估計後，結果發現在答對 17 題之前，大都低於或等於學測之量尺分數，只有答對 2 或 3 題例外，在答對 17 題之後則相反，若以變化率而言，此種估計與學測只有 30.3%(10/33)相同，真分數範圍為 .97~29.47。
3. 第二次學測樣本資料經過 1PL 模式校準後，結果發現除了答對 6 與 30 題高於學測所公布的量尺外，餘皆等於或小於學測所公布的量尺，因此量尺分數若是以此種情形呈現，將使得大部份的學生的量尺分數低於學測的量尺分數，若以變化率而言，此種估計與學測只有 28.13%(9/32)相同，真分數範圍為 .97~29.47。
4. 第二次學測以 3PL 模式校準估計後，結果發現除了答對 1、3、19、20 及 21 題高於學測所公布的量尺外，餘皆等於或小於學測所公布的量尺，因此量尺分數若是以此種情形呈現，將使得大部份的學生的量尺分數低於學測的量尺

分數，若以變化率而言，此種估計與學測只有 15.63%(5/32)相同，真分數範圍為.97~29.47。

5. 根據學測所提供之數學科統計資料，發現第二次考試平均難度高於第一次，但第二次學測的整體原始與量尺平均分數都高於第一次學測成績。

第二節 建議

根據以上的結論，提出下列建議：

一、測驗應用上的建議

1. 因為是從同一量尺的題庫中選取題目，建議除了符合雙向細目表的內容外，亦應事先考慮兩次考試測驗平均難度及總字數，希望能相近。
2. 兩次考試的時間既然相同，試題的題數也能相同，可避免兩次估出能力值範圍差距太大的問題，而且在國際上大型的考試，如托福、GRE 與 GMAT 的考試，相同類型的考試題數似乎也都一樣。
3. 試題的型式皆是選擇題，因此猜測因素是無法避免，而且所有試題要做到鑑別度相同，似乎也不容易，將來題庫的建立與資料的校準可以朝 3PL 模式的方向建立，較符合試題的型態，但相對的採用 3PL 模式所面對的問題亦有待克服。
4. 為了等化的精確性，建議可以在考生第二次報告時，請考生加註第一次考試

編號，以利追蹤比較兩次成績或作為等化之用。

二、未來研究的建議

1. 在可以取得有效且足夠大的樣本後，以 Rasch 與 3PL 模式校準估計後所產生量尺，對考生影響程度的比較。
2. 可以擴大到其它科目進行研究分析。
3. 兩次間隔時間長短對學生考試成績的影響。