

26	44	0.80	42	26.08	0.72	41	24.86
27	47	1.06	46	27.28	1.06	46	27.31
28	49	1.27	49	28.01	1.28	49	28.46
29	51	1.46	51	28.51	1.37	51	28.83
30	54	1.72	55	29.02	1.54	53	29.39
31	60	2.10	60	29.47	2.00	60	30.3

表八 第二次國中學測母體與樣本量尺分數統計表

	量尺平均分數	量尺分數的 SD
母體	32.54	13.12
樣本_1PL	31.03	13.42
樣本_3PL	30.65	14.01

### 第三節 綜合討論

本研究的量尺分數計算，是基於所有題目皆來自題庫，且題庫中的題目參數都具有相同的量尺的假設，即所有試題的參數皆已知並具相同的量尺，在此假設下，其實兩個測驗間即不需做等化，因將此兩測驗施測於學生，獲得作答資料後，即可利用軟體求得學生的能力估計值，一般的測驗軟體，都會將考生的能力估計設定於平均數為 0、標準差為 1 的分佈。所以，所估得之能力量尺再

經過平均數 30、全部答對題數設為 60 的直線轉換，即可轉換為學測的量尺分數。但是在本研究因母群體之資料無法獲得，且研究者認為 3-PL 模式較符合四選一之選擇題，因此，擬用所獲得之樣本重新校準題目之參數，卻發生因為兩次考試題數不同，透過 BILOG-MG 估計後所得到之能力值範圍亦不同的情形，用 1PL 模式的估計，第一次學測的範圍為-3.05~2.41，第二學測的範圍則縮小為-2.82~2.10 之間；利用 3PL 模式的估計，第一次學測的範圍為-2.28~2.38，第二學測的範圍則縮小為-2.01~2.00 之間。為了等化量尺，因此將第二次學測的能力範圍皆調成與第一次的相等後，並將滿分調為 60 後呈現出。

原擬用的真分數等化法，亦因兩次考試不同，而無法做等化，由表三與表七的結果顯示，無論 1PL 模式或 3PL 模式呈現之量尺分數都與學測中心有不小的差異，這是因為項目與學生能力參數校準估計不同所使然，歸納其原因如下：

1. 項目與學生能力參數的校準估計與資料的內容有很大相關，以第一次學測的樣本而言，其樣本數不到母體的 2%，第二次學測的樣本數也不及 3%，樣本在這麼小的比例下，產生誤差的機率自然大增。
2. 項目與學生能力參數的校準估計和所使用的模式有關，本研究所採用之 1PL 模式或 3PL 模式皆與學測中心所採用的 Rasch 模式不同，因此，估計出來的項目參數與能力值也不相同，量尺分數自然有差異。
3. 項目與學生能力參數的校準估計和所使用的測驗軟體有關，本研究所採用的測驗軟體為 BILOG-MG，統計估計是用 EAP 法，若不同的軟體或統計估計法所

求出的參數，亦會不同。

由以上的經驗，我們得知作答資料的大小會影響題目參數的估計，進而影響測驗的等化，雖然學測中心報告中顯示，題庫中每道題都經過 240~300 位學生預試估計而得，但比起真正考試的母群體約 30 萬人，還是微不足道。因此，真正所得資料校準後估計的題目參數與原先題庫的參數值間的差異值得研究，若兩次考試中有任何一次產生顯著差異，則等化的問題即產生，在此狀況下，若兩次考試間又無定錨問題，或兩次都參加考試的人又無法篩選出，而另外蒐集做估算，則等化就值得商榷。為了避免此問題發生，而此考試牽涉到事後試題的公開，因此定錨問題似乎不可行。考試的學生有一現象，即幾乎所有參加第二次學測的學生都有參加第一次學測，若能請學生於報考第二次時加註第一次學測的編號，以利追蹤比對兩次的結果。但若不再重新校準估計題目參數，直接估計能力值也是一種方法。

另外由學測所給的統計資料中顯示第一次測驗難度平均為-0.45，而第二次測驗難度平均為-0.34，因此第二次的題目較第一次難些，但第二次雖只考 31 題，可是原始平均分數 18.67 確比第一次 18.45 高，量尺平均分數也呈現同樣情況，第一次量尺平均為 30，第二次為 32.54，其中原因是學生因間隔 70 天能力增加了，還是成績較不好的大都放棄第二次考試的機會，或是尚有其他因素，值得探討。