

## 第四章 研究結果與討論

本章分為三節，第一節民國九十年第一次國中基本學力測驗數學科校準後所呈現量尺分數之研究結果，第二節民國九十年第二次國中基本學力測驗數學科校準後所呈現量尺分數之研究結果，第三節綜合討論。

### 第一節 民國九十年第一次國中基本學力測驗數學科校準後所呈現量尺分數之研究結果

為了解所有考生與本研究所獲得資料間的差異，茲將敘述性的統計列於表一，民國九十年第一次國中基本學力測驗數學科所參加的總人數為 299368 人，若以每題答對得一分來計算，平均分數為 18.45，分數的標準差為 7.26，此次考試的信度  $\alpha$  為 0.89，而此研究所獲得的樣本數為 4999 人，平均分數為 18.62，分數的標準差為 6.74，信度  $\alpha$  為 0.88。因此樣本平均數略高於母體 0.17，而標準差的差異則較母體少了 0.52，因此樣本分數的分布將較母體的分數集中。

表一 第一次國中學測母體與樣本資料統計表

	母體	樣本
總人數	299368	4999
原始平均分數	18.45	18.62
原始分數的 SD	7.26	6.74
信度 $\alpha$	0.89	0.88

第一次國中學測所取得的樣本資料經過 BILOG-MG 以 1PL 及 3PL 模式估計後，和學測所題供之 b 值呈現於表二中，本研究經 1PL 模式校準後估計所得 b 值與學測所提供之 Rasch 模式 b 值頗為相近，平均難度相差 0.01，其中以第 26 題差異最大為 0.14，第 8、10、25 和 29 幾乎相等，標準差相差為 0.02。

表二 第一次國中學測 Rasch、1PL 及 3PL 模式之試題參數估計值

題號	學測估計		樣本估計		
	Rasch_b	1PL_b	3PL_a	3PL_b	3PL_c
1	-1.52	-1.45	0.85	-0.55	0.39
2	-2.04	-1.99	1.46	-1.15	0.20
3	-1.51	-1.50	1.09	-0.82	0.23
4	-2.06	-2.02	0.83	-1.49	0.20
5	-1.62	-1.57	0.81	-0.71	0.38
6	-1.70	-1.75	0.86	-1.26	0.17
7	-1.00	-0.97	1.63	-0.18	0.30
8	-1.55	-1.55	1.85	-0.81	0.17
9	-1.46	-1.38	1.00	-0.79	0.21
10	-0.35	-0.35	1.18	0.32	0.28
11	-0.49	-0.45	1.09	0.04	0.20
12	-0.39	-0.43	1.73	-0.10	0.11
13	-0.80	-0.85	1.45	-0.23	0.23
14	-0.08	-0.07	1.06	0.63	0.27
15	-1.11	-1.14	0.83	-0.59	0.23
16	0.03	0.06	0.77	1.23	0.33
17	-0.51	-0.55	1.59	0.09	0.26
18	0.14	0.09	1.56	0.29	0.11
19	-0.43	-0.42	0.49	-0.10	0.14
20	0.32	0.29	1.22	0.65	0.18
21	0.19	0.14	1.12	0.69	0.23
22	0.13	0.09	1.23	0.51	0.19
23	-0.19	-0.12	1.26	0.46	0.25

24	0.43	0.40	0.85	1.47	0.29
25	-0.41	-0.41	1.85	0.10	0.21
26	0.06	0.20	1.18	0.96	0.28
27	0.29	0.28	0.90	0.56	0.13
28	0.17	0.25	1.17	1.14	0.31
29	0.70	0.70	0.94	1.07	0.15
30	0.74	0.70	1.01	1.20	0.19
31	1.12	1.09	1.50	1.73	0.22
32	0.62	0.57	1.06	1.52	0.28
平均	-0.45	-0.44	1.17	0.18	0.23
SD	0.88	0.86	0.34	0.87	0.07

樣本資料經過 1PL 模式校準後，所得答對題數之能力估計如表三，1PL 所估得之能力範圍為-3.05 至 2.41 之間，其所採用的量尺平均為 0、標準差為 1，轉換為平均 30、滿分為 60 後的量尺分數呈現於表三的第 4 欄，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。

若以 1PL 模式校準後，再將估得之能力配合題目難度估出真分數，所呈現之真分數分佈如表三之第 5 欄，真分數範圍為 0.98~31.46。此與原始答對題數範圍(0~32)不同，這是因為除非考生的能力值為 $-\infty$ 或 $\infty$ 才有可能真分數達到 0 或 32。

因為學測所有的題目皆是四選一的選擇題，以試題反應理論而言，若鑑別度  $a$  無法控制在一小範圍內(0.8~1.25)(王寶墉，民 85)，且猜測度  $c$  又無法避

免，則最好的模式應是 3PL 模式(Kolen, & Brennan, 1995)。因此，本研究亦採用 3PL 模式來校準估計題目及學生能力參數，所得之結果如表二第 4, 5, 6 欄與表三第 7 欄，由表二之估計結果可以看出，事實上鑑別度及猜測度是存在的。鑑別度從 0.49~1.85，難度範圍為-1.49~1.73，猜測度為 0.11~0.39，其平均分別為 1.17, 0.18, 0.23。能力估計值的範圍則是-2.28~2.38。以 3PL 模式校準估計，對於答對題數相同，但其答對題目不同者，會產生能力估計不同的情形發生，甚至於有可能會有答對題數較多反而能力估計值較低的情況出現，如本研究答對 3 題之能力-2.03 大於答對 4 題之能力-2.04 的情況。為了以答對題數的方式來呈現能力估計值，因此，研究者將答對相同題數的能力作加權平均後呈現出。最後再藉由此能力分布轉換成量尺分數如表三第 8 欄，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。若再配合估得之題目參數，可計算得配合答對題數的 3PL 真分數分佈，如表三第 9 欄。我們發現其範圍為 8.19~30.87，與 0~32 有很大的差異，這是因為我們採用 3PL 模式，猜測度存在的原因，即使學生能力再低，他都有猜對該題的機率。

表三 第一次學測答對題數與量尺分數、1PL 與 3PL 模式之能力、真分數對照表

答對題數	學測	1PL 能力估計	1PL 量尺	1PL 真分數	3PL 能力估計	3PL 量尺	3PL 真分數
0	1	-3.05	1	0.98	-2.28	1	8.19
1	1	-2.75	1	1.37	-2.28	1	8.19
2	1	-2.49	1	2.15	-2.07	4	8.47
3	3	-2.30	1	2.78	-2.03	4	8.53

4	6	-2.14	3	3.42	-2.04	4	8.51
5	8	-1.94	6	4.36	-1.91	6	8.73
6	10	-1.71	9	5.63	-1.76	8	9.03
7	11	-1.52	11	6.83	-1.63	9	9.33
8	13	-1.41	12	7.58	-1.54	11	9.57
9	15	-1.33	13	8.15	-1.42	12	9.94
10	17	-1.23	15	8.89	-1.24	14	10.57
11	18	-1.08	17	10.05	-1.07	17	11.28
12	20	-0.87	19	11.79	-0.87	19	12.25
13	21	-0.67	22	13.54	-0.69	21	13.25
14	23	-0.54	23	14.73	-0.57	23	13.99
15	24	-0.47	24	15.37	-0.49	24	14.51
16	26	-0.42	25	15.84	-0.40	25	15.13
17	27	-0.34	26	16.59	-0.30	26	15.85
18	29	-0.21	27	17.82	-0.09	29	17.47
19	30	0.00	30	19.8	0.12	32	19.19
20	32	0.20	33	21.64	0.30	34	20.66
21	33	0.34	34	22.87	0.40	35	21.47
22	35	0.43	35	23.63	0.44	36	21.78
23	36	0.50	36	24.19	0.48	36	22.09
24	38	0.61	38	25.04	0.57	37	22.78
25	40	0.79	40	26.31	0.75	39	24.07
26	42	1.02	43	27.7	1.02	43	25.8
27	44	1.22	45	28.69	1.15	45	26.55
28	46	1.38	47	29.34	1.33	47	27.49
29	48	1.55	49	29.92	1.44	48	28.01
30	51	1.79	52	30.55	1.67	51	28.96
31	54	2.09	56	31.09	2.02	55	30.09
32	60	2.41	60	31.46	2.38	60	30.87

根據學測提供之資料，若作答資料以 Rasch 模式估計，以 0~60 量尺分數呈

現，則平均量尺分數為 30，量尺分數的標準差為 12.3；以 1PL 模式估計，則平均量尺分數為 30，量尺分數的標準差為 11.81；以 3PL 模式估計，則平均量尺分數為 30，量尺分數的標準差為 11.91。統計結果如表四。

表四 第一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	30	12.3
樣本_1PL	30	11.81
樣本_3PL	30	11.91

## 第二節 民國九十年第二次國中基本學力測驗數學科校準後所呈現量尺分數之研究結果

為了解所有考生與本研究所獲得資料間的差異，茲將敘述性的統計列於表五，民國九十年第二次國中基本學力測驗數學科所參加的總人數為 167440 人，若以每題答對得一分來計算，平均分數為 18.67，分數的標準差為 7.40，此次考試的信度學測中心並無提供。而此研究所獲得第二次的樣本數為 4737 人，平均分數為 18.76，分數的標準差為 7.35，與母體的差異均很小，分別只有 0.09 與 0.05，信度 alpha 為 0.90。

若作答資料改以 Rasch 模式估計，以量尺分數呈現，則平均量尺分數為 30，

量尺分數的標準差為 12.3，

表五 第二次國中學測母體與樣本資料統計表

	母體	樣本
總人數	167440	4737
原始平均分數	18.67	18.76
原始分數的 SD	7.4	7.35
信度 alpha	無	0.9

第二次國中學測所取得的樣本資料經過 BILOG-MG 以 1PL 及 3PL 模式估計後，和學測所題供之 Rasch 模式 b 值呈現於表六中，本研究經 1PL 模式校準後估計所得 b 值與學測所提供之 b 值差異大於第一次學測，平均難度相差達 0.16，其中以第 16 題差異最大為 1.00，第 23 題最小差異為 0.01，標準差相差也較第一次學測大，相差達 0.37。

表六 第二次國中學測 Rasch、1PL 及 3PL 模式之試題參數估計值

題號	學測估計		樣本估計		
	Rasch_b	1PL_b	3PL_a	3PL_b	3PL_c
1	-1.61	-1.24	1.62	-0.67	0.20
2	-0.35	-0.50	1.18	0.23	0.31
3	0.13	-0.29	0.91	0.29	0.24
4	-1.80	-1.03	1.40	-0.51	0.20
5	-1.01	-0.92	1.39	-0.47	0.17
6	-1.27	-0.98	2.05	-0.33	0.26
7	-0.15	-0.55	1.48	0.12	0.29
8	0.04	-0.24	0.91	0.02	0.10

9	-2.15	-1.61	1.25	-0.73	0.42
10	-0.65	-0.54	0.92	0.15	0.29
11	-0.76	-1.16	1.87	-0.13	0.45
12	-0.87	-0.82	0.84	-0.46	0.18
13	0.66	0.15	1.60	0.52	0.18
14	0.07	-0.50	1.71	-0.14	0.13
15	-1.57	-1.33	1.45	-0.46	0.39
16	-2.82	-1.82	1.17	-1.22	0.26
17	-0.34	-0.53	1.15	-0.10	0.18
18	-0.80	-0.70	1.44	-0.23	0.19
19	0.14	-0.17	1.56	0.06	0.08
20	-1.24	-1.47	1.10	-0.85	0.29
21	0.03	-0.30	1.59	0.29	0.26
22	0.71	0.12	2.15	0.68	0.25
23	0.35	0.36	2.10	0.60	0.15
24	-0.39	-0.44	1.58	0.05	0.21
25	0.41	-0.14	1.89	0.86	0.38
26	0.42	0.04	1.43	0.63	0.25
27	1.02	0.38	2.06	1.02	0.27
28	0.76	0.19	1.24	1.46	0.37
29	1.04	0.37	1.64	0.68	0.16
30	0.67	0.16	1.18	1.20	0.33
31	0.64	0.10	0.94	0.78	0.24
平均	-0.34	-0.50	1.44	0.11	0.25
SD	0.99	0.62	0.38	0.64	0.09

樣本資料經過 IPL 模式校準後，所得答對題數之能力估計如表七，IPL 所估得之能力範圍為-2.82 至 2.10 之間，為了配合第一次學測的能力範圍，因此將其拓寬使其相符後，再轉為滿分為 60 的量尺分數，結果呈現於表七的第 4 欄，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。答對 2 題空白，乃所得資料並無恰好答對 2 題者。另以 3PL 模式估計，結果答對 1 者，其能力竟高於 3 題與 4 題者，此乃樣本資料答對 1 題的只有 1 人，而答對 3 與 4 題者其平均反而小的緣故。

若以 1PL 模式校準後，再將估得之能力配合題目難度估出真分數，所呈現之真分數分布如表七之第 5 欄，真分數範圍為 0.97~29.47。

同樣地，本研究亦採用 3PL 模式來校準估計題目及學生能力參數，所得之結果如表六第 4, 5, 6 欄與表七第 6 欄，由表六之估計結果可以看出，事實上鑑別度及猜測度亦是存在的。鑑別度從 0.84~2.15，難度範圍為-1.22~1.46，猜測度為 0.08~0.45，其平均分別為 1.44, 0.11, 0.25。能力估計值的範圍則是 -2.01~2.00。若先轉換成與第一次學測所估的能力範圍，再轉成 0~60 的量尺，結果呈現於表七的第 7 欄，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。若再配合估得之題目參數，可計算得配合答對題數的 3PL 真分數分布，如表七第 8 欄，真分數範圍為 8.22~30.3。

表七 第二次學測答對題數與量尺分數、1PL 與 3PL 模式之能力、真分數對照表

答對題數	學測	1PL 能力估計	1PL 量尺	1PL 真分數	3PL 能力估計	3PL 量尺	3PL 真分數
0	1	-2.82	1	0.97	-2.01	1	8.22
1	1	-2.50	1	1.58	-1.74	5	8.55
2	1						
3	4	-2.13	2	2.7	-1.85	5	8.4
4	6	-1.92	4	3.6	-1.76	5	8.52
5	8	-1.66	8	5	-1.67	6	8.67
6	10	-1.47	11	6.25	-1.60	7	8.8
7	12	-1.38	11	6.91	-1.52	8	8.96
8	14	-1.34	12	7.21	-1.46	9	9.1
9	16	-1.29	13	7.6	-1.32	11	9.47
10	18	-1.19	15	8.42	-1.16	14	10.01
11	20	-0.99	17	10.19	-0.98	16	10.77
12	21	-0.74	21	12.61	-0.85	18	11.45
13	23	-0.57	23	14.33	-0.62	22	12.91
14	25	-0.49	24	15.15	-0.52	23	13.65
15	26	-0.46	25	15.46	-0.47	24	14.05
16	28	-0.44	25	15.67	-0.43	25	14.37
17	29	-0.41	25	15.97	-0.30	26	15.48
18	31	-0.33	26	16.8	-0.13	29	17.03
19	32	-0.15	29	18.61	0.12	33	19.39
20	34	0.10	32	21	0.31	35	21.19
21	36	0.30	35	22.72	0.40	37	22.03
22	37	0.40	37	23.5	0.44	37	22.4
23	39	0.44	37	23.8	0.45	37	22.49
24	41	0.49	38	24.16	0.46	38	22.58
25	43	0.59	39	24.84	0.52	38	23.12

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。