

## 五、 模擬研究與探討

『智勝鮮師網』的適性測驗除了僅支援文字試題、單一貝氏能力估計引擎…(表 1: 系統功能延伸前後比較)等缺點以外,經研究者實際操作後發現系統尚有 IRT 能力估算不精確的缺點。除了全對和全錯的反應組型以外,多數反應組型的適性測試長度均為 20 題。換言之,『智勝鮮師網』適性測驗的終止是因為符合最大測驗長度。它類似大多數線上電腦適性測驗系統基於測驗效率、測驗目的、測驗時間、…等因素所採用的測驗終止條件之一。最大測驗長度終止適性測驗的代價即為 IRT 能力估算引擎輸出的能力估計值具有某種程度的誤差。如何改善能力估計精確度與效率不可兼得的問題?解決策略設定為在系統延伸部分內建適性模擬功能,以模擬真實電腦適性測驗運作,比較不同能力估計引擎的效能。另外,文獻也指出當受試者的反應組型特異(aberrant),例如:答對一些很難的題目,但一些很簡單的題目卻答錯時,牛頓迭代過程會發散(diverge)(R. K. Hambleton et al., 1991[14]的第 35、36 頁),最大牛頓迭代次數限制雖可避免無窮迭代迴路,但極大值不存在,造成最大概似估計法仍是無法收斂,因此必須仰賴最大測驗長度終止適性測驗。模擬研究進一步探討其它的能力估計引擎是否也存在無法收斂的反應組型呢?相同反應組型是否可能造成四種能力估計引擎無法收斂呢?

模擬研究是使用以 ASP 發展的 CAT 模擬程式所主導的。為了避免混淆效果(confounding effects),模擬施測限制在一個單純的電腦適性測驗(pure CAT)環境下,換言之,並不考慮試題主題內容均衡(content balancing)和曝光度控制(exposure control)。能力估計引擎使用四種方法,分別是 OWEN、MLE、EAP 和 WLE。至於受試反應組型則包括 (1) all “1”s、(2) all “0”s、(3) “0” bits shifted right to all “1”s、(4) “1” bits shifted right to all “0”s、(5) consecutive multiple bits toggling、和 (6) walking “0” bits 等六大類別,共計 150 種反應組型。

### 5.1 方法

首先,建立虛擬適性測驗題庫。以智勝公司「編輯手」軟體製作 160 題多媒體試題,上傳所有試題至系統資料庫,接著開啟 160 題 IRT 試題參數的 Excel 檔案和系統資料庫的 Access 檔案,將所有試題參數以複製貼上的方式,將 160 題 IRT 試題 a、b、c 參數分別加到 TestItem 資料表的 pa、pb、pc 欄位,將 160 題試題的 cat 欄位勾選。任意給定一個系統尚未存在的 nSampleExamid 和 nIRTEExamid 值,接著設定 nSampleExamid 和 nIRTEExamid 資料表的紀錄(record)內容,如表所示。步驟至此即完成 160 題虛擬適性測驗題庫建立,可供模擬測驗使用。

有關於設定 SampleExam 資料表和 IRTEExam 資料表有關模擬取樣測驗部分紀錄(recordset)可能的內容如表 14 和表 15 所示。

表 14 SampleExam 資料表模擬取樣測驗紀錄部分內容

nSampleExamid	nSampleExamName	nSnoid	nTeacherID	nNum
45	多媒體數學模擬取樣	6.1	3.323603	160

表 15 IRTEExam 資料表模擬適性測驗紀錄部分內容

nIRTEExamid	nIRTEExamName	nSnoid	nTeacherID	nNum
5	多媒體數學模擬適性	6.1	3.323603	160

50 組受試反應組型儲存成一個本文檔，每種受試反應組型長 160 位元，代表可能的 0 與 1 反應組型。根據虛擬題庫的 IRT 參數值，系統的四種能力估計(OWEN、EAP、MLE 和 WLE)模擬程式分別讀取受試反應組型文字檔，以模擬程式估計能力值，測驗終止時輸出每一種反應組型使用的測驗長度和終止時能力估計值。

## 5.2 題庫

本研究所使用的題庫包含 160 題虛擬試題。試題參數則是引用陳新豐，民 87[10] 研究結果，該研究所建立 CAT 題庫的 IRT 試題參數敘述統計量如表 16 所示。完整試題參數表列請參考附件 5。表 17 題庫參數資料表為示意。該研究所產生的 IRT 試題參數對本系統模擬的適性測驗中試題選擇、試題反應的產生和能力估計被視為真實。

表 16 題庫的試題參數敘述統計量

	a	b	c
平均數	1.036231	-0.07557	0.272688
變異數	0.144668	1.503791	0.003127
最大值	2.5	3	0.44
最小值	0.4	-3	0.08

表 17 題庫參數示意資料表

題目序號	參數 a	參數 b	參數 c
1	0.763	-3	0.27
2	0.542	-2.218	0.28
3	1.252	0.665	0.33
.	.	.	.
.	.	.	.
.	.	.	.
160	1.17	0.124	0.26

資料來源：陳新豐，「多媒體線上適性測驗系統發展及其相關研究」，國立台南師範學院，碩士論文，民 88，附錄三：題庫參數估計結果資料(95-99 頁)

### 5.3 反應組型

本研究的主要目的並不在於從任意反應組型中確認受試者行為特性(作弊、猜題、不明白題目語意、事先知道題庫中的一些試題、.....)所造成特異反應組型(aberrant response patterns)。這些模擬測驗反應組型是用來測試在 3-PL IRT 模式下四種能力估計法的能力估計效能與特定反應組型是否造成 IRT 能力即時估算 (IRT 重點) 誤差？進而導致能力估算無法收斂或快速收斂，最終影響 IRT 估算的可靠度 (reliability) ？

模擬測驗反應組型的建構是基於下列因素考量：

1. 規律性
2. 實際 CAT 環境會產生的
3. 探討同一類反應組型能力估計收斂與不收斂邊界條件
4. 有些學生作答猜題的習慣，全部選擇特定答案
5. 試題正確答案分佈
6. 能力值估算能夠呈現電腦化適性測驗不同於古典測驗的優點

本研究答題反應組型包含六大類別如下，類別 1~5 反應組型如表 17 所示，至於第 6 類反應組型則列於附件 6.

反應組型類別：

1. all "1" bits : 111.....1111
2. all "0" bits : 000.....0000
3. "0" bits shifted right into all "1"s : 0...01111...1 et al.
4. "1" bits shifted right into all "0"s : 1...1000....0 et al.
5. consecutive multiple bits toggling :
  - (1)、11...100...011...100...0.....
  - (2)、00...011...100...011...1.....

6. walking “0” bits :

- (1) 、 one “0” : one “0” bit walking through the left most 10 bits
- (2) 、 two “0”s : two “0” bits walking through the left most 14 bits
- (3) 、 three “0”s : three “0” bits walking through the left most 16 bits
- (4) 、 four “0”s : four “0” bits walking through the left most 19 bits
- (5) 、 five “0”s : five “0” bits walking through the left most 20 bits
- (6) 、 six “0”s : six “0” bits walking through the left most 24 bits
- (7) 、 seven “0”s : seven “0” bits walking through the left most 21bits

表 18 1~5 類受試反應組型

反應組型 編號	反應組型內容，長度為 160 位元 123...160
1	01...1
2	001...1
3	0001...1
4	00001...1
5	000001...1
6	0000001...1
7	00000001...1
8	000000001...1
9	0000000001...1
10	00000000001...1
11	000000000001...1
12	0000000000001...1
13	00000000000001...1
14	000000000000001...1
15	0000000000000001...1
16	10...0
17	110...0
18	1110...0
19	11110...0
20	111110...0
21	1111110...0
22	11111110...0
23	111111110...0
24	1111111110...0
25	11111111110...0
26	111111111110...0

27	111111111110...0
28	111111111110...0
29	111111111110...0
30	111111111110...0
31	111111111110...0
32	111111111110...0
33	111111111110...0
34	111111111110...0
35	10...
36	01...
37	1100...
38	0011...
39	111000...
40	000111...
41	11110000...
42	00001111...
43	1111100000....
44	0000011111...
45	111111000000...
46	000000111111...
47	11111110000000...
48	00000001111111...
49	1...1
50	0...0

#### 5.4 四種能力估計法之 CAT 模擬結果

對任何受試反應組型而言，由於 OWEN 和 EAP 模擬適性測驗的第一題是從整個題庫以亂數方式選取，因此測驗長度和能力估計值可能不同。表 19 ~ 表 22 是四種能力估計法饋送 50 組 1~5 類答題反應組型的模擬結果，其中平均測驗長度和能力估計平均值是將 10 次相同能力估計法模擬結果利用 Excel 求平均值。至於第 6 類 100 組答題反應組型的單一模擬結果則列於附件 7 供研究參考。

表 19 Owen's Bayesian Sequential 能力估計法模擬結果

反應組型編號	平均測驗長度 (10 次)	能力估計平均值 (10 次)	受試反應組型
1	21.8	3	01...1

2	42.3	3	001...1
3	70.8	3	0001...1
4	116.9	2.99	00001...1
5	151	2.985	000001...1
6	142.6	2.365	0000001...1
7	75.4	0.03	00000001...1
8	58.8	-0.22	000000001...1
9	7.5	-3	0000000001...1
10	8.2	-3	00000000001...1
11	8.2	-3	000000000001...1
12	8.1	-3	0000000000001...1
13	7.8	-3	00000000000001...1
14	7.7	-3	000000000000001...1
15	7.7	-3	0000000000000001...1
16	11	-3	10...0
17	13.3	-3	110...0
18	15.4	-3	1110...0
19	18.1	-3	11110...0
20	20.2	-3	111110...0
21	23.2	-3	1111110...0
22	26.1	-3	11111110...0
23	27.1	-3	111111110...0
24	23.9	-1.8	1111111110...0
25	17.6	0.6	11111111110...0
26	9.9	3	111111111110...0
27	10.2	3	1111111111110...0
28	9.9	3	11111111111110...0
29	10.2	3	111111111111110...0
30	10.3	3	1111111111111110...0
31	10.5	3	11111111111111110...0
32	10.1	3	111111111111111110...0
33	10.3	3	1111111111111111110...0
34	10.1	3	11111111111111111110...0
35	130.3	-1.215	10...
36	132.2	-1.41111	01...
37	160	-1.345	1100...
38	160	-1.62	0011...

39	160	-1.3	111000...
40	160	-1.585	000111...
41	160	-1.405	11110000...
42	160	-1.635	00001111...
43	160	-1.455	1111100000....
44	144.5	-1.765	0000011111...
45	160	-1.51	111111000000...
46	144.6	-1.725	000000111111...
47	160	-1.575	11111110000000...
48	98.7	-2.155	00000001111111...
49	10.4	3	All '0'
50	7.9	-3	All '1'

表 20 期望後驗估計法(EAP)模擬結果

反應組型編號	平均測驗長度 (10 次)	能力估計平均值 (10 次)	受試反應組型
1	20.5	2.813138737	01...1
2	32.8	2.799110503	001...1
3	58.3	2.776596592	0001...1
4	75.9	2.754059963	00001...1
5	94.7	2.741118327	000001...1
6	130.4	2.169351832	0000001...1
7	6.9	-2.740431952	00000001...1
8	7	-2.737083179	000000001...1
9	7	-2.738554443	0000000001...1
10	6.9	-2.741133975	00000000001...1
11	7	-2.739490783	000000000001...1
12	7	-2.73802124	0000000000001...1
13	7	-2.739972255	00000000000001...1
14	6.9	-2.737514816	000000000000001...1
15	6.7	-2.736416393	00000000000000001...1
16	12.2	-2.73912113	10...0
17	14.9	-2.73807901	110...0
18	17.2	-2.737624389	1110...0
19	18.6	-2.735986373	11110...0
20	20.3	-2.736548792	111110...0

21	20.3	-2.738975896	111110...0
22	20.1	-2.737104662	1111110...0
23	20.9	-2.738013422	11111110...0
24	20.9	-2.73818342	111111110...0
25	12.6	1.758823586	1111111110...0
26	10.6	2.886962787	11111111110...0
27	10.4	2.885447664	111111111110...0
28	10.7	2.88747604	1111111111110...0
29	10.5	2.885531224	11111111111110...0
30	10.2	2.884750028	111111111111110...0
31	10.7	2.888223759	1111111111111110...0
32	10.5	2.886549503	11111111111111110...0
33	10.9	2.888392655	111111111111111110...0
34	10.4	2.886079981	1111111111111111110...0
35	160	-1.341786832	10...
36	160	-1.775128409	01...
37	160	-1.363207141	1100...
38	160	-1.650274271	0011...
39	160	-1.527541047	111000...
40	160	-1.802418673	000111...
41	160	-1.490668803	11110000...
42	160	-2.010156756	00001111...
43	160	-1.539256226	1111100000....
44	160	-1.766020416	0000011111...
45	160	-1.537470228	111111000000...
46	160	-1.912469359	000000111111...
47	160	-1.46637061	11111110000000...
48	6.8	-2.739248982	00000001111111...
49	10.5	2.887449763	All '0'
50	7	-2.738951412	All '1'

表 21 最大概似估計法( MLE )模擬結果，

反應組 型編號	測驗長度	能力估計值	受試反應組型
1	7.9	1.140459	01...1
2	11.7	0.270501	001...1
3	15.4	-0.0219	0001...1



4	17.5	-0.4623	00001...1
5	19.1	-0.73248	000001...1
6	20.5	-0.96511	0000001...1
7	27	-1.081	00000001...1
8	29	-1.13705	000000001...1
9	31.7	-1.21308	0000000001...1
10	35	-1.27638	00000000001...1
11	37	-1.30806	000000000001...1
12	40	-1.34547	0000000000001...1
13	23	-3	00000000000001...1
14	23	-3	000000000000001...1
15	23	-3	0000000000000001...1
16	22.3	-3	10...0
17	24.3	-3	110...0
18	23.6	-2.45084	1110...0
19	12.3	-1.00719	11110...0
20	8.4	0.626983	111110...0
21	31.2	-2.22801	1111110...0
22	11	1.158391	11111110...0
23	12.1	1.250299	111111110...0
24	13.3	1.326948	1111111110...0
25	15	1.233276	11111111110...0
26	14	2.680514	111111111110...0
27	16	-3	1111111111110...0
28	17	-3	11111111111110...0
29	18	3	111111111111110...0
30	18	3	1111111111111110...0
31	19	3	11111111111111110...0
32	20	3	111111111111111110...0
33	21	3	1111111111111111110...0
34	22	3	11111111111111111110...0
35	8.6	-0.62862	10...
36	11	-2.36326	01...
37	9.5	-0.26732	1100...
38	16.1	-2.75744	0011...
39	14.5	-0.4272	111000...
40	34.5	-2.38905	000111...

41	9.2	0.285764	11110000...
42	16.2	-1.74592	00001111...
43	8.4	0.634165	1111100000....
44	16.2	-1.9401	0000011111...
45	16.3	0.726589	111111000000...
46	13	-1.6942	000000111111...
47	11	1.15212	11111110000000...
48	15	-2.23212	00000001111111...

表 22 權值概似估計法( WLE )模擬結果，

反應組 型編號	測驗長度	能力估計值	備註
1	12.3	0.65578	01...1
2	13.5	0.17866	001...1
3	17.3	-0.05338	0001...1
4	18.6	-0.3995	00001...1
5	19.5	-0.64608	000001...1
6	20.2	-0.89257	0000001...1
7	20.1	-1.08338	00000001...1
8	21.4	-1.12127	000000001...1
9	22.3	-1.2117	0000000001...1
10	24.9	-1.29733	00000000001...1
11	28	-1.34897	000000000001...1
12	25.5	-1.3545	0000000000001...1
13	25.4	-1.42882	00000000000001...1
14	26.8	-1.50772	000000000000001...1
15	28	-1.71596	0000000000000001...1
16	12.5	-1.12311	10...0
17	7.5	-1.02142	110...0
18	9.8	-1.09997	1110...0
19	12.5	-0.3592	11110...0
20	13.9	0.08402	111110...0
21	16.1	1.007075	1111110...0
22	17	0.45178	11111110...0
23	14	-0.13407	111111110...0
24	14.7	0.22279	1111111110...0
25	15.1	0.5616	11111111110...0

26	18.6	0.62286	11111111110...0
27	20	0.84884	111111111110...0
28	21.8	0.163478	1111111111110...0
29	21.4	1.18876	11111111111110...0
30	18	-0.42201	111111111111110...0
31	21.8	-0.27106	1111111111111110...0
32	26	-2.08413	11111111111111110...0
33	25.2	0.48196	111111111111111110...0
34	22	0.22599	1111111111111111110...0
35	10.2	-0.37001	10...
36	9.6	-0.63635	01...
37	11.5	-0.18634	1100...
38	13.1	-0.97093	0011...
39	11.8	-0.24291	111000...
40	11.2	-1.15895	000111...
41	14	0.28377	11110000...
42	10.8	-1.16386	00001111...
43	16	0.15075	1111100000....
44	12	-0.75924	0000011111...
45	19.2	0.62069	111111000000...
46	14.6	-1.34596	000000111111...
47	19.7	1.61745	11111110000000...
48	16.1	-1.52996	00000001111111...

將上述模擬結果能力估計值部分整理成表，如表 23~表 25 所示。並以 Excel 畫出不同能力估計法對 48 組反應組型能力估計值圖形，如圖 41 所示。

表 23 饋送反應組型 1~10 之四種能力估計法的模擬測驗能力估計值

反應組型	MLE 能力估計值	OWEN 能力估計值	WLE 能力估計值	EAP 能力估計值
01...1	1.140459	3	0.65578	2.813138737
001...1	0.270501	3	0.17866	2.799110503
0001...1	-0.0219	3	-0.05338	2.776596592
00001...1	-0.4623	2.99	-0.3995	2.754059963
000001...1	-0.73248	2.985	-0.64608	2.741118327
0000001...1	-0.96511	2.365	-0.89257	2.169351832
00000001...1	-1.081	0.03	-1.08338	-2.740431952
000000001...1	-1.13705	-0.22	-1.12127	-2.737083179
0000000001...1	-1.21308	-3	-1.2117	-2.738554443

00000000001...1	-1.27638	-3	-1.29733	-2.741133975
-----------------	----------	----	----------	--------------

表 24 饋送反應組型 11~40 之四種能力估計法的模擬測驗能力估計值

反應組型	MLE 能力估計值	OWEN 能力估計值	WLE 能力估計值	EAP 能力估計值
000000000001...1	-1.30806	-3	-1.34897	-2.739490783
0000000000001...1	-1.34547	-3	-1.3545	-2.73802124
00000000000001...1	-3	-3	-1.42882	-2.739972255
0000000000000001...1	-3	-3	-1.50772	-2.737514816
00000000000000001...1	-3	-3	-1.71596	-2.736416393
10...0	-3	-3	-1.12311	-2.73912113
110...0	-3	-3	-1.02142	-2.73807901
1110...0	-2.45084	-3	-1.09997	-2.737624389
11110...0	-1.00719	-3	-0.3592	-2.735986373
111110...0	0.626983	-3	0.08402	-2.736548792
1111110...0	-2.22801	-3	1.007075	-2.738975896
11111110...0	1.158391	-3	0.45178	-2.737104662
111111110...0	1.250299	-3	-0.13407	-2.738013422
1111111110...0	1.326948	-1.8	0.22279	-2.73818342
11111111110...0	1.233276	0.6	0.5616	1.758823586
111111111110...0	2.680514	3	0.62286	2.886962787
1111111111110...0	3	3	0.84884	2.885447664
11111111111110...0	3	3	0.163478	2.88747604
111111111111110...0	3	3	1.18876	2.885531224
1111111111111110...0	3	3	-0.42201	2.884750028
11111111111111110...0	3	3	-0.27106	2.888223759
111111111111111110...0	3	3	-2.08413	2.886549503
1111111111111111110...0	3	3	0.48196	2.888392655
11111111111111111110...0	3	3	0.22599	2.886079981
10...	-0.62862	-1.215	-0.37001	-1.341786832
01...	-2.36326	-1.41111	-0.63635	-1.775128409
1100...	-0.26732	-1.345	-0.18634	-1.363207141
0011...	-2.75744	-1.62	-0.97093	-1.650274271
111000...	-0.4272	-1.3	-0.24291	-1.527541047
000111...	-2.38905	-1.585	-1.15895	-1.802418673

表 25 饋送反應組型 41~48 之四種能力估計法的模擬測驗能力估計值

反應組型	MLE 能力估計值	OWEN 能力估計值	WLE 能力估計值	EAP 能力估計值
11110000...	0.285764	-1.405	0.28377	-1.490668803
00001111...	-1.74592	-1.635	-1.16386	-2.010156756
1111100000....	0.634165	-1.455	0.15075	-1.539256226
0000011111...	-1.9401	-1.765	-0.75924	-1.766020416
111111000000...	0.726589	-1.51	0.62069	-1.537470228
000000111111...	-1.6942	-1.725	-1.34596	-1.912469359
11111110000000...	1.15212	-1.575	1.61745	-1.46637061
00000001111111...	-2.23212	-2.155	-1.52996	-2.739248982

四種能力估計法之能力估計值 vs 48組反應組型

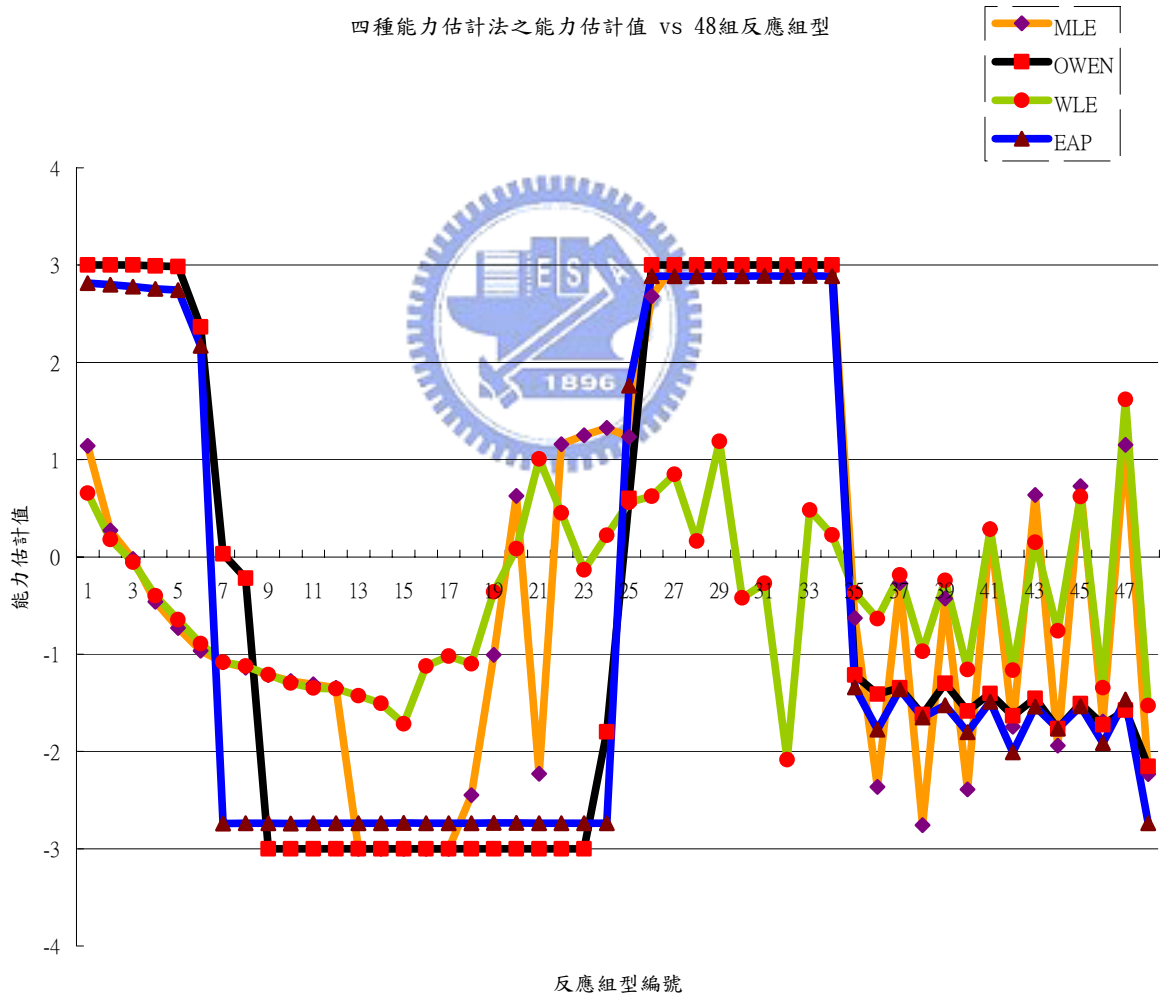


圖 41 48 種受試反應組型 vs 四種能力估計法能力估計值圖形

## 5.5 模擬研究相關討論

檢視模擬結果表 19~表 22 和圖 41 後，由測驗所需題數 (test length) 發現反應組型 2~8 和 35~48 將造成 OWEN 能力估計法慢速收斂 (slowly converges) 甚至發生整個題庫用盡也無法收斂 (diverges) 的情況。同樣的，反應組型 2~6 和 35~47 將造成 EAP 能力估計法慢速收斂甚至發生整個題庫用盡也無法收斂的情況。因此，本研究將挑選組型 2、6、35、41、47，研究饋送以上五種反應組型時，四種能力估計法在 CAT 運作過程中能力估計值的變化 (behavior of provisional ability,  $\hat{\theta}$ ) (圖 42~圖 61)，作進一步觀察與結果分析。

圖 42~圖 49 的圖形顯示組型 2 和組型 6 的 MLE 和 WLE 能力估計法快速收斂，能力估計值標準誤小於預定值，符合 CAT 固定標準誤 (fixed-SE) 終止規則。OWEN 和 EAP 能力估計法則發生慢速收斂 (slowly converges) 甚至發生整個題庫用盡也無法收斂 (diverges) 的情況。更重要的是，當測驗長度為 20、25 或 30 時，其暫時能力估計值與最終收斂值或題庫用盡時的能力估計值有相當大的誤差。這證實 CAT 系統能力估計引擎採用貝氏定理能力估計法實作，並且其測驗終止規則包含最大測驗長度為 20、25 或 30 時，貝氏引擎的能力估計值精確度是不可接受的，因此電腦適性測驗會喪失的能力估計可靠度高的基本優點。問題解決的方法是改用 MLE 或 WLE 能力估計引擎重新估算。至此獲得研究結論之一，當 CAT 系統運作偵測到組型 2~組型 6 時，最好改採 MLE 或 WLE 能力估算引擎重新估算能力。

圖 50~圖 57 的上下擺盪圖形顯示組型 35 和組型 41 的 MLE 和 WLE 能力估計法快速收斂 (收斂速度大於組型 2 和組型 6)，能力估計值標準誤小於預定值，符合 CAT 固定標準誤 (fixed-SE) 終止規則。OWEN 和 EAP 能力估計法則發生整個題庫用盡也無法收斂 (diverges) 的情況。更重要的是，當測驗長度為 20、25 或 30 時，其暫時能力估計值與最終收斂值或題庫用盡時的能力估計值有相當大的誤差。其誤差程度甚至大於組型 2 和組型 6 之貝氏能力估計法，因此，上述改採 MLE 或 WLE 能力估算引擎重新估算能力的建議更適用於組型 35 和組型 41。

總結，反應組型 2~6 和反應組型 35~48 適用 MLE 或 WLE 能力估算引擎重新估算能力的建議，故研究目的二已達成。

此外，檢視模擬結果如附件 7 (饋送第 6 類反應組型-- "walking "0" bits") 的附表所示，對於第 6 類反應組型 OWEN 和 EAP 能力估計法子類 1~子類 5 的反應組型 1~69，OWEN 和 EAP 能力估計法則發生能力估計值幾乎相同的結果，這證實貝氏定理具有向先前驗 (priori) 分配平均數迴歸的本質缺點 (inherent drawback)，結果是造成估計的偏差。相反的，對同一子類的反應組型而言，MLE 能力估計法的效能表現則展現電腦適性測驗勝過古典測驗理論之處，相同的錯誤題數，卻因為錯誤發生於 CAT 過程的不同試題順序而有不同能力估計值。

前面文獻探討論述中，提到在實際 CAT 系統運作，系統採用固定標準誤 (fixed-SE) 終止規則時，則 MLE 的估計偏差會呈現兩端往內偏差 (inward bias)，也就是在  $\theta$  量尺上較低能力端的能力值高估，較高能力端的能力值低估 (Yi et al., 2001[26])。這將部分解釋組型 1~6 的兩類估計法的能力估計值的差距甚大。另外，從實際模擬操作中發

現，相同反應組型之模擬測驗之 MLE 能力估計值變化相當大。

1. 組型 2 (001...1) 能力估計值的變化的細部觀察

組型2 (001..1) 的OWEN能力估計法的CAT能力值變化

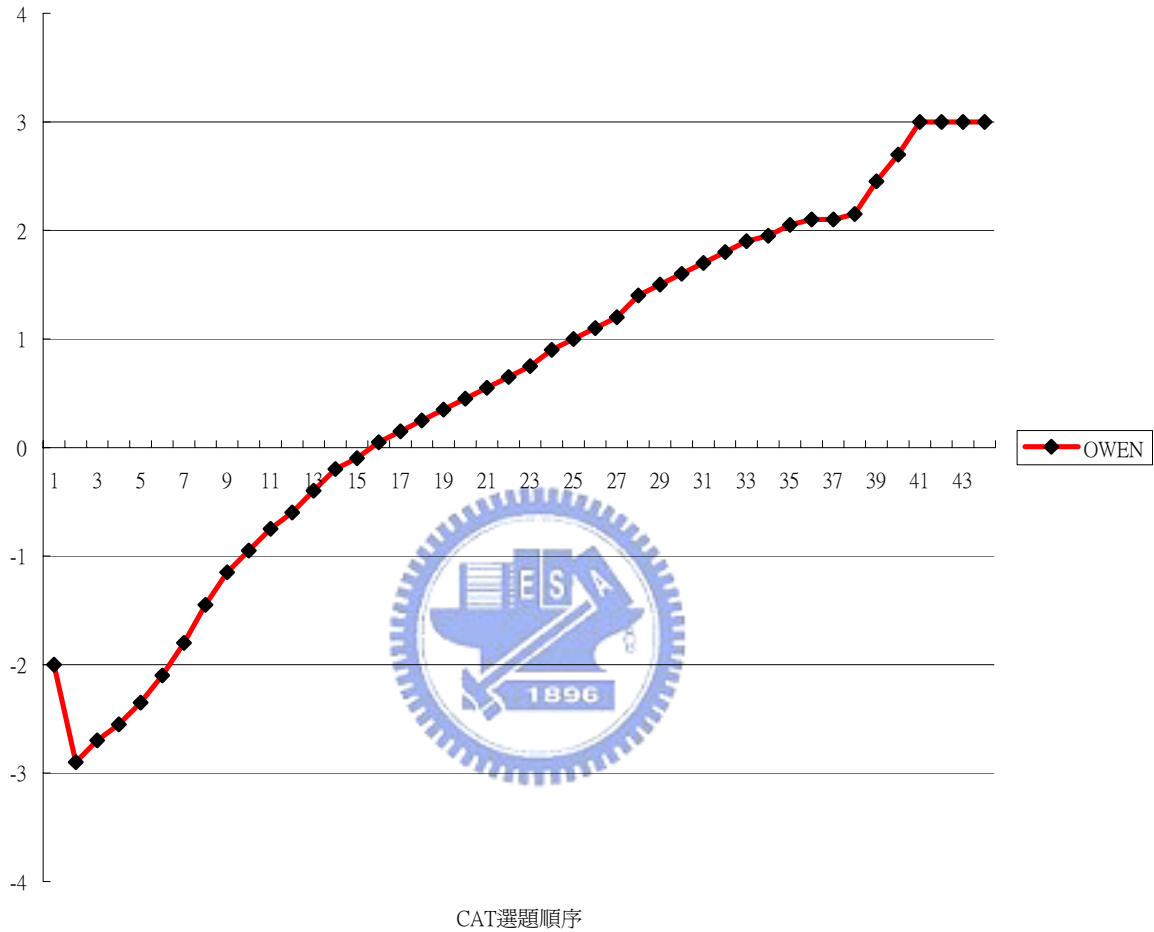


圖 42 組型 2 (001..1) 的 OWEN 能力估計法的 CAT 能力值變化

組型2 (001..1) 的EAP能力估計法的CAT能力值變化

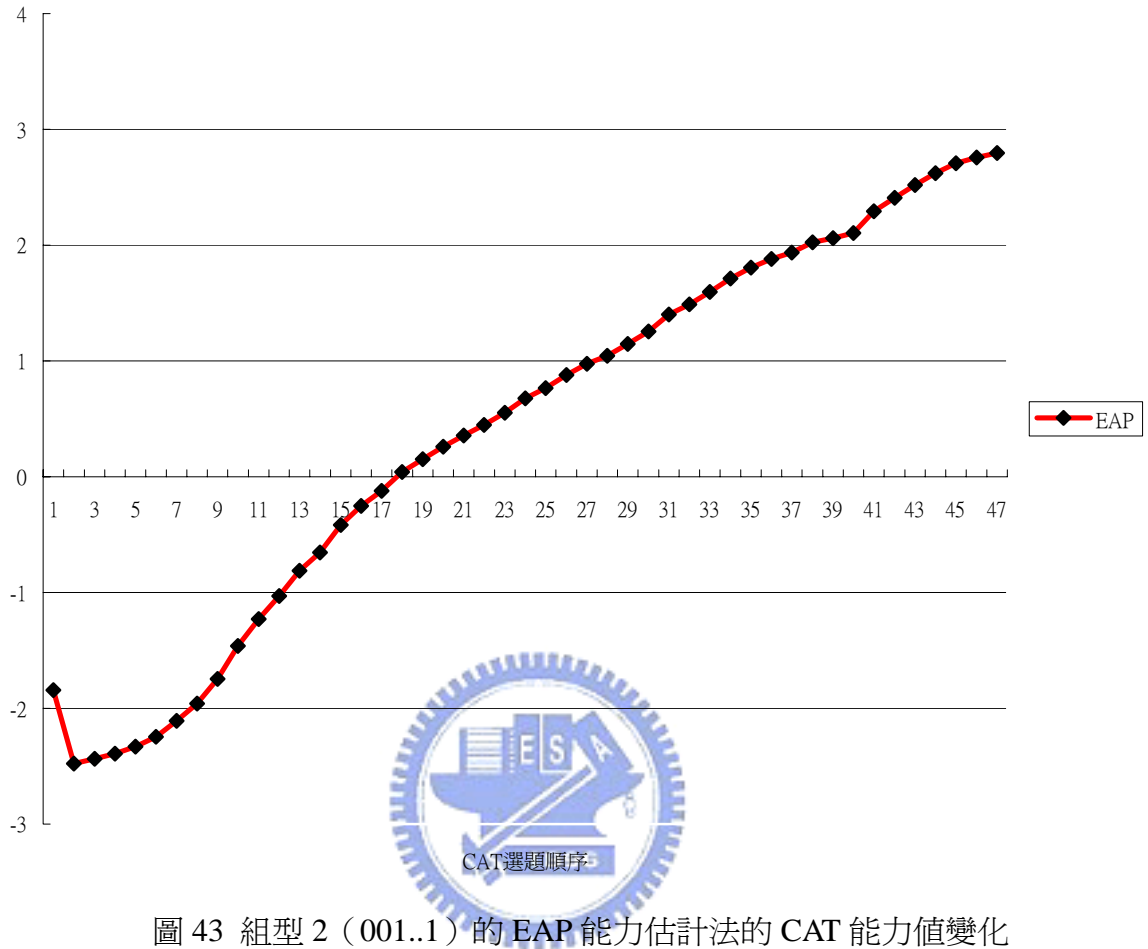


圖 43 組型 2 (001..1) 的 EAP 能力估計法的 CAT 能力值變化

組型2 (001..1) 的MLE能力估計法的CAT能力值變化

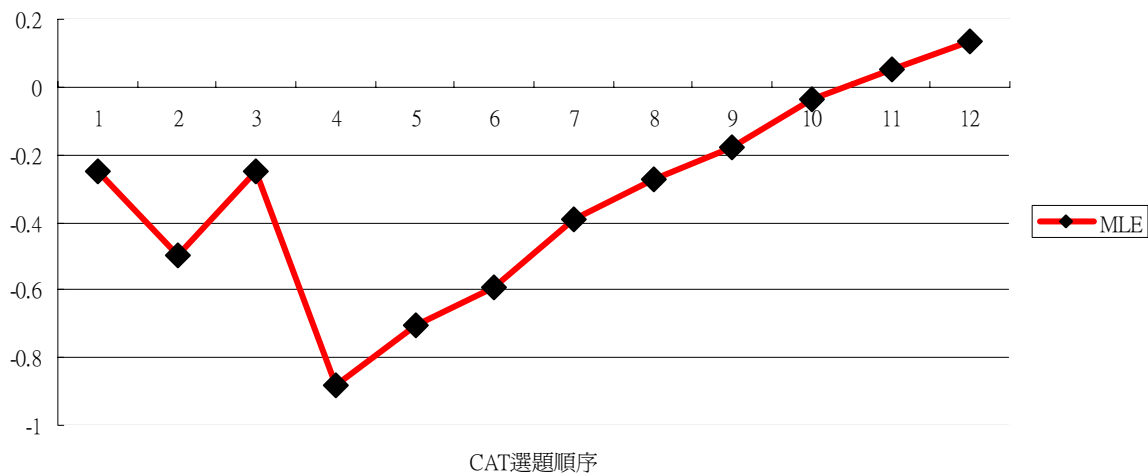


圖 44 組型 2 (001..1) 的 MLE 能力估計法的 CAT 能力值變化  
 註：MLE 能力估計是從反應組型位元 4 開始，反應組型位元 1~3 代表 step rule 能力值變化



組型2 (001..1) 的WLE能力估計法的CAT能力值變化

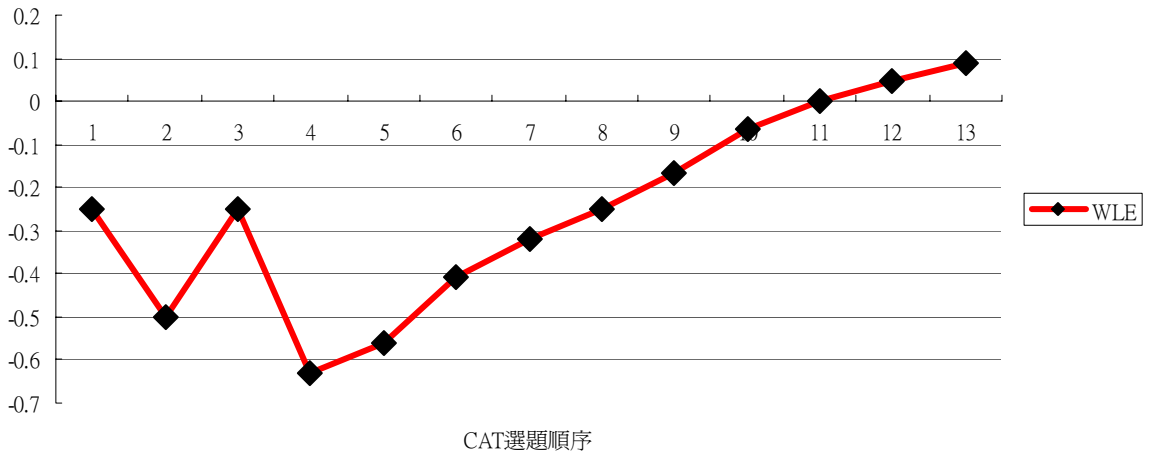
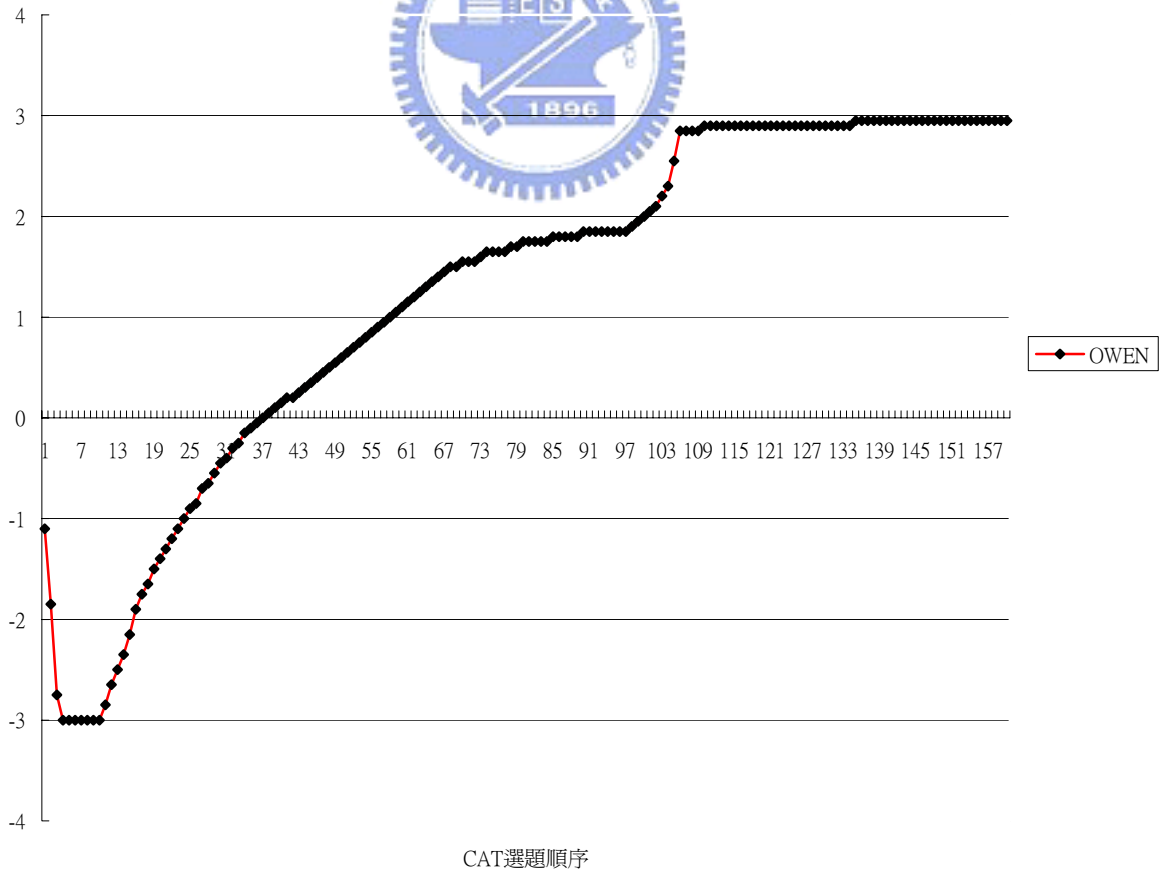


圖 45 組型 2 (001..1) 的 WLE 能力估計法的 CAT 能力值變化  
 註：WLE 能力估計是從反應組型位元 4 開始，反應組型位元 1~3 代表 step rule 能力值變化

2. 組型 6 (0000001...1) 能力估計值的變化的細部觀察

組型6 (0000001..1) 的OWEN能力估計法的CAT能力值變化



組型6 (0000001..1) 的EAP能力估計法的CAT能力值變化

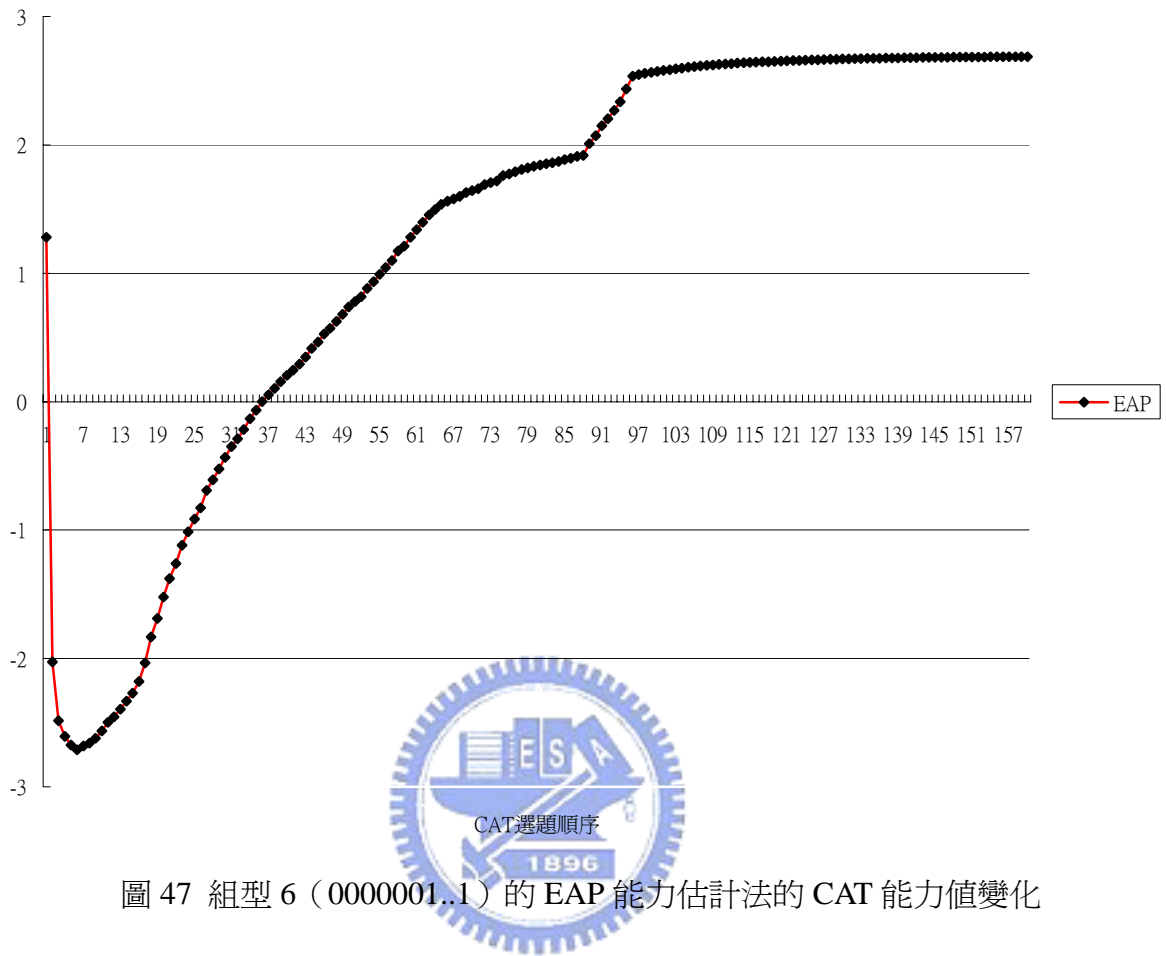


圖 47 組型 6 (0000001..1) 的 EAP 能力估計法的 CAT 能力值變化

組型6 (0000001..1) 的MLE能力估計法的CAT能力值變化

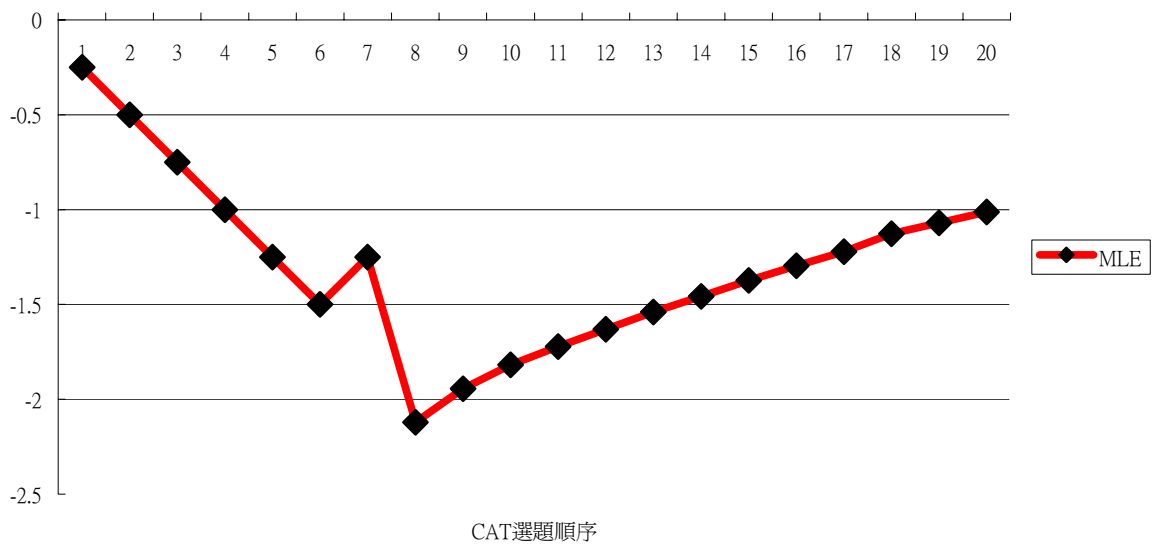


圖 48 組型 6 (0000001..1) 的 MLE 能力估計法的 CAT 能力值變化

註：MLE 能力估計是從反應組型位元 8 開始，反應組型位元 1~7 代表 step rule 能力值變化

組型6 (0000001..1) 的WLE能力估計法的CAT能力值變化

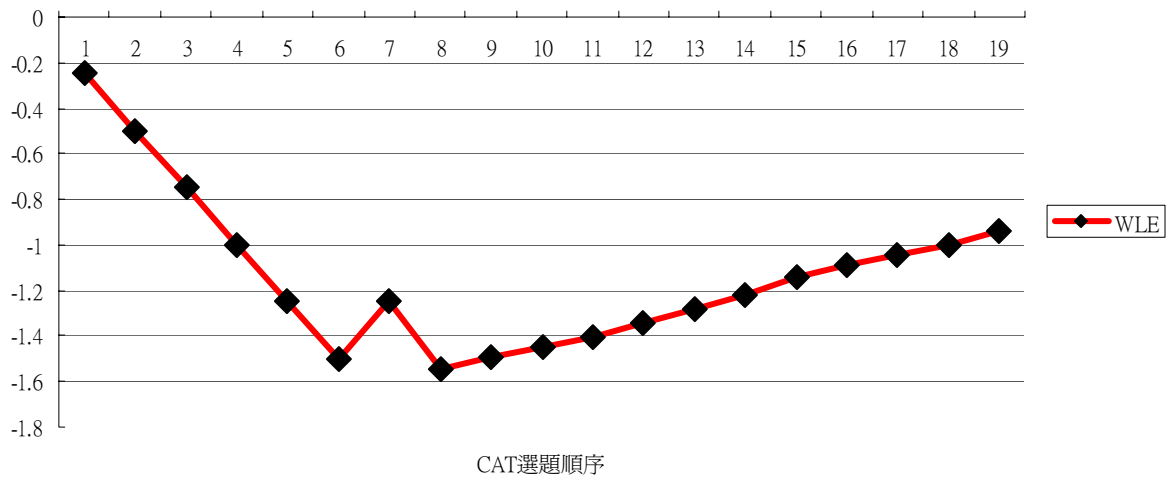


圖 49 組型 6 (0000001..1) 的 WLE 能力估計法的 CAT 能力值變化  
 註：WLE 能力估計是從反應組型位元 8 開始，反應組型位元 1~7 代表 step rule 能力值變化

3. 組型 35 (1010...) 能力估計值的變化的細部觀察

組型35 (1010...) 的OWEN能力估計法的CAT能力值變化

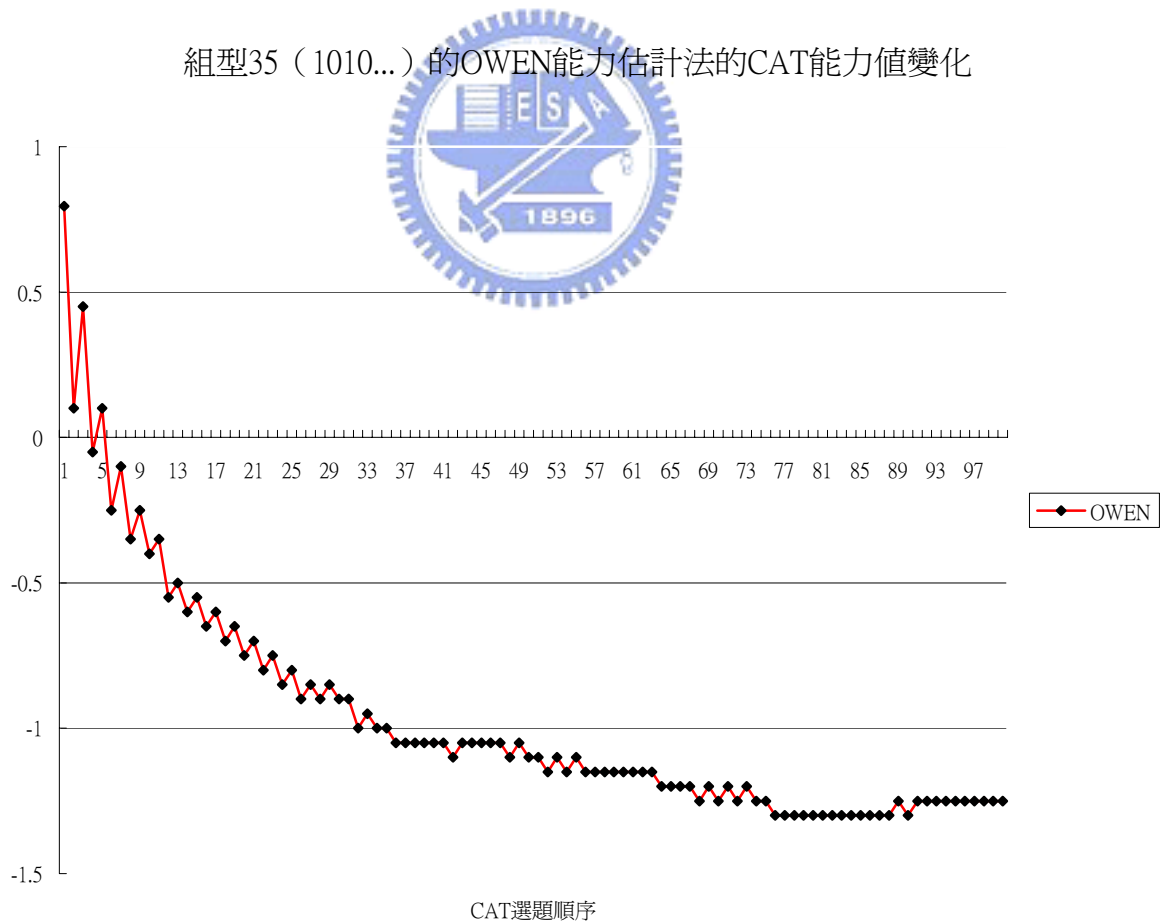


圖 50 組型 35 (1010...) 的 OWEN 能力估計法的 CAT 能力值變化

組型35 (1010...) 的EAP能力估計法的CAT能力值變化

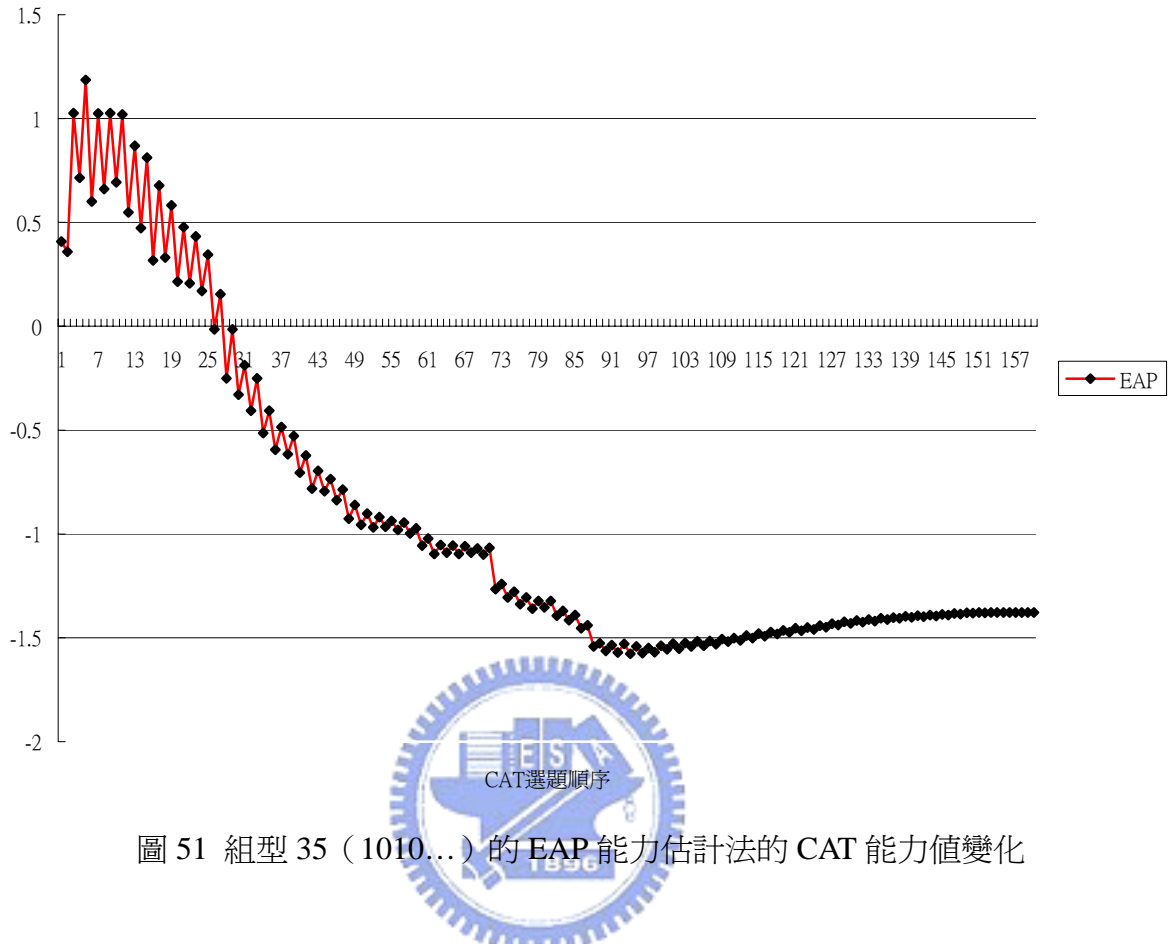


圖 51 組型 35 (1010...) 的 EAP 能力估計法的 CAT 能力值變化

組型35 (1010...) 的MLE能力估計法的CAT能力值變化

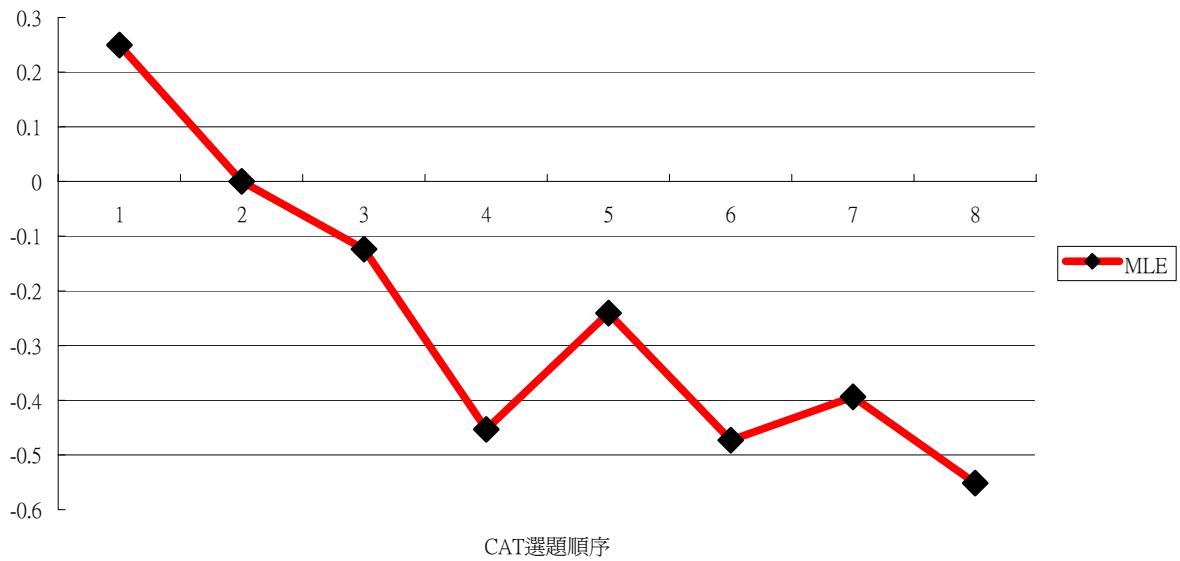


圖 52 組型 35 (1010...) 的 MLE 能力估計法的 CAT 能力值變化

註：MLE 能力估計是從反應組型位元 3 開始，反應組型位元 1~2 代表 step rule 能力值變化

組型35 (1010...) 的WLE能力估計法的CAT能力值變化

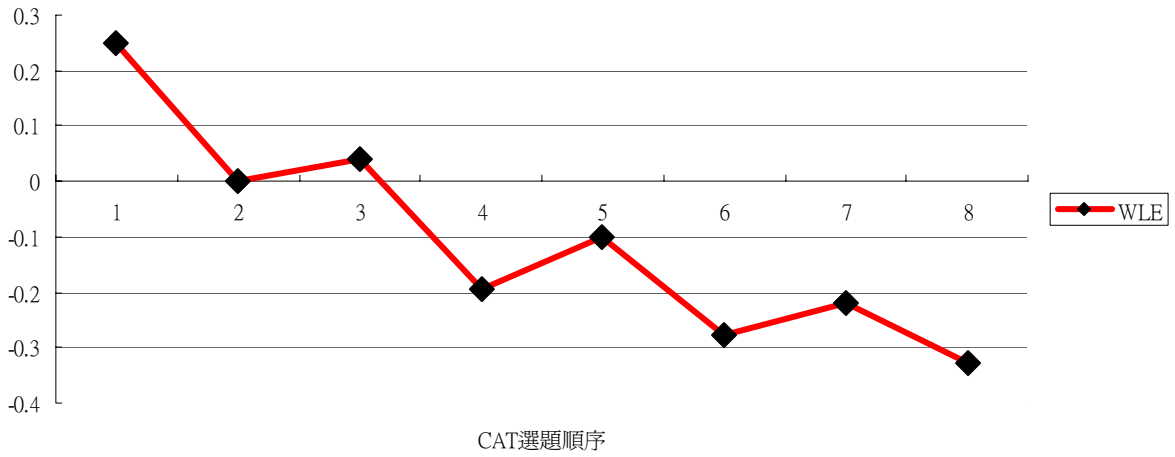


圖 53 組型 35 (1010...) 的 WLE 能力估計法的 CAT 能力值變化  
 註：WLE 能力估計是從反應組型位元 3 開始，反應組型位元 1~2 代表 step rule 能力值變化

4. 組型 41 (11110000...) 能力估計值的變化的細部觀察

組型41(11110000...)OWEN能力估計法的CAT能力估計值變化

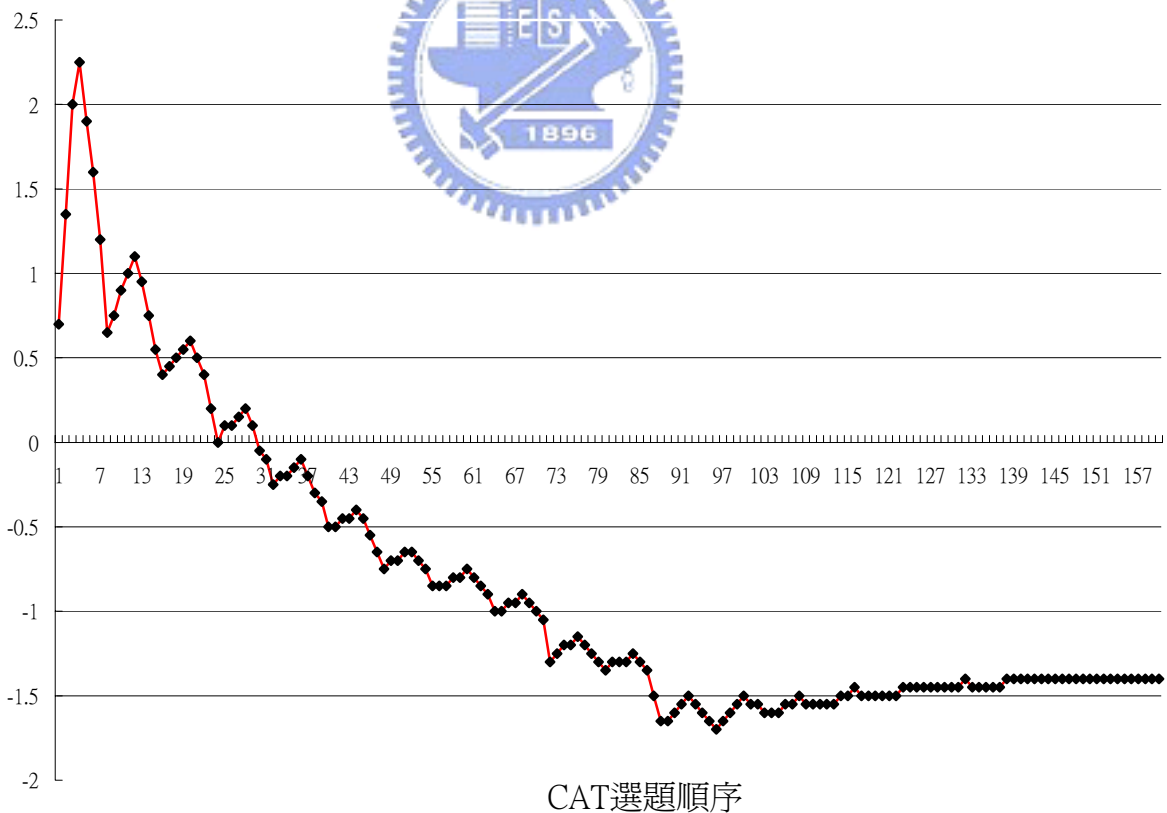


圖 54 組型 41 (11110000...) 的 OWEN 能力估計法的 CAT 能力值變化

組型41(11110000···)EAP能力估計法的CAT能力估計值變化

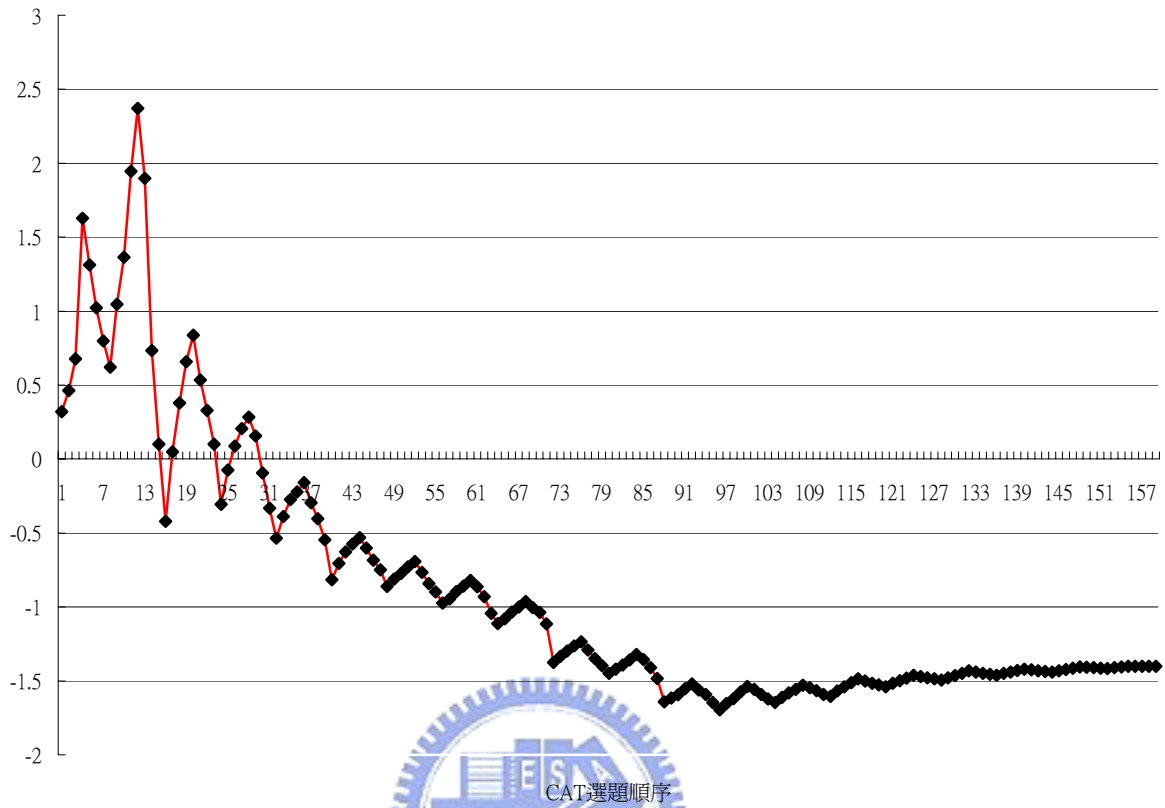


圖 55 組型 41 ( 11110000... ) 的 EAP 能力估計法的 CAT 能力值變化

組型41(11110000···)MLE能力估計法的CAT能力估計值變化

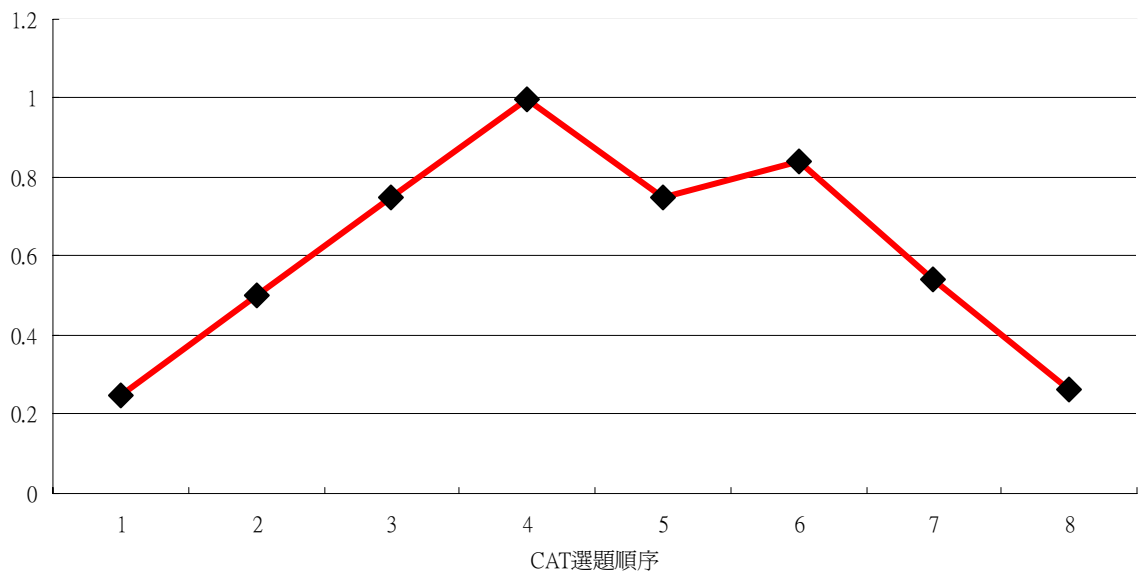


圖 56 組型 41 ( 11110000... ) 的 MLE 能力估計法的 CAT 能力值變化  
 註：MLE 能力估計是從反應組型位元 6 開始，反應組型位元 1~5 代表 step rule 能力值變化

組型41(11110000...)WLE能力估計法的CAT能力估計值變化

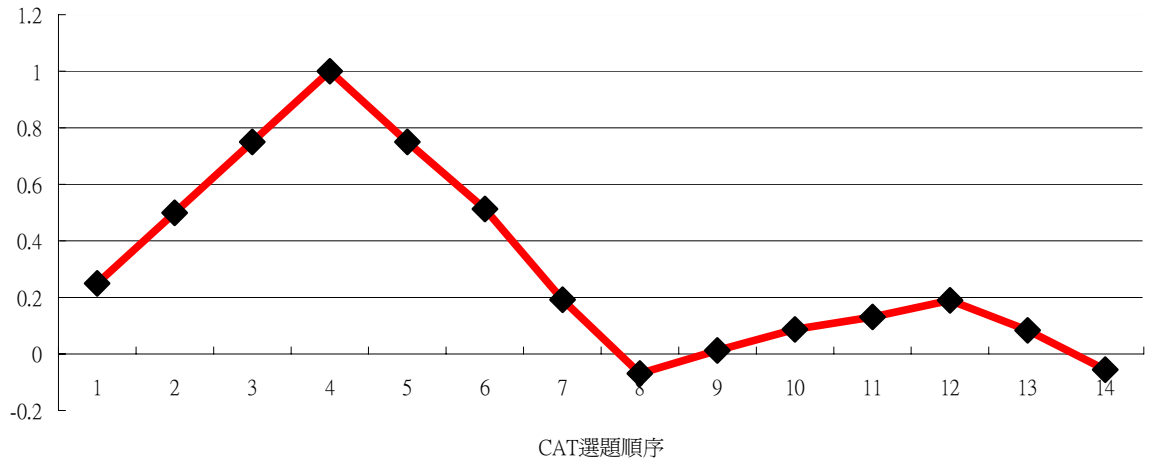


圖 57 組型 41 (11110000...) 的 WLE 能力估計法的 CAT 能力值變化  
 註：WLE 能力估計是從反應組型位元 6 開始，反應組型位元 1~5 代表 step rule 能力值變化

5. 組型 47 (11111110000000...) 能力估計值的變化的細部觀察

組型47(11111110000000...)OWEN能力估計法的CAT能力估計值變化

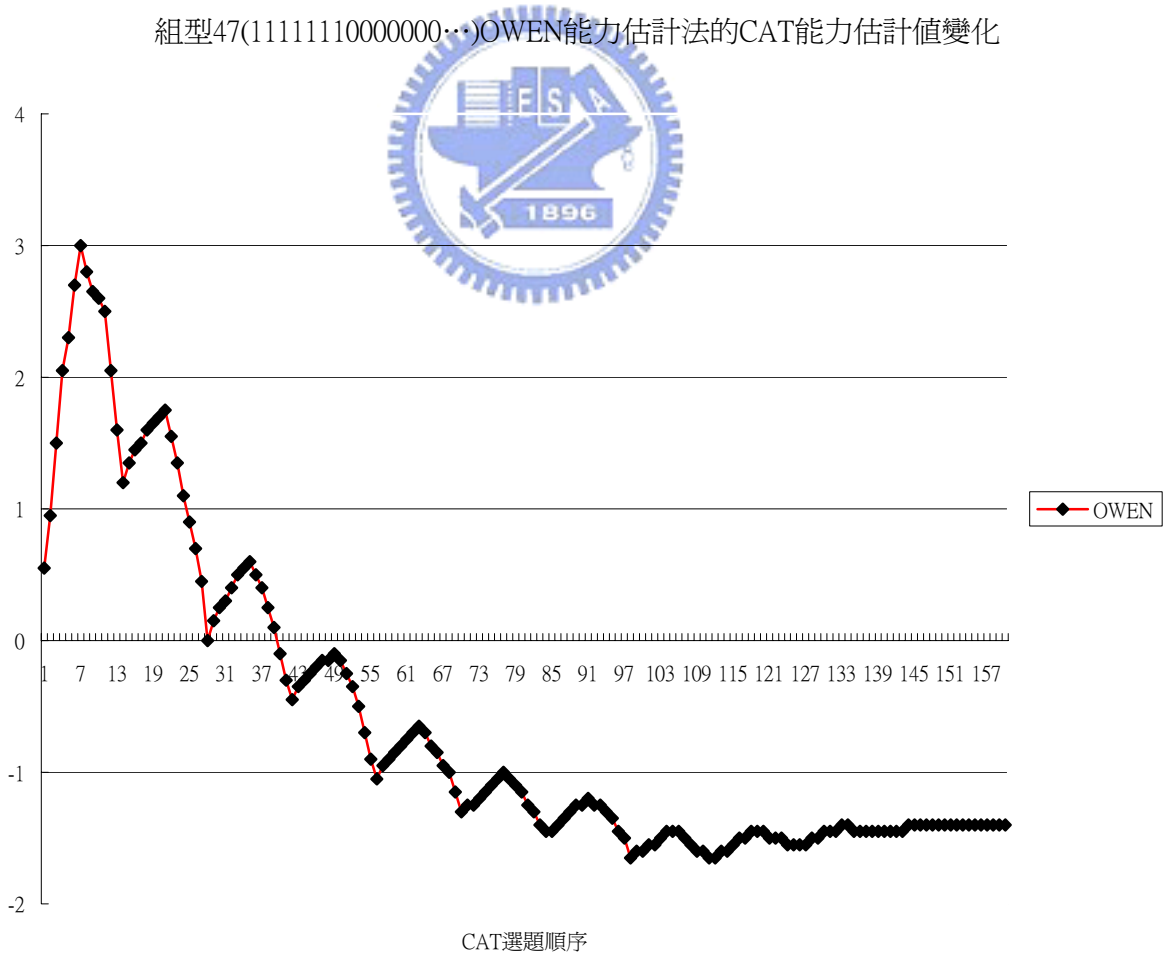


圖 58 組型 47 (11111110000000...) 的 OWEN 能力估計法的 CAT 能力值變化

組型47(11111110000000...)EAP能力估計法的CAT能力估計值變化

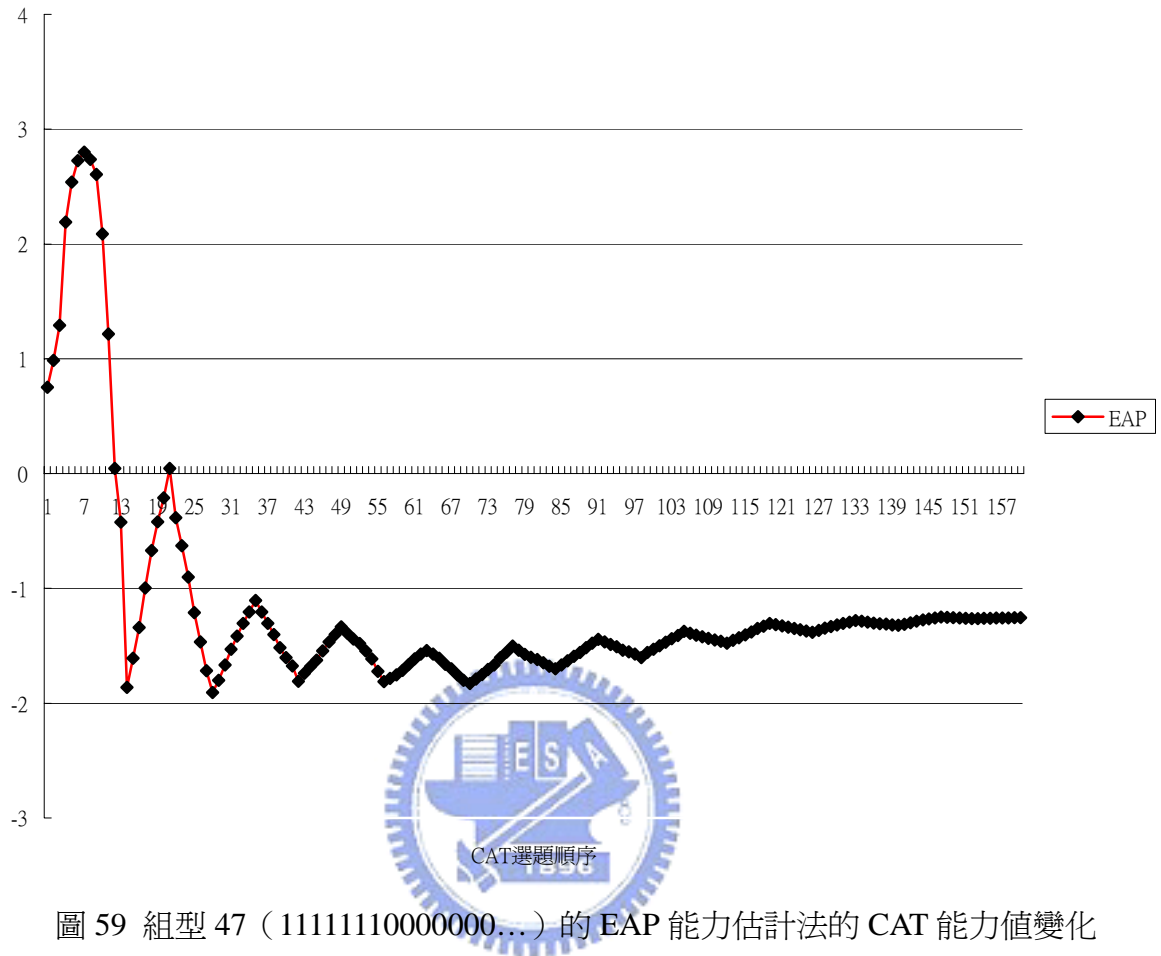


圖 59 組型 47 (11111110000000...) 的 EAP 能力估計法的 CAT 能力值變化

組型47(11111110000000...)MLE能力估計法的CAT能力估計值變化

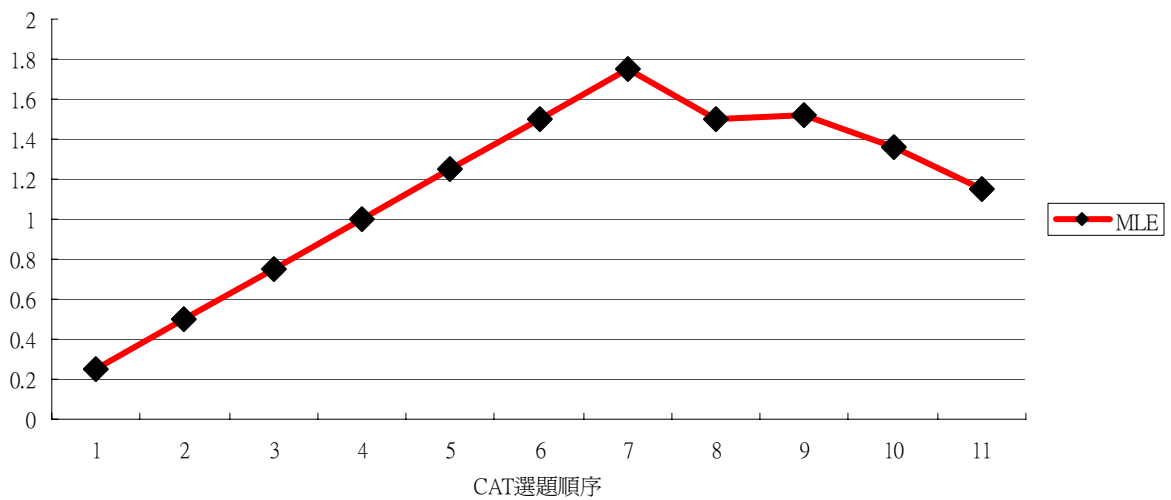


圖 60 組型 47 (11111110000000...) 的 MLE 能力估計法的 CAT 能力值變化  
 註：MLE 能力估計是從反應組型位元 9 開始，反應組型位元 1~8 代表 step rule 能力值變化



組型47(11111110000000...)WLE能力估計法的CAT能力估計值變化

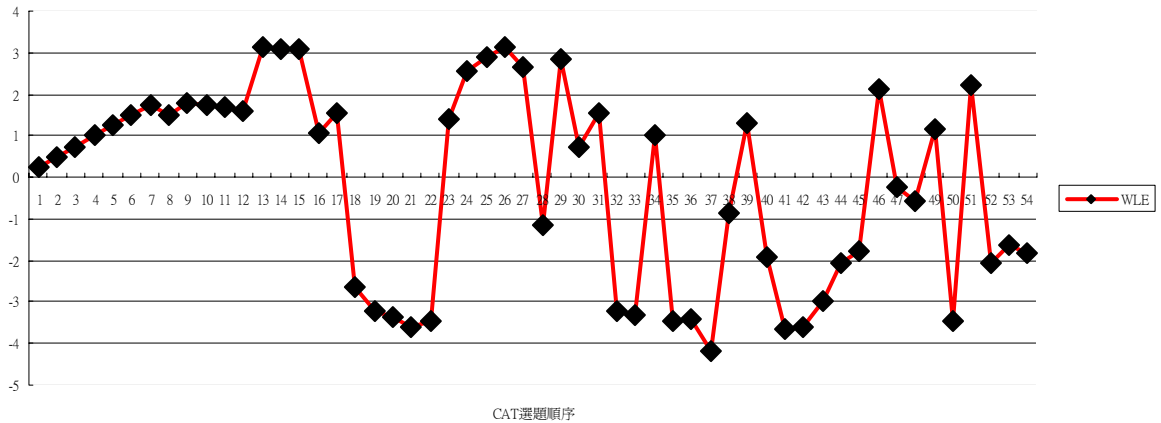


圖 61 組型 47 (11111110000000...) 的 WLE 能力估計法的 CAT 能力值變化  
 註：WLE 能力估計是從反應組型位元 9 開始，反應組型位元 1~8 代表 step rule 能力值變化

在貝氏順序估計法(OWEN)和期望後驗估計法(EAP)模擬研究中，系統設定適性測驗終止的條件之一為 PSD 值小於 0.25；而在最大概似估計法(MLE)和權值概似估計法(WLE)模擬研究中，系統設定適性測驗終止的條件之一為標準誤變動值小於 0.01。若考慮測驗效率，而使用最大測驗長度為 20、25 或 30 取代題庫用盡作為另一適性測驗終止的條件時，本研究分別建表（如表 26~表 28）歸納模擬研究中四種能力估計法對 50 種受試反應組型個別使用哪一種終止條件，並以○(收斂)代表系統能力估計能夠在小於或等於最大測驗長度收斂。換言之，PSD 小於 0.25 或標準誤變動值小於 0.01。以×(不收斂)代表在最大測驗長度時能力估計無法正常收斂，必須選擇最大測驗長度終止條件結束測驗。最後以表 29 比較不同最大測驗長度之四種能力估計法收斂與不收斂的組型數目。從表 29 的結果亦可看出 MLE 和 WLE 能力估計法的結果比較容易收斂。

表 26 最大測驗長度為 20，各種能力估計法使用的終止條件，  
 ×(不收斂)：最大測驗長度，○(收斂)：PSD 小於 0.25 或標準誤變動值小於 0.01

反應組型	貝氏順序估計法	期望後驗估計法	最大概似估計法	權值概似估計法
01...1	×	×	○	○
001...1	×	×	○	○
0001...1	×	×	○	○
00001...1	×	×	○	○
000001...1	×	×	○	○
0000001...1	×	×	○	×
00000001...1	×	○	○	×
000000001...1	×	○	×	×
0000000001...1	○	○	×	×

00000000001...1	○	○	×	×
000000000001...1	○	○	×	×
0000000000001...1	○	○	×	×
00000000000001...1	○	○	×	×
000000000000001...1	○	○	×	×
10...0	○	○	×	○
110...0	○	○	○	○
1110...0	○	○	○	○
11110...0	○	○	○	○
111110...0	×	×	○	○
1111110...0	×	×	○	○
11111110...0	×	×	○	○
111111110...0	×	×	○	○
1111111110...0	○	○	○	○
11111111110...0	○	○	○	○
111111111110...0	○	○	○	○
1111111111110...0	○	○	○	×
11111111111110...0	○	○	○	×
111111111111110...0	○	○	○	○
1111111111111110...0	○	○	×	×
11111111111111110...0	○	○	×	×
111111111111111110...0	○	○	×	×
10...	×	×	○	○
01...	×	×	○	○
1100...	×	×	○	○
0011...	×	×	○	○
111000...	×	×	○	○
000111...	×	×	○	○
11110000...	×	×	○	○
00001111...	×	×	○	○
1111100000....	×	×	○	○
0000011111...	×	×	○	○
111111000000...	×	×	○	○
000000111111...	×	×	○	○

1111110000000...	×	×	○	○
00000001111111...	×	○	○	○
1...1	○	○	×	×
0...0	○	○	×	×

表 27 最大測驗長度為 25，各種能力估計法使用的終止條件，  
 × (不收敛): 最大測驗長度，○ (收敛): PSD 小於 0.25 或標準誤變動值小於 0.01

反應組型	貝氏順序估計法	期望後驗估計法	最大概似估計法	權值概似估計法
01...1	×	○	○	○
001...1	×	×	○	○
0001...1	×	×	○	○
00001...1	×	×	○	○
000001...1	×	×	○	○
0000001...1	×	×	○	○
00000001...1	×	○	○	○
000000001...1	×	○	○	○
0000000001...1	○	○	×	○
00000000001...1	○	○	×	○
000000000001...1	○	○	×	×
0000000000001...1	○	○	×	×
00000000000001...1	○	○	×	×
000000000000001...1	○	○	×	×
10...0	○	○	○	○
110...0	○	○	○	○
1110...0	○	○	○	○
11110...0	○	○	○	○
111110...0	○	○	○	○
1111110...0	×	○	○	○
11111110...0	×	○	○	○
111111110...0	○	○	○	○
1111111110...0	○	○	○	○
11111111110...0	○	○	○	○
111111111110...0	○	○	○	○
1111111111110...0	○	○	○	○
11111111111110...0	○	○	○	○

111111111111110...0	○	○	○	○
111111111111110...0	○	○	○	○
111111111111110...0	○	○	○	○
111111111111110...0	○	○	○	×
111111111111110...0	○	○	○	×
111111111111110...0	○	○	○	○
10...	×	×	○	○
01...	×	×	○	○
1100...	×	×	○	○
0011...	×	×	○	○
111000...	×	×	○	○
000111...	×	×	○	○
11110000...	×	×	○	○
00001111...	×	×	○	○
1111100000...	×	×	○	○
0000011111...	×	×	○	○
111111000000...	×	×	○	○
000000111111...	×	×	○	○
11111110000000...	×	×	○	○
00000001111111...	×	○	○	○
1...1	○	○	×	×
0...0	○	○	×	×

表 28 最大測驗長度為 30，各種能力估計法使用的終止條件，  
 × (不收敛): 最大測驗長度，○ (收敛): PSD 小於 0.25 或標準誤變動值小於 0.01

反應組型	貝氏順序估計法	期望後驗估計法	最大概似估計法	權值概似估計法
01...1	×	○	○	○
001...1	×	×	○	○
0001...1	×	×	○	○
00001...1	×	×	○	○
000001...1	×	×	○	○
0000001...1	×	×	○	○
00000001...1	×	○	○	○
000000001...1	×	○	○	○
0000000001...1	○	○	○	○
00000000001...1	○	○	○	○

000000000001...1	○	○	×	○
0000000000001...1	○	○	×	○
00000000000001...1	○	○	×	○
000000000000001...1	○	○	×	○
0000000000000001...1	○	○	×	○
10...0	○	○	○	○
110...0	○	○	○	○
1110...0	○	○	○	○
11110...0	○	○	○	○
111110...0	○	○	○	○
1111110...0	○	○	○	○
11111110...0	○	○	○	○
111111110...0	○	○	○	○
1111111110...0	○	○	○	○
11111111110...0	○	○	○	○
111111111110...0	○	○	○	○
1111111111110...0	○	○	○	○
11111111111110...0	○	○	○	○
111111111111110...0	○	○	○	○
1111111111111110...0	○	○	○	○
11111111111111110...0	○	○	○	○
111111111111111110...0	○	○	○	○
10...	×	×	○	○
01...	×	×	○	○
1100...	×	×	○	○
0011...	×	×	○	○
111000...	×	×	○	○
000111...	×	×	○	○
11110000...	×	×	○	○
00001111...	×	×	○	○
1111100000....	×	×	○	○
0000011111...	×	×	○	○
111111000000...	×	×	○	○
000000111111...	×	×	○	○
11111110000000...	×	×	○	○

00000001111111...	×	○	○	○
1...1	○	○	×	×
0...0	○	○	×	×

表 29 不同最大測驗長度 vs 四種能力估計法：×（不收敛）和○（收敛）統計表

能力估計法 終止條件 最大測驗長度	貝氏順序估計法		期望後驗估計法		最大概似估計法		權值概似估計法	
	×	○	×	○	×	○	×	○
20 題	27 組	23 組	24 組	26 組	15 組	35 組	18 組	32 組
25 題	24 組	26 組	18 組	32 組	9 組	41 組	9 組	41 組
30 題	22 組	28 組	18 組	32 組	7 組	43 組	2 組	48 組

