

### 第三節 試題特性

試題特性分析對於國內外大型測驗評比資料庫是資料處理上相當重要的一環，藉由測量模式或一般性描述統計分析試題各項參數的穩定性可以提高測驗的品質，確實瞭解學生學習成就表現的變化。本節探討內容重點在綜合討論比較 NAEP、PISA、TIMSS 三個國際大型測驗評比資料庫與 TASA 在試題特性分析上之差異，並提出具體建議，期望整理出一套普遍一致的標準化測驗試題分析流程，供 TASA 後續研究分析使用。

NAEP、PISA、TIMSS 等國際大型測驗評比資料庫為了確保測驗的信度及效度，對於測驗過程中所使用的試題皆會進行不同的統計分析及信、效度檢測，其中包含了一般性試題統計描述、試題參數估計、評分者一致性效應檢測及差異試題功能分析 (DIF) 等內容。同樣的，TASA 身為國內第一個大型測驗評比資料庫，對於測驗中所使用的試題依然會進行分析、檢測與篩選。底下即分別針對 NAEP、PISA、TIMSS 及 TASA 資料庫對於試題特性所採用的分析方法進行比較，期望獲得普遍一致的標準化分析流程，以提升 TASA 測驗的信度及效度。

#### 壹、試題統計描述

##### 一、NAEP

分為二元計分試題及多元計分試題兩種，所呈現的相關數據如下：

##### (一) 二元計分試題 (dichotomously scored items)

人數、選項百分比、答對率、遺漏未作答百分比、無效試卷百分比、答對率、難度指標轉換為平均=13、標準差=4、二系列相關、點二系列相關。

##### (二) 多元計分試題 (polytomously scored items)

無效試卷百分比、難度指標轉換為平均=13、標準差=4、多系列相關、Pearson 相關。

##### 二、PISA

##### (一) 國家報告呈現格式

##### 1. 以表格的形式呈現個別試題統計描述

表 4-3-1 為 PISA 2006 技術報告有關試題描述性統計分析之表格呈現方式，第一行標籤部分，1, 2, 3, 4 為作答反應選項，8 為無效值，9 為遺漏值，其餘分析包含了人數、鑑別度、閾值、試題不合適之均方根的值(item

infit mean square)、選項百分比、答對率、遺漏未作答百分比、點二系列相關 (t 檢定)、PV1 Avg:1 與 PV1 SD:1 等等。

表 4-3-1 試題之描述性統計分析 (表格化)

Example of item statistics in Report 1

Item:70 (S423Q01)									
Cases for this item		355	Discrimination		0.13				
Item Threshold(s):		0.49	Weighted MNSQ		1.17				
Item Delta(s):		0.49							
Label	Score	Count	% of tot	Pt Bis	t	(p)	PV1Avg:1	PV1 SD:1	
0		0	0.00	NA	NA	(.000)	NA	NA	
1	0.00	65	18.31	-0.16	-3.02	(.003)	-0.30	0.78	
2	1.00	138	38.87	0.13	2.54	(.011)	0.12	0.89	
3	0.00	115	32.39	0.09	1.76	(.080)	0.07	0.83	
4	0.00	26	7.32	-0.08	-1.44	(.152)	-0.41	0.83	
5		0	0.00	NA	NA	(.000)	NA	NA	
6		0	0.00	NA	NA	(.000)	NA	NA	
8	0.00	4	1.13	-0.06	-1.19	(.233)	-0.62	0.79	
9	0.00	7	1.97	-0.15	-2.87	(.004)	-0.76	0.58	

註：PV1 Avg:1 和 PV1 SD:1 是學生反應在每一類別的平均，以上表而言，正確答對的學生的平均可能值是 0.12，作答反應 1 之學生的平均可能值是-0.3，依此類推。

資料來源：PISA 2006 技術報告 (頁 148)

## 2. 以圖的形式呈現個別試題統計描述

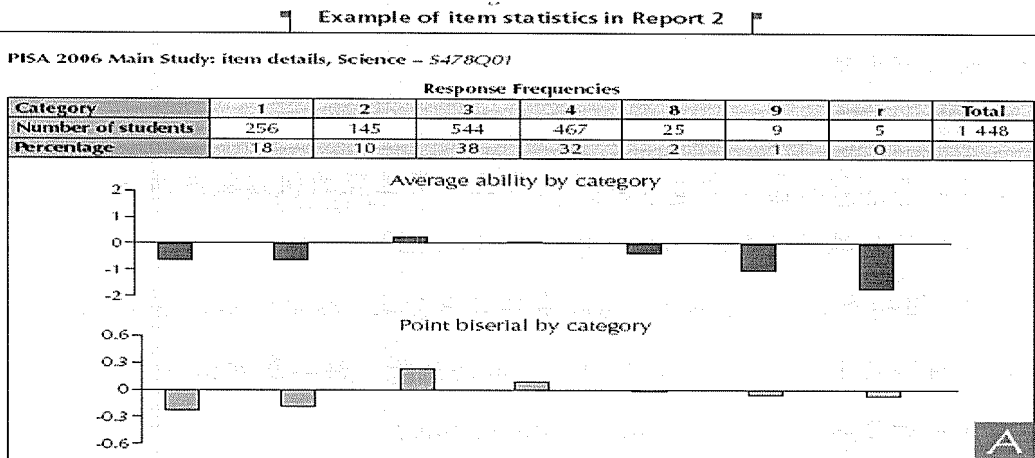


圖 4-3-1 試題選項之平均能力值與點二系列相關值

資料來源：PISA 2006 技術報告 (頁 149)

### 3. 試題在國家與國際間之模式適合度與鑑別度參數比較

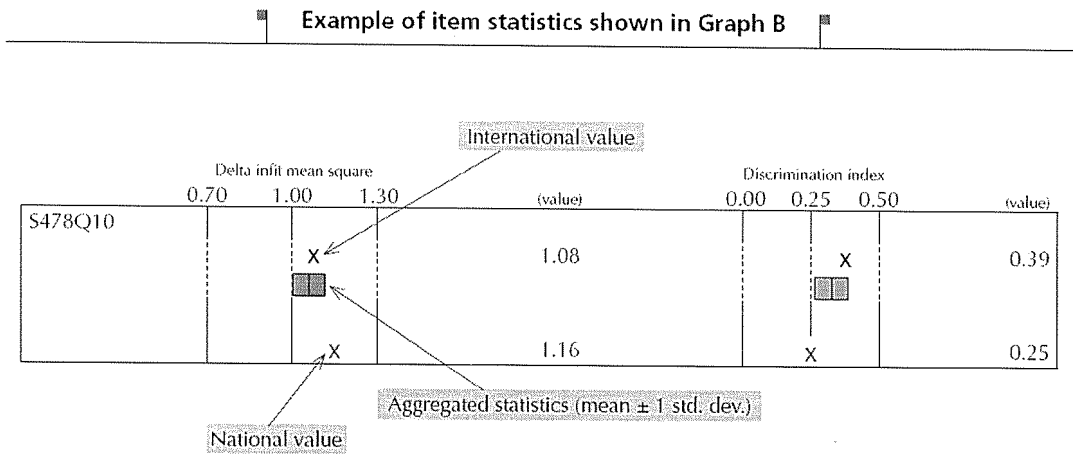


圖 4-3-2 試題在國家與國際間之模式適合度與鑑別度參數比較  
資料來源：PISA 2006 技術報告（頁 150）

### 4. 試題在國家與國際間之難度與閾值比較

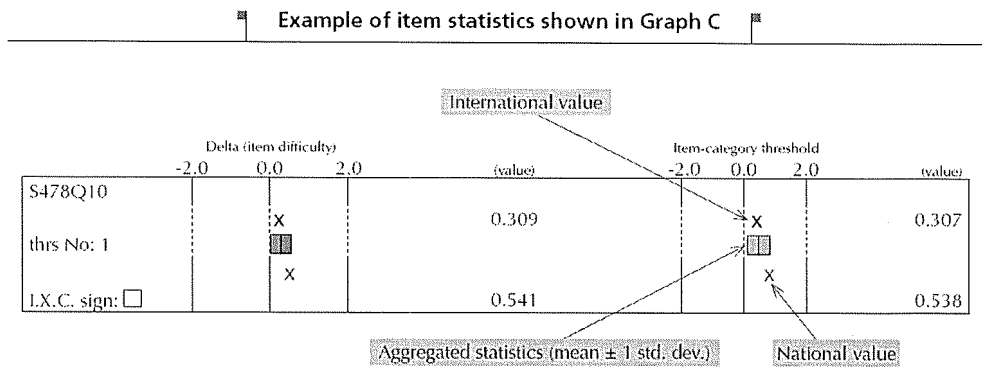


圖 4-3-3 試題在國家與國際間之難度與閾值比較  
資料來源：PISA 2006 技術報告（頁 151）

## 5. 期望分數曲線

觀察試題在單一國家的實際表現、國際間表現與預期分數表現之差異，如圖 4-3-4。

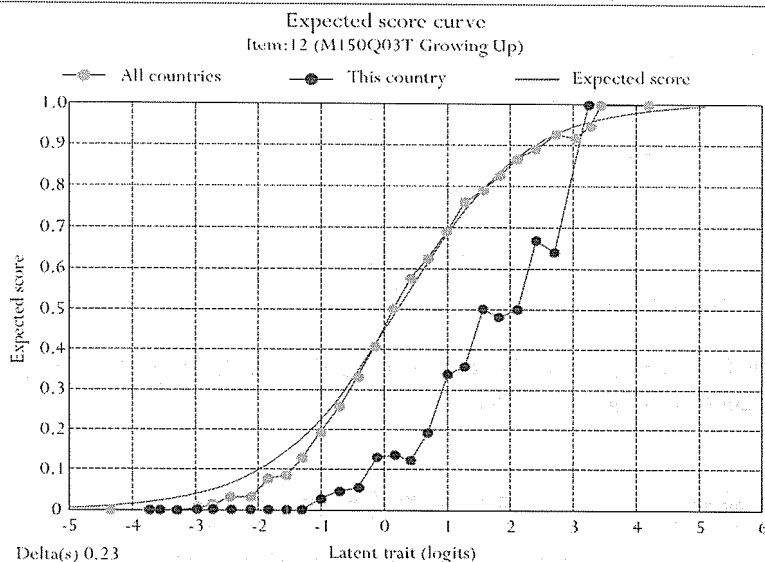


圖 4-3-4 單一試題在單一國家、國際間與期望之分數表現曲線比較  
資料來源：PISA 2003 技術報告（頁 127）

### 三、TIMSS

分為選擇題及開放性試題兩種，另外針對定錨試題會呈現年度間的數據比較，並利用試題難度值的在各國間的不同檢視是否存在試題效應，所呈現的相關數據如下：

#### （一）選擇題

表 4-3-2 為人數、答對率（古典難度）、鑑別度（點二系列相關）、選項百分比、缺失遺漏未作答百分比、相關、單參難度。

表 4-3-2 TIMSS 2007 選擇題描述性統計分析

Trends in International Mathematics and Science Study - TIMSS 2007 Assessment Results  
 International Item Statistics (Unweighted) - Review Version - 8th Grade  
 For Internal Review Only; DO NOT CITE OR CIRCULATE  
 Mathematics: Algebra / Knowledge (M02077 - M10\_05)  
 Item: 10040 (Equivalent to 41304)  
 Type: MC\_Resp\_C

Country	N	Diff	Disc	Det_A	Det_B	Det_C	Pct_A	Pct_B	Pct_C	Pct_0	Pct_1	Pct_2	Pct_3	Pct_4	Pct_5	Pct_6	Pct_7	Pct_8	Pct_9	Point Biserials				RDIFF	Flags
																				PR_A	PR_B	PR_C	PR_D		
Algeria	752	16.2	0.11	46.2	19.8	16.2	19.3	4.5	2.0	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.36	D, E	
Armenia	583	73.2	0.46	35.4	5.0	73.2	3.7	2.8	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	-1.01	E
Australia	578	31.1	0.40	32.0	19.0	31.1	15.6	2.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	-1.01	E
Austria	571	46.1	0.51	22.5	4.6	46.1	11.1	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.11	E
Bahrain and Kazakhstan	566	34.2	0.36	30.9	38.3	34.2	15.4	1.2	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.97	E
Bulgaria	576	73.6	0.55	34.6	5.7	73.6	3.6	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.13	C, E
China Taipei	572	85.8	0.61	7.5	2.3	85.8	3.7	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.44	E
Colombia	62	26.0	0.25	36.2	17.1	26.0	11.1	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.95	E
Cyprus	583	59.3	0.49	20.3	5.2	59.3	15.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.73	E
Czech Republic	935	56.9	0.42	19.3	11.9	56.8	11.3	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.38	E
El Salvador	568	24.6	0.32	34.9	19.9	24.6	19.0	1.6	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.28	E
England	571	46.1	0.59	20.8	18.4	46.1	13.7	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.60	E
Georgia	715	57.7	0.41	23.5	7.8	57.7	13.8	1.6	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.19	E
Hong Kong SAR	502	87.5	0.51	5.6	2.8	87.5	3.5	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.45	E
Hungary	599	63.4	0.60	15.7	6.5	63.4	13.4	3.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.39	E
Indonesia	581	36.9	0.47	28.4	13.7	36.9	20.3	6.7	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.45	E
Ireland	574	40.9	0.45	30.7	4.1	40.9	29.3	0.7	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.52	E
Israel	626	48.1	0.36	26.0	6.9	48.1	17.0	1.1	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.40	E
Jordan	781	53.9	0.58	18.3	14.5	53.9	12.5	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.21	E
Korea, Rep. of	603	85.6	0.58	7.5	3.6	85.6	3.3	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.02	E
Kuwait	571	23.5	0.32	34.2	20.8	23.5	21.9	2.9	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.89	E
Lebanon	571	23.5	0.32	34.2	20.8	23.5	21.9	2.9	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.89	E
Lithuania	571	63.9	0.59	16.1	7.7	63.9	12.6	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.52	E
Malaysia	634	41.6	0.50	27.3	13.1	41.6	14.4	0.6	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.65	E
Malta	671	60.4	0.51	16.8	6.9	60.4	10.5	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.10	E
Mongolia	610	57.4	0.49	18.7	7.9	57.4	14.4	1.6	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.14	E
Morocco	634	40.9	0.42	27.9	18.5	40.9	14.7	3.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.56	E
Norway	628	41.7	0.49	20.5	20.8	41.7	16.1	0.9	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.36	E
Palau	653	37.1	0.44	23.6	22.4	37.1	15.3	1.7	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.08	E
Palau Nat'l Aut.	1017	31.7	0.26	27.7	23.1	31.7	15.7	3.6	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.07	E
Romania	601	61.6	0.53	20.6	9.8	61.6	7.5	0.5	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.22	E
Russian Federation	627	72.3	0.37	30.2	30.2	72.3	13.7	1.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.27	E
Saudi Arabia	627	72.3	0.37	30.2	30.2	72.3	13.7	1.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.27	E
Scotland	564	43.1	0.54	19.3	22.2	43.1	23.4	1.6	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.47	C
Serbia	572	69.9	0.61	16.3	4.0	69.9	8.6	1.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.20	E
Singapore	655	86.3	0.58	8.1	1.8	86.3	3.5	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.22	E
Slovenia	576	37.5	0.41	38.5	4.4	37.5	11.6	3.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.13	E
Sweden	670	46.1	0.45	23.3	13.9	46.1	15.4	1.3	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.36	E
Sri Lanka	763	37.0	0.55	27.0	18.1	37.0	16.9	1.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.02	E
Thailand	580	34.9	0.48	14.5	31.2	34.9	18.4	1.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.00	E
Turkey	553	23.5	0.60	21.0	14.9	23.5	9.6	0.8	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.25	E
Ukraine	604	61.1	0.51	26.5	9.6	61.1	15.3	0.9	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.35	E
United States	1049	56.3	0.53	26.5	9.6	56.3	15.3	0.9	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.14	E
International Avg.		56.5	0.46	22.6	12.5	56.5	13.1	1.3	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.61	
Basque Country, Spain	331	66.3	0.40	16.7	2.1	66.3	11.7	1.6	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.88	E
Belgium	782	56.0	0.56	25.2	2.8	56.0	14.5	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.39	E
Minnesota, US	247	42.1	0.52	35.6	5.7	42.1	16.2	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.79	E
Ontario, Canada	486	24.1	0.26	44.6	11.3	24.1	20.2	2.6	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.88	E
Quebec, Canada	590	56.1	0.44	24.6	5.4	56.1	12.7	1.2	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.91	E

Keys: Diff: Percent correct score; Disc: Item discrimination; Det\_A, B: Percent choosing option; Det\_C, OH, PR: Percent Omitted and Not Reached; PR\_A, B, C, D, E: Point Biserials for option; PR\_OH: Point Biserial for Omitted; RDIFF: Rasch difficulty.  
 Flags: A= Ability not ordered/Attractive distractor; C= Difficulty less than chance; D= Negative flow discrimination; E= Higher than average; F= Distractor chosen by less than 10%; H= Harder than average; I= Scoring reliability < 80%; M= Difficulty greater than 95.

資料來源：TIMSS 2007 技術報告（頁 195）



### (三) 定錨試題

表 4-3-4 為針對歷年等化過程中所使用的定錨試題，呈現年度、人數、選項答對百分比、答對率（古典難度）、無效未作答遺漏百分比、女生答對率、男生答對率等等相關數據。

表 4-3-4 TIMSS 2003 及 TIMSS 2007 試題參數比較（定錨題）

COUNTRY	Year	N	10	70	71	79	99	V1	NOT REACHED	OMIT	1.GIRL % Right	2.BOY % Right
Armenia	2003	464	33.2	18.6	16.8	11.7	19.6	33.2	0.8	18.9	33.0	33.5
	2007	273	48.7	8.8	21.2	1.8	19.4	48.7	0.7	18.7	48.1	49.1
Australia	2003	375	64.0	17.5	11.0	5.5	2.1	64.0	0.0	2.1	65.3	62.7
	2007	291	64.6	17.2	14.1	3.4	0.7	64.6	0.0	0.7	63.5	65.6
Chinese Taipei	2003	380	76.5	3.3	13.7	0.3	1.3	76.5	0.0	1.3	76.5	76.4
	2007	300	64.0	6.7	26.0	3.0	0.3	64.0	0.0	0.3	63.3	64.6
England	2003	291	64.2	13.0	15.4	6.1	1.2	64.2	0.0	1.2	63.8	64.8
	2007	305	69.2	14.8	10.8	4.6	0.7	69.2	0.0	0.7	70.8	67.5
Hong Kong SAR	2003	373	69.8	7.9	18.9	3.0	0.4	69.8	0.0	0.4	72.5	67.7
	2007	268	69.8	5.2	22.0	1.9	1.1	69.8	0.0	1.1	69.9	69.6
Hungary	2003	268	70.5	15.4	9.9	3.6	0.6	70.5	0.0	0.6	75.9	65.1
	2007	288	72.2	11.1	11.8	3.8	1.0	72.2	0.0	1.0	73.5	71.1
Iran, Islamic Rep. of	2003	352	41.7	22.8	20.3	6.7	8.5	41.7	0.0	8.5	48.2	37.7
	2007	274	39.4	25.9	20.1	8.4	6.2	39.4	0.0	6.2	36.9	41.4
Italy	2003	353	72.4	9.2	14.3	3.7	0.5	72.4	0.0	0.5	75.4	69.8
	2007	323	66.3	10.5	15.2	3.1	5.0	66.3	0.0	5.0	65.2	67.0
Latvia	2003	295	61.0	19.5	13.5	4.8	1.1	61.0	0.0	1.1	62.0	60.1
	2007	277	68.2	15.2	11.6	3.6	1.4	68.2	0.0	1.4	71.1	65.5
Lithuania	2003	371	58.3	23.2	9.2	7.3	1.9	58.3	0.0	1.9	59.3	56.3
	2007	285	54.7	22.8	14.4	7.0	1.1	54.7	0.0	1.1	54.7	54.8
Morocco	2003	339	14.8	32.1	12.5	28.2	12.5	14.8	0.0	12.5	10.2	13.9
	2007	300	12.0	16.7	7.7	47.7	16.0	12.0	0.7	15.3	10.5	13.7
Netherlands	2003	242	54.7	15.4	19.9	10.0	0.0	54.7	0.0	0.0	57.3	51.8
	2007	237	63.3	15.2	14.8	5.5	1.3	63.3	0.0	1.3	62.5	64.1
New Zealand	2003	354	61.1	17.8	14.2	4.9	2.1	61.1	0.0	2.1	59.9	62.3
	2007	349	56.7	18.3	15.8	8.3	0.9	56.7	0.0	0.9	56.3	57.2
Norway	2003	361	58.6	20.5	12.8	6.3	1.8	58.6	0.0	1.8	61.7	55.9
	2007	290	55.9	16.9	20.7	4.1	2.4	55.9	0.0	2.4	57.4	54.5
Russian Federation	2003	325	53.9	16.6	8.3	17.1	4.2	53.9	0.0	4.2	51.7	56.1
	2007	323	62.8	15.5	11.5	7.4	2.8	62.8	0.0	2.8	60.1	65.0
Scotland	2003	330	54.5	19.1	17.3	6.8	2.3	54.5	0.4	1.9	54.2	54.9
	2007	286	59.1	21.7	10.1	7.7	1.4	59.1	0.0	1.4	50.0	66.3
Singapore	2003	562	78.1	9.7	9.1	2.4	0.7	78.1	0.0	0.7	80.7	75.5
	2007	360	78.3	10.3	9.4	1.7	0.3	78.3	0.0	0.3	81.4	75.1
Slovenia	2003	280	58.9	18.4	13.6	6.4	2.7	58.9	0.0	2.7	63.2	56.4
	2007	316	66.5	14.6	9.5	8.5	0.9	66.5	0.0	0.9	69.7	63.4
Tunisia	2003	354	18.5	26.2	20.3	27.2	7.8	18.5	0.0	7.8	16.4	20.6
	2007	290	18.6	29.3	13.1	30.0	9.0	18.6	0.0	9.0	20.7	16.6
United States	2003	809	60.7	20.0	11.5	6.7	1.0	60.7	0.0	1.0	62.9	58.6
	2007	566	61.3	19.3	12.0	6.0	1.4	61.3	0.2	1.2	59.2	63.3
International Avg.	2003	.	56.3	17.3	14.4	8.4	3.6	56.3	0.1	3.6	57.5	55.2
	2007	.	57.6	15.8	14.6	8.4	3.7	57.6	0.1	3.6	57.2	57.8
Ontario, Canada	2003	359	55.1	21.3	14.6	7.1	2.0	55.1	0.0	2.0	49.1	61.9
	2007	255	50.6	22.7	12.9	8.6	5.1	50.6	0.0	5.1	49.3	52.1
Quebec, Canada	2003	373	51.9	25.2	15.6	7.0	0.3	51.9	0.0	0.3	47.7	55.5
	2007	276	61.2	15.2	13.4	6.9	3.3	61.2	0.0	3.3	64.5	58.0

V1 = Percent scoring 1 or better      V2 = Percent scoring 2 or better  
percent right for boys and girls corresponds to the percent obtaining the maximum score on the item.  
Because of missing gender information, some totals may appear inconsistent.

資料來源：TIMSS 2007 技術報告（頁 201）

### (四) 試題與國家之交互作用（item-by-country interaction）





**TIMSS 2007 Bridge – Plot of Difference in Rasch Difficulties**

ItemName=S11\_03\_UH1QUEID=S031233 Label=Main features of four animals shown

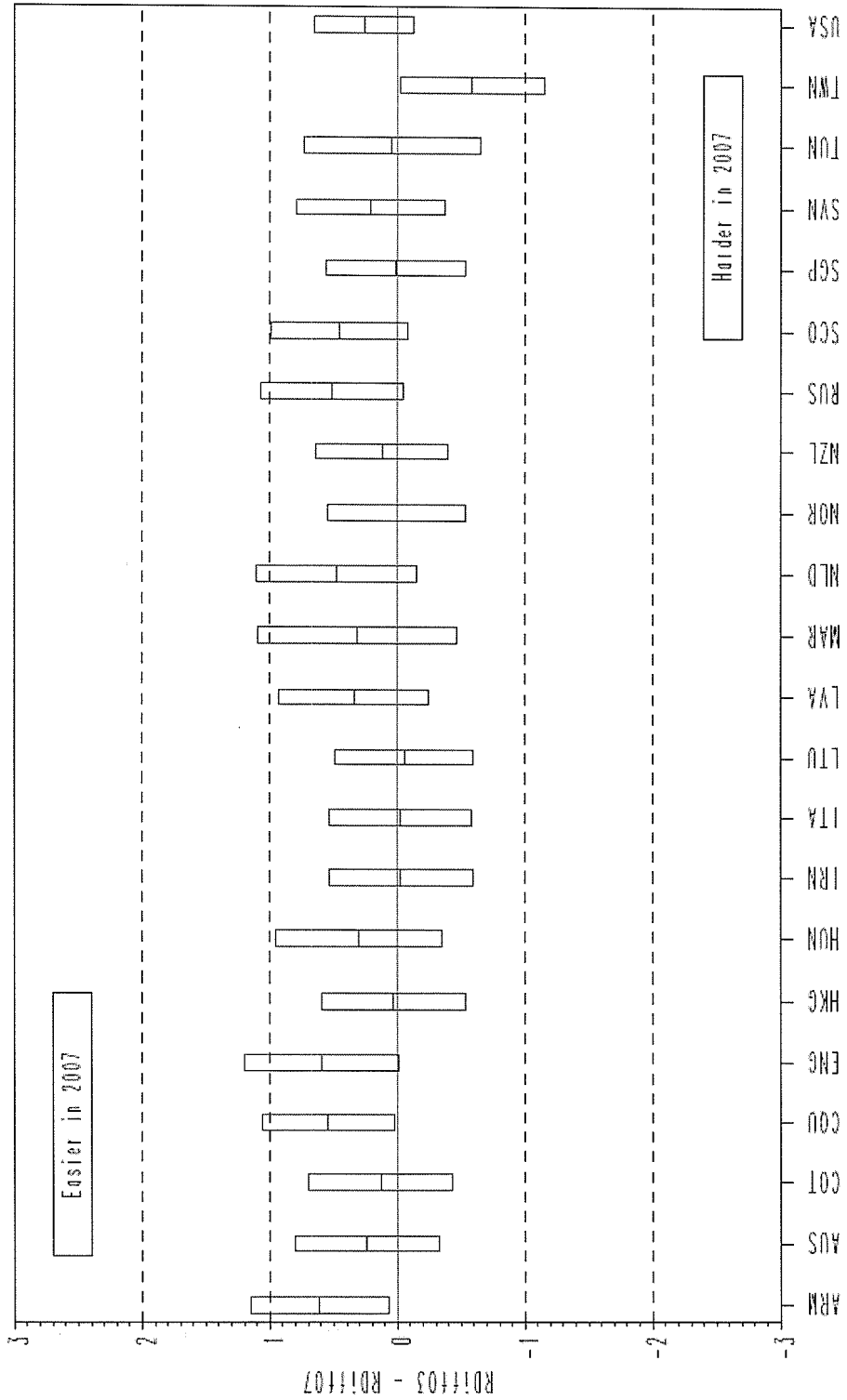


圖 4-3-6 定錨題在各國不同年度之難度狀況比較 (1)

資料來源：TIMSS 2007 技術報告 (頁 206)

# TIMSS 2007 Bridge – Plot of Rasch Difficulties by Country

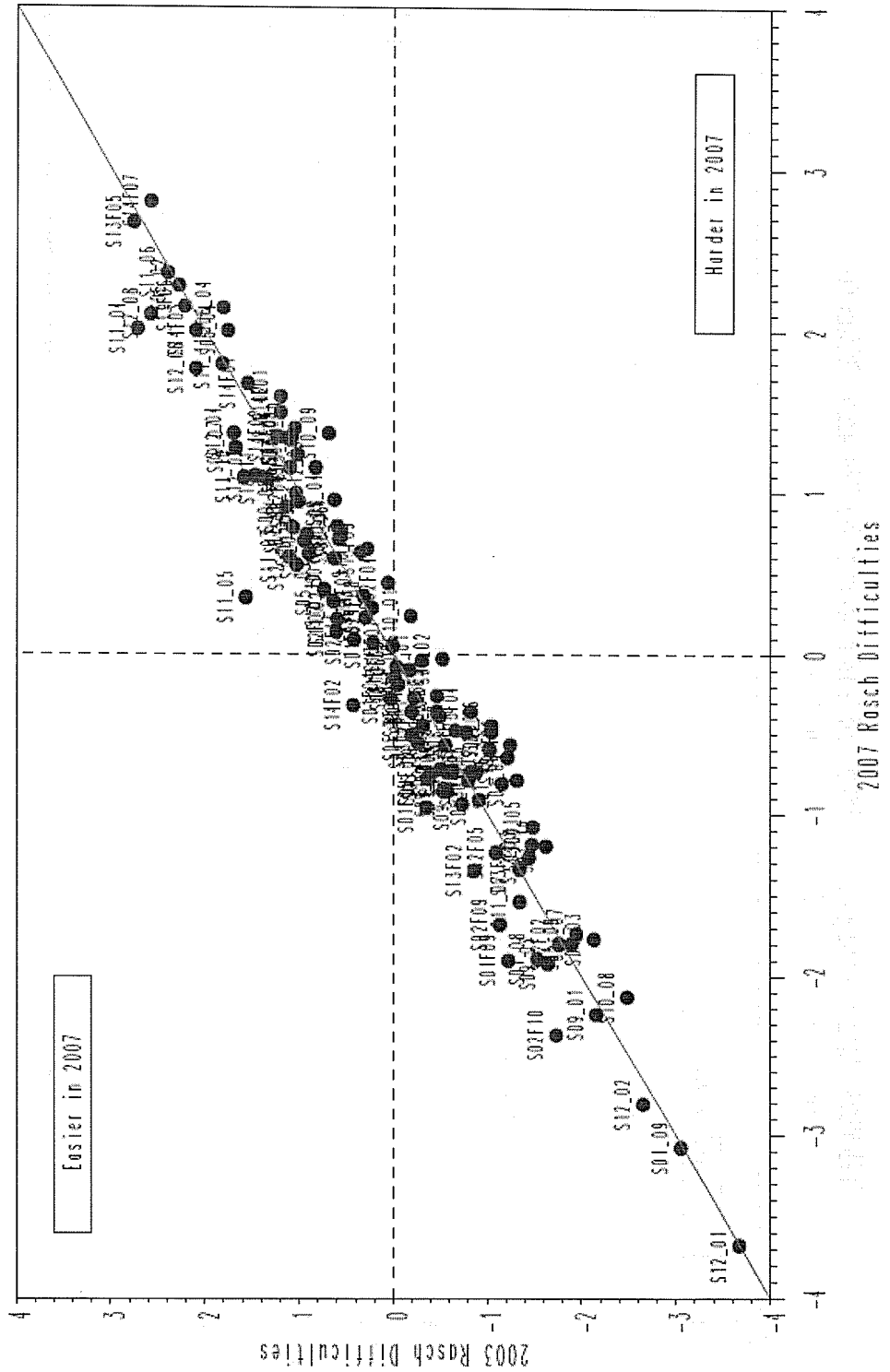


圖 4-3-7 定錨題在各國不同年度之難度狀況比較 (2)

資料來源：TIMSS 2007 技術報告 (頁 207)

#### 四、TASA

##### (一) 選擇題

呈現試題之作答人數、答對率、鑑別度、難度、猜測度、點二系列相關、單參難度、選項百分比、難度指標轉換為平均=250.標準差=50 的量尺分數等相關數據，如表 4-3-5。

##### (二) 開放性試題

呈現試題之作答人數、答對率、選項次數分配、缺失遺漏未作答次數分配等相關數據。

表 4-3-5 TASA 試題參數估計與描述性統計分析

科別	年級	區塊	題號				
數學	06	M1	2				
題目	下面哪一個比的比值跟其他三個不同？ (1) 12 : 16 (2) 15 : 21 (3) 24 : 32 (4) 30 : 40。						
能力指標	略						
IRT 參數估計：鑑別度= <input type="checkbox"/> 難度= <input type="checkbox"/> 猜測度= <input type="checkbox"/> 點二系列相關= <input type="checkbox"/>							
答案	2	認知層次	略				
	編號	略					
古典理論 (CTT)選項分析	選 項	1	2*	3	4	其他	
	選項率	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	通過率： <input type="checkbox"/>
	高分組	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	難 度： <input type="checkbox"/>
	低分組	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	鑑別度： <input type="checkbox"/>

資料來源：TASA 線上測驗數學科 <http://tasa.naer.edu.tw/3online-1-detail.asp?id=3>。

## 五、綜合討論與建議

表 4-3-6 為 TASA 與國際性大型資料庫，有關試題一般性統計描述、1PL 或 3PL 參數估計等作一綜合分析比較其中異同。由表 1 可以發現，NAEP、TIMSS 及 TASA 對於選擇題的分析所選用的測量模式為三參數對數模式，PISA 則為多向度隨機係數多項洛基模式，因此在參數的估計上，或呈現的資料上皆包含了人數、選項百分比、答對率、鑑別度、難度、點二系列相關等數據，有差異且較重要的部份就在 PV 值的計算應用與缺失遺漏未作答反應的描述。

開放性試題部分，TASA 僅呈現人數、無效遺漏未作答百分比或次數分配、選項百分比或次數分配等描述性統計資料，並沒有像 NAEP 及 TIMSS 使用二參數對數模式及一般化部份給分模式或 PISA 所使用的多向度隨機係數多項洛基模式，可見選用適當的測量模式對填充題及開放性試題進行分析為 TASA 日後資料處理應增加討論的部份。

表 4-3-6 試題之描述性統計分析比較

		NAEP	PISA	TIMSS	TASA
	相同	人數、選項百分比、答對率、鑑別度、難度、點二系列相關			
選擇題	相異	遺漏未作答百分比、無效試卷百分比、難度指標轉換為平均=13、標準差=4	試題不合適之均方根的值、遺漏未作答百分比、PV 值平均、PV 值標準差、期望分數曲線	缺失遺漏未作答百分比	無此資料
	相同	人數、無效遺漏未作答百分比或次數分配、選項百分比或次數分配			
開放性試題	相異	難度指標轉換為平均=13、標準差=4、polyserial 相關、Pearson 相關	單參難度 (Item Map)	鑑別度、單參難度、試題在國家間的交互作用。	無此資料

針對表 4-3-6 的差異比較可以發現，TASA 在 PV 值的使用、缺失遺漏未作答的資料處理等議題上並無著墨，因此提出底下幾點建議：

1. 增加 PV 值之計算。
2. 增加缺失、遺漏及未作答選項之描述性統計值。
3. 增加開放性試題難度、鑑別度等參數的估計。
4. 增加不同年度間定錨題之參數比較。

## 貳、評分者間一致性檢定

### 一、NAEP

對於開放性試題，在閱讀、寫作、公民測驗中，至少抽取 25% 的試題；而閱讀、寫作在每個州（郡），至少抽取 6% 進行兩次評分，表 4-3-7 為所抽取試題之二次評分一致性狀況比較。

表 4-3-7 試題二次評分評分者一致性分析

*Interrater Reliability Ranges for the NAEP 1998 Assessment*

Grade	Total Number of Unique Items	Number and Percentage of Items in Percentage Exact Agreement Range								
		60-69%		70-79%		80-89%		Above 90%		
		Number	Percent	Number	Percent	Number	Percent	Number	Percent	
<b>Reading</b>										
4	46	—	—	3	6.5	16	34.7	27	58.6	
8	69	1	1.4	4	5.8	28	40.6	36	52.2	
12	76	1	1.3	4	5.2	36	47.4	35	46.1	
<b>Writing</b>										
4	20	4	20.0	16	80.0	—	—	—	—	
8	23	18	78.3	4	17.4	—	—	—	—	
12	23	10	43.5	9	39.1	3	13.0	—	—	
<b>Civics</b>										
4	21	—	—	3	14.3	11	52.4	7	33.3	
8	28	1	3.6	6	21.4	17	60.7	4	14.3	
12	29	—	—	8	27.6	20	70.0	1	3.4	

資料來源：NAEP 1998 技術報告（頁 125）

表 4-3-8 填充題二次評分評分者一致性分析

Item	Block	Range of Response Codes	Correct Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
R012102	R4	1-2	2	1,923	98	0.970
R012104	R4	1-2	2	1,900	96	0.910
R012106	R4	1-2	2	1,862	93	0.859
R012108	R4	1-2	2	1,761	97	0.920
R012109	R4	1-2	2	1,752	97	0.922
R012112	R4	1-2	2	1,299	94	0.870
R012201	R6	1-2	2	1,925	96	0.923
R012206	R6	1-2	2	1,697	98	0.956
R012208	R6	1-2	2	1,547	93	0.852
R012210	R6	1-2	2	1,452	95	0.820
R012503	R10	1-2	2	1,921	96	0.922
R012504	R10	1-2	2	1,897	98	0.969
R012506	R10	1-2	2	1,865	97	0.949
R012508	R10	1-2	2	1,794	98	0.956
R012511	R10	1-2	2	1,637	97	0.941
R012601	R5	1-2	2	1,897	90	0.759
R012604	R5	1-2	2	1,855	93	0.834
R012611	R5	1-2	2	1,475	89	0.779
R012702	R7	1-2	2	1,908	97	0.913

資料來源：NAEP 1998 技術報告（頁 565）

表 4-3-9 多點計分試題二次評分評分者間一致性分析

Item	Block	Range of Response Codes <sup>1</sup>	Sample Size	Percent Agreement	Intraclass Correlation
R012111	R4	1-4	1,555	91	0.946
R012204	R6	1-4	1,894	81	0.906
R012512	R10	1-4	1,591	90	0.957
R012607	R5	1-4	1,770	85	0.867
R012708	R7	1-4	1,637	87	0.908
R015702	R8	1-3	2,036	87	0.841
R015703	R8	1-3	2,017	89	0.862
R015704	R8	1-3	1,978	84	0.870
R015705	R8	1-3	1,963	90	0.942
R015707	R8	1-4	1,834	89	0.904
R015709	R8	1-3	1,558	88	0.881
R015803	R9	1-3	1,891	88	0.841
R015804	R9	1-4	1,845	83	0.873
R015806	R9	1-3	1,706	87	0.884
R015807	R9	1-3	1,548	87	0.880

資料來源：NAEP 1998 技術報告（頁 568）

## 二、TIMSS

針對開放性試題，為了提高測驗成績的信度，在每個國家的測驗題本中，選取約 25% 的試題進行兩次評分，並在每個國家當中，計算每個試題被評分兩次的次數及前後兩次評分之一致性。

### (一) 各國國內二次評分評分者一致性分析 (within-country scoring reliability)

TIMSS2007 計 14 個題本，每個試題至少出現在兩個題本中，每個題本抽 100 個學生做二次評分，每個試題進行評分者間一致性檢定的樣本數為 200，共 1400 個學生接受二次評分，約佔總受試者的 25%。

表 4-3-10 各國國內二次評分評分者間一致性分析

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Min	Max		Min	Max
Algeria	92	58	99	85	54	98
Armenia	99	94	100	97	91	100
Australia	100	98	100	99	95	100
Austria	99	95	100	99	94	100
Chinese Taipei	98	84	100	97	83	100
Colombia	99	93	100	97	89	100
Czech Republic	98	90	100	96	77	100
Denmark	97	83	100	93	74	99
El Salvador	99	96	100	98	85	100
England	99	91	100	98	89	100
Georgia	97	88	100	94	68	100
Germany	97	75	100	95	71	100
Hong Kong SAR	100	98	100	100	98	100
Hungary	100	97	100	99	95	100
Iran, Islamic Rep. of	99	96	100	96	84	100
Italy	99	94	100	99	79	100
Japan	99	94	100	98	84	100
Kazakhstan	99	96	100	99	94	100
Kuwait	100	98	100	98	95	100
Latvia	95	41	100	92	39	100
Lithuania	98	88	100	97	50	100
Morocco	95	33	100	88	29	98
Netherlands	97	86	100	95	72	100
New Zealand	99	95	100	97	90	100
Norway	99	92	100	97	88	100
Qatar	99	91	100	95	78	100
Russian Federation	100	98	100	99	96	100
Scotland	99	91	100	98	87	100
Singapore	99	93	100	97	90	100
Slovak Republic	99	92	100	98	90	100
Slovenia	100	99	100	99	94	100
Sweden	98	89	100	97	87	100
Tunisia	98	86	100	93	77	99
Ukraine	100	98	100	100	98	100
United States	98	83	100	96	72	100
Yemen	98	83	100	93	80	99
International Avg.	98	88	100	96	81	100
<b>Benchmark Participants</b>						
Alberta, Canada	99	93	100	98	90	100
British Columbia, Canada	99	96	100	99	91	100
Dubai, UAE	97	87	100	94	78	100
Massachusetts, US	98	82	100	96	72	100
Minnesota, US	98	79	100	96	68	100
Ontario, Canada	99	88	100	98	88	100
Quebec, Canada	98	90	100	97	86	100

資料來源：TIMSS 2007 技術報告 (頁 212)

(二) 定錨題在各國國內二次評分評分者間一致性分析 (trend item scoring reliability)

表 4-3-11 定錨題在各國國內二次評分評分者間一致性分析

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Min	Max		Min	Max
Armenia	96	90	100	93	65	99
Australia	97	84	100	96	83	100
Chinese Taipei	97	93	100	96	88	100
England	98	92	100	97	87	100
Hong Kong SAR	99	93	100	98	87	100
Hungary	99	96	100	97	92	100
Iran, Islamic Rep. of	98	95	100	96	86	100
Japan	98	93	100	96	88	100
Lithuania	97	88	100	94	74	100
Netherlands	97	90	99	95	88	99
New Zealand	98	95	100	97	90	100
Norway	98	96	100	97	93	100
Russian Federation	99	95	100	98	92	100
Scotland	96	91	100	95	90	100
Singapore	95	86	100	93	83	100
Slovenia	96	68	99	93	47	99
Tunisia	98	97	100	95	81	100
United States	98	92	100	96	88	100
International Avg.	97	91	100	96	83	100

資料來源：TIMSS 2007 技術報告 (頁 216)

(三) 國際間二次評分評分者間一致性分析 (cross-country scoring reliability study)

TIMSS2003 為了檢測不同國家的評分者在做二次評分時是否有所差異，從南半球四個國家抽 150 個學生，每位學生測驗 20 題數學與 21 題自然科學；共有 6150 個作答反應。從北半球是二十個國家中找 37 位評分者，二次評分有 666 個組合。每個組合都評分 150 個學生，共有 99900 種評分反應。

表 4-3-12 國際間二次評分評分者間一致性分析 (TIMSS 2007 小四數學)

Item Label	Total Valid Comparisons	Exact Percent Agreement	
		Correctness Score Agreement	Diagnostic Score Agreement
M04_02 - M041056	265200	98	96
M04_04 - M041076	265200	99	98
M04_07 - M041146	265200	92	92
M04_09 - M041258A	265200	96	94
M04_09 - M041258B	265200	86	74
M04_11 - M041275	265200	85	85
M05_02 - M031309	265200	99	99
M05_04 - M031242A	265200	98	97
M05_04 - M031242B	265200	97	96
M05_05 - M031247	265200	94	91
M11_02 - M031009	265200	100	99
M11_04 - M031316	265200	99	99
M11_06 - M031079B	261579	99	99
M11_06 - M031079C	261579	97	97
M11_09 - M031325	265200	97	92
M12_04 - M041059	265200	99	95
M12_13 - M041276A	265200	98	98
M12_13 - M041276B	265200	83	79
Average Percent Agreement		95	93

資料來源：TIMSS 2007 技術報告 (頁 221)

三、綜合討論與建議



表 4-3-13 為 NAEP、TIMSS 對開放性試題進行二次評分之評分者間一致性檢定的比較，NAEP 及 TIMSS 分別針對國內及國際間部分開放性試題二次評分資料進行評分者間一致性檢定，藉以提高測驗評量數據之信度，分析方法僅抽樣有部分差異，檢測則皆以一致性百分比為依據，輔以 Kappa 一致性係數。TASA 在國語文、英語文及數學某些年度或年段中亦有多元計分試題之測驗，因此 TASA 往後對於多元計分試題可增加評分者間一致性檢定，以提高學生學習成就描述之信度。

表 4-3-13 評分者間一致性分析比較

NAEP	TIMSS	TASA
對於開放性試題，在閱讀、寫作、公民測驗中，至少抽取25%的試題；而閱讀、寫作在各州（郡）至少抽取6%的試題，進行二次評分的評分者一致性分析。	針對開放性試題在各國測驗題本中，選取約25%的試題進行各國國內、定錨題在各國國內及國際間二次評分的評分者一致性分析。	無此資料

表 4-3-14 為 TASA 實徵資料分析，觀察評分者間一致性檢測是否適用於 TASA 資料。選用數學科應用題進行分析，因為 TASA 資料庫數學科應用題在 2006 及 2007 並無二次評分資料，因此以 2009 年數學科小四資料為例，可以發現在總題數 26 題中，二次評分資料呈現的一致性達 96% 以上，Cohen's Kappa 值也在 0.93 以上。建議 TASA 往後對於多元計分試題在閱卷者人數設計上可要求至少兩人，以便進行評分者間一致性檢定，提高測驗評量之信度。

表 4-3-14 2009 年數學科小四應用題評分者間一致性檢測結果整理

Item	Block	Sample Size	Percent Agreement	Cohen's Kappa	
1	S1	S26	656	98.93%	0.976
2	S1	S2	660	97.73%	0.954
3	S2	S3	666	99.25%	0.969
4	S3	S4	665	99.10%	0.972
5	S4	S5	655	98.47%	0.969
6	S5	S6	639	98.59%	0.937
7	S6	S7	634	98.26%	0.959
8	S7	S8	642	99.22%	0.982
9	S8	S9	645	98.91%	0.974
10	S9	S10	650	96.77%	0.947
11	S10	S11	663	99.10%	0.979
12	S11	S12	671	98.36%	0.97
13	S12	S13	661	97.58%	0.954
14	S13	S14	644	97.98%	0.963
15	S14	S15	642	98.29%	0.969
16	S15	S16	645	98.76%	0.961
17	S16	S17	639	99.37%	0.986
18	S17	S18	637	97.17%	0.949
19	S18	S19	645	98.45%	0.965
20	S19	S20	648	99.69%	0.99
21	S20	S21	650	98.77%	0.979
22	S21	S22	656	96.65%	0.951
23	S22	S23	652	97.85%	0.957
24	S23	S24	637	96.55%	0.939
25	S24	S25	632	99.21%	0.981
26	S25	S26	646	98.76%	0.972

參、試題標記與刪題標準

一、PISA

(一) 不良試題

1. 假如某一試題的特徵經過 10 個以上國家的分析都是不良的，則此試題會被刪除，此種試題又被稱為“不良試題”。

2. 有些試題可能在某些國家中沒有被施測，因為這些試題的參數在這些國家分析的結果是不良，但在其他主要的國家這些試題卻表現良好
3. 有些試題具有良好的參數特性，但卻顯示試題和國家具有交互作用，即所謂具有差異性的試題，或試題難度對於不同國家而言是不同的。

上述第二類和第三類的試題都會對國家間的比較造成影響。

## (二) 第二類和第三類不良試題標記

PISA2006 針對所謂的不良試題進行標記，藉以區辨出試題不良的原因，以便進行修改、訂正或刪除的作業，如表 4-3-15 所示。

表 4-3-15 不良試題在國家間之標記

Example of summary of dodgy items for a country in Report 3a

*PISA 2006 Main Study, Report 3a: Science dodgy items*

	Item by country interactions			Discrimination				PISA 2003 link items	
	Number of valid responses	Easier than expected	Harder than expected	Non-key PB is positive	Key PB is negative	Low discrimination	Ability not ordered	Link items	Requires checking
S456Q02	1 437	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S476Q01	1 482	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S477Q04	1 442	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S478Q01	1 443	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S493Q01	1 452	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S495Q01	1 442	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S495Q02	1 440	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S508Q02	1 435	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S510Q04	1 459	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S519Q01	1 438	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
S524Q06	1 427	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

資料來源：PISA 2006 技術報告（頁 152）

## 二、TIMSS

(一) 使用各種標記藉以突顯試題的特性，底下為整理 TIMSS 針對試題進行標記的各種充分條件：

1. 難度>95%
2. 難度<25% (4 選 1 試題)
3. 誘答力<10%
4. 錯誤選項的誘答力>正確選項的答對率或錯誤選項的點二系列相關>0
5. 鑑別度<0.2

6. 開放性試題鑑別度無遞增
7. 針對開放性試題，1PL 難度 > 總平均
8. 針對開放性試題，1PL 難度 < 總平均
9. 男性與女性在試題難度上有差異
10. 開放性試題兩次評分的一致性 < 70%

(二) 符合底下刪題標準之不良試題予以刪除，藉以提高測驗的信度。

1. 已知試題有翻譯上的錯誤，在測試前無及時修正之試題。
2. 選擇題在不同國家間無統一答案。
3. 選擇題正確選項的相關呈現負值或開放性試題鑑別度無遞增。
4. 試題在高（低）成就國家有低（高）表現。
5. 開放性試題，二次評分的一致性（信度） < 70%。
6. 試題在 1999 年與 2003 年的測驗表現上產生差異或試題在 1999 年在某特殊國家已刪除。

### 三、TASA

(一) 不良試題檢測

1. 依據古典測驗理論（classical test theory, CTT）訂定標準如下：

- (1) 試題通過率低於 0.25， $P_{CTT} < 25\%$ ；
- (2) 試題鑑別度指數介於 0 ~ 0.2 之間， $0 < C.T.T. \text{的鑑別度} < 0.2$ 。

試題鑑別度指數使用點二系列相關，公式如下：

$$r_j = \frac{(\mu_j - \mu_x)}{\sigma_x} \sqrt{\frac{P_j}{1 - P_j}} \quad (1)$$

其中， $r_j$  指第  $j$  題鑑別度；

$\mu_j$  指答對第  $j$  題之受試者之測驗總分的平均數；

$\mu_x$  指所有受試者之測驗總分的平均數；

$\sigma_x$  指所有受試者之測驗總分的標準差。

2. 依據試題反應理論 (item response theory, IRT) 訂定標準如下：

- (1) 試題鑑別度參數介於 0~0.4 之間， $0 < a < 0.4$  ；
- (2) 試題難度參數小於或等於-3， $b \leq -3$  ；
- (3) 試題難度參數大於或等於 3， $b \geq 3$  ；
- (4) 試題猜測度參數大於或等於 0.3， $c \geq 0.3$  。

(二) 符合底下刪題標準之不良試題予以刪除

1. 估計後的試題參數無法收斂；
2. 試題鑑別度指數小於 0，C.T.T.的鑑別度 $<0$ 。
3. 點二系列相關 $<0$ 。

TASA 使用的軟體為 BILOG-MG (Zimowski, Muraki, Mislevy, & Book, 2003) 與 SCORIGHT 3.0 (Wang, Bradlow, & Wainer, 2004)。其中，刪題標準 (1) 是指使用上述兩軟體估計試題參數後，試題參數出現異常之情形 (例如  $a = 8.88$ 、 $b = 13.27$  等)；刪題標準 (2) 是指該試題估計之試題鑑別度指數小於 0 ( $r_j < 0$ )。

#### 四、綜合討論與建議

表 4-3-16 為綜合分析比較 TASA 與各大型測驗資料庫之間對於不良試題分析上的差異，可以發現因為試題參數估計原則上皆採用 1PL 或 3PL，因此區辨是否為不良試題之參考標準相近，增加對不良試題各項參數的描述或標記，如表 4-3-15，使讀者更清楚了解其產生不良效應背後的意涵，或許是 TASA 往後進行不良試題分析可以著墨之處。

表 4-3-16 試題標記與刪題標準比較

	PISA	TIMSS	TASA
不良試 題標記	1. 試題在國際 間比較過易 或過難	1. $P > 95\%$ 2. $P < 25\%$ (4 選 1 試題) 3. 誘答力 $< 10\%$ 4. 錯誤選項的誘答力 $>$ 正確選項 的答對率或錯誤選項的點二 系列相關 $> 0$	1. 依據古典測驗 理論訂定標準 如下： $P < 25\%$ 、鑑別度指 數 $D$ 介於 $0 \sim 0.2$ 。
	2. 錯誤選項的 點二系列相 關 $> 0$	5. $D < 0.2$	2. 依據試題反應 理論訂定標準 如下：
	3. 正確選項的 點二系列相 關 $< 0$	6. C-R 試題鑑別度無遞增 7. $1-PL$ 難度 $>$ 總平均 8. $1-PL$ 難度 $<$ 總平均	$0 < a < 0.4$ 、 $b \leq$ $-3$ 、 $b \geq 3$ 、 $c \geq$ $0.3$
	4. 鑑別度過低	9. 性別 DIF	
	5. C-R 試題鑑 別度無遞增	10. C-R 試題兩次評分的一致性 $< 70\%$	
刪題 標準	某一試題的特徵 經過 10 個以上 國家的分析都是 不良的，則此試 題會被刪除	1. 有翻譯上錯誤，測試前無及時 修正 2. M-C 試題在不同國家無統一 答案 3. M-C 試題正確選項的相關呈 現負值或 C-R 試題鑑別度無 遞增 4. 試題在高（低）成就國家有低 （高）表現 5. C-R 試題二次評分一致性 $<$ $70\%$ 6. 試題在 1999 年與 2003 年的測 驗表現上產生差異或試題在 1999 年某特殊國家已刪除	1. 估計後的試題 參數無法收斂 2. 試題鑑別度指 數 C.T.T. 的鑑 別度 $< 0$ 3. 點二系列相關 $< 0$

肆、差異試題功能分析 (**differential item functioning analysis**，簡稱 **DIF**)

### 一、NAEP

對照組及實驗組的選取分三類：男性/女性、白人/黑人及白人/西班牙裔，所使用的分析方法有三種：

#### (一) The Mantel-Haenszel Procedure

針對選擇題和開放性試題使用不同的分析流程，選擇題使用 Mantel-Haenszel chi-square 分析流程 (Mantel & Haenszel, 1959)。開放性試題則使用 Mantel-Haenszel ordinal 分析流程 (Mantel, 1963)，它的分析流程乃是比較兩群體間各得分群人數百分比是否存在差異而得。

#### (二) SIBTEST Procedure

NAEP1998 首次針對所有試題同時進行 Mantel-Haenszel (Mantel & Haenszel, 1959) 及 SIBTEST (Shealy & Stout, 1993) 差異試題功能分析，如同 M-H 法，SIBTEST 也是利用兩群體間的答對率或選項作答率進行分析，另外再考慮取樣權重的影響，計算標準平均差異 (SMD) 及效果量係數 (SMD/std. dev.)，若試題滿足以下兩個條件，則被評為具嚴重試題差異之試題：

1. SMD 顯著的偏離 0。

2. 效果量係數 (SMD/std. dev.) 的絕對值大於 0.25。

結果發現對於是否存在試題差異的檢測結果一致，M-H 法認為存在試題差異之試題，SIBTEST 法依然會將該試題判定存在試題差異，並沒有任何存在嚴重試題差異之試題單獨被 SIBTEST 法所判定。

#### (三) Standardization Method

針對寫作試題，若單獨使用 Mantel-Haenszel 法或 SIBTEST 法進行試題差異功能分析，所獲取的資訊稍嫌不足，因此 NAEP 結合 M-H 法及 SIBTEST 法中部份分析方法，合併使用稱為 Standard NAEP DIF Method (Dorans and Kulick, 1986)。在分析上，寫作試題被視為多點計分，前半段分析流程與 M-H 法一致，接著計算各得分群兩兩之標準平均差異 (SMD)，將獲得的所有 SMD 取平均再除以標準偏差 (std. dev.) 得到效果量係數 (SMD/std. dev.)，效果量係數的絕對值大於 0.1 即被標記出，如表 4-3-17 所示。

表 4-3-17 兩族群間的試題差異功能分析

*Items With Absolute SMD (Standardized Mean DIF) > .10*

Group	Grade	SMD	ID
NonAcc/Acc	4	-.106	W005402
B/W	4	-.108	W005302
B/W	12	-.129	W009802
B/W	12	.127	W010402
H/W	4	-.101	W004602
H/W	12	-.112	W009202

**LEGEND**

NonAcc/Acc Nonaccommodated versus accommodated students  
 B/W Black versus White students  
 H/W Hispanic versus White students

資料來源：NAEP 1998 技術報告（頁 369）

二、綜合討論與建議

表 4-3-18 是大型測驗之差異試題功能分析比較，相較 NAEP 及 PISA，TASA 問卷調查亦包含男女性別及國籍別資料，可評估是否進行 DIF 檢測，表 4-3-19 及表 4-3-20 分別就 M-H 法及 SIBTEST 法，以 2006TASA 數學科小四試題資料進行 DIF 之實徵資料分析，提供參考之依據，因本土及其他各國籍學生數量相差過大，因此僅以男女性別資料為例，紅色標記部分表示在顯著水準 0.01 下，性別間存在 DIF 現象之試題。建議 TASA 往後可針對所有試題，增加性別、城鄉及語言族群間之 DIF 分析，提高試題品質的控管。

表 4-3-18 差異試題功能分析比較

NAEP	PISA	TASA
對照組及實驗組的選取分三類：男性/女性、白人/黑人及白人/西班牙裔，所使用的分析方法有底下三種：The Mantel-Haenszel Procedure、SIBTEST Procedure、Standardization Method	性別、國家間的DIF檢測	無此資料



表 4-3-19 2006TASA 數學科小四試題 DIF 分析 M-H 法結果整理

ITEM	Alpha	X <sup>2</sup>	P-Value	MH D-DIF	SE (MH D-DIF)
A B1 01***	0.695	8.442	0.00	0.855	0.289
A B1 02	1.185	1.986	0.16	-0.399	0.272
A B1 03*	0.756	6.285	0.01	0.656	0.256
A B1 04	1.059	0.266	0.61	-0.134	0.238
A B1 05*	0.754	5.983	0.01	0.662	0.265
A B1 06	0.86	1.929	0.17	0.353	0.245
B B1 07***	1.755	24.476	0.00	-1.322	0.266
A B1 08	1.298	4.536	0.03	-0.612	0.28
A B2 01***	0.662	13.96	0.00	0.969	0.256
A B2 02	1.032	0.036	0.85	-0.074	0.292
A B2 03	0.991	0.001	0.98	0.022	0.269
A B2 04	0.98	0.012	0.91	0.048	0.287
A B2 05**	1.428	6.829	0.01	-0.837	0.313
A B2 06	1.059	0.147	0.70	-0.134	0.3
A B2 07	1.087	0.486	0.49	-0.195	0.26
A B2 08	1.077	0.381	0.54	-0.174	0.259
A B3 01	0.918	0.392	0.53	0.202	0.293
A B3 02	1.102	0.638	0.42	-0.227	0.266
A B3 03	1.361	5.713	0.02	-0.725	0.296
A B3 04	1.048	0.171	0.68	-0.11	0.237
A B3 05	0.805	4.136	0.04	0.508	0.244
A B3 06	0.902	0.837	0.36	0.242	0.251
A B3 07	0.993	0	1.00	0.017	0.267
A B3 08	0.944	0.242	0.62	0.136	0.249
A B4 01	0.995	0	0.99	0.013	0.292
A B4 02	1.01	0.001	0.97	-0.023	0.254
A B4 03*	0.765	6.062	0.01	0.629	0.251
A B4 04	0.876	1.321	0.25	0.311	0.258
A B4 05	1.128	1.063	0.30	-0.283	0.26
A B4 06	1.213	2.834	0.09	-0.454	0.261
A B4 07	1.184	1.29	0.26	-0.397	0.33
A B4 08	0.876	0.688	0.41	0.311	0.345
A B5 01	1.273	4.453	0.04	-0.568	0.262
A B5 02	0.974	0.026	0.87	0.061	0.276
A B5 03	0.719	5.105	0.02	0.775	0.333
CR B5 04***	1.952	35.281	0.00	-1.571	0.265
A B5 05**	0.732	7.535	0.01	0.733	0.262
A B5 06**	1.362	7.406	0.01	-0.726	0.262
A B5 07	1.067	0.192	0.66	-0.151	0.301
A B5 08	0.897	0.886	0.35	0.254	0.256

表 4-3-19 2006TASA 數學科小四試題 DIF 分析 M-H 法結果整理(續)

ITEM	Alpha	X <sup>2</sup>	P-Value	MH D-DIF	SE (MH D-DIF)
A B6_01	0.887	0.916	0.34	0.282	0.278
A B6_02	0.9	0.782	0.38	0.247	0.263
A B6_03	1.037	0.074	0.79	-0.085	0.258
A B6_04	1.181	2.023	0.16	-0.39	0.264
A B6_05	0.884	0.991	0.32	0.29	0.276
A B6_06**	1.364	7.804	0.01	-0.729	0.256
A B6_07***	0.675	9.894	0.00	0.925	0.289
A B6_08	0.997	0	0.98	0.007	0.241
A B7_01	0.791	4.324	0.04	0.55	0.258
A B7_02	0.979	0.02	0.89	0.051	0.258
A B7_03***	1.508	13.175	0.00	-0.966	0.262
A B7_04	0.8	3.687	0.06	0.524	0.265
A B7_05	1.223	2.93	0.09	-0.474	0.269
A B7_06	1.199	2.504	0.11	-0.426	0.261
A B7_07	0.869	1.519	0.22	0.331	0.257
A B7_08	0.861	1.802	0.18	0.351	0.252
A B8_01	0.894	0.783	0.38	0.262	0.278
A B8_02***	1.403	7.996	0.01	-0.797	0.276
A B8_03	1.099	0.629	0.43	-0.221	0.26
A B8_04	0.988	0.003	0.95	0.027	0.246
A B8_05	1.066	0.211	0.65	-0.151	0.289
A B8_06	0.791	3.799	0.05	0.55	0.274
A B8_07	1.243	3.383	0.07	-0.511	0.27
A B8_08	0.815	2.691	0.10	0.48	0.282
A B9_01*	1.337	6.317	0.01	-0.683	0.266
A B9_02**	0.667	6.924	0.01	0.953	0.353
A B9_03	0.778	4.924	0.03	0.59	0.26
A B9_04	0.956	0.119	0.73	0.105	0.261
A B9_05	0.961	0.085	0.77	0.093	0.267
A B9_06***	1.42	8.335	0.00	-0.825	0.28
A B9_07	1.098	0.739	0.39	-0.219	0.241
A B9_08	0.83	2.8	0.09	0.439	0.254
A B10_01	0.918	0.614	0.43	0.2	0.239
A B10_02	1.153	1.368	0.24	-0.336	0.273
A B10_03	1.101	0.461	0.50	-0.227	0.304
A B10_04	0.817	2.723	0.10	0.475	0.278
A B10_05	0.899	0.821	0.37	0.25	0.261
A B10_06	0.876	1.224	0.27	0.311	0.267
A B10_07	1.222	3.061	0.08	-0.47	0.261
A B10_08	1.02	0.014	0.91	-0.047	0.266

表 4-3-19 2006TASA 數學科小四試題 DIF 分析 M-H 法結果整理(續)

ITEM	Alpha	X <sup>2</sup>	P-Value	MH D-DIF	SE (MH D-DIF)
A B11_01	1.258	4.262	0.04	-0.54	0.255
A B11_02	0.922	0.47	0.49	0.191	0.257
A B11_03	1.154	1.598	0.21	-0.338	0.256
A B11_04	0.996	0	0.99	0.01	0.267
A B11_05	1.115	0.405	0.52	-0.256	0.36
A B11_06	0.823	3.719	0.05	0.458	0.232
A B11_07	0.936	0.331	0.57	0.156	0.249
A B11_08	0.939	0.308	0.58	0.148	0.244
A B12_01	0.941	0.18	0.67	0.144	0.295
A B12_02	1.17	1.441	0.23	-0.369	0.292
A B12_03	1.034	0.042	0.84	-0.079	0.295
A B12_04	0.993	0.003	0.96	0.016	0.425
A B12_05	1.218	2.882	0.09	-0.464	0.265
A B12_06	0.961	0.07	0.79	0.094	0.289
A B12_07	0.709	5.068	0.02	0.809	0.35
A B12_08	0.938	0.308	0.58	0.149	0.246
A B13_01	1.032	0.061	0.81	-0.073	0.245
A B13_02	1.303	4.611	0.03	-0.623	0.282
A B13_03	0.993	0	1.00	0.016	0.293
A B13_04	0.798	3.866	0.05	0.529	0.262
A B13_05	0.792	4.673	0.03	0.547	0.247
A B13_06	1.297	5.117	0.02	-0.612	0.264
A B13_07	0.967	0.082	0.78	0.079	0.236
A B13_08	1.028	0.043	0.84	-0.064	0.246

表 4-3-20 2006TASA 數學科小四試題 DIF 分析 SIBTEST 法結果整理

SIBTEST-pooled weighting								
Run	Suspect Subtest	Beta	standard		p-	elim	MS	FLAG
no.	Item Numbers	estimate	error	p-value	R	F	SSD	
1	B1_01	-0.08	0.02	0.000 E	0.11	0.11	0.09	0
2	B1_02	0.03	0.021	0.211 E	0.12	0.11	0.04	0
3	B1_03	-0.06	0.024	0.014 E	0.09	0.09	0.09	0
4	B1_04	0.01	0.024	0.657 E	0.12	0.12	0.05	0
5	B1_05	-0.06	0.022	0.011 E	0.1	0.11	0.09	0
6	B1_06	-0.03	0.024	0.267 E	0.11	0.1	0.07	0
7	B1_07	0.11	0.023	0.000 E	0.1	0.11	-0.01	0
8	B1_08	0.05	0.021	0.020 E	0.11	0.11	0.03	0
1	B2_01	-0.09	0.025	0.001 E	0.2	0.19	0.13	0
2	B2_02	0.00	0.02	0.991 E	0.23	0.2	0.09	0
3	B2_03	0.01	0.024	0.639 E	0.17	0.15	0.09	0
4	B2_04	-0.01	0.023	0.628 E	0.16	0.15	0.09	0
5	B2_05	0.06	0.019	0.001 E	0.22	0.2	0.07	0
6	B2_06	0.02	0.02	0.251 E	0.21	0.22	0.09	0
7	B2_07	0.00	0.024	0.881 E	0.18	0.17	0.08	0
8	B2_08	0.02	0.025	0.483 E	0.18	0.17	0.08	0
1	B3_01	-0.02	0.02	0.414 E	0.15	0.15	0.1	0
2	B3_02	0.02	0.023	0.420 E	0.12	0.11	0.08	0
3	B3_03	0.05	0.02	0.015 E	0.16	0.15	0.07	0
4	B3_04	0.03	0.025	0.209 E	0.13	0.13	0.09	0
5	B3_05	-0.03	0.025	0.175 E	0.14	0.14	0.12	0
6	B3_06	-0.01	0.024	0.687 E	0.12	0.11	0.11	0
7	B3_07	-0.01	0.024	0.760 E	0.11	0.1	0.1	0
8	B3_08	-0.01	0.024	0.656 E	0.13	0.11	0.1	0
1	B4_01	0.01	0.019	0.593 E	0.12	0.1	0.14	0
2	B4_02	0.02	0.023	0.414 E	0.08	0.06	0.14	0
3	B4_03	-0.06	0.024	0.010 E	0.08	0.06	0.17	0
4	B4_04	-0.01	0.023	0.580 E	0.09	0.07	0.15	0
5	B4_05	0.04	0.022	0.065 E	0.09	0.08	0.13	0
6	B4_06	0.05	0.023	0.020 E	0.07	0.06	0.11	0
7	B4_07	0.03	0.017	0.105 E	0.09	0.09	0.13	0
8	B4_08	-0.01	0.017	0.756 E	0.09	0.09	0.16	0
1	B5_01	0.04	0.023	0.116 E	0.08	0.07	0.12	0
2	B5_02	0.00	0.022	0.837 E	0.1	0.08	0.14	0
3	B5_03	-0.04	0.017	0.029 E	0.13	0.14	0.17	0
4	B5_04	0.14	0.022	0.000 E	0.1	0.08	0.05	0
5	B5_05	-0.06	0.023	0.009 E	0.09	0.09	0.19	0
6	B5_06	0.07	0.022	0.001 E	0.09	0.08	0.1	0
7	B5_07	0.02	0.019	0.308 E	0.12	0.14	0.14	0
8	B5_08	-0.02	0.023	0.350 E	0.09	0.07	0.16	0

表 4-3-20 2006TASA 數學科小四試題 DIF 分析 SIBTEST 法結果整理(續)

Run no.	Suspect Item Numbers	SIBTEST-pooled weighting		p-value	p-R	elim F	MS SSD	FLAG
		Beta estimate	standard error					
1	B6_01	-0.02	0.022	0.297 E	0.15	0.15	0.06	0
2	B6_02	-0.02	0.024	0.380 E	0.14	0.13	0.06	0
3	B6_03	0.01	0.024	0.582 E	0.15	0.13	0.04	0
4	B6_04	0.02	0.023	0.400 E	0.14	0.13	0.03	0
5	B6_05	-0.01	0.022	0.577 E	0.17	0.16	0.06	0
6	B6_06	0.07	0.024	0.002 E	0.15	0.13	0.01	0
7	B6_07	-0.05	0.02	0.009 E	0.18	0.19	0.08	0
8	B6_08	0.02	0.025	0.549 E	0.16	0.14	0.05	0
1	B7_01	-0.05	0.024	0.053 E	0.17	0.16	0.17	0
2	B7_02	-0.01	0.024	0.620 E	0.18	0.16	0.15	0
3	B7_03	0.09	0.023	0.000 E	0.19	0.16	0.1	0
4	B7_04	-0.04	0.023	0.087 E	0.2	0.18	0.17	0
5	B7_05	0.03	0.022	0.247 E	0.2	0.17	0.12	0
6	B7_06	0.03	0.024	0.260 E	0.17	0.14	0.12	0
7	B7_07	-0.03	0.024	0.285 E	0.17	0.14	0.16	0
8	B7_08	-0.04	0.025	0.096 E	0.17	0.15	0.16	0
1	B8_01	-0.04	0.024	0.151 E	0.22	0.19	0.19	0
2	B8_02	0.06	0.024	0.016 E	0.22	0.22	0.15	0
3	B8_03	0.03	0.025	0.224 E	0.23	0.22	0.17	0
4	B8_04	-0.01	0.026	0.861 E	0.25	0.23	0.18	0
5	B8_05	0.01	0.023	0.715 E	0.2	0.2	0.17	0
6	B8_06	-0.06	0.024	0.012 E	0.22	0.2	0.2	0
7	B8_07	0.01	0.025	0.646 E	0.21	0.2	0.16	0
8	B8_08	-0.02	0.022	0.314 E	0.24	0.23	0.19	0
1	B9_01	0.06	0.022	0.010 E	0.16	0.14	0.08	0
2	B9_02	-0.03	0.016	0.062 E	0.17	0.16	0.14	0
3	B9_03	-0.05	0.024	0.029 E	0.14	0.09	0.14	0
4	B9_04	-0.01	0.024	0.792 E	0.12	0.1	0.12	0
5	B9_05	0.00	0.023	0.927 E	0.16	0.11	0.12	0
6	B9_06	0.06	0.023	0.007 E	0.13	0.09	0.08	0
7	B9_07	0.04	0.025	0.147 E	0.15	0.11	0.1	0
8	B9_08	-0.01	0.024	0.635 E	0.15	0.12	0.14	0
1	B10_01	0.00	0.026	0.901 E	0.22	0.22	0.03	0
2	B10_02	0.03	0.023	0.232 E	0.27	0.25	0.01	0
3	B10_03	0.01	0.022	0.574 E	0.2	0.18	0.01	0
4	B10_04	-0.04	0.024	0.110 E	0.2	0.21	0.04	0
5	B10_05	0.00	0.025	0.922 E	0.19	0.21	0.03	0
6	B10_06	-0.02	0.024	0.438 E	0.23	0.25	0.04	0
7	B10_07	0.03	0.023	0.202 E	0.23	0.21	0	0
8	B10_08	0.00	0.023	0.890 E	0.2	0.21	0.02	0

表 4-3-20 2006TASA 數學科小四試題 DIF 分析 SIBTEST 法結果整理(續)

SIBTEST-pooled weighting								
Run no.	Suspect Subtest Item Numbers	Beta estimate	standard error	p-value	p-R	elim F	MS SSD	FLAG
1	B11_01	0.04	0.024	0.080 E	0.15	0.17	0.09	0
2	B11_02	-0.02	0.024	0.545 E	0.18	0.22	0.12	0
3	B11_03	0.03	0.024	0.215 E	0.18	0.2	0.1	0
4	B11_04	0.00	0.023	0.867 E	0.14	0.17	0.11	0
5	B11_05	0.01	0.016	0.365 E	0.15	0.18	0.11	0
6	B11_06	-0.04	0.026	0.156 E	0.18	0.24	0.14	0
7	B11_07	-0.01	0.025	0.736 E	0.16	0.2	0.12	0
8	B11_08	-0.01	0.025	0.664 E	0.17	0.2	0.12	0
1	B12_01	-0.02	0.02	0.377 E	0.08	0.06	0	0
2	B12_02	0.03	0.02	0.100 E	0.05	0.04	-0.04	0
3	B12_03	0.02	0.019	0.443 E	0.05	0.04	-0.03	0
4	B12_04	-0.01	0.013	0.688 E	0.13	0.1	-0.01	0
5	B12_05	0.05	0.022	0.021 E	0.05	0.04	-0.05	0
6	B12_06	0.00	0.019	0.854 E	0.06	0.04	-0.02	0
7	B12_07	-0.05	0.016	0.005 E	0.12	0.1	0.02	0
8	B12_08	-0.01	0.023	0.571 E	0.06	0.04	-0.01	0
1	B13_01	0.00	0.024	0.854 E	0.14	0.11	0.01	0
2	B13_02	0.04	0.021	0.091 E	0.15	0.12	-0.01	0
3	B13_03	-0.01	0.021	0.600 E	0.14	0.11	0.01	0
4	B13_04	-0.06	0.024	0.008 E	0.12	0.1	0.04	0
5	B13_05	-0.07	0.024	0.006 E	0.13	0.11	0.04	0
6	B13_06	0.04	0.022	0.086 E	0.16	0.12	-0.01	0
7	B13_07	-0.04	0.025	0.151 E	0.15	0.11	0.02	0
8	B13_08	-0.02	0.024	0.368 E	0.14	0.12	0.01	0