

七、題目反應理論模式適合度研究

(一)單一向度性分析

本研究所採用的IRT模式假定所測量的能力有相當程度的單一向度性，因此，在使用IRT分析之前需探討學術性向測驗第一個主要因素所佔比重。

表4.30呈現有關學術性向測驗單一向度性檢驗之有關訊息。首先呈現各題目與第一個因素之相關係數，也就是每個題目在第一個因數之負荷量。此項係數與二系列相關有類似的意義。所有負荷量均為正值，除了乙式的第14及19以及甲式的第2，14及30題外，其餘均在.20以上，顯示所有題目均與學術性向測驗甲式或乙式所測量的主要特質（學術性向）有相當程度的關聯。再就第一個特徵值（ λ_1 ）所佔比率來看，乙式為26.088%，甲式為20.142%（以四分相關矩陣進行因素分析）， λ_1 所佔百分比均超過Reckase（1979）所提出的單一向度評鑑標準（ λ_1 佔20%以上）。另外，第一個特徵值（ λ_1 ）與第二個特徵值（ λ_2 ）之比，乙式為4.601，甲式為3.711，乙式超過Reckase所設定的評鑑標準（ λ_1/λ_2 大於4），甲式則稍低於此標準。此結果顯示學術性向測驗大致適合採用單一向度IRT模式來分析。

表4.30 國小學術性向測驗各題目在第一因素之負荷量
及前二因素之特徵值

題號	各題在第一因素的負荷量	
	六年級 乙 式	五年級 甲 式
1	.447	.398
2	.399	.151
3	.497	.510
4	.536	.449
5	.652	.585
6	.592	.607
7	.596	.550
8	.220	.448
9	.688	.424
10	.536	.475
11	.437	.323
12	.654	.477
13	.500	.434
14	.142	.129
15	.238	.272
16	.520	.283
17	.451	.365
18	.633	.271
19	.470	.236
20	.264	.350
21	.617	.684
22	.623	.515
23	.527	.485
24	.657	.469
25	.403	.273
26	.445	.434
27	.574	.491
28	.580	.585
29	.411	.246
30	.302	.138
31	.389	.239
32	.608	.305
33	.518	.395

題號	各題在第一因素的負荷量	
	六年級 乙 式	五年級 甲 式
34	.457	.300
35	.544	.496
36	.538	.363
37	.500	.419
38	.535	.507
39	.560	.303
40	.271	.399
41	.278	.503
42	.507	.446
43	.476	.569
44	.529	.458
45	.511	.423
46	.323	.461
47	.378	.317
48	.402	.582
49	.523	.378
50	.472	.459
51	.193	.590
52	.589	.566
53	.561	.530
54	.484	.458
55	.527	.477
56	.516	.513
57	.523	.579
58	.680	.274
59	.395	.390
60	.596	.342
最大特徵值	15.653	12.085
第二特徵值	3.402	3.256
最大特徵值 解釋的變異%	26.088	20.142
前二特徵值之比	4.601	3.711

(二)相等鑑別力分析

本研究採三參數logistic模式，主要原因有二：1. 學術性向測驗所有題目全部都是五選一選擇題，必然有某種程度的猜對率，因此 c 參數是必要的；2. 無論是甲式或乙式測驗，各題目的鑑別力 (r_{bis}) 均有相當程度的差異，而 r_{bis} 又與 a 參數有函數關係，因此不能將 a 視為常數，也就是 a 參數是必要的。由於 a 和 c 參數均是必要的，因此需採用三參數模式來分析。各題目鑑別力之差異情形，請參見附錄一所附各表之資料。

(三)適合度考驗

使用題目反應理論來發展測驗需考慮所選用的IRT模式與實際資料之適合度。一般IRT軟體多使用 χ^2 統計數來考驗模式之適合度。 χ^2 值很容易受到受試樣本大小的影響，不適合以有無達顯著水準來決定題目的適合度。最好是以值相對大小來判斷，如果某幾個題目的 χ^2 值明顯高出一般題目數倍，那這幾個題目可能就不太適合所選用的IRT模式。適合度不佳的題目，雖然也可以得到題目參數的估計值，但這些參數值的估計誤差必然很大，不能採用。

附錄五附有各題目適合度考驗之 χ^2 值，雖然有極少數題目之 χ^2 值明顯偏高，但整體而言，大致還可以接受。

結 語

本研究主要目的是在發展一套適用於國內國小高年級的學術性向測驗，並希望以該測驗為基礎，逐漸擴充題庫範圍，並將施測方式電腦適性化，以提昇分類決策的效率和品質。本研究報告屬於第一階段研究成果，已發展具有臺灣區常模的甲、乙兩式高年級學術性向測驗。

爲了要使學術性向測驗更能有效預測未來學業成就，本研究的學術性向測驗內容雖然與一般的IQ測驗有頗多相同之處，但在取材時，特別著重與學校的學習有關。受試在學術性向測驗的得分同時代表目前已學得的能力及天賦能力，不可認為完全是先天能力或是不可改變的能力。此外，在解釋學術性向測驗結果時還需注意本報告第二章最後一部份所列的解釋IQ分數之注意事項，方能儘可能避免對測驗分數作錯誤的解釋。