

以BLM分析琉球國中二年級學生英語科段考試題

~by Roger J. J. Lin

壹、研究設計

一、研究對象

本研究之對象以屏東縣琉球國中的八年級全體同學為樣本，針對其 97 學年度第一學期第二次段考英語科試題進行試題與受試者之能力分析，有效樣本為 117 人。

二、研究內容

本研究的內容為英語成就測驗，以研究者任職的學校 8 年級本學期第二次段考英語科考卷之選擇題共計 28 題為研究內容，進行分析。

三、資料分析

以 Bilog-mg 軟體估計試題之鑑別度、難度與猜中機率及受試者之能力值。

四、研究設計與程序

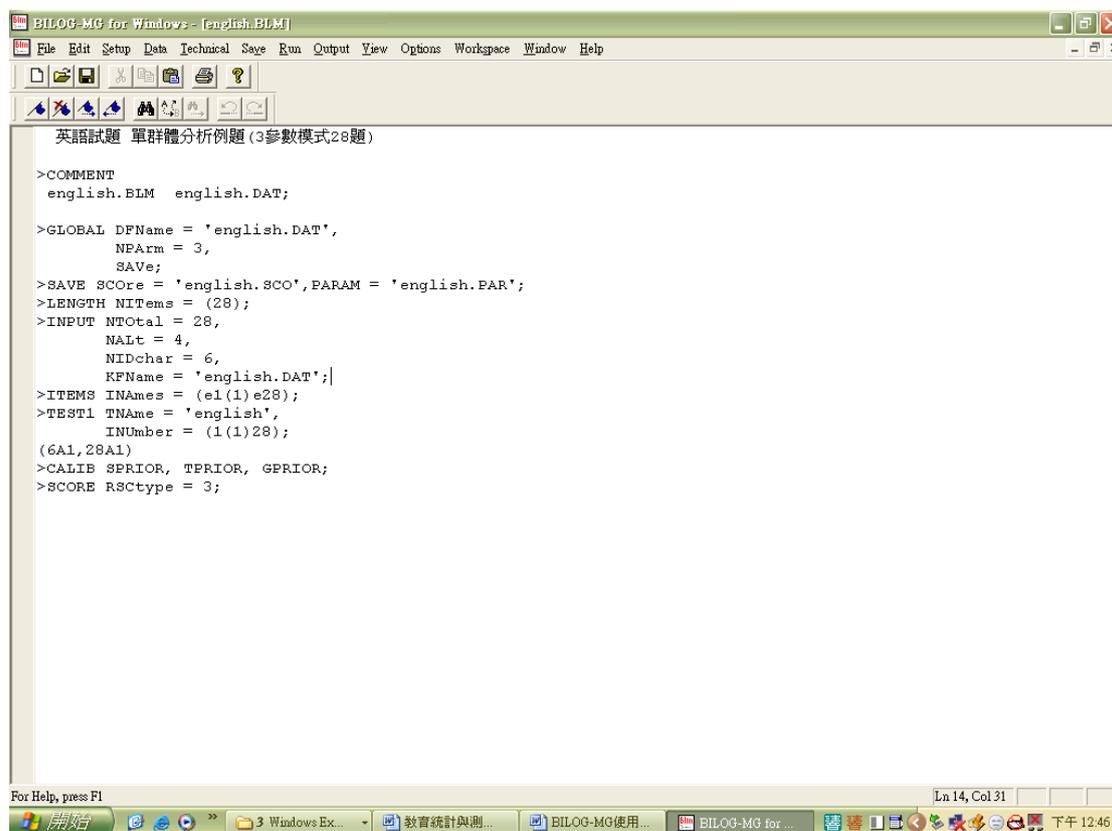
呈現出 Bilog-mg 軟體所得到試題分析結果之數據以及各項資料，做成本學期之研究報告。

貳、研究結果

一、測驗题目的統計結果資料(ITEM STATISTICS FOR SUBTEST ENGLISH)

測驗题目的統計程式資料，如下圖一所示。輸出報表包括前述的 PH1、PH2 及 PH3。圖二是 PH1 的報表，這是傳統題目分析所得結果。報表內依序為題號、題目名稱、考生人數、答對人數、答對率(PERCENT, PCT, 即傳統的難度 P)，對數勝算除以 1.7 「LOGIT/1.7」。LOGIT 是對數勝算， $\log_e(q/p)$ ， p 是答對率， q 是答錯率。題目與測驗相關 (ITEM*TEST CORRELATION) 包括 PEARSON (就是點二系列相關) 及二系列相關 (BISERIAL)。「LOGIT/1.7」就相當於前述等距量尺的難度指數 z ，從分析結果來看，二者極為接近。

從第一階段(English. PH1)的統計資料(傳統題目分析結果)來看,在全體 117 位受試者中,以 E17 提的答對率最高,達到 76.9%,答對人數為 90 人;以 E24 提的答對率最低,僅達到 23.1%,答對人數為 27 人。其統計相關資訊如下圖二所示。



```
BILOG-MG for Windows - [english.BLM]
File Edit Setup Data Technical Save Run Output View Options Workspace Window Help

英語試題 單群體分析例題 (3參數模式28題)

>COMMENT
english.BLM english.DAT;

>GLOBAL DFName = 'english.DAT',
NPArm = 3,
SAVE;
>SAVE SCORe = 'english.SCO',PARAM = 'english.PAR';
>LENGTH NITems = (28);
>INPUT NTOTal = 28,
NALt = 4,
NIDchar = 6,
KFName = 'english.DAT';|
>ITEMS INames = (e1(1)e28);
>TEST1 TName = 'english',
INUmber = (1(1)28);
(6A1,28A1)
>CALIB SPRIOR, TPRIOR, GPRIOR;
>SCORE RSctype = 3;

Ln 14, Col 31
For Help, press F1
開始 3 Windows Ex... 教育統計與測... BILOG-MG使用... BILOG-MG for ... 下午 12:46
```

圖一 測驗题目的統計程式資料

ITEM	NAME	#TRIED	#RIGHT	PCT	ITEM*TEST CORRELATION		
					LOGIT/1.7	PEARSON	BISERIAL
1	E01	117.0	76.0	65.0	-0.36	0.560	0.721
2	E02	117.0	79.0	67.5	-0.43	0.563	0.732
3	E03	117.0	79.0	67.5	-0.43	0.462	0.601
4	E04	117.0	66.0	56.4	-0.15	0.479	0.603
5	E05	117.0	64.0	54.7	-0.11	0.595	0.748
6	E06	117.0	79.0	67.5	-0.43	0.710	0.924
7	E07	117.0	71.0	60.7	-0.26	0.675	0.857
8	E08	117.0	66.0	56.4	-0.15	0.677	0.852
9	E09	117.0	80.0	68.4	-0.45	0.576	0.753
10	E10	117.0	63.0	53.8	-0.09	0.467	0.586
11	E11	117.0	64.0	54.7	-0.11	0.687	0.863
12	E12	117.0	78.0	66.7	-0.41	0.653	0.846
13	E13	117.0	77.0	65.8	-0.39	0.546	0.705
14	E14	117.0	76.0	65.0	-0.36	0.764	0.984
15	E15	117.0	86.0	73.5	-0.60	0.706	0.951
16	E16	117.0	81.0	69.2	-0.48	0.692	0.909
17	E17	117.0	90.0	76.9	-0.71	0.515	0.714
18	E18	117.0	78.0	66.7	-0.41	0.283	0.367
19	E19	117.0	82.0	70.1	-0.50	0.417	0.550
20	E20	117.0	50.0	42.7	0.17	0.557	0.703
21	E21	117.0	50.0	42.7	0.17	0.424	0.535
22	E22	117.0	66.0	56.4	-0.15	0.689	0.867
23	E23	117.0	45.0	38.5	0.28	0.319	0.406
24	E24	117.0	27.0	23.1	0.71	0.014	0.020
25	E25	117.0	68.0	58.1	-0.19	0.612	0.773
26	E26	117.0	64.0	54.7	-0.11	0.508	0.639
27	E27	117.0	59.0	50.4	-0.01	0.626	0.785
28	E28	117.0	50.0	42.7	0.17	0.564	0.711

圖二 第一階段(English. PH1)輸出檔—傳統題目分析結果

從第二階段(English. PH2)的統計資料 (IRT 題目參數) 來看, 各題項分析後的參數值, 如下圖三所示 (以 E1 到 E10 為例)。

輸出報表的第二部份(PH2)是題目的 IRT 參數(見圖三)。包括題號、截距 (INTERCEPT)及標準誤(S.E.)、斜率(SLOPE)及標準誤、門檻 (THRESHOLD)及標準誤、負荷量 (LOADING) 及標準誤、漸近線 (ASYMPTOTE)及標準誤、 χ^2 (CHISQ)及概率、自由度 (DF) "SLOPE" 即為 a 參數 (鑑別度), "THRESHOLD" 即為 b 參數 (難度), "ASYMPTOTE" 即 c 參數 (猜測率)。"INTERCEPT" 是 $-ab$ (即 $-(0.947 \times (1.758))$), "DISPERSN" 是 $1/SLOPE$ 。"LOADING" 是因素負荷量, 也就是每個試題與該測驗所要測量的最主要成份或因素 (只抽取一個因素) 之關係, 其數據接近傳統的二系列相關係數。

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
E01	-0.054 0.372*	2.043 0.585*	0.027 0.176*	0.898 0.257*	0.283 0.067*	3.6 (0.6089)	5.0
E02	0.389 0.235*	1.263 0.352*	-0.308 0.214*	0.784 0.218*	0.213 0.080*	7.3 (0.1983)	5.0
E03	-0.586 0.626*	2.483 1.369*	0.236 0.168*	0.928 0.511*	0.382 0.065*	3.7 (0.5917)	5.0
E04	-0.143 0.258*	1.006 0.243*	0.142 0.236*	0.709 0.171*	0.200 0.073*	3.8 (0.8076)	7.0
E05	-0.205 0.293*	1.383 0.366*	0.148 0.186*	0.810 0.215*	0.179 0.063*	6.0 (0.4203)	6.0
E06	0.947 0.573*	2.896 1.686*	-0.327 0.144*	0.945 0.550*	0.187 0.066*	0.7 (0.9570)	4.0
E07	0.087 0.310*	2.247 0.878*	-0.039 0.140*	0.914 0.357*	0.193 0.061*	2.0 (0.8443)	5.0
E08	-0.613 0.571*	3.472 1.613*	0.177 0.124*	0.961 0.446*	0.207 0.051*	2.2 (0.8243)	5.0
E09	0.435 0.256*	1.435 0.392*	-0.303 0.205*	0.820 0.224*	0.225 0.079*	5.4 (0.3711)	5.0
E10	-0.427 0.326*	1.111 0.358*	0.384 0.225*	0.743 0.240*	0.237 0.073*	7.4 (0.3885)	7.0

圖三 第二階段(English. PH2)的統計資料—IRT 題目參數(以 E1 到 E10 為例)

表一 第二階段輸出檔(所有試題 IRT 題目參數之基本統計)

PARAMETER	MEAN	STN DEV	
ASYMPTOTE	0.211	0.051	c參數(猜測率)
SLOPE	1.810	0.715	a參數(鑑別度)
LOG(SLOPE)	0.515	0.412	
THRESHOLD	0.112	0.569	b參數(難度)

報表(表一)右下角也提供所有題目的總 χ^2 (CHISQ), 總 DF 及實際顯著水準。另外, 在報表左下角提供最後一次疊代時試題參數最大改變值。表一是本研究樣本所有試題題目參數之平均數及標準差。a 的平均為 1.810, b 平均為 0.112, c 的平均為 0.211。

從第三階段(English. PH3)的統計資料(IRT 能力分數參數)來看, 各題項分析後的能力分數參數, 如下圖四、五、六、七所示(以本校八年級各班為例)。

GROUP WEIGHT	SUBJECT TEST	IDENTIFICATION		PERCENT	ABILITY	S. E.	MARGINAL PROB
		TRIED	RIGHT				
1 01 001	1.00 ENGLISH	28	28	100.00	1.8920	0.5458	0.030332
1 01 002	1.00 ENGLISH	28	22	78.57	0.4691	0.0524	0.000017
1 01 003	1.00 ENGLISH	28	28	100.00	1.8920	0.5458	0.030332
1 01 004	1.00 ENGLISH	28	5	17.86	-1.4811	0.4779	0.000000
1 01 005	1.00 ENGLISH	28	20	71.43	0.4669	0.0299	0.000011
1 01 006	1.00 ENGLISH	28	12	42.86	-0.3894	0.0313	0.000000
1 01 007	1.00 ENGLISH	28	7	25.00	-0.5503	0.3785	0.000000
1 01 008	1.00 ENGLISH	28	9	32.14	-0.4188	0.1634	0.000000
1 01 009	1.00 ENGLISH	28	16	57.14	-0.2049	0.3524	0.000000
1 01 010	1.00 ENGLISH	28	19	67.86	0.2435	0.3753	0.000000
1 01 011	1.00 ENGLISH	28	9	32.14	-0.6249	0.4254	0.000000
1 01 012	1.00 ENGLISH	28	22	78.57	0.4767	0.0955	0.000027
1 01 013	1.00 ENGLISH	28	9	32.14	-1.1684	0.4755	0.000000
1 01 014	1.00 ENGLISH	28	5	17.86	-1.3832	0.3796	0.000001
1 01 015	1.00 ENGLISH	28	4	14.29	-1.4608	0.4273	0.000005
1 01 016	1.00 ENGLISH	28	27	96.43	1.5268	0.3873	0.032582

圖四 第三階段輸出檔(八年一班 IRT 能力分數參數)

GROUP WEIGHT	SUBJECT TEST	IDENTIFICATION		PERCENT	ABILITY	S. E.	MARGINAL PROB
		TRIED	RIGHT				
1 02 001	1.00 ENGLISH	28	28	100.00	1.8920	0.5458	0.030332
1 02 002	1.00 ENGLISH	28	25	89.29	1.2717	0.2512	0.001303
1 02 003	1.00 ENGLISH	28	25	89.29	1.2472	0.2791	0.000195
1 02 004	1.00 ENGLISH	28	26	92.86	1.2835	0.2548	0.000519
1 02 005	1.00 ENGLISH	28	26	92.86	1.3673	0.2522	0.012799
1 02 006	1.00 ENGLISH	28	23	82.14	0.4951	0.1553	0.000032
1 02 007	1.00 ENGLISH	28	21	75.00	0.4668	0.0293	0.000004
1 02 008	1.00 ENGLISH	28	23	82.14	0.5151	0.1993	0.000011
1 02 009	1.00 ENGLISH	28	7	25.00	-1.4418	0.4403	0.000000
1 02 010	1.00 ENGLISH	28	24	85.71	0.8853	0.4286	0.000197
1 02 011	1.00 ENGLISH	28	9	32.14	-1.3308	0.4232	0.000000
1 02 012	1.00 ENGLISH	28	6	21.43	-1.4535	0.4327	0.000000
1 02 013	1.00 ENGLISH	28	2	7.14	-1.5426	0.4741	0.000007
1 02 014	1.00 ENGLISH	28	9	32.14	-1.0987	0.5363	0.000000
1 02 015	1.00 ENGLISH	28	2	7.14	-1.4929	0.4412	0.000017
1 02 016	1.00 ENGLISH	28	8	28.57	-1.2589	0.6188	0.000000
1 02 017	1.00 ENGLISH	28	6	21.43	-1.3870	0.4058	0.000000
1 02 018	1.00 ENGLISH	28	5	17.86	-1.1115	0.5974	0.000000

圖五 第三階段輸出檔(八年二班 IRT 能力分數參數)

Item ID	Subject	Total Items	Correct Items	Observed Score	IRT Parameter 1	IRT Parameter 2	IRT Parameter 3
1.00	ENGLISH	28	11	39.29	-0.4499	0.2489	0.000000
1 03 001							
1.00	ENGLISH	28	24	85.71	0.9368	0.4269	0.000103
1 03 002							
1.00	ENGLISH	28	27	96.43	1.5006	0.3859	0.001661
1 03 003							
1.00	ENGLISH	28	26	92.86	1.2835	0.2548	0.000519
1 03 004							
1.00	ENGLISH	28	26	92.86	1.1768	0.3452	0.000057
1 03 005							
1.00	ENGLISH	28	20	71.43	0.4661	0.0156	0.000004
1 03 006							
1.00	ENGLISH	28	22	78.57	0.4667	0.0272	0.000006
1 03 007							
1.00	ENGLISH	28	15	53.57	-0.3896	0.0510	0.000000
1 03 008							
1.00	ENGLISH	28	9	32.14	-0.5253	0.3296	0.000000
1 03 009							
1.00	ENGLISH	28	2	7.14	-1.5709	0.4910	0.000008
1 03 010							
1.00	ENGLISH	28	8	28.57	-1.3010	0.3618	0.000000
1 03 011							
1.00	ENGLISH	28	13	46.43	-0.3973	0.0842	0.000000
1 03 012							
1.00	ENGLISH	28	15	53.57	-0.0615	0.4174	0.000000
1 03 013							
1.00	ENGLISH	28	5	17.86	-1.5650	0.4889	0.000000
1 03 014							
1.00	ENGLISH	28	5	17.86	-1.3054	0.3676	0.000001
1 03 015							
1.00	ENGLISH	28	4	14.29	-1.5818	0.4948	0.000000
1 03 016							
1.00	ENGLISH	28	26	92.86	1.4041	0.2916	0.003932
1 03 017							
1.00	ENGLISH	28	25	89.29	1.2142	0.3015	0.001522
1 03 018							

圖六 第三階段輸出檔(八年三班 IRT 能力分數參數)

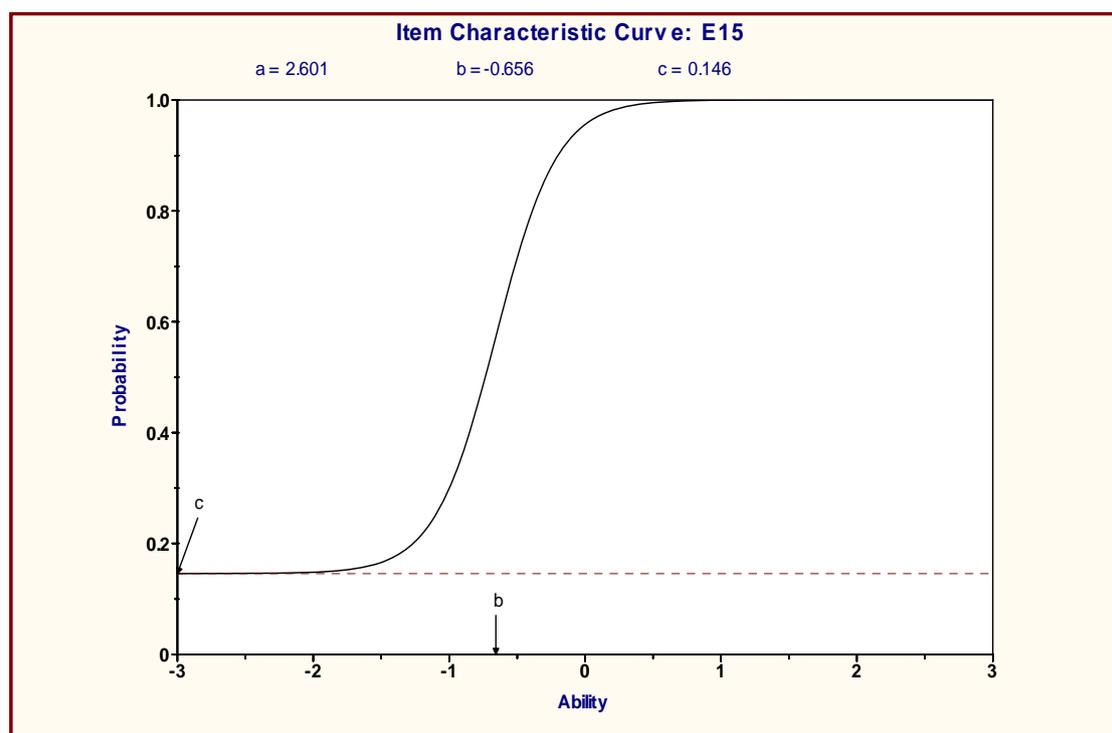
Item ID	Subject	Total Items	Correct Items	Observed Score	IRT Parameter 1	IRT Parameter 2	IRT Parameter 3
1 04 001							
1.00	ENGLISH	28	26	92.86	1.1856	0.3426	0.000087
1 04 002							
1.00	ENGLISH	28	25	89.29	0.8842	0.4292	0.000021
1 04 003							
1.00	ENGLISH	28	27	96.43	1.4005	0.3159	0.001649
1 04 004							
1.00	ENGLISH	28	20	71.43	0.4657	0.0151	0.000001
1 04 005							
1.00	ENGLISH	28	15	53.57	-0.3883	0.0462	0.000000
1 04 006							
1.00	ENGLISH	28	22	78.57	0.4747	0.0865	0.000012
1 04 007							
1.00	ENGLISH	28	20	71.43	0.4661	0.0159	0.000004
1 04 008							
1.00	ENGLISH	28	25	89.29	1.2856	0.2472	0.000461
1 04 009							
1.00	ENGLISH	28	21	75.00	0.4660	0.0119	0.000002
1 04 010							
1.00	ENGLISH	28	17	60.71	0.1004	0.4232	0.000000
1 04 011							
1.00	ENGLISH	28	4	14.29	-1.4581	0.4266	0.000002
1 04 012							
1.00	ENGLISH	28	9	32.14	-1.0547	0.5210	0.000000
1 04 013							
1.00	ENGLISH	28	10	35.71	-0.8932	0.5534	0.000000
1 04 014							
1.00	ENGLISH	28	8	28.57	-1.3312	0.3931	0.000000
1 04 015							
1.00	ENGLISH	28	7	25.00	-1.3575	0.3777	0.000000
1 04 016							
1.00	ENGLISH	28	27	96.43	1.5006	0.3859	0.001661
1 04 017							
1.00	ENGLISH	28	26	92.86	1.3673	0.2522	0.012799
1 04 018							
1.00	ENGLISH	28	26	92.86	1.2798	0.2494	0.002365

圖七 第三階段輸出檔(八年四班 IRT 能力分數參數)

輸出報表的第三部份(PH3)是受試者的得分狀況如上圖四、五、六、七所示(以本校八年級各班為例),包括受試的 ID 及加權分測驗、作答題數(TRIED)、答對題數(RIGHT)、答對率%(PERCENT)、能力(ABILITY)、標準誤(S.E.)及邊緣機率(PROB)。例如圖七第一位受試答對 26 題,答對率為 92.86,能力估計值為 1.1856,表示比平均數高 1.1856 個標準差,能力估計誤差為 0.3426 個標準差。

三個階段全部分析完畢之後,接下來執行"Plot"(繪圖功能,參見圖 1"RUN"選單的"Plot")。執行"Plot"可得到「題目特徵曲線」、「題目訊息曲線」、「測驗訊息曲線」、「測驗標準誤曲線」、「所有題目試題特徵曲線矩陣」、「由答對率預測能力估計值之回歸線」、「能力分布與常態分布比較」等。本研究僅以其中部分題目為例,撰寫報告。

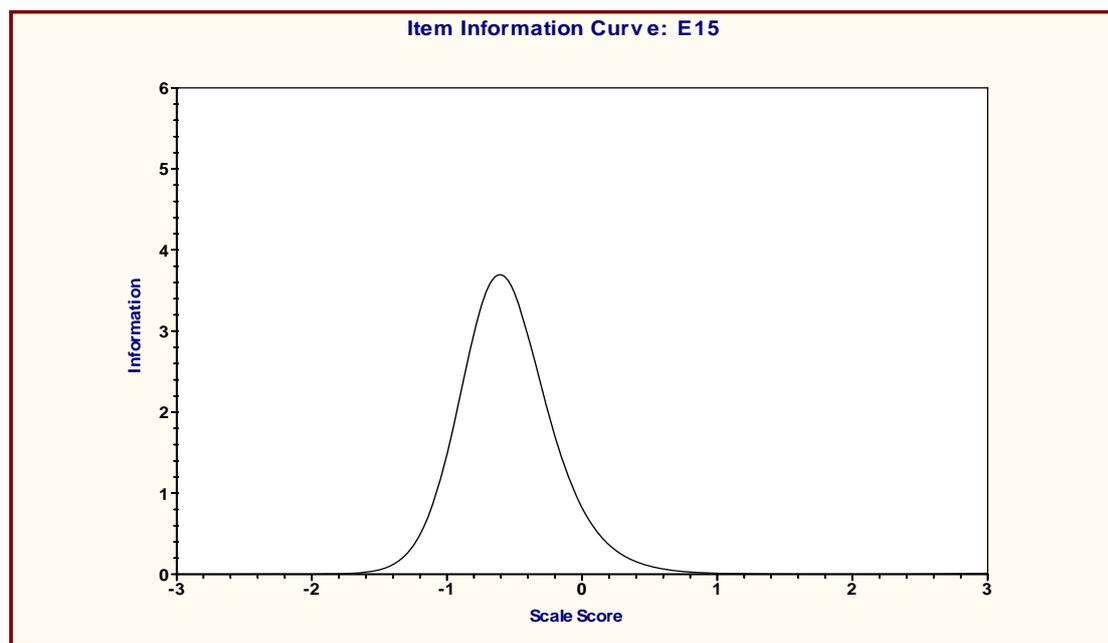
選 ICC 至 Histogram 可分別得到圖八至圖十三。以下圖八為例,是第 15 題試題特徵曲線,也就是依據該題的 a,b,c 所繪製的函數圖。圖八的橫座標是能力,縱座標是答對率,由於該題難度低,因此特徵曲線偏左。除了函數圖之外,圖中也列出該題的 a,b,c 值,以及 b,c 所在位置。



圖八 第 15 題試題特徵曲線

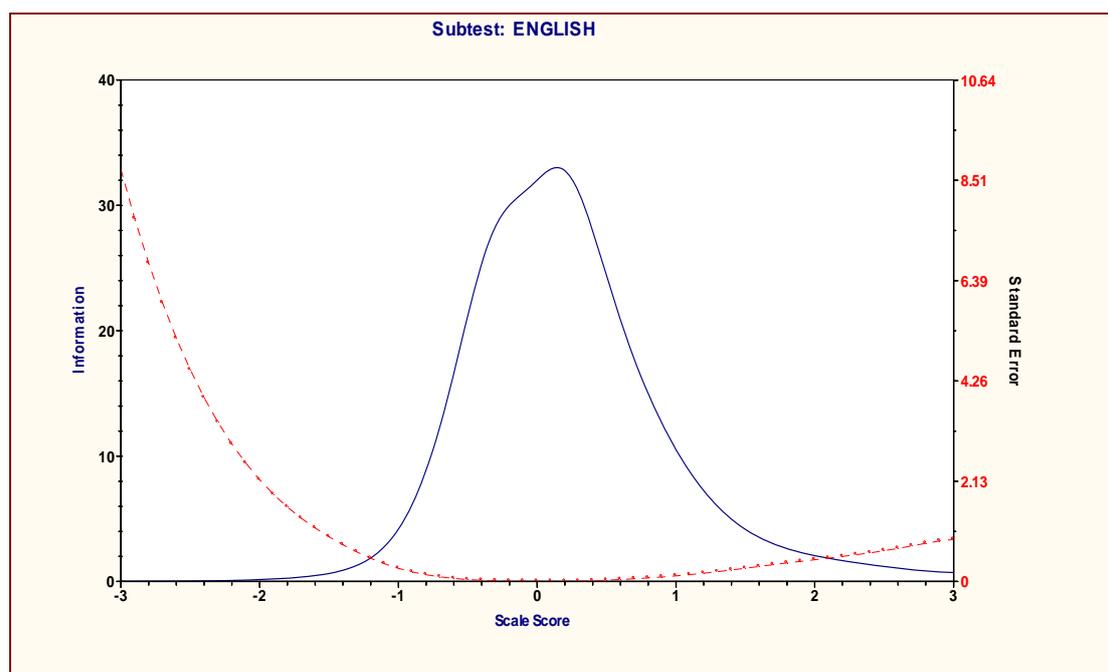
下圖九是第 15 題的題目訊息曲線,橫座標是能力,縱座標是訊息量。訊息量最大的地方,就是題目特徵曲線的「斜率」最大的地方。由於第 15 題的 $b = -0.656$, $c = 0.146$ 。如果 $c = 0$,那該題的訊息量最高點的位置是在橫座標(能

力值) = -0.656 處。本題之 $c > 0$ ，因此訊息量最高點在比 -0.656 稍高之處。



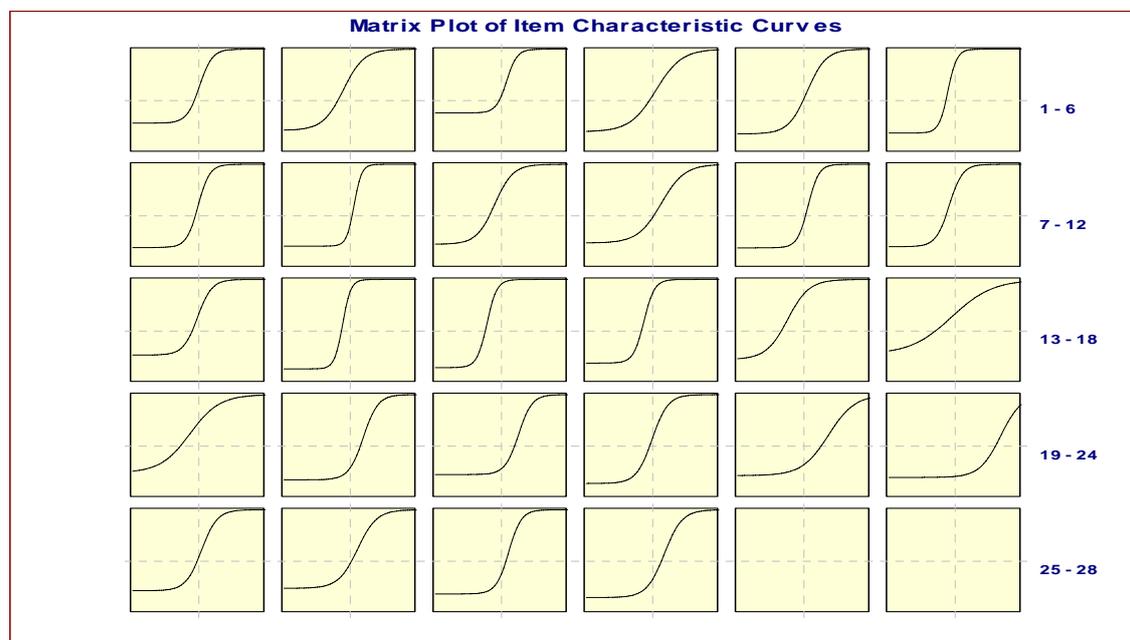
圖九 第 15 題試題訊息曲線

下圖十是「全測驗」的訊息曲線 (實線) 及標準誤曲線 (虛線)，圖形左邊的量尺是訊息量，圖形右邊的量尺是標準誤的數值。全測驗的訊息曲線，是所有個別题目的訊息曲線之總和。「訊息量最大的地方就是標準誤最小的地方」，二者之關係已如前述內容。



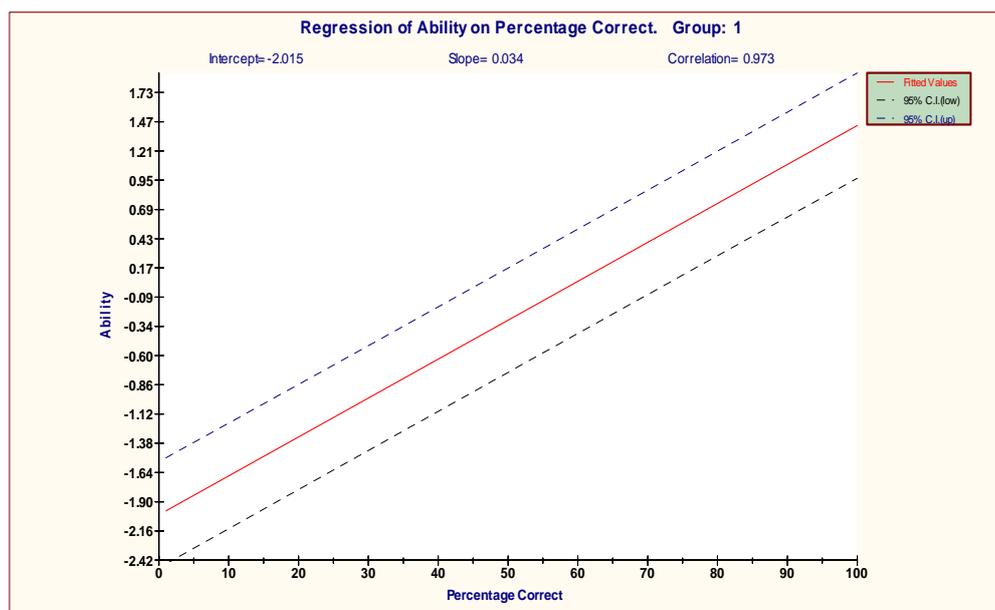
圖十 分數運算測驗訊息及標準誤曲線

下圖十一是以矩陣形式列出所有題目試題特徵曲線，使用者可以同時觀看並比較所有題目之特徵。例如第 18 題之曲線斜率顯然較其他題目低 ($a = 0.629$ 最小，表示鑑別度低)；第 15、17 題之曲線偏左邊，難度明顯低於其他曲線較偏右之題目；第 23、24 題之曲線偏右上角，難度明顯高於其他曲線較偏右之題目。另一方面，第 8、14 題之鑑別度，顯然較其他題目高。



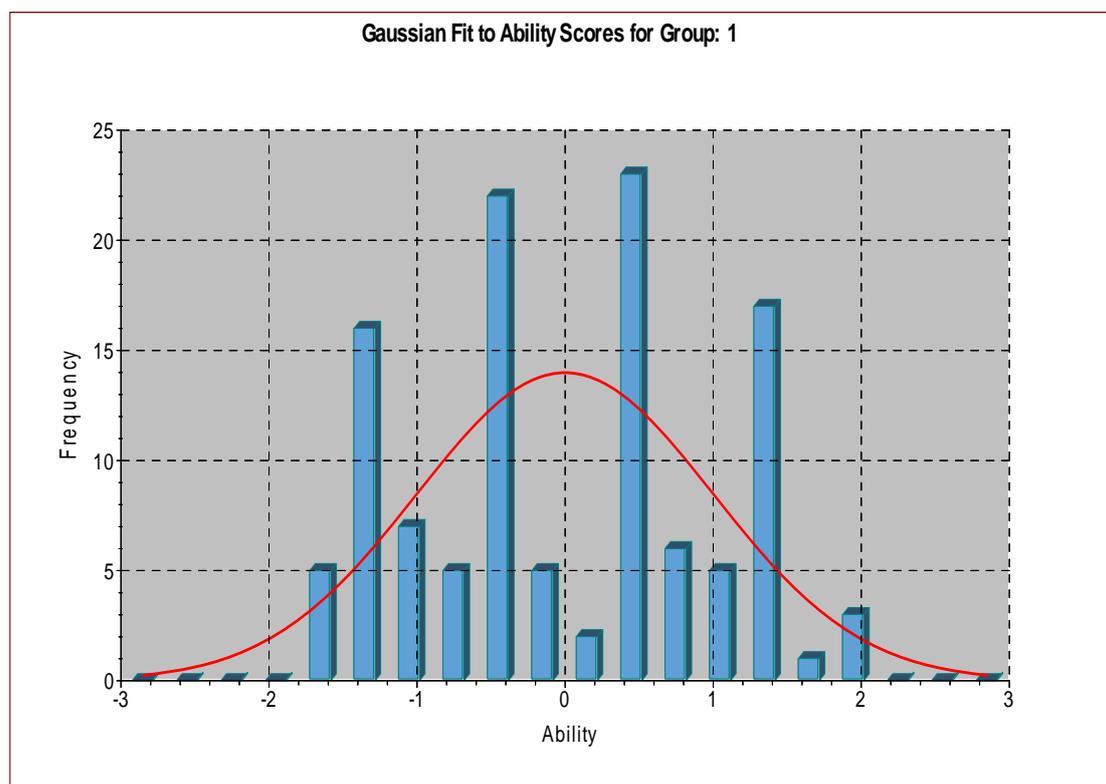
圖十一 英語測驗所有題目試題特徵曲線矩陣

下圖十二是由「答對率」預測「能力估計值」之迴歸線，中心之實線是迴歸線 (預測值或估計值)，上下各有一條曲線，分別是迴歸線 95%信賴水準之上下界線。圖中也列出迴歸線之截距及斜率，以及答對率與能力估計值之相關係數。



圖十二 英語測驗由答對率預測能力估計值之迴歸線

下圖十三是受試樣本能力分布 (直方圖) 與常態分布 (曲線) 之比較。如果二者很接近, 那有無"RSC=3"對題目參數估計值之影響很小(如果 BILOG-MG 指令的"SCORE"有加入"RSC=3", 即">SCORE,RSC = 3;"那在第 3 階段的報表會將「受試樣本能力分數」重新量尺化為平均數 = 0, 標準差 = 1。如未指定"RSC = 3", 則是將「母群能力分布標準」化為平均數 = 0, 標準差 = 1。將受試樣本能力分數重新量尺化為平均數 = 0, 標準差 = 1, 也會影響試題的"a"及"b"參數)



圖十三 受試樣本能力分布 (直方圖) 與常態分布 (曲線) 之比較

參考資料

吳裕益 (2004)。測驗理論。高雄師大教育系博士班 97 學年度上學期「教育統計與測驗專題研究」上課講義，未出版。