

台灣教育長期追蹤資料庫

心理計量報告：TEPS 2001 分析能力測驗

楊孟麗 譚康榮 黃敏雄

民國九十年第一波高中樣本

(高中/高職/五專/國中)

第一版

(民國九十二年十二月)¹

¹ 感謝中央研究院統計所研究員劉長萱教授、國立台灣師範大學科學教育研究所譚克平副教授及國立中正大學王文中教授對本報告初稿的意見與建議。

目 錄

第一章 導言.....	1
第二章 測驗的設計過程與原則.....	3
第三章 各題項的古典測驗理論指標.....	5
學生是否都可以在限定的時間內做完 (UNSPEEDEDNESS)	5
答對率 (P_i , 也稱難度)	6
題項與測驗總分的點二系列相關係數 (POINT-BISERIAL CORRELATION, r_{pbi})	6
3.1 高中職五專的結果	7
3.2 國中的結果	10
第四章 項目反應理論 (IRT) 的分析	14
IRT (項目反應理論) 簡介	14
單參數 IRT	15
二參數 IRT	17
三參數 IRT	20
受試者能力估算法.....	22
EAP 的實證信度 (EMPIRICAL RELIABILITY)	22
4.1 高中職五專樣本的結果	23
4.2 國中樣本的結果	26
模型適合度.....	28
4.3 高中職五專樣本的模型適合度	29
4.4 國中樣本的模型適合度	31
第五章 題目的內部效度 (internal validity)	39
DIF: 受試者其他的特質是否影響答對機率	39
分析方法.....	39
5.1 高中職五專	41
5.2 國中.....	43
學生的答題意願.....	47
附錄一 高中職五專 IRT 分析結果圖示	49
附錄二 國中 IRT 分析結果圖示	81

圖 表

表 1-1、參加高中職五專與國中分析能力測驗的學校數、班級數、及人數統計表.....	1
表 3-1、高中職五專學生在五種領域下，最後一至五題的沒有回答比率（%）.....	6
表 3-2、國中學生在五種領域下，最後一至五題的沒有回答比率（%）.....	6
表 3-3、高中職五專，綜合分析能力題的古典測量指標.....	8
表 3-4、高中職五專，一般分析能力題的古典測量指標.....	9
表 3-5、高中職五專，數學分析能力題的古典測量指標.....	9
表 3-6、高中職五專，數學或數字型分析能力題的古典測量指標.....	10
表 3-7、國中綜合分析能力題的古典測量指標.....	11
表 3-8、國中一般分析能力題的古典測量指標.....	12
表 3-9、國中數學分析能力題的古典測量指標.....	12
表 3-10、國中數學或數字型分析能力題的古典測量指標.....	13
圖 4-1、1PL 模式下的題項答對曲線圖與題項訊息圖.....	16
圖 4-2、1PL 模式下的題項答對曲線圖與題項訊息圖.....	16
圖 4-3、1PL 模式下的題項答對曲線圖與題項訊息圖.....	17
圖 4-4、2PL 模式下的題項答對曲線圖與題項訊息圖 --- 鑑別度 = .739.....	18
圖 4-5、2PL 模式下的題項答對曲線圖與題項訊息圖 --- 鑑別度 = .422.....	19
圖 4-6、2PL 模式下的題項答對曲線圖與題項訊息圖 --- 鑑別度 = 1.098.....	19
圖 4-7、3PL 的題項答對曲線圖與題項訊息圖.....	20
圖 4-8、3PL 模式下的題項答對曲線圖與題項訊息圖 --- 鑑別度 = 1.297.....	21
圖 4-9、3PL 模式下的題項答對曲線圖與題項訊息圖 --- 鑑別度 = .753.....	21
表 4-1、各分項測驗中，1PL 模式下，各題項難度估計值 b_i 之基本描述.....	23
表 4-2、各分項測驗中，3PL 模式下，各題項難度估計值 b_i 之基本描述.....	24
表 4-3、各分項測驗中，3PL 模式下，各題項猜對機會 c_i 之基本描述.....	24
表 4-4、各分項測驗中，3PL 模式下，各題項鑑別度 a_i 之基本描述.....	24
表 4-5、1PL 模式下與 3PL 模式下難度的相關係數.....	24
表 4-6、各分項測驗下能力估計值間的相關係數.....	25
表 4-7、1PL 模式下的實證信度.....	25
表 4-8、3PL 模式下的實證信度.....	25
表 4-9、國中各分項測驗中，1PL 模式下，各題項難度估計值 b_i 之基本描述.....	26
表 4-10、國中各分項測驗中，3PL 模式下，各題項難度估計值 b_i 之基本描述.....	26

表 4-11、國中各分項測驗中，3PL 模式下，各題項猜對機會 C_i 之基本描述	26
表 4-12、國中各分項測驗中，3PL 模式下，各題項鑑別度 a_i 之基本描述	26
表 4-13、國中部分，1PL 模式下與 3PL 模式下難度的相關係數	27
表 4-14、各分項測驗下能力估計值間的相關係數	27
表 4-15、國中 1PL 模式下的實證信度	28
表 4-16、國中 3PL 模式下的實證信度	28
圖 4-10 高中職五專語文分析能力第二題 理論上的答對機率曲線和實際答對機率的比較圖	30
圖 4-11 高中職五專一般分析能力第二十四題 理論上的答對機率曲線和實際答對機率的比較圖	31
圖 4-12 國中數學分析能力第八題 理論上的答對機率曲線和實際答對機率的比較圖	32
圖 4-13 國中一般分析能力第二十一題 理論上的答對機率曲線和實際答對機率的比較圖	33
圖 4-14 國中一般分析能力第十一題 理論上的答對機率曲線和實際答對機率的比較圖	34
圖 4-15 國中一般分析能力第三題 理論上的答對機率曲線和實際答對機率的比較圖	35
圖 4-16 國科學分析能力第一題 理論上的答對機率曲線和實際答對機率的比較圖	36
圖 4-17 科學分析能力第八題 理論上的答對機率曲線和實際答對機率的比較圖	37
圖 4-18 語文分析能力第十四題 理論上的答對機率曲線和實際答對機率的比較圖	38
圖 5-1 控制總答對題數後，高中職五專男女生在各題答對機率之差異 (b_{1j})	41
圖 5-2 控制總答對題數後，國中男女生在各題答對機率之差異 (b_{1j})	43
圖 5-3 控制總答對題數後，國中樣本中，都市或城鎮的學生，相對於鄉村學生，在各題答對機率上之差異 (b_{1j})	45
圖 5-4 控制總答對題數後，國中樣本中，都市的學生，相對於城鎮或鄉村學生，在各題答對機率上之差異 (b_{1j})	46
表 5-1、高中職五專學生在 73 題中答對題數的累積比率 (%)	47
表 5-2、國中學生在 71 題中答對題數的累積比率 (%)	47
圖 附一-1、綜合分析能力題 1PL 模式下，學生能力分佈估計值分佈長條圖	50
圖 附一-2、綜合分析能力題 1PL 模式下的測驗訊息圖	51
圖 附一-3、綜合分析能力題 1PL 模式下，各題項的答對機率曲線圖	52
圖 附一-4、綜合分析能力題 答對題數與 1PL 模式下能力估計值的散佈圖， $R = .997$	53
圖 附一-5、綜合分析能力題 3PL 模式下，學生能力估計值分佈長條圖	54
圖 附一-6、綜合分析能力題 3PL 模式下的測驗訊息圖	55
圖 附一-7、綜合分析能力題 3PL 模式下，各題項的答對機率曲線圖	56
圖 附一-8、綜合分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $R = .984$	57
圖 附一-9、一般分析能力題，1PL 模式下的能力估計值分佈長條圖	58

圖 附一-10、	一般分析能力題 1PL 模式下的測驗訊息圖	59
圖 附一-11、	一般分析能力題，1PL 模式下各題項的答對機率曲線圖	60
圖 附一-12、	一般分析能力題，答對題數與 1PL 模式能力估計值的散佈圖， $R = .998$	61
圖 附一-13、	一般分析能力題，3PL 模式下的能力估計值分佈長條圖	62
圖 附一-14、	一般分析能力題，3PL 模式下的測驗訊息圖	62
圖 附一-15、	一般分析能力題，3PL 模式下各題項的答對機率曲線圖	63
圖 附一-16、	一般分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $R = .952$	64
圖 附一-17、	數學分析能力題，1PL 模式下的能力估計值分佈長條圖	65
圖 附一-18、	數學分析能力題，1PL 模式下的測驗訊息圖	66
圖 附一-19、	數學分析能力題，1PL 模式下的各題項答對機率圖	67
圖 附一-20、	數學分析能力題 1PL 模式下，能力估計值與答對比率的迴歸線， $R = .996$	68
圖 附一-21、	數學分析能力題，3PL 模式下的能力估計值分佈長條圖	69
圖 附一-22、	數學分析能力題，3PL 模式下的測驗訊息圖	70
圖 附一-23、	數學分析能力題，3PL 模式下的各題項答對機率曲線圖	71
圖 附一-24、	數學分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $R = .971$	72
圖 附一-25、	數學或數字型分析能力題，1PL 模式下的能力估計值分佈長條圖	73
圖 附一-26、	數學或數字型分析能力題，1PL 模式下的測驗訊息圖	74
圖 附一-27、	數學或數字型分析能力題，1PL 模式下的各題項答對機率圖	75
圖 附一-28、	數學或數字型分析能力題，答對題數與 1PL 能力估計值的散佈圖， $R = .992$	76
圖 附一-29、	數學或數字型分析能力題，3PL 模式下的能力估計值分佈長條圖	77
圖 附一-30、	數學或數字型分析能力題，3PL 模式下的測驗訊息圖	78
圖 附一-31、	數學或數字型分析能力題，3PL 模式下各題項的答對機率曲線圖	79
圖 附一-32、	數學或數字型分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $R = .979$	80
圖 附二-1、	國中綜合分析能力題 1PL 模式下，學生能力分佈估計值分佈長條圖	82
圖 附二-2、	國中綜合分析能力題 1PL 模式下的測驗訊息圖	83
圖 附二-3、	國中綜合分析能力題 1PL 模式下，各題項的答對機率曲線圖	84
圖 附二-4、	國中綜合分析能力題 答對題數與 1PL 模式下能力估計值的散佈圖， $R = .999$	85
圖 附二-5、	國中綜合分析能力題 3PL 模式下，學生能力估計值分佈長條圖	85
圖 附二-6、	國中綜合分析能力題 3PL 模式下的測驗訊息圖	86
圖 附二-7、	國中綜合分析能力題 3PL 模式下，各題項的答對機率曲線圖	87
圖 附二-8、	國中綜合分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $R = .984$	88
圖 附二-9、	國中一般分析能力題，1PL 模式下的能力估計值分佈長條圖	88
圖 附二-10、	國中一般分析能力題 1PL 模式下的測驗訊息圖	89
圖 附二-11、	國中一般分析能力題，1PL 模式下各題項的答對機率曲線圖	90
圖 附二-12、	國中一般分析能力題，答對題數與 1PL 模式能力估計值的散佈圖， $R = 1.000$	91
圖 附二-13、	國中一般分析能力題，3PL 模式下的能力估計值分佈長條圖	91
圖 附二-14、	國中一般分析能力題，3PL 模式下的測驗訊息圖	92
圖 附二-15、	國中一般分析能力題，3PL 模式下各題項的答對機率曲線圖	93

圖 附二-16、	國中一般分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $R = .934$	94
圖 附二-17、	國中數學分析能力題，1PL 模式下的能力估計值分佈長條圖	94
圖 附二-18、	國中數學分析能力題，1PL 模式下的測驗訊息圖	95
圖 附二-19、	國中數學分析能力題，1PL 模式下的各題項答對機率圖	96
圖 附二-20、	國中數學分析能力題 1PL 模式下，能力估計值與答對比率的迴歸線， $R = .997$	97
圖 附二-21、	國中數學分析能力題，3PL 模式下的能力估計值分佈長條圖	97
圖 附二-22、	國中數學分析能力題，3PL 模式下的測驗訊息圖	98
圖 附二-23、	國中數學分析能力題，3PL 模式下的各題項答對機率曲線圖	99
圖 附二-24、	國中數學分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $R = .975$	100
圖 附二-25、	國中數學或數字型分析能力題，1PL 模式下的能力估計值分佈長條圖	100
圖 附二-26、	國中數學或數字型分析能力題，1PL 模式下的測驗訊息圖	101
圖 附二-27、	國中數學或數字型分析能力題，1PL 模式下的各題項答對機率圖	102
圖 附二-28、	數學或數字型分析能力題，答對題數與 1PL 能力估計值的散佈圖， $R = .997$	103
圖 附二-29、	國中數學或數字型分析能力題，3PL 模式下的能力估計值分佈長條圖	103
圖 附二-30、	國中數學或數字型分析能力題，3PL 模式下的測驗訊息圖	104
圖 附二-31、	數學或數字型分析能力題，3PL 模式下各題項的答對機率曲線圖	105
圖 附二-32、	國中數學或數字型分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $R = .980$	106

第一章 導言

「台灣教育長期追蹤資料庫」(以下簡稱「資料庫」)是由中央研究院、教育部與國科會共同主辦，中研院社會所研究員張荳雲主持的研究。主要是蒐集國高中階段的學生各方面的成長情形(包括認知能力、心理健康、行為)，和他們所處的各种環境的資料(包括家庭、班級與學校的氣氛及其所提供的資源)。資料庫藉著長期持續觀察同一批學生，提供學者各種資訊，以了解在大環境(教育政策與傳統文化)及小環境(家庭、教室及學校)的影響下，有就學的台灣孩子的成長軌跡。因此它不僅針對學生蒐集資料，也詢問家長及導師、國文、英文及數學三科的老師有關學生在家及在學校所處的環境。有關本資料庫較詳盡的介紹請見張荳雲(2002)。

「資料庫」總共調查了約四萬名國中、高中職及五專的學生，家長也約四萬名，大約有八千多位老師也受到訪問。這些學生則散佈在將近六百三十所學校中。而這些樣本學校及學生和家長的選取方法，在國中與高中(職)的部份，主要是先將學校按照行政區域、公私立學校及城鄉分佈等面向加以分類(也就是所謂的「分層」)，從各種類別中抽出學校後，再從每所學校平均抽出四個班級，最後從每個班級抽出大約十五名學生；在五專部分，則因五專學校數目不多，僅按照地理區域(南北)及公私立分層，抽出二十七所(完成二十六所)。而調查期間任教於該班的導師及國英數三科的科任老師，則成為資料庫中的老師樣本。扣除了拒答的學校，拒答或填答題數過少¹的學生之後，本篇報告分別使用了19232位高中職五專及19957位國中學生的測驗資料。表1-1列出各學程所抽樣的班級數與人數。

本篇報告所分析的資料裡，其中70%的隨機樣本，已經在中央研究院蔡元培人文社會科學研究中心之調查研究專題中心所建置的「學術調查研究資料庫」(<http://srda.sinica.edu.tw>)釋出，釋出的資料還包括使用者手冊及國高中的題項對照表。研究者可經由申請成為會員取得這些資料。

表1-1列出國高中職及五專所抽樣的班級數與人數。本篇報告在扣除了拒答的學校，拒答或填答題數過少²的學生之後，分別使用了19232位高中職五專的學生，及19957位國中學生的測驗資料。其中因填答題數過少而不列入分析者，高中有23人，高職25人，五專20人，國中27人。

表1-1、參加高中職五專與國中分析能力測驗的學校數、班級數、及人數統計表

	高中	高職	五專	國中
學校數	184(67.65%)	62(22.79%)	26(9.56%)	333
班級數	704(66.42%)	260(24.53%)	96(9.06%)	1244
學生數	10764(55.77%)	4061(21.04%)	4475(23.19%)	19984

¹所謂填答題數過少，在操作上定義為：四種能力測驗中，任一種分項測驗完全沒有作答。

「資料庫」針對學生所蒐集的資料分為兩部分，一部份是學生問卷，另一部份則是綜合分析能力測驗。綜合分析能力測驗，顧名思義，是以測驗的方式取得學生在分析能力方面的資料。本篇報告的目的是說明第一波的高中職、五專及國中學生所使用的的能力測驗，包括其設計構想及實證特質。因此，本篇報告將先說明該測驗的設計過程與原則；接著從各種角度探討測驗的特質，包括古典測驗理論中的參考指標，新測驗理論(也就是「項目反應理論」(Item Response Theory, 簡稱IRT))的各種指標。最後，並探討測驗的內部效度(internal validity)。

第二章 測驗的設計過程與原則

綜合分析能力測驗的設計工作，開始於民國八十九年初。「資料庫」召開了十幾次測驗諮詢小組會議，在會議中確立測驗的測量原則、方式與標準，並確定測驗必須反映學生的學習成就及學習的成長情形，因此決定：測驗內容必須測量學生解決問題的能力，而非一般的學科成就測驗。最後，測驗小組擬定了以下大致的測驗原則：

在內容方面

1. 測驗的目的在瞭解學生一般的學習能力，因此應避免死背記憶題，而且要強調學生思考活用知識的能力。
2. 測驗題項應該盡量跳脫課程的限制(也就是盡量使題目是 curriculum free)，避免涉及過於專門的領域，盡可能使所有題項的內容都是各種學校的學生有接觸過的。
3. 測驗題項必須多元化，包含語文，自然科學與數學，及類似一般智力測驗的題型，以瞭解學生在各種領域下思考及解決問題的綜合分析能力。

在形式方面

4. 在資料處理成本的考量之下，所有題項都以選擇題的方式出現，每個題項都是四個選項，以利電腦掃描讀卡的作業。
5. 測驗題項必須包含各種難度，並且學生都可以在限定的時間內做完。
6. 題項的呈現，先列出較容易者，較難者出現於後，以免使學生一開始就遭受挫折，而無意繼續嘗試解題，因而影響測驗的效度。

方向確定以後，測驗小組即著手建立測驗題庫。題庫的建立方法，包括尋找並改編國外大型測驗或調查曾使用過的題目，及聘請國內相關科目的教師及專家擬定題目。改編國外曾使用過的題目之好處是，這些題目已經過謹慎的設計及實證的檢驗，具有良好的信度與效度。所有題項都經過該科目(領域)專家(高中教師)的多次修改或刪減，評量每個題項適合哪一個年級程度，並做最後的鑑定。民國八十九年秋並將所有的題項向將近二千名國高中生施測，進行預試。預試的題目中，屬於一般分析能力的題項有 48 題；科學領域的題項有 47 題；數學領域的題項有 79 題；中文領域的題項有 42 題；英文領域的題項則有 19 題。所獲得的結果，則以傳統的測驗理論指標及 IRT 模式觀察各個題項的「實際表現」。再依據分析結果將「表現不佳」的題項請專家再次檢討，瞭解原因，修改或刪除。最後，所有預試過的題目都保留在題庫中，且大多數的題目之選項次序或數字都經過變動。

這些不同領域的測驗題庫中，比較特殊的是所謂的「一般分析能力」題庫。題庫中的「一般分析能力」題，所測量的分析能力包含三方面的智能：分析能力、生活應用的能力及創造力；而每一種智能的測量方式都又各以三種呈現方式測量：語文型、圖像型及數字型。

民國九十年底開始第一波高中高職、五專及國中學生的資料蒐集。由於時間的限制，只

能有約 100 分鐘的測驗時間，於是在高中職五專部分共安排了 75 道測驗題，時間 94 分鐘；國中則安排 71 題，時間 88 分鐘。各分項測驗的題數與測驗時間如下：

高中職五專

一般分析能力	27 題	40 分鐘
英文	6 題	7 分鐘
數學	20 題	27 分鐘
中文	10 題	10 分鐘
自然科學	12 題	10 分鐘

國中

一般分析能力	27 題	45 分鐘
數學	20 題	27 分鐘
語文	14 題	8 分鐘
自然科學	10 題	8 分鐘

所有題項的計分方法都是答對該題得一分，答錯得零分，不倒扣；施行測驗時，訪員則明確表達請學生不要猜題。國中部分的語文題包括 10 題中文與 4 題英文。此外，高中職五專的題目中，英文領域裡的 5 個題項，數學領域的 4 個題項，及科學領域的 2 個題項都是從別的管道獲得，而非出自題庫。國中題目中英文領域裡的 4 題，也是從別的管道獲得，並非出自題庫。這些新加入的題目都改編自國外大型計畫的題目或出自專業機構之手。

本篇報告的內容即是以上九十學年度學生樣本資料的分析結果。分析的方式，除了將所有題目看作一項綜合分析能力測驗之外，也將一般分析能力的題目及數學領域下的題目（數學領域的 20 題，及一般分析能力測驗中 9 道數字型的題目）都當作分項測驗，另行分析。而由於語文及自然科學領域的題項過少，將不另行分析；因為若將其視為分項測驗，可能會因題項過少而有測驗信度過低之虞。最後，由於高中職五專的一般分析能力測驗中，生活應用能力部分的文字題型中有兩題涉及社會譚許度（social desirability）且具爭議性，決定不計入本報告的分析範圍；因此本篇報告中，高中職五專部分之綜合分析能力題數共有 73 題，一般分析能力題數則有 25 題，其他題數與上述相同。

第三章 各題項的古典測驗理論指標

資料庫釋出的分析能力估計值來自分析能力測驗的四種組合，而國中與高中職五專之間的差異只有在所使用的題數上稍有差異：(1)綜合分析能力：所使用的測驗包括一般分析能力測驗 (curriculum-free analytical ability subtest)、數學測驗、語文測驗，及科學測驗。高中職五專共 73 題，國中有 71 題；(2)一般分析能力：使用一般分析能力測驗，高中職五專有 25 題，國中有 27 題；(3)數學分析能力：數學測驗，高中職五專及國中兩者皆是 20 題；(4)數學或數字型分析能力：除了數學測驗的 20 題外，另加入一般分析能力測驗中的 9 題數字型題項，故共 29 個題項，國高中都相同。每種組合下都有三種能力指標，其一是答對題數，第二與第三種則是項目反應理論 (Item Response Theory, 簡稱 IRT) 模式下的能力估計值，分別是以單參數 (1-item parameter, 簡稱 1PL) 及三參數 (3-item parameter, 簡稱 3PL) 的模式估計而得。之所以使用 1PL 與 3PL，而不使用 2PL，是因為 3PL 模型是 IRT 模式中最複雜的模式 (比 2PL 模型多估算了一個項目參數)，1PL 模型則是 IRT 中最簡約的模式 (比 2PL 模型少估算了一個項目參數)，但事實上兩者所估計的受試者能力相關頗高 (見第四章)。因此我們認為，從受試者能力估計值的觀點而言，三者其實差異很小，但為了滿足研究者的好奇心，我們分別作了 1PL 和 3PL 的分析，並在第四章稍微比較兩者的結果。

「能力估計值」的計算方式將在第四章討論。本章先探討測驗是否達到「學生都可以在限定的時間內做完」的目標，再報告這些題項的古典測驗理論使用的指標，包括：每一題的答對率 (難度) 及每一題與該測驗總分的點二系列相關係數 (Pearson correlation, 也是 point biserial correlation)。報告這些指標時，我們先介紹它們的含義後，再依照各分項測驗列出測驗題項在這些指標的值。高中職五專的結果放在 3.1，國中的結果在 3.2。

學生是否都可以在限定的時間內做完 (Unspeededness)

我們以學生是否回答該項測驗的最後五題，作為觀察學生是否能在限定的時間作完所有題目的依據。然而，我們承認，這個依據其實很薄弱，因為學生是否能在限定的時間內做完所有的題目，其實很難從他們的答題模式分辨出來；並非所有的學生一定都從第一題開始作答，直到最後一題。因此除非學生回答了每一題，否則即使他回答了最後一題，並不一定表示他嘗試過了所有的題目；因為可能有些學生是從最後一題開始作答，或者可能先挑看來簡單的題目回答，而恰巧覺得最後一題還算簡單，如此一來，即使他們真的來不及作完整份測驗，我們也無從得知。但也許「大多數的人習慣從第一題開始寫，直到最後一題」這樣的假定還是可以被接受。這裡，我們在這樣的假定條件下，觀察學生是否完成了每個分項測驗中的最後五題，以探討學生是否能在限定時間內作完本測驗的所有題目。表 3-1 與 表 3-2 分別列出高中職五專及國中學生，在五 (四) 種領域下的測驗中，沒有回答最後一題的比率；沒有回答最後第一及第二題的比率；沒有回答最後第一、二、三題的比率，直至最後五題都沒有回答的比率。觀察表 3-1 與 表 3-2，這些百分比非常低，表示學生幾乎都可以在限定的時間內把所有的題目做完；沒有回答的原因，很可能是無法回答 (不會)；例如，高中職五專一

般分析能力的最後一題，沒有回答的百分比較高（1.23%），反映了當初排列題目時，將最難的題目放在該類測驗的最後的設計。而國中的語文領域部分，最後四題沒有回答百分比比其他地方高，但連續五題都沒有回答者驟降，可能也是因為英文對國中一年級學生而言較為困難，而最後四題恰巧都是英文領域的題目，所以有較高比例的學生無法作答。因此以下所有的分析都將沒有回答的題目直接視為零分，而不以沒有回答或遺漏值計算。

表 3-1、高中職五專學生在五種領域下，最後一至五題的沒有回答百分比（%）

連續題數	一般分析能力	英文	數學	中文	科學
最後 1 題	1.23	0.48	0.26	0.06	0.12
最後 2 題	0.37	1.67	0.18	0.08	0.13
最後 3 題	0.16	0.14	0.12	0.01	0.06
最後 4 題	0.07	0.10	0.05	0.01	0.02
最後 5 題	0.03	0.08	0.03	0	0.02

表 3-2、國中學生在五種領域下，最後一至五題的沒有回答百分比（%）

連續題數	一般分析能力	數學	語文	科學
最後 1 題	.58	.33	.42	.27
最後 2 題	.21	.10	.26	.07
最後 3 題	.05	.12	.71	.05
最後 4 題	.04	.03	.66	.01
最後 5 題	.02	.02	.06	.01

答對率（ P_i ，也稱難度）

古典測驗理論中的難度(item difficulty)，是該題的答對率，也就是所有受試者中答對該題的百分比。以數學式表示：

$$P_i = \frac{n_i}{N}$$

其中 P_i 是答對題項 i 的百分比， N 是所有受測人數， n_i 是答對該題的人數。因此 P_i 值愈大表示愈多人答對該題，即該題愈簡單。答對率深受當次受測者的平均能力之影響：如果該次受測者的平均能力相對於題項 i 普遍較高，則 P_i 會較高（顯得容易），反之則 P_i 會顯得較低。由於答對率只考慮受測的總人數及答對該題的人數，因此無論如何劃分各種分項測驗都不會影響它的值。

題項與測驗總分的點二系列相關係數（Point-Biserial Correlation, r_{pbi} ）

r_{pbi} 是題項 i 與測驗總分之間的相關係數，是從 Pearson product moment 簡化而來。它探討題項 i 與總分之間的關係。更詳細而言， r_{pbi} 要瞭解的是，答對題項 i 者，是否傾向在整個測驗得到高分，因此間接顯示該題項是否能區別出得高分者（能力較佳者）與得分低者（能

力較弱者)。若 r_{pbi} 較高，表示答對該題者，其總分也較高的機會較大，因此該題頗能區辨受測者在測驗內容上能力高低之間的差異；若 r_{pbi} 低，則表示受測者在該題的表現與其在該測驗的得分高低之間關係不大，也就是該題在所測量的內容的區辨力不高；如果 r_{pbi} 是負值，則表示該題若非計分錯誤即是所測量的特質與整個測驗不符。通常認為測驗中大多數題項的 r_{pbi} 在 .3 與 .6 之間為最佳。若 $r_{pbi} \leq .25$ 時，該題項會減低測驗整體的信度，減少測驗總分真實反應受測者之間能力差異的機會。其數學式是

$$r_{pbi} = \frac{M_p - M_q}{S_t} \sqrt{pq}$$

其中， M_p 表示所有答對該題者總分的平均數；

M_q 表示所有答錯該題者總分的平均數，因此該題沒有回答者則不列入計算（但本報告中直接將沒有回答視為答錯，因此都計算在內）；

S_t 是所有回答該題者的總分之間的標準差；

p 是所有回答該題者中答對該題的比率，而 q 是答錯該題的比率。

由於同一個題項可能被歸類於幾種不同的分項測驗中，因此同一題項會因被編列在不同的分項測驗中，而有不同的 r_{pbi} 值。

3.1 高中職五專的結果

表 3-3 至表 3-6 列出高中職五專各分項測驗之答對率及點二系列相關係數 (r_{pbi})，並將點二系列相關係數小於 .25 者以粗體字標示。以表 3-3 而言，一般分析能力 (1 至 25 題) 中共有 6 題的點二系列相關係數較差；英文領域 (26 至 31 題) 則有 4 題；數學領域 (32 至 51 題) 有 2 題；中文 (52 至 61 題) 有 4 題；自然科學領域 (62 至 73 題) 則有 6 題。其他分項測驗中出現鑑別度不高者，也都是以上相同的題項，因此不再贅述。

表 3-3、高中職五專，綜合分析能力題的古典測量指標

題號	答對率 (%)	點二系列相關 (r_{pbi})	題號	答對率 (%)	點二系列相關 (r_{pbi})
1	63.6	0.25	38	78.9	0.29
2	54.6	0.23	39	54.7	0.32
3	57.1	0.28	40	67.9	0.55
4	68.7	0.32	41	70.1	0.43
5	65.3	0.47	42	66.0	0.30
6	64.2	0.52	43	61.7	0.44
7	59.7	0.27	44	55.6	0.48
8	66.4	0.19	45	63.9	0.42
9	51.5	0.31	46	93.9	0.33
10	91.1	0.11	47	35.0	0.21
11	80.4	0.35	48	53.4	0.45
12	63.1	0.32	49	42.5	0.38
13	69.9	0.42	50	44.9	0.16
14	47.3	0.11	51	55.9	0.43
15	53.3	0.20	52	86.1	0.31
16	44.1	0.28	53	58.0	0.20
17	35.9	0.32	54	81.7	0.23
18	29.4	0.33	55	59.5	0.21
19	28.8	0.40	56	66.7	0.42
20	88.1	0.31	57	60.5	0.43
21	86.7	0.38	58	61.0	0.42
22	64.9	0.40	59	48.0	0.40
23	70.6	0.35	60	41.7	0.32
24	54.6	0.40	61	37.0	0.24
25	45.3	0.26	62	88.8	0.18
26	66.6	0.24	63	78.9	0.20
27	69.5	0.06	64	81.0	0.42
28	35.4	0.31	65	71.3	0.50
29	71.3	0.43	66	66.7	0.19
30	31.3	0.23	67	55.0	0.27
31	18.6	0.17	68	58.2	0.44
32	95.4	0.31	69	39.5	0.14
33	72.8	0.27	70	33.4	0.25
34	91.0	0.25	71	32.2	0.04
35	87.4	0.39	72	27.6	0.04
36	88.2	0.33	73	53.6	0.42
37	79.6	0.35			

表 3-4、高中職五專，一般分析能力題的古典測量指標

題號	答對率 (%)	點二系列相關 (r_{pbi})	題號	答對率 (%)	點二系列相關 (r_{pbi})
1	63.6	0.22	14	47.3	0.12
2	54.6	0.21	15	53.3	0.20
3	57.1	0.25	16	44.1	0.27
4	68.7	0.32	17	35.9	0.35
5	65.3	0.44	18	29.4	0.35
6	64.2	0.48	19	28.8	0.40
7	59.7	0.25	20	88.1	0.31
8	66.4	0.19	21	86.7	0.37
9	51.5	0.29	22	64.9	0.37
10	91.1	0.11	23	70.6	0.34
11	80.4	0.33	24	54.6	0.38
12	63.1	0.32	25	45.3	0.25
13	69.9	0.41			

表 3-5、高中職五專，數學分析能力題的古典測量指標

題號	答對率 (%)	點二系列相關 (r_{pbi})	題號	答對率 (%)	點二系列相關 (r_{pbi})
32	95.4	.30	42	66.0	.28
33	72.8	.26	43	61.7	.42
34	91.0	.24	44	55.6	.44
35	87.4	.38	45	63.9	.41
36	88.2	.31	46	93.9	.32
37	79.6	.35	47	35.0	.20
38	78.9	.28	48	53.4	.45
39	54.7	.31	49	42.5	.35
40	67.9	.54	50	44.9	.15
41	70.1	.41	51	55.9	.41

表 3-6、高中職五專，數學或數字型分析能力題的古典測量指標

題號	答對率 (%)	點二系列相關 (r_{pbi})	題號	答對率 (%)	點二系列相關 (r_{pbi})
4	68.7	0.34	38	78.9	0.30
5	65.3	0.49	39	54.7	0.33
6	64.2	0.54	40	67.9	0.55
11	80.4	0.35	41	70.1	0.44
12	63.1	0.31	42	66.0	0.29
13	69.9	0.42	43	61.7	0.43
20	88.1	0.33	44	55.6	0.46
21	86.7	0.40	45	63.9	0.41
22	64.9	0.41	46	93.9	0.34
32	95.4	0.33	47	35.0	0.20
33	72.8	0.27	48	53.4	0.44
34	91.0	0.26	49	42.5	0.34
35	87.4	0.40	50	44.9	0.15
36	88.2	0.33	51	55.9	0.42
37	79.6	0.36			

3.2 國中的結果

表 3-7 至表 3-10 是國中測驗题目的結果。也將鑑別度較差者（點二系列相關係數小於 .25）以粗體字標出。一般分析能力中（題號 1 至 27）有 15 題的鑑別度較差，比高中職五專（6 題）多了一倍半，而且這些題目大多答對率不高（小於 40%），顯示國中一年級的學生似乎還不太熟習課堂以外的解決問題的思考方法，因而答對率低，致使相關係數也較低；另一方面則可能是因為一般分析能力所測量的三種面向——分析、生活應用與創造——不但與課堂教授內容不同，相互之間也不同，使得相關係數顯得更低。數學領域題中（題號 28 至 47）則只有 3 題的鑑別度較差。語文領域（題號 48 至 61）有 3 題，且其中 2 題是英文領域，答對率也偏低。科學領域題的鑑別度則都達到標準以上。

表 3-7、國中綜合分析能力題的古典測量指標

題號	答對率 (%)	點二系列 相關(r_{pbi})	題號	答對率 (%)	點二系列 相關(r_{pbi})
1	35.6	.17	37	56.8	.27
2	39.7	.24	38	72.3	.55
3	30.0	.03	39	66.2	.40
4	27.2	.26	40	47.4	.33
5	29.3	.32	41	49.9	.38
6	26.5	.20	42	62.5	.48
7	53.7	.20	43	51.0	.22
8	39.7	.28	44	53.1	.21
9	28.1	.18	45	47.1	.41
10	75.8	.29	46	41.3	.31
11	88.1	.16	47	41.5	.15
12	81.1	.20	48	90.1	.25
13	57.6	.37	49	79.7	.34
14	45.1	.21	50	68.5	.32
15	42.0	.34	51	73.0	.34
16	41.2	.07	52	74.1	.35
17	43.8	.16	53	61.1	.18
18	25.9	.17	54	62.4	.31
19	79.9	.26	55	51.7	.38
20	42.7	.21	56	54.3	.44
21	7.8	.09	57	55.8	.40
22	72.8	.34	58	79.5	.36
23	66.0	.44	59	35.7	.27
24	50.8	.45	60	34.4	.12
25	57.0	.39	61	37.0	.02
26	29.4	.25	62	87.4	.27
27	31.9	.22	63	83.3	.27
28	89.4	.30	64	87.4	.42
29	80.6	.39	65	74.1	.42
30	79.1	.48	66	72.9	.28
31	76.6	.41	67	67.5	.36
32	86.5	.35	68	61.7	.37
33	80.2	.42	69	55.6	.22
34	71.5	.51	70	50.4	.29
35	72.0	.46	71	50.7	.30
36	63.6	.46			

表 3-8、國中一般分析能力題的古典測量指標

題號	答對率 (%)	點二系列 相關(r_{pbi})	題號	答對率 (%)	點二系列 相關(r_{pbi})
1	35.6	.15	15	42.0	.31
2	39.7	.20	16	41.2	.06
3	30.0	.04	17	43.8	.14
4	27.2	.25	18	25.9	.16
5	29.3	.32	19	79.9	.20
6	26.5	.20	20	42.7	.19
7	53.7	.19	21	7.8	.10
8	39.7	.27	22	72.8	.32
9	28.1	.17	23	66.0	.40
10	75.8	.24	24	50.6	.42
11	88.1	.12	25	57.0	.35
12	81.1	.16	26	29.4	.23
13	57.6	.32	27	31.9	.20
14	45.1	.17			-

表 3-9、國中數學分析能力題的古典測量指標

題號	答對率 (%)	點二系列 相關 (r_{pbi})	題號	答對率 (%)	點二系列 相關(r_{pbi})
28	89.4	.29	38	72.3	.54
29	80.6	.36	39	66.2	.39
30	79.1	.48	40	47.4	.31
31	76.6	.42	41	49.9	.37
32	86.5	.35	42	62.5	.48
33	80.2	.43	43	51.0	.20
34	71.5	.51	44	53.1	.18
35	72.0	.46	45	47.1	.38
36	63.6	.43	46	41.3	.29
37	56.8	.26	47	41.5	.14

表 3-10、國中數學或數字型分析能力題的古典測量指標

題號	答對率 (%)	點二系列 相關(r_{pbi})	題號	答對率 (%)	點二系列 相關(r_{pbi})
4	27.2	.26	34	71.5	.51
5	29.3	.33	35	72.0	.46
6	26.5	.22	36	63.6	.44
13	57.6	.36	37	56.8	.27
14	45.1	.20	38	72.3	.54
15	42.0	.34	39	66.2	.39
22	72.8	.35	40	47.4	.32
23	66.0	.44	41	49.9	.38
24	50.6	.46	42	62.5	.49
28	89.4	.29	43	51.0	.21
29	80.6	.36	44	53.1	.19
30	79.1	.49	45	47.1	.41
31	76.6	.41	46	41.3	.31
32	86.5	.34	47	41.5	.15
33	80.2	.43			

第四章 項目反應理論 (IRT) 的分析

本資料庫將長期追蹤受訪學生，以觀察學生的學習成就與能力的成長。然而，古典測驗理論中的指標會因當次受測者的能力不同而改變，無法讓研究者觀察受測者能力的變化，因此對於長期研究而言不合適。項目反應理論模式 (Item Response Theory, 簡稱 IRT) 則因為容許研究者固定住題目的「難度」，以反應同一受測者在兩個時間點之間所產生的變化，或比較不同波次的受測者在相同年齡時的表現，而可免除這方面的缺點。此外，IRT 能力估計值相對於答對題數，也有另一個優點：即使答對題數之間的差異相同 (例如：答對 45 題相較於答對 40 題，和答對 65 題相較於答對 60 題，都是相差五題)，IRT 會將樣本整體的答題模式及題目的特性列入考量，使得兩組能力估計值之間的差異有所不同，較精緻反映不同能力表現之間的差異。所以若要觀察樣本長期的能力成長情形，都必須使用項目反映理論模式作分析。

本次報告也以 IRT 作分析，待獲得更多波次的資料後，就能充分利用 IRT 的優點，理解學生分析能力的成長軌跡。本章先介紹 IRT 的概念，再介紹常見的三種估計受測者能力的方法，包括最大概似估算法 (Maximum Likelihood Estimation, 簡稱 ML) 貝氏估算法 (Bayes Estimation) 中的 EAP 估算法 (Expected A Posteriori) 及 Maximum A Posterior (簡稱 MAP)。由於資料庫中的學生能力估計值是採用 EAP 估算法，因此我們也說明 EAP 所適用的實證信度 (empirical reliability) 之計算方法。

分析的結果則包括各分項測驗在 1PL 和 3PL 的 IRT 模式下，各題項的難度估計值及學生能力估計值的描述性統計值及相關係數，及各分項測驗在兩種模式下的實證信度。高中職五專部分的結果都放在 4.1，國中樣本的結果則放在 4.2。各種圖示結果，包括：(1) 各分項測驗下所有題項的答對機率圖 (Item Characteristic Curve, 簡稱 ICC)；(2) 測驗的訊息圖 (Test information)；(3) 學生在各種分項測驗下的能力分佈長條圖 (histogram)；(4) 學生在各分項測驗下的答對題數與能力估計值之間的散佈圖 (scatterplot)，則分別收錄在附錄一 (高中職五專) 與附錄二 (國中)。

最後，也許讀者會覺得，「多面向的 IRT 模式」 (multi-dimensional IRT) 會使模型適合度更佳，何不使用之？然而基於種種因素考量，尤其經過實際估算過多面向模式後，發覺其與單面向 IRT (也就是本文所用者) 模式所產生的能力估計值相關極高；且單面向 IRT 模式的結果與一些常見而重要的共變項之間的相關明確而易於解釋，因此我們決定維持使用單面向 IRT。本章最後稍微討論這些單面向的 1PL 和 3PL 的模型適合度，4.3 討論高中職五專樣本的模型適合度，4.4 討論國中樣本的模型適合度。

IRT (項目反應理論) 簡介

目前使用最廣的 IRT 模式包括單參數的 IRT (1PL) 模式及三參數的 IRT (3PL) 模式。

IRT 的主要假定是：受測者答對某一題的機會，決定於其（被測量的）能力及該題項的特質（如測量的內容，型式等）。而目前發展的 IRT 理論最多考量了三種特質：（1）題項的難度、（2）題項的鑑別度、（3）可猜對的機會。只考量第一種特質（題項難度）的 IRT 稱為「單參數」（簡稱 1PL）的 IRT；考量第一及第二種特質者則是「二參數」（簡稱 2PL）的 IRT；三種都考量的則是「三參數」（簡稱 3PL）的 IRT。

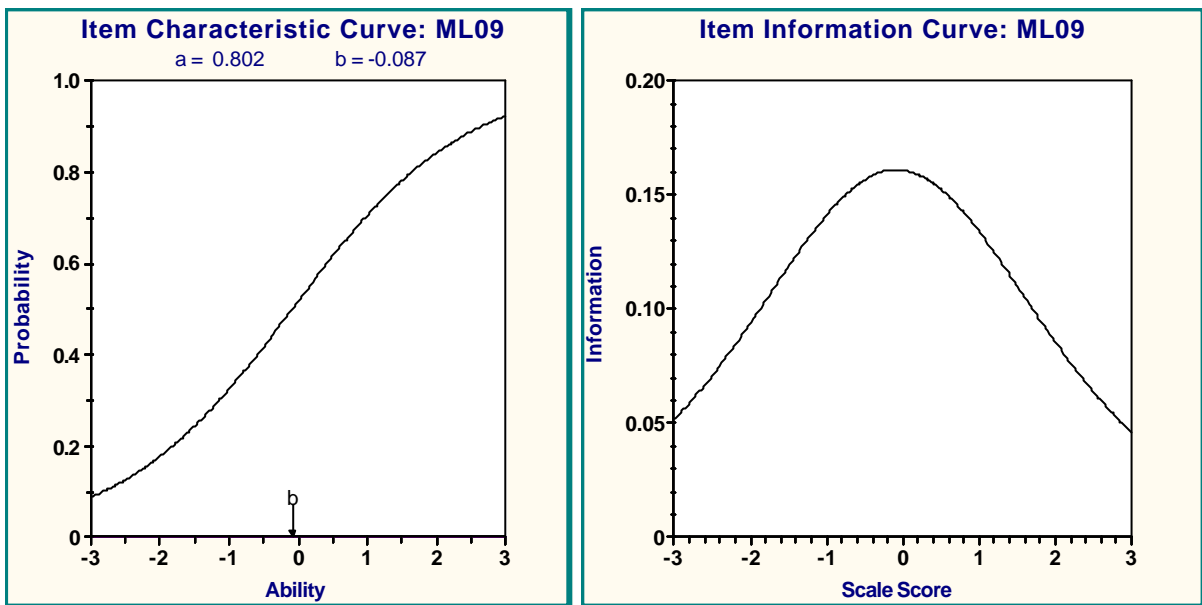
單參數 IRT

最早發展出來的，是 1PL。它只考慮題項難度(difficulty)與能力對於答對機率的影響。其數學式如下：

$$P_i(q) = \frac{1}{1 + \exp[-(q - b_i)]}$$

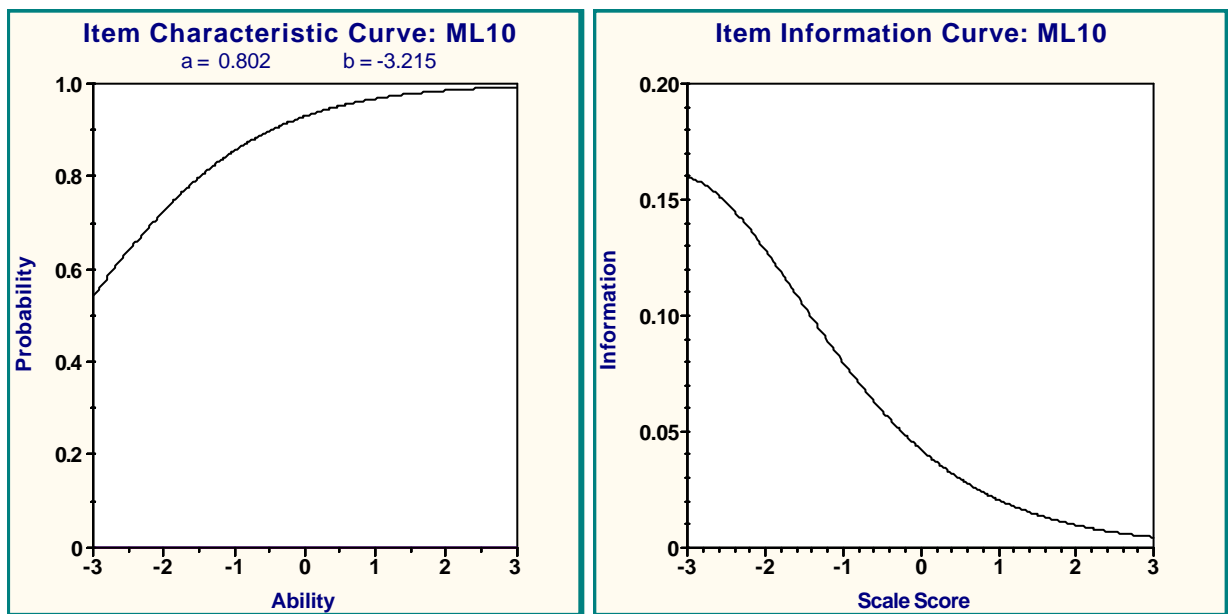
其中， q 是受試者的能力， $P_i(q)$ 是能力 q 的受試者答對題項 i 的機率， b_i 是題項的難度，與 q 在同一量尺上。 $b_i = 0$ 表示難度中等， b_i 值愈小難度愈低，題目愈簡單， b_i 值愈大難度愈高。 q 值愈大（受試者能力愈高），或 b_i 值愈小（題目愈簡單），都使 $P_i(q)$ 值愈大，也就是答對題項 i 的機率愈高。 $\exp(\)$ 是表示自然對數的底， $\exp(0)=1$ ， $\exp(-1)=.3679$ ， $\exp(1)=2.7183$ ，因此 $(q - b_i)$ 的值愈大（正數），則分母愈小而 $P_i(q)$ 愈大；當能力與難度相等時，答對機率等於 .50。

以下圖 4-1 至 圖 4-3 是三個不同的題項，利用 1PL 模式分析的結果。1PL 模式下，在運算時仍需要估算鑑別度，但估算時整個測驗中所有的鑑別度都被設定相同。在此處，圖 4-1 到 圖 4-3 的鑑別度都是 .802。圖左是題項的答對機率曲線，圖右則是該題的訊息曲線（Item Information Curve）。兩種圖的橫軸都代表受測者的能力。答對機率圖呈現能力與答對機會之間的關係。由於當能力與難度相等時，答對機率是 .50，因此 IRT 的答對機率曲線圖，機率等於 .50 時，所對應的能力（也就是 scale score）值等於該題的難度。訊息曲線圖則呈現能力與該題項所能提供的訊息間的關係：若該題項在某能力範圍之內能提供豐富的訊息，則表示該題項對於該能力範圍之內的受試者較有鑑別力。訊息曲線愈高之處，表示該題在該能力範圍較能精準地反應受試者能力；訊息曲線最高之處，也是該題項的難度值所在。在這個例子中，各題項的鑑別度是都 .802，這三題的難度 b_i 都不高，唯有圖 4-1 的難度屬於中等。圖 4-1 到 4-3 的訊息曲線，雖然因為各題項難度不同而致高峰處所對應的能力值不同，但因為鑑別度預設為相同，它們上升下降的弧度都相同。



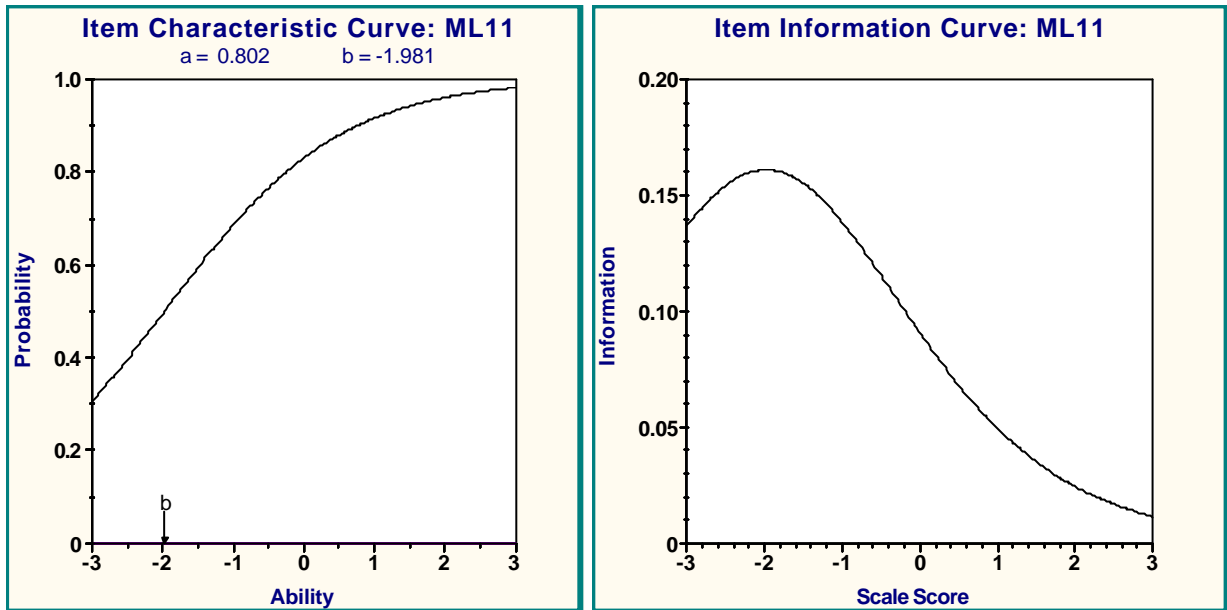
1-Parameter Model, Logistic Metric Item: 9
 The parameter a is the item discriminating power, the reciprocal ($1/a$) is the item dispersion, and the parameter b is an item location parameter.

圖 4-1、1PL 模式下的題項答對曲線圖與題項訊息圖



1-Parameter Model, Logistic Metric Item: 10
 The parameter a is the item discriminating power, the reciprocal ($1/a$) is the item dispersion, and the parameter b is an item location parameter.

圖 4-2、1PL 模式下的題項答對曲線圖與題項訊息圖



1-Parameter Model, Logistic Metric Item: 11

The parameter a is the item discriminating power, the reciprocal ($1/a$) is the item dispersion, and the parameter b is an item location parameter.

圖 4-3、1PL 模式下的題項答對曲線圖與題項訊息圖

二參數 IRT

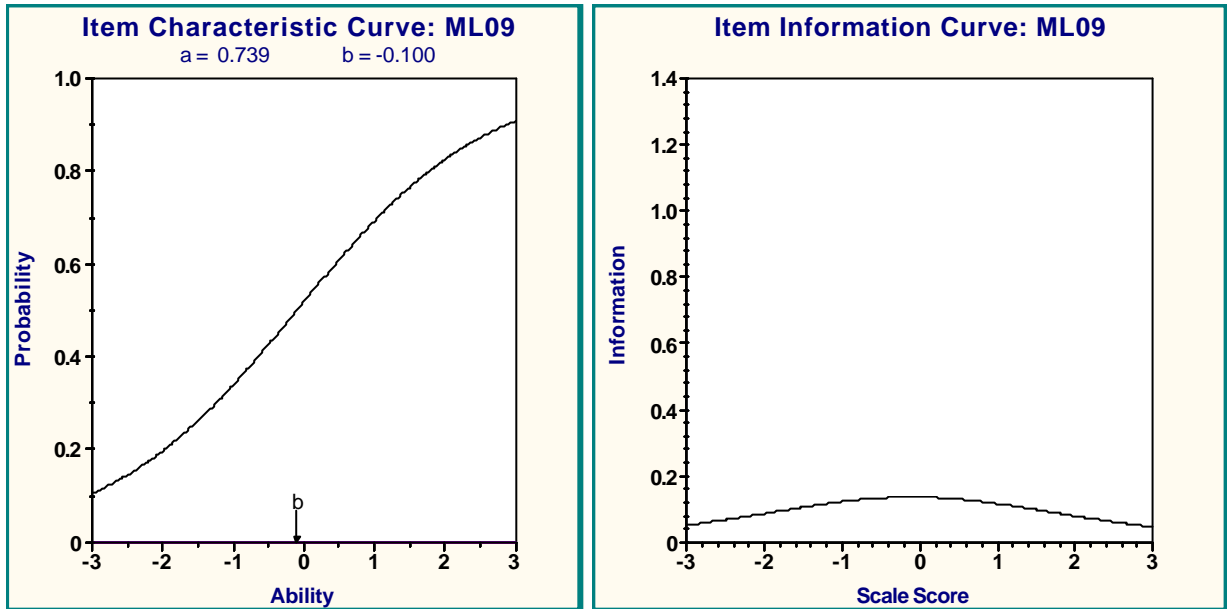
繼單參數之後，有研究者認為，各題項在不同能力範圍內的鑑別度(discrimination)可能不一，例如某一題項可能在某一個能力範圍內鑑別度極強，在該範圍內不同能力之間細微的差別，都可由該題偵測得知；但在該範圍以外，鑑別度低落，受試者能力的增減對於答對的機會幾乎沒有影響；或者某些題項在各個能力的鑑別度都平平，區分每種能力的鑑別度都差不多。因此在 1PL 模式中多加了鑑別度 a_i ，成為 2PL。其數學式如下：

$$P_i(\mathbf{q}) = \frac{1}{1 + \exp[-a_i(\mathbf{q} - b_i)]}$$

由 $a_i(\mathbf{q} - b_i)$ ，可以瞭解，題項的鑑別度與能力程度之間是倍率的關係，而能力與難度仍在同一量尺，可以相加減；難度愈低，或能力愈高，都會使答對的機會增高。鑑別度則使兩者的差異值或放大 ($a_i > 1$) 或縮小 ($0 < a_i < 1$)。

圖 4-4 至 圖 4-6 中 2PL 模式所描繪的題項與 圖 4-1 至 圖 4-3 的題項相同；也就是：圖 4-4 所使用的題項與 圖 4-1 相同；圖 4-5 所使用的與 圖 4-2 相同，而 圖 4-6 所使用的則與 圖 4-3 所使用的題項相同。由於多考慮了鑑別度，題項難度也有所改變。例如 圖 4-5 的難度在 2PL 中大幅下降(從 圖 4-2 的 -3.215 降至 圖 4-5 的 -5.692)，而圖 4-6 的難度則稍微上升(從 圖 4-3 的 -1.981 升至 圖 4-6 的 -1.574)。而由於每一題的鑑別度不同，答對曲線和訊息曲線都變得頗為不同；鑑別度愈高，答對曲線和訊息曲線都相對較為陡峭，反之則較為平緩。例

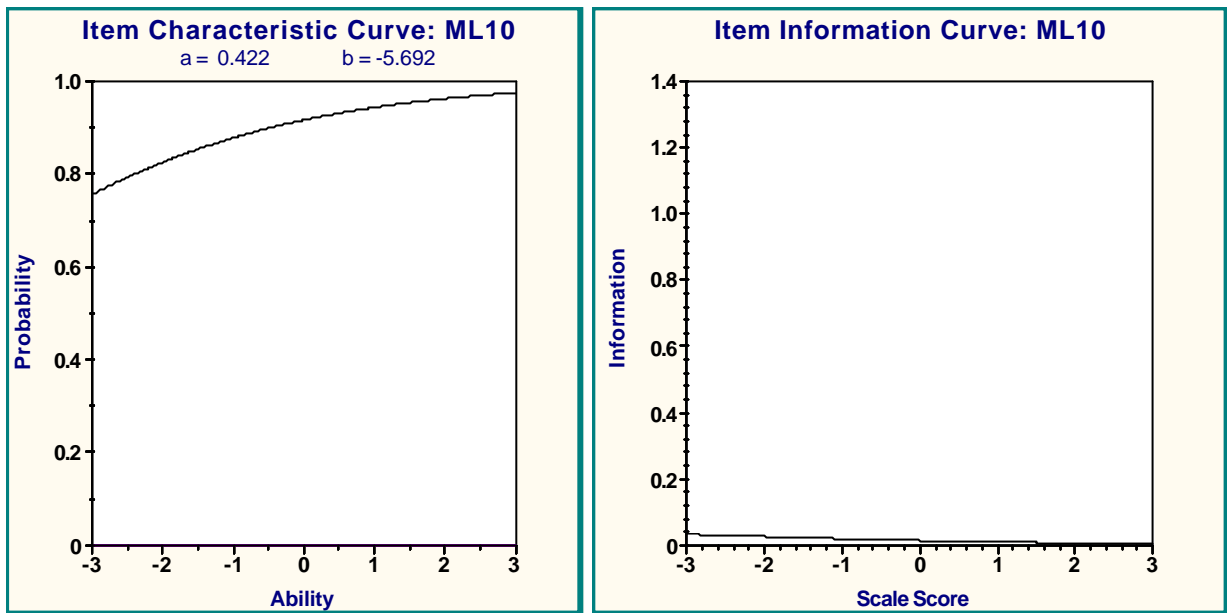
如圖 4-4 的難度雖然屬於中等，但鑑別度不高 (< 1)，因此即使訊息曲線的極高點是在 -0.10 ，但只形成小丘，坡度並不高；圖 4-5 的鑑別度更低，再加上極低的難度，使該題項在一般常見的能力範圍內所能提供的訊息幾乎為零；圖 4-6 的鑑別度稍高一些，它所能提供的訊息又比圖 4-4 稍高。



2-Parameter Model, Logistic Metric Item: 9

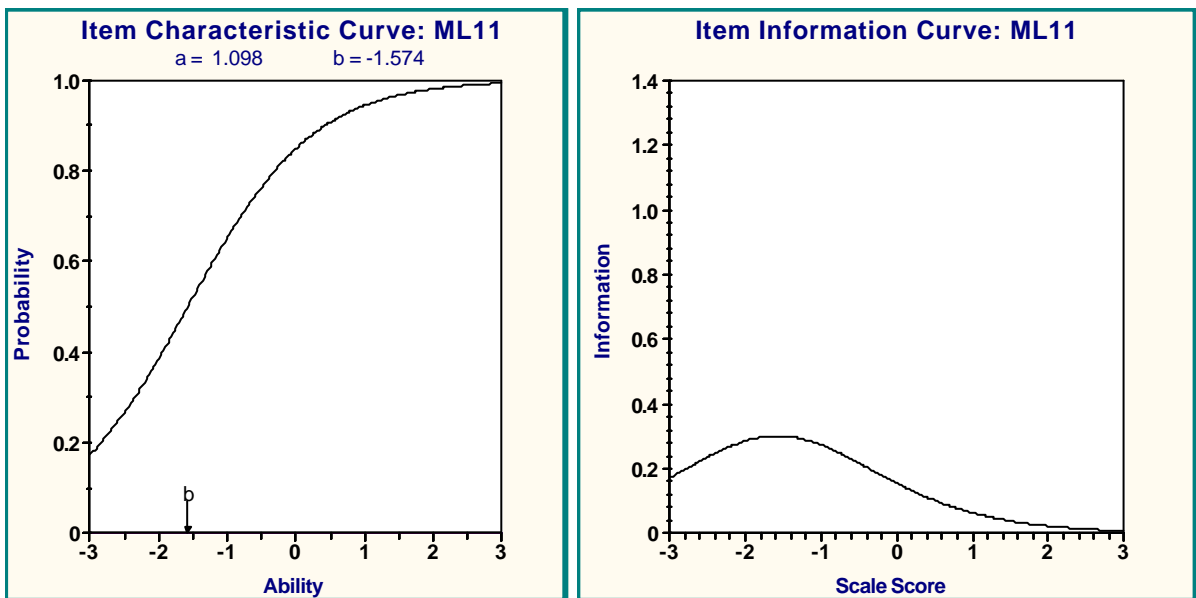
The parameter a is the item discriminating power, the reciprocal ($1/a$) is the item dispersion, and the parameter b is an item location parameter.

圖 4-4、2PL 模式下的題項答對曲線圖與題項訊息圖 --- 鑑別度 = .739



2-Parameter Model, Logistic Metric Item: 10
 The parameter a is the item discriminating power, the reciprocal ($1/a$) is the item dispersion, and the parameter b is an item location parameter.

圖 4-5、2PL 模式下的題項答對曲線圖與題項訊息圖 --- 鑑別度 = .422



2-Parameter Model, Logistic Metric Item: 11
 The parameter a is the item discriminating power, the reciprocal ($1/a$) is the item dispersion, and the parameter b is an item location parameter.

圖 4-6、2PL 模式下的題項答對曲線圖與題項訊息圖 --- 鑑別度 = 1.098

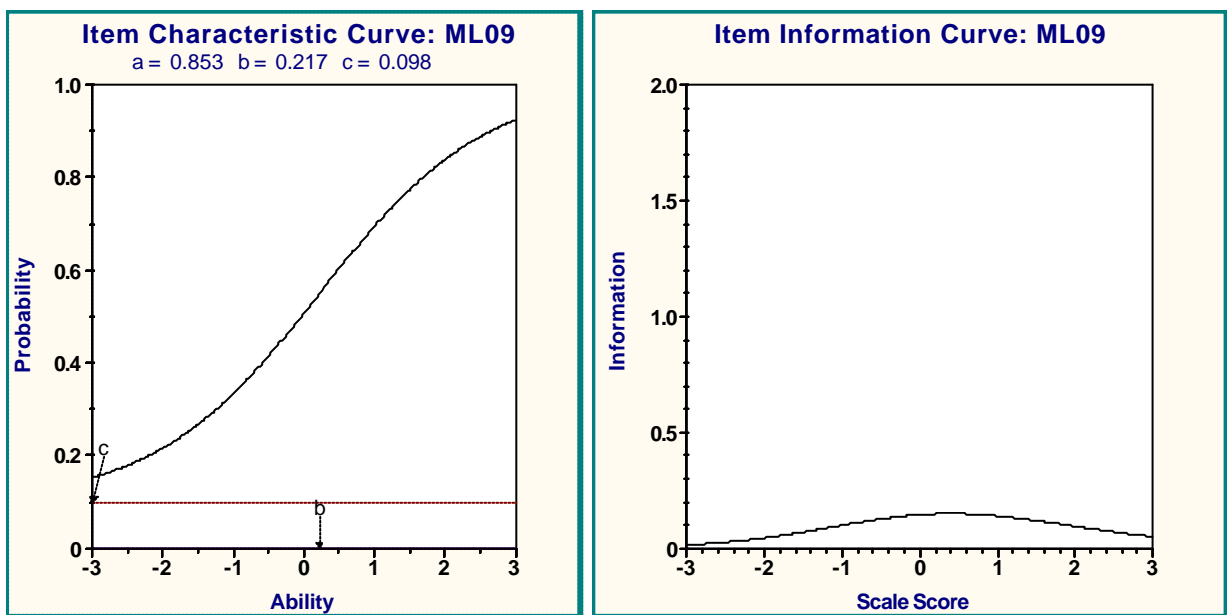
三參數 IRT

最後，又有研究者認為每個題項可猜對 (guessing) 的機會不同，因此也必須將「可猜對」的機會納入考慮，而在 2PL 模型多加了一項「可猜對」的機率 c_i ，變成了 3PL 模式。因此 3PL 模式中「可猜對」的機率成為答對機率的底線；即使能力非常低，對於該題也有 c_i 的答對率。3PL 模式以數學式表示如下：

$$P_i(\mathbf{q}) = c_i + \frac{(1 - c_i)}{1 + \exp[-a_i(\mathbf{q} - b_i)]}$$

答對機率的高低除了因每位受試者自身的能力而不同外，還受該題項本身特質的影響。在上式中可以看出，若可猜對的機會愈大，則答對的機會愈高，且能力或鑑別度對於答對率的影響愈小。

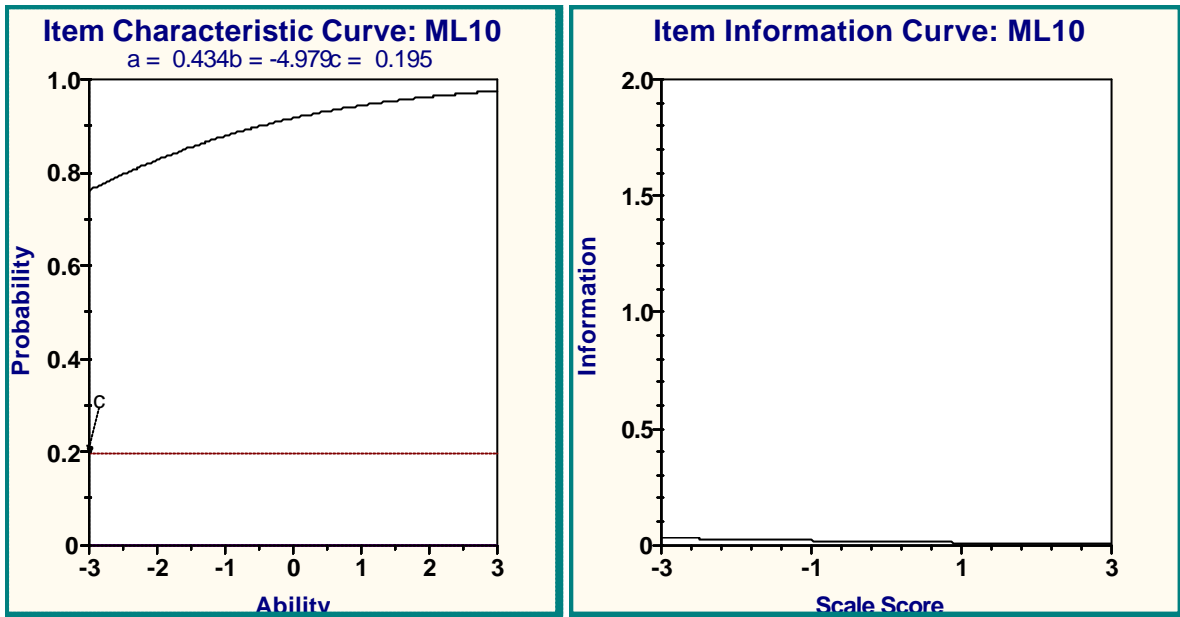
圖 4-7 至 圖 4-9 所使用的題項也與 圖 4-1 至 圖 4-3 相同。比較 圖 4-4 至 4-6 與 圖 4-7 至 4-9，加入了猜對機會後，在 3PL 模式下，各題難度都稍微增高了一些。第二題 (ML10) 與第三題 (ML11) 之間的差異在 3PL 中愈見明顯：在 1PL 時，兩者對於此次受試者而言都是很簡單的題項，雖然第二題更簡單一些；但在 3PL 中，第二題由於難度極低，猜對機會大，鑑別度也低，使得該題項所能提供的訊息幾乎等於零，而第三題的訊息則相對較多。但整體而言，這三題所能提供的訊息還是偏低。



3-Parameter Model, Logistic Metric Item: 9

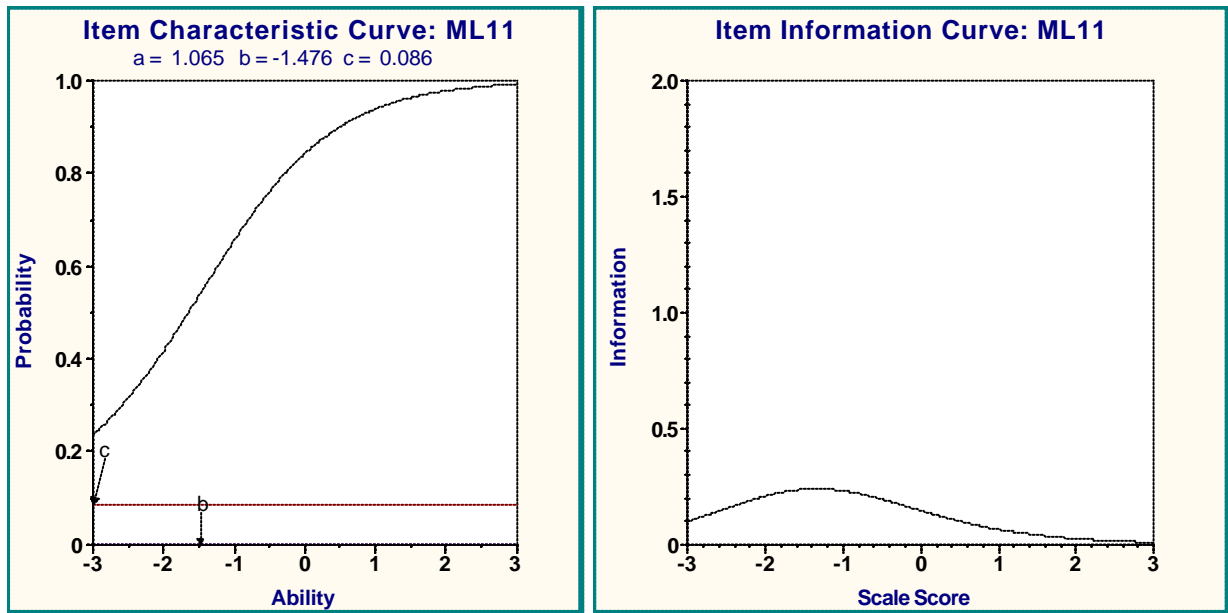
The parameter a is the item discriminating power, the reciprocal ($1/a$) is the item dispersion, b is an item location parameter and c the guessing parameter.

圖 4-7、3PL 的題項答對曲線圖與題項訊息圖



3-Parameter Model, Logistic Metric Item: 10
 The parameter a is the item discriminating power, the reciprocal ($1/a$) is the item dispersion, b is an item location parameter and c the guessing parameter.

圖 4-8、3PL 模式下的題項答對曲線圖與題項訊息圖 --- 鑑別度 = 1.297



3-Parameter Model, Logistic Metric Item: 11
 The parameter a is the item discriminating power, the reciprocal ($1/a$) is the item dispersion, b is an item location parameter and c the guessing parameter.

圖 4-9、3PL 模式下的題項答對曲線圖與題項訊息圖 --- 鑑別度 = .753

由於 2PL 與 3PL 的估算，還必須考量每一題項的鑑別度（及可猜對的機會），因此所得的能力估計值與受測者的答對題數之間並無一定明確的關係。而 1PL 中，答對總題數即是 IRT 的「充分統計量」(sufficient statistic)，故答對題數相同者，不論是答對哪些題項，其能力估計值都會相同。

受試者能力估算法

要計算受測者的能力，IRT 模式的軟體中可供選擇的估算法有三種。權衡各種方法的優劣之後，本報告採用 EAP 的能力估計值，也就是貝氏估算法的一種。以下大略說明這三種估算法(參考自 Bilog-MG3.0 使用者手冊)：

1. 最大概似估算法 (Maximum Likelihood Estimation, 簡稱 ML) 是純粹根據資料，估算出「使資料最有可能出現的」參數（在此為能力）。也就是依據現有的資料所呈現的模式，利用統計方法回推「使這種答題模式最有可能出現」的受訪者的能力。所謂「答題模式」是指答對幾題或答對哪些題。當受測者的能力恰巧與測驗中許多題項的難度相近時，其所估計的能力的標準誤較小；反之，若和受測者的能力相當的題數較少時，該受測者能力的測量標準誤就會較大。ML 的一項缺點是：當某一受測者全部答對或全部答錯該測驗中的所有試題時，ML 將無法估算其能力程度。（在三參數的 IRT 中，如果資料所顯現的模式幾近亂答時，偶爾也會出現這種情形）。

貝氏估算法 (Bayes Estimation) 有兩種方法：

2. Expected A Posteriori 估算(簡稱 EAP) 是利用測驗的答題模式來「修改」受測者先前假定的能力分佈 (prior distribution) (以上是「貝氏定理」的概念) 後，得到的新分佈 (posterior distribution) 中的平均值。EAP 估算法適用於任何一種答題模式，因此即使受測者答對或答錯所有的題目，其能力都可由 EAP 估計。而 EAP 的缺點是：它所估得的值稍微過於集中在母群體的平均值附近 (shrink toward the mean), 但若其標準誤很小時，這個誤差也很小。較為靠近平均值的傾向也使得 EAP 的標準差較小，但如果所分析的資料是使用相同的測驗，這不會是大問題。
3. Maximum A Posterior 估算(簡稱 MAP) 也屬於貝氏估算法，但所用的是新分佈的眾數 (mode) (Bayes Modal estimator)，而且它在求「使資料中的答題模式最有可能出現的」參數（能力）時，除了最大概似估算法 (ML) 中的資料分佈之外，也將參數原有（或假定）的分佈納入考量，因此得名。MAP 與 EAP 相同，即使得滿分或得零分，也可以估算其相對的能力；一般而言，MAP 也稍微偏向集中於母群體的平均值，但誤差比 EAP 稍大。

EAP 的實證信度 (empirical reliability)

信度的概念是：真正能力分佈的變異數 (true score variance) 佔所有觀察到的分佈之變異數 (observed variance) 的比率；觀察到的能力分佈之變異量則包括測量誤差 (measurement error) 的變異量 (error variance) 及真正能力分佈的變異量。真正能力分佈的變異量所佔的比重愈

大，表示該測驗的信度愈高，所獲得的結果愈穩定。而在 EAP 中，測量誤差的變異量(error variance)的估計值，是所有能力估計值的誤差變異數（也就是能力估計值的標準誤之平方）之平均值；而真正能力分佈的變異量（true score variance）的估計值，則是所有能力估計值的變異量（BILOG-MG3.0 使用者手冊）。以數學式表示較清楚：

$$\hat{s}_{error}^2 = \frac{\sum_j^N (S.E.\hat{q}_j)^2}{N}, \quad \hat{s}_{true-score}^2 = \frac{\sum_j^N (\hat{q}_j - \bar{\hat{q}})^2}{N},$$

其中 \hat{s}_{error}^2 是測量誤差的變異數估計值， $S.E.\hat{q}_j$ 是 \hat{q}_j 的標準誤； $\hat{s}_{true-score}^2$ 是真正能力分佈的

變異數估計值， \hat{q}_j 是受測者 j 的能力估計值，共有 N 位受測者；而 $\bar{\hat{q}} = \frac{\sum_j^N \hat{q}_j}{N}$ ，是所有受測

者的能力估計值之平均值。則 EAP 的實證信度 $r_{xx'} = \frac{\hat{s}_{true-score}^2}{\hat{s}_{true-score}^2 + \hat{s}_{error}^2}$ 。

以下依序列出各分項測驗的題項特質在 1PL 及 3PL 之下的描述統計值，學生在各分項測驗中 1PL 及 3PL 模式下的能力估計值之相關係數，及四個分項測驗在 1PL 及 3PL 模式下的實證信度。高中職五專的結果在 4.1，國中的結果在 4.2。

4.1 高中職五專樣本的結果

IRT 模式將受測者的能力平均值固定於 0.00，因此若題項難度的平均值也接近 0.00，則題項所測量的程度與受測者的平均程度較為近似；如果難度平均值愈比零大，則表示題目愈難；難度平均值愈比零小，則題目愈簡單。由表 4-1 與表 4-2 可看出，第一波高中職與五專的題目，大致而言，難度稍低於學生的平均能力，尤其是與數學相關的兩個分項測驗難度都又更低。而當初測驗的設計，即是希望絕大多數的學生都有機會學過相關的內容；因此難度稍低是反映了當初題目的選定原則。表 4-1 所列是 1PL 模式下的難度估計值之平均值。由於答對率是 1PL 的難度之充分統計量，因此表 4-1 中，難度是極大值/極小值的題項即分別是該分項測驗中答對率最低/最高的題項。

表 4-1、各分項測驗中，1PL 模式下，各題項難度估計值 b_i 之基本描述

	題數	平均值	標準差	極大值	極小值
綜合分析能力題	73	-.751	1.290	2.076	-4.126
一般分析能力題	25	-.653	1.093	1.235	-3.104
數學分析能力題	20	-1.055	1.139	.719	-3.322
數學及數字型分析能力題	29	-1.078	.987	.710	-3.268

表 4-2、各分項測驗中，3PL 模式下，各題項難度估計值 b_i 之基本描述

	題數	平均值	標準差	極大值	極小值
綜合分析能力題	73	-.139	1.802	9.597	-4.979
一般分析能力題	25	-.351	1.319	2.269	-4.920
數學題分析能力題	20	-.579	1.212	1.697	-2.452
數學及數字型題分析能力題	29	-.664	1.051	1.658	-2.481

比較表 4-1 與表 4-2 發現，3PL 模式下，難度估計值的範圍較大（極大值與極小值之間的差距），標準差亦較大，難度 b_i 值顯得較高，但題項平均而言仍然顯得偏易。表 4-1 與表 4-2 中，擁有極大或極小難度估計值之題項不盡相同，這是因為 3PL 並非單純以答對率為充分統計量，還須要考慮可猜對的機會與鑑別度。

表 4-3、各分項測驗中，3PL 模式下，各題項猜對機會 c_i 之基本描述

	題數	平均值	標準差	極大值	極小值
綜合分析能力題	73	.165	.089	.394	.026
一般分析能力題	25	.112	.075	.287	.016
數學分析能力題	20	.155	.094	.350	.050
數學及數字型分析能力題	29	.149	.088	.355	.023

表 4-4、各分項測驗中，3PL 模式下，各題項鑑別度 a_i 之基本描述

	題數	平均值	標準差	極大值	極小值
綜合分析能力題	73	1.293	.627	3.253	.120
一般分析能力題	25	1.071	1.569	6.382	.255
數學分析能力題	20	1.747	.851	4.133	.705
數學及數字型分析能力題	29	1.594	.681	3.613	.708

表 4-5、1PL 模式下與 3PL 模式下難度的相關係數

	綜合分析 能力題	一般分析 能力題	數學分析 能力題	數學及數字型 分析能力題
相關係數	.800	.871	.935	.935

表 4-6、各分項測驗下能力估計值間的相關係數

	綜合分析 能力 1PL	綜合分析 能力 3PL	數學及數 字型 1PL	數學及數 字型 3PL	數學 1PL	數學 3PL	一般分析 能力-1PL
綜合分析 能力 1PL	1.000						
綜合分析 能力 3PL	0.983	1.0000					
數學及數 字型 1PL	0.917	0.934	1.0000				
數學及數 字型 3PL	0.916	0.949	0.985	1.000			
數學 1PL	0.882	0.896	0.957	0.939	1.000		
數學 3PL	0.884	0.918	0.943	0.960	0.975	1.000	
一般分析 能力 1PL	0.885	0.863	0.806	0.806	0.692	0.704	1.000
一般分析 能力 3PL	0.858	0.860	0.813	0.821	0.688	0.702	0.951

表 4-5 與 表 4-6 顯示，即使 3PL 模式比 1PL 複雜，但兩者之間的難度估計值相關性還是頗高（表 4-5）；而受測者在同一個分項測驗中，不論是以 1PL 和 3PL 估計，其能力估計值的相關更高，達到 .95 以上（表 4-6 中以粗體字顯示者）。這兩個表顯示：即使使用複雜的模型，即使題項難度看似有變動，受測者的能力估計值高低之間的變動差異並不大；所以，如果研究者希望使用簡潔的模型來描述受試者的能力，1PL 頗為足夠。

表 4-6 中，數學能力估計值與一般分析能力估計值之間的相關最低，大約在 .70 左右。事實上，一般分析能力估計值與其他分項測驗的能力估計值之間的相關都相對較低一些，可能是一般分析能力測驗中有些題項是一般學科或課堂很少強調練習的。

表 4-7、 1PL 模式下的實證信度

分項測驗	綜合分析能 力題	一般分析能 力題	數學分析能 力題	數學或數字型分析 能力題
實證信度	.893	.763	.766	.828

表 4-8、 3PL 模式下的實證信度

分項測驗	綜合分析能 力題	一般分析能 力題	數學分析能 力題	數學或數字型分析 能力題
實證信度	.919	.813	.814	.859

比較 表 4-7 與 表 4-8，3PL 下的信度比 1PL 的信度稍高，而題數愈多，信度也相對提高。大致而言，本次測驗信度還算良好可接受，即使是 1PL 模式下，各分項測驗的信度都還

是在可接受的範圍之內。

4.2 國中樣本的結果

表 4-9 與表 4-10 顯示，相較於高中職五專測驗的難度而言，國中生所接受的測驗，對國中生而言難度稍高（表 4-9 中 b_i 的平均值都較表 4-1 中的高），尤其一般分析能力題，在 1PL 下顯得比國中學生的平均能力稍微高 .27 個標準差，在 3PL 下則需要平均 1.020 的能力。此外，數學題及數學或數字型題的難度也是國中測驗中難度較低的。

表 4-9、國中各分項測驗中，1PL 模式下，各題項難度估計值 b_i 之基本描述

	題數	平均值	標準差	極大值	極小值
綜合分析能力題	71	-.521	1.356	3.52	-3.18
一般分析能力題	27	.273	1.651	4.16	-3.396
數學分析能力題	20	-.773	.841	.416	-2.414
數學及數字型分析能力題	29	-.499	.986	1.239	-2.528

表 4-10、國中各分項測驗中，3PL 模式下，各題項難度估計值 b_i 之基本描述

	題數	平均值	標準差	極大值	極小值
綜合分析能力題	71	.242	1.969	8.668	-3.747
一般分析能力題	27	1.020	2.277	8.472	-3.738
數學分析能力題	20	-.231	1.003	1.557	-2.126
數學及數字型分析能力題	29	.038	1.021	1.757	-2.156

表 4-11、國中各分項測驗中，3PL 模式下，各題項猜對機會 c_i 之基本描述

	題數	平均值	標準差	極大值	極小值
綜合分析能力題	71	.150	.091	.442	.035
一般分析能力題	27	.139	.077	.384	.051
數學分析能力題	20	.163	.103	.346	.024
數學及數字型分析能力題	29	.161	.104	.422	.021

表 4-12、國中各分項測驗中，3PL 模式下，各題項鑑別度 a_i 之基本描述

	題數	平均值	標準差	極大值	極小值
綜合分析能力題	71	1.215	.573	2.765	.087
一般分析能力題	27	1.081	.739	2.873	.135
數學題	20	1.650	.615	2.736	.698
數學及數字型題	29	1.549	.536	2.633	.478

表 4-13、國中部分，1PL 模式下與 3PL 模式下難度的相關係數

	綜合分析 能力題	一般分析 能力題	數學分析 能力題	數學及數字型 分析能力題
相關係數	.783	.750	.954	.952

表 4-13 是 1PL 與 3PL 所估算出的難度之間的相關。一般分析能力題的相關係數最低，且比高中職五專樣本中的一般分析能力題之 1PL 與 3PL 相關係數小了.121。而綜合分析能力題因為包括一般分析能力題之故，相關也稍低。這可能是因為國中部分的一般分析能力題中有不少題目的答對率不高，在 3PL 模式中加入了其他參數（ a_i 和 c_i ）後，使得 b_i 值產生較大的變化。其他類型的題目，在兩種模式下的 b_i 值之相關係數都與高中職五專樣本中的值相近。

表 4-14、各分項測驗下能力估計值間的相關係數

	綜合分析 能力 1PL	綜合分析 能力 3PL	數學及數 字型 1PL	數學及數 字型 3PL	數學 1PL	數學 3PL	一般分析 能力-1PL
綜合分析 能力 1PL	1.000						
綜合分析 能力題 3PL	.985	1.000					
數學及數 字型 1PL	.927	.937	1.000				
數學及數 字型 3PL	.921	.951	.982	1.000			
數學 1PL	.894	.911	.954	.947	1.000		
數學 3PL	.891	.925	.942	.967	.979	1.000	
一般分析 能力 1PL	.869	.824	.804	.785	.682	.689	1.000
一般分析 能力 3PL	.846	.835	.829	.825	.685	.695	.934

儘管一般分析能力題在不同模式下的難度估計值之相關顯得較小，但其能力估計值在兩個模式下的相關還是頗高（.934，表 4-14）因此，與高中職五專的狀況相同的是：即使 3PL 模式比 1PL 複雜，同一個分項測驗在 1PL 和 3PL 下的能力估計值還是很高（.93 以上）；即使使用複雜的模型，受測者的能力估計值高低之間的變動差異並不大。

其他分項測驗在 1PL 和 3PL 下的能力估計值相關也很高。而不同的分項測驗能力估計值之間，以數學和一般分析能力的相關最低，約 .70，與高中職五專的情況相同；其他兩種分項測驗則因包含（部分）一般分析能力測驗的題目，因此相關較高。

表 4-15、國中 1PL 模式下的實證信度

分項測驗	綜合分析能力題	一般分析能力題	數學分析能力題	數學或數字型分析能力題
實證信度	.884	.675	.781	.829

表 4-16、國中 3PL 模式下的實證信度

分項測驗	綜合分析能力題	一般分析能力題	數學分析能力題	數學或數字型分析能力題
實證信度	.911	.738	.822	.860

表 4-15 與表 4-16 列出兩種模式下的實證信度，除了一般分析能力題的實證信度稍弱外，其他分項測驗的信度都還算良好。一般分析能力題的實證信度較弱，可能是因為這項測驗所涵蓋的內容較多而廣，包括分析、生活應用與創造三種能力，使用的途徑不但包括數字、文字，還包括圖表及圖形；測驗本身內各題之間的異質性比較高，加上國中生不熟習這些題目的思考方式，而致答對率低，亂猜的機會增高，因此信度相對顯得較低。

模型適合度

在進行模型適合度檢驗之前，想提醒讀者的是：3PL 由於多了兩個題項參數（鑑別度與可猜對的機會），更能依據樣本答題模式的變化，較準確估計受測者的能力，預測每一題的答對模式；但即使 1PL 和 3PL 的模型適合度相差甚遠，兩者能力估計值的相關係數非常高，共變量也很高（見 4.1 與 4.2），因此 1PL 還是不失為一個簡潔的模型。

我們從統計的角度來檢驗兩種模型的適合度。IRT 的模型適合度主要是看每一題項估計後的卡方值(chi-square)。然而卡方值會隨著樣本數而增大；高中職五專及國中的樣本都將近兩萬人，即使模型適合度頗佳，卡方值還是可能很高。由於沒有可比較的絕對標準，我們將 3PL 下的卡方值與 1PL 下的卡方值相比較，先初步瞭解兩種模型適合度的大略情形，並呈現 3PL 下，卡方值偏大的題項之「理論上的答對機率曲線和實際答對機率的比較圖」。

所謂「理論上的答對機率」，是依照 IRT 的數學模型（第四章 1PL 與 3PL 的數學模型），代入估算所得的題項參數與能力估計值而得。然而，另一方面，由於受測者在某一題的答題行為，不是答對就是答錯，無法判斷其在該題的「實際答對機率」，因此無法為每一位受測者估計他的「實際答對機率」，只能將能力非常類似的受測者併成一組，再以該組在該題的答對百分比作為該組的「實際答對機率」，該組的能力則是該組的能力估計平均值。因此，即使計算「理論上的答對機率」時，可以採用每位受測者自己的能力估計值，為遷就「實際答對機率」的運算，我們必須以每一組的能力估計值來代表該組所有受測者的能力估計值。而這樣做的結果是：不論原來的能力估計值有多少個不同的數值，比較圖上的單位將不是這些不同的能力估計值，而是這些能力組。因此，不但組數不能太少，各能力組內的差異尤其要盡量縮小，才能確實反應理論與實際兩者之間的差異。由於 IRT 所獲得的能力估計值呈常態分佈 (normal distribution)，我們利用常態分佈中 Z 分數的值與累積面積（人數）百分比的關係，

將能力估計值分成 100 組，每組在理論上佔樣本百分之一的人數（經觀察，每組在樣本中所佔的實際比率也約佔百分之零點八至百分之一點二之間），而該組的能力值就以該組人數在理論上達到百分之五十時的能力值為代表。實際答對率，則是這組能力相當的樣本中，實際答對該題的人數百分比。

由於根據比較圖看來，卡方值大約 100 的題項，模型適合度似乎還不錯，因此以下呈現 3PL 中，卡方值超過 100 的題項之「理論上的答對機率曲線和實際答對機率的比較圖」。呈現的順序，也是先報告高中職五專樣本的模型適合度，再報告國中樣本的情形。

4.3 高中職五專樣本的模型適合度

高中職五專樣本中，1PL 的卡方值比 3PL 的卡方值大了許多。在 1PL 模型下，高中職五專的題項中，有 23 題的卡方值超過 500，其中 7 個題項的卡方值超過 1000，只有 14 個題項的卡方值小於 100；相較之下，3PL 的卡方值就顯得很小，只有兩個題項的卡方值超過 100，而最大值是 298.9。因此，3PL 模型顯得比 1PL 模型更適合資料。我們觀察 3PL 中適合度稍差（卡方值大於 100）的題項。

圖 4-10 是利用英文分析能力第二題所繪製的，它的卡方值是高中職五專樣本中，3PL 模式下卡方值最高的，298.9。理論上的答對率與實際答對率之間，差異的確頗大，但主要是因為這一題的答題情況很特別，即使 3PL IRT 也無法有效推估它的實際答對機率。這一題的實際答對（散亂的許多點）機率情形，能力估計值極低者答對的機率低，能力估計值極高者答對的機率也高，但是能力估計值在 -1.0 左右時，答對機率開始下滑，直至能力估計值在 .50 左右才又開始上升；表示這個能力範圍（中等能力）的受測者，在此題項的答題表現模式與預期殊異。檢查題項內容，發現這個題項是一個閱讀測驗之一部分，答案在短文本身即可尋得。因此能力估計值在 -1.0 左右而不至於太差者，可由短文直接找到答案；但可能因為對語法的不夠熟練，中等能力的受試者反而以為答案（日期）必須另行推算，因此造成這部分能力範圍者的答對機率下降。

模型適合度檢驗：語文分析能力第二題

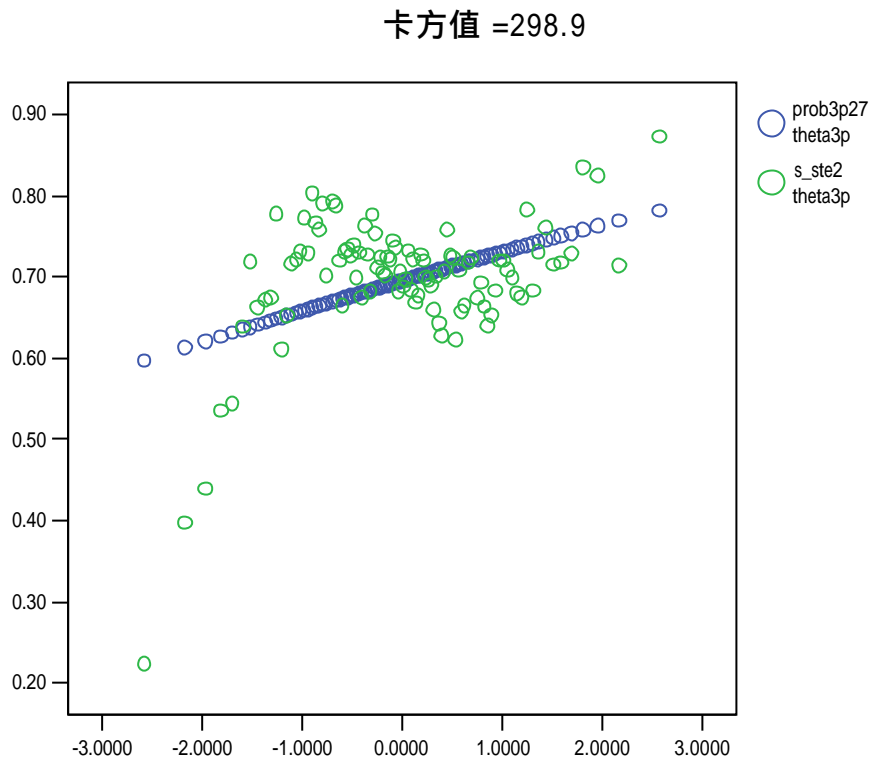


圖 4-10 高中職五專語文分析能力第二題 理論上的答對機率曲線和實際答對機率的比較圖

圖 4-11 是根據一般分析能力第二十四題所繪製。值得注意的是，圖 4-11 看似比圖 4-12 情況好很多，固然是適合度的確較佳，另一方面則是：圖 4-11 每一格所代表的比率是圖 4-10 每一格所代表比率的兩倍，因此圖 4-11 中實際答對機率顯得比較密集，而圖 4-10 的實際答對機率顯得比較疏散，似乎離理論上的答對曲線較遠。圖 4-11 顯示，3PL 的預測與實際情況差距不大，只是在能力估計值極低時稍微高估了答對率，能力估計值為中等時又稍微低估了答對率。能力估計值在 1.0 左右的受測者在這一題的答對率則有點偏低。檢查這一題項的內容發現，它是以文字代替篩選及運算的規則，需要受測者靈活應變，其他並無特殊會影響答題行為之處。

模型適合度檢驗：一般分析能力第二十四題

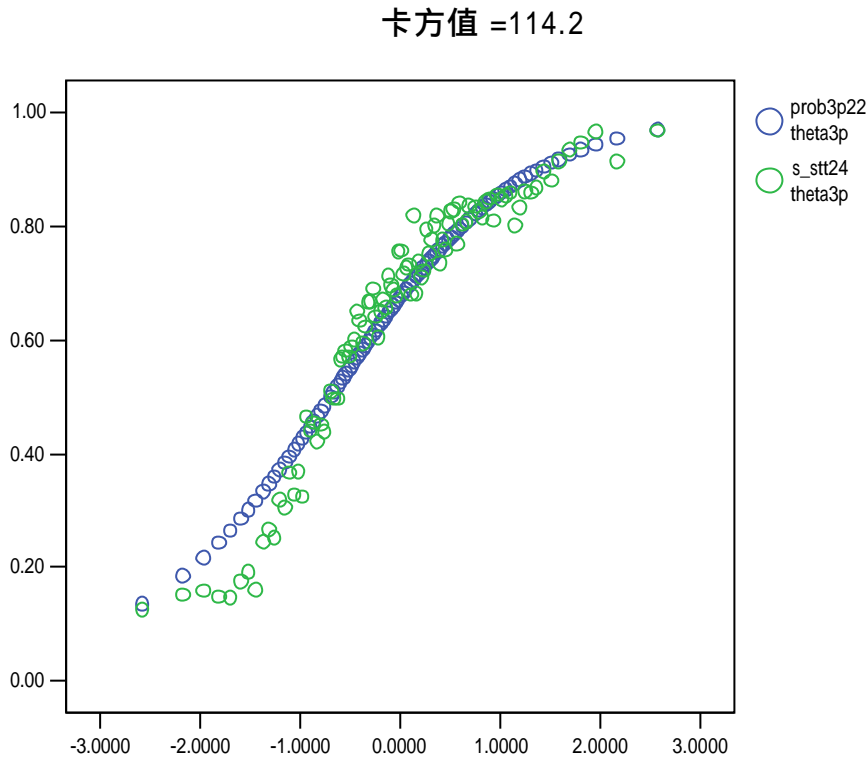


圖 4-11 高中職五專一般分析能力第二十四題 理論上的答對機率曲線和實際答對機率的比較圖

4.4 國中樣本的模型適合度

國中樣本也有類似高中職五專樣本的情形，1PL 的模型適合度都比 3PL 為不佳，卡方值都顯得頗大。1PL 中，卡方值超過 500 的題項有 22 個，其中 7 個超過 1000，最大值是 2164.5；卡方值小於 100 的只有 13 個題項。相對地，3PL 中，卡方值超過 100 的有 7 個，其中最大值是 146.7，卡方值都不高，因此國中測驗的 3PL 模型適合度也算大致良好。以下列出這七個題項理論上的答對機率曲線和實際答對機率的比較圖。

國中樣本中，卡方值最高的是數學分析能力第八題，圖 4-12 即是此題的比較圖。由於每一格代表 .20 的機率，使得 圖 4-12 的實際答對率顯得密集於理論上的答對率曲線上，看似比圖 4-13 的模型適合度較佳；但這是因為 圖 4-13 中每個刻度代表.10 的機率，使得資料點之間的差距顯得較大。但儘管如此，數學分析能力第八題的答題行為模式大致還符合 3PL 的預測。圖 4-13 則顯示，模型稍微低估了極低與極高能力者的答對機會，而稍微高估了中等能力者的答對機會。圖 4-14 與圖 4-16 情況類似，都顯示，由於中等能力者在這個題項（一般分析能力第十一題與科學分析能力第一題）的答對機率變異量頗大，致使模型高估了極低能力者的答對機會，而稍微低估了極高能力者的答對機率。

圖 4-15（一般分析能力第三題）的情形則與高中職五專語文分析能力第二題（圖 4-10）的情形類似，除了能力在兩極端者之外，其他能力程度者答對的機率並不一定隨著能力升高而升高。一般分析能力第三題的內容，是要求受測者從一段文章中猜測出隱藏的詞。可能由於選項中的其中一個詞與正確答案稍微類似，因此必須平均能力為中上（能力估計值約為 1.5 以上）者，答對的機率才確定節節往上升。

圖 4-17 顯示，科學分析能力第八題對於中等以上能力（約 .5 以上）者頗具區辨力，但中等以下能力者的答對率卻與能力呈現類似曲線（curvilinear）相關。圖 4-18 顯示，語文分析能力第十四題的答對率並不隨綜合分析能力增加而增加，而且它算是頗為困難的題目：絕大部分能力組的受測者，答對率都小於 .50。察看題目內容，發現這一題在考驗動名詞的用法。也許對於國中一年級學生而言，對於英文還限於字義與文句上的掙扎瞭解，對於較深入的文法還無法掌握，因此造成整體能力與這一題的答題表現之間沒有預期的正相關。

模型適合度檢驗：數學分析能力第八題

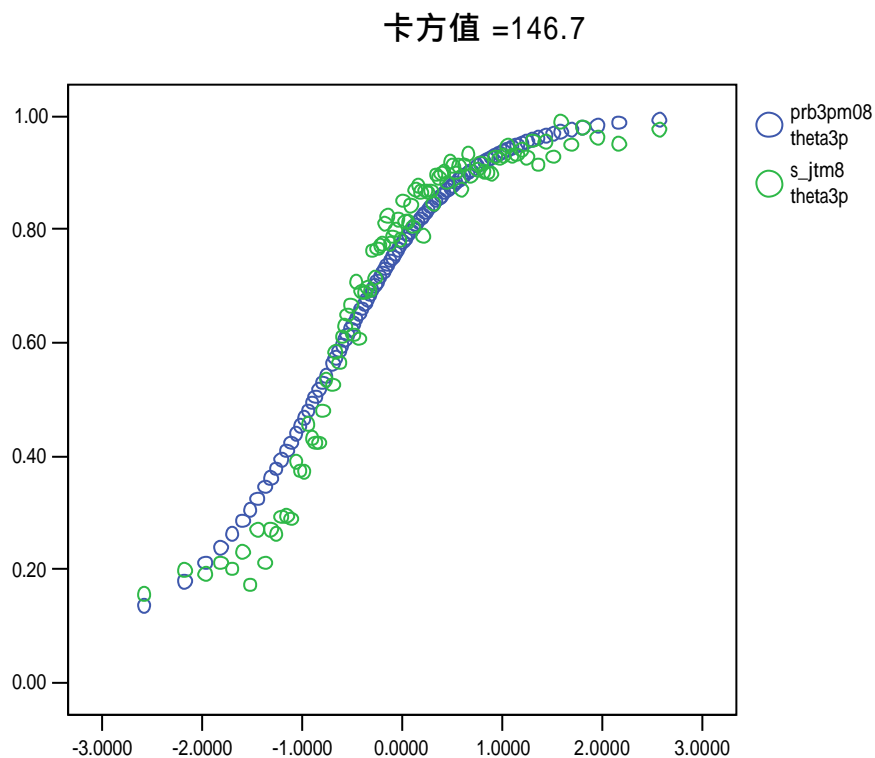


圖 4-12 國中數學分析能力第八題 理論上的答對機率曲線和實際答對機率的比較圖

模型適合度檢驗：一般分析能力第二十一題

卡方值 =137.5

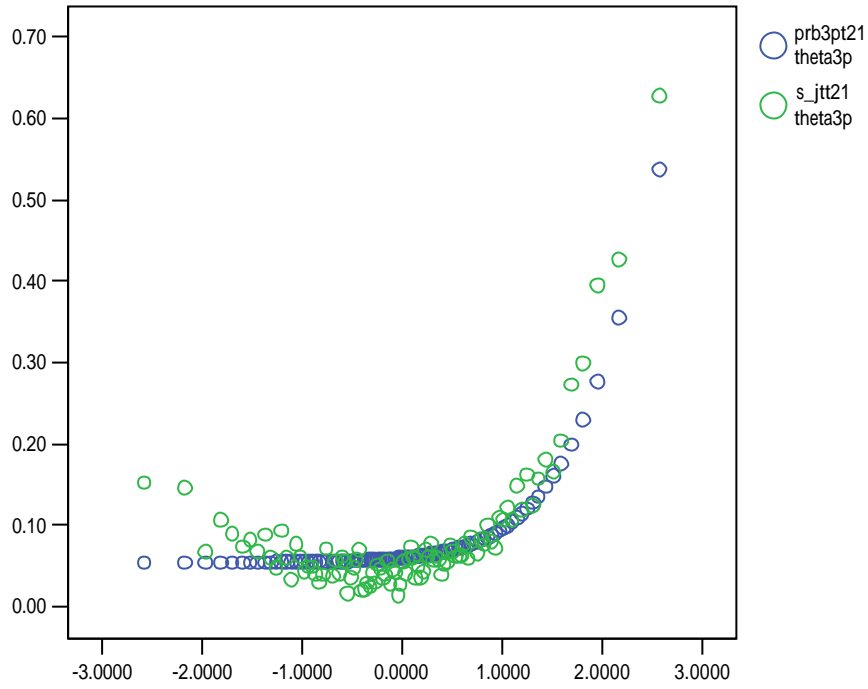


圖 4-13 國中一般分析能力第二十一題 理論上的答對機率曲線和實際答對機率的比較圖

模型適合度檢驗：一般分析能力第十一題

卡方值 =130.3

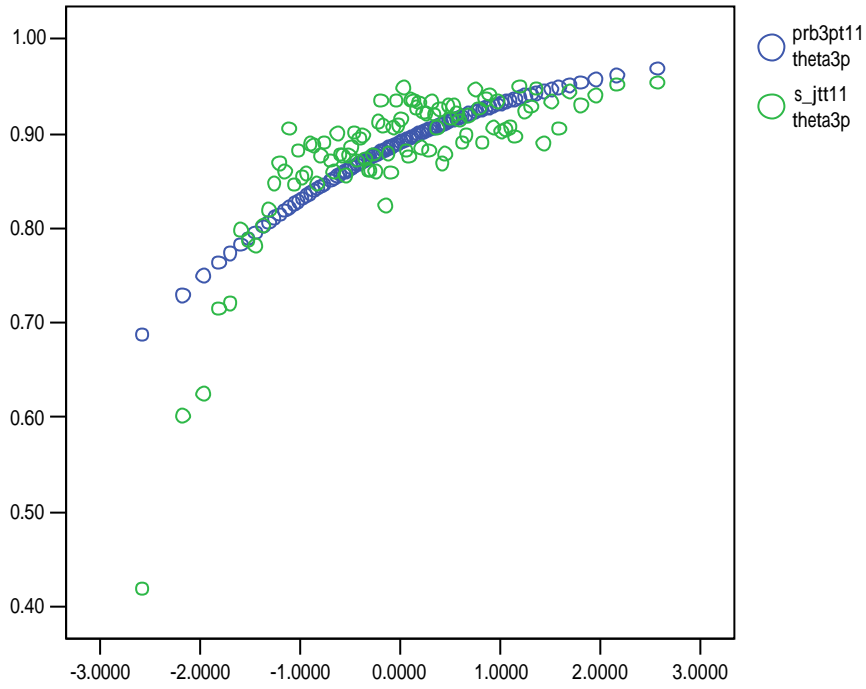


圖 4-14 國中一般分析能力第十一題 理論上的答對機率曲線和實際答對機率的比較圖

模型適合度檢驗：一般分析能力第三題

卡方值 =124.1

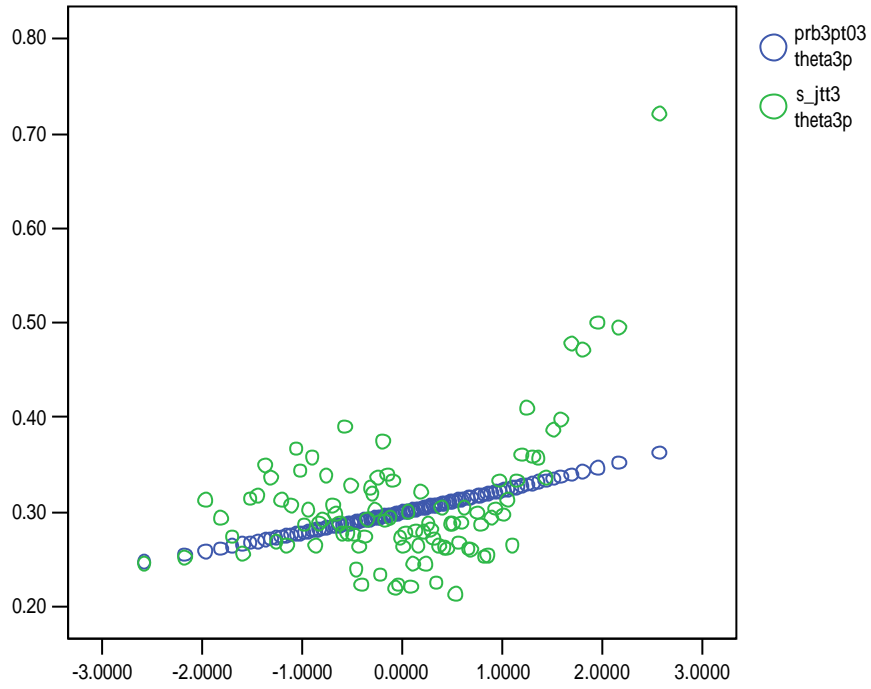


圖 4-15 國中一般分析能力第三題 理論上的答對機率曲線和實際答對機率的比較圖

模型適合度檢驗：科學分析能力第一題

卡方值 =132.4

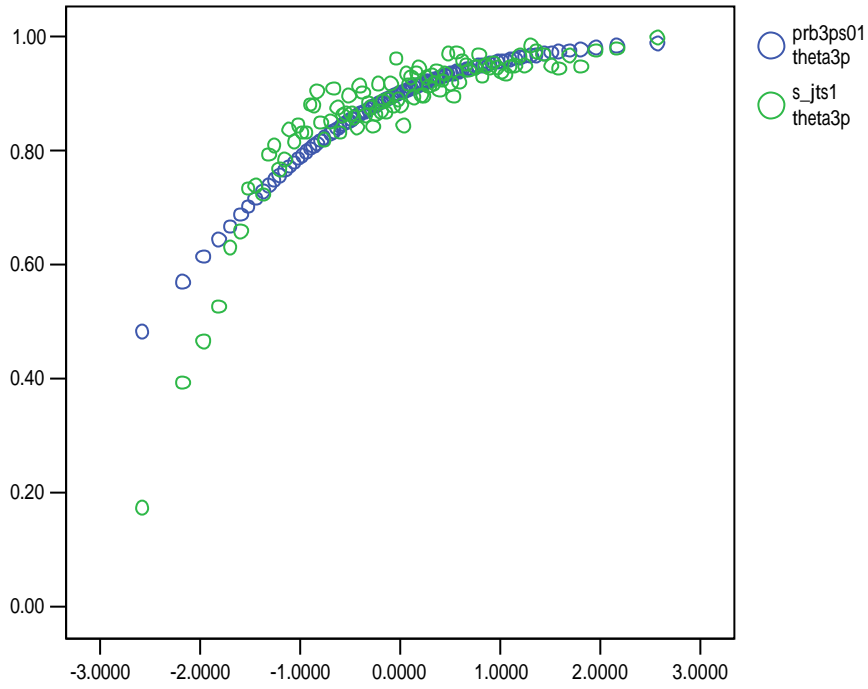


圖 4-16 國中科學分析能力第一題 理論上的答對機率曲線和實際答對機率的比較圖

模型適合度檢驗：科學分析能力第八題

卡方值 =121.0

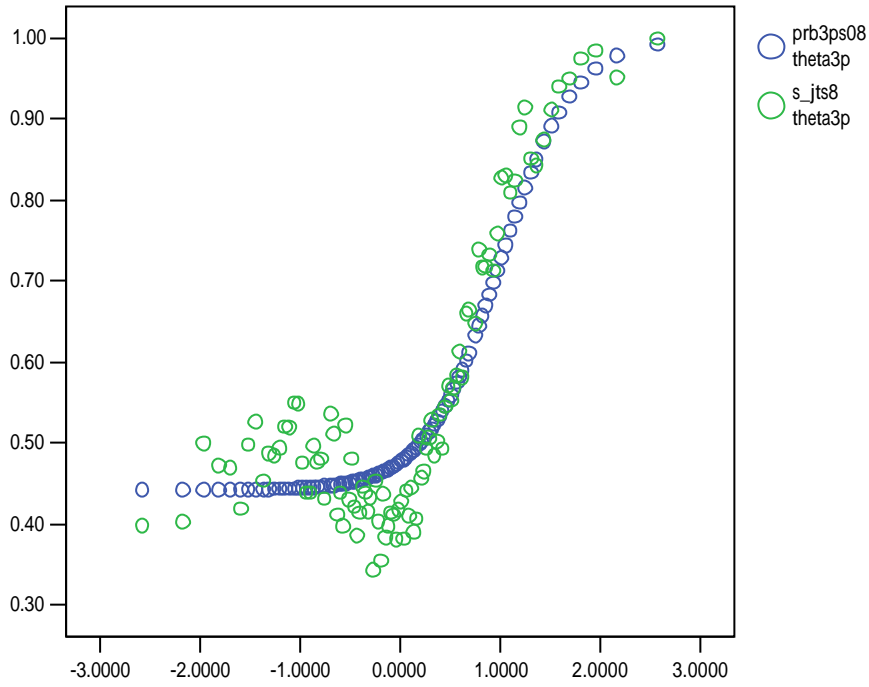


圖 4-17 科學分析能力第八題 理論上的答對機率曲線和實際答對機率的比較圖

模型適合度檢驗：語文分析能力第十四題

卡方值 =108.1

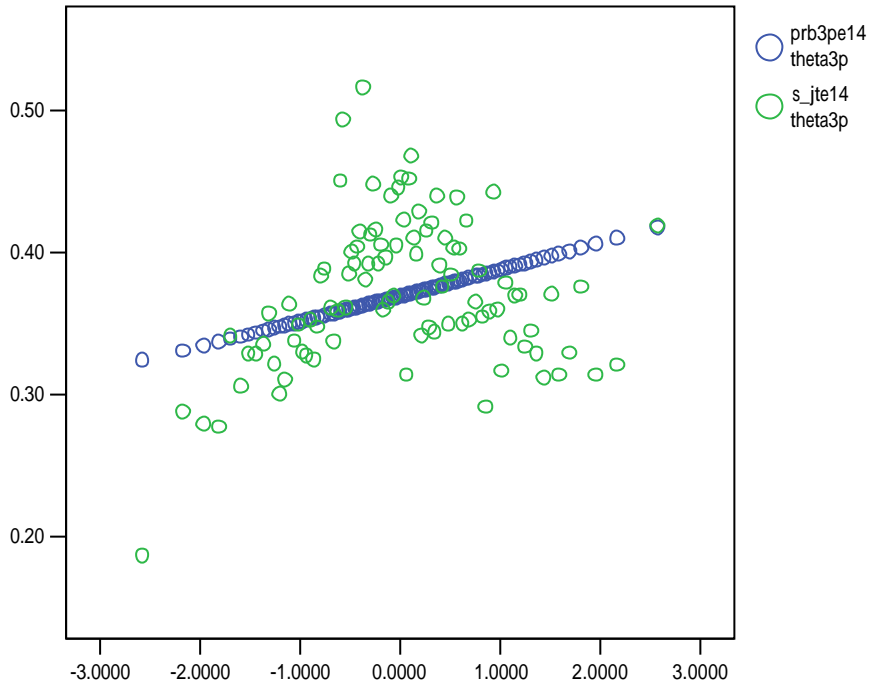


圖 4-18 語文分析能力第十四題 理論上的答對機率曲線和實際答對機率的比較圖

第五章 題目的內部效度 (internal validity)

具有內部效度(internal validity)的測驗，才能有效反應學生在所測量的領域之能力程度。本章討論兩個可能影響資料庫分析能力測驗的效度之因素：(1) 受試者分析能力以外的特質是否影響題項的答對機率 (DIF)；(2) 學生的答題意願。

當然，絕大部分的學生必須有足夠的時間完成所有的題目，也是內部效度的構成的重要因素之一，而我們在第二章已經確定，百分之九十八以上的學生都可以在限定的時間之內作完題目，因此速度不會對本次測驗的內部效度造成威脅。此外，內容效度 (content validity) 也是內部效度的必要條件，但是內容效度的確認則必須經由相關領域專家作內容上的檢核；而本次所使用的題目，若非在建立題庫時，即請該領域老師出題，並經預試測試；否則即改寫自國際性質的學生能力測量題庫，因此內容效度應不致成為問題，故本章亦不再討論。

本章的組織為：先介紹 DIF 的概念及我們的分析方法，再分別介紹高中職五專 (5.1) 及國中 (5.2) 在這方面的結果；最後簡單介紹答題意願方面的分析結果，也是分為高中職五專 (5.3) 與國中 (5.4) 兩部分。

DIF: 受試者其他的特質是否影響答對機率

題目的情境內容，必須對同等能力的每一種學生群體 (男女生、城鄉別等) 而言，都是同樣熟習，才不致因為不熟習題目中某些與主要測量目標無干的詞句，而影響某群體答對該題的機會。如果由於與主要測量目標無關的詞句之干擾，而影響了某群體的受試者在某一題的答對機會，測驗的內部效度就會降低。測量目標的能力相當的群體，在某題項的表現不相等時，該題項所測量反應的，就不止於測量者心中的測量目標，而可能摻雜了其他的「雜音」。例如：測量目標是數學分析能力，但題目包含球賽的用詞，這可能使男生在答題上佔優勢，因為男生一般對體育活動比較有興趣，比較瞭解球賽的規則。這種因某些非關測量目標本身的内容，而使能力相同但背景不同的受試者有不相等的答對機率之情形，稱為 DIF(differential item functioning)。DIF 會降低測驗的內部效度，是必須盡可能避免的。

然而，有時不同的群體的確在某些能力 (測量目標) 上有所不同，例如男生在數學或數字型問題的分析能力較強，而女生在語文方面較強，是一些跨國研究都發現的情況(UNESCO, 2000)。因此不同的學生群體在相同題項上的表現不一，可能因為題目的確含有「雜音」，有 DIF，但也可能是不同群體在該題內容的能力的確有差異，而不應以題目有「雜音」視之。因此，DIF 的判斷，除了統計數字的判讀之外，還必須瞭解實證研究中，某些群體是否的確在某些能力上有差異，也必須探討題項內容本身是否的確含有「雜音」。

分析方法

受測者在測量目標上的能力高低，很難有客觀而公正的評斷。測驗本身的總答對題數，則是最貼近事實而方便的標準；而且利用總答對題數來作統計上的控制，也符合一般探討 DIF

的方式。

我們使用「固定效果模式」(fixed-effects model)，將答對總分相同者視為同一個群體，納入固定效果中，並利用 robust variance 解決總分相同的群聚效應 (clustering effect) 而使標準誤較小的問題。我們以探討同等能力的男女生在題項 j 的表現是否不同 (DIF) 為例，說明本篇報告用以討論 DIF 的分析方法。所使用的方程式如下：

$$Y_{ijk} = b_{0j} + b_{1j}(\text{girl})_i + u_{jk} + e_{ijk} \quad (5-1)$$

其中 Y_{ijk} 是在第 k 個總分層的受試者 i 在題項 j 的得分 (答對得 1 分, 答錯得 0 分) 因為 Y_{ijk} 的值只有兩種, 非 0 即 1, 理想上最適宜使用邏輯迴歸 (logistic regression)。然而邏輯迴歸的實際依變項是 logit, 不但使係數 b_{1j} 不易詮釋, 而且不同模型間 (例如第 j 題與第 $j+1$ 題的分析模型) 的係數無法比較 (譚康榮, 2002)。此外, 使用邏輯迴歸在作估算時會產生一個大問題: 當某個總分層的所有人都答對或答錯某一題時, 該題在該總分層的所有樣本都必須排除, 不能納入計算, 致使較容易的題目所能使用的樣本不足所有樣本的一半, 而使各題項所得出的估算值難以比較。所以我們使用線性機率的模型, 直接以線性迴歸 (OLS) 作分析, 但以 robust 標準誤 取代線性迴歸中的標準誤。線性機率模型的係數易於詮釋: 其截距值即是控制了所有其他變項之後, $Y_{ijk} = 1$ 的平均機率。而自變項的係數則是: 當該自變項改變一個單位時, 依變項為 1 的相對機率之改變。我們也作了邏輯迴歸, 其所產生的係數與標準誤的模式與本文所用的線性機率模型所產生的模式極端相似。

方程式右方的其他變項與係數意義如下: 如果受試者是女生則 $(\text{girl})_{ik} = 1$, 否則 $(\text{girl})_{ik} = 0$; u_{jk} 即是所謂的固定效果, 吸納了所有答對總題數對於該題答對機率的影響; e_{ijk} 則是實際觀察所得的值與模型估計值之間的誤差。因此, b_{0j} 是男生答對題項 j 的預期機率 (predicted probability), 也就是男生答對該題的平均機率, 而 b_{1j} 則是女生比男生, 在答對該題的期望機率上的差異; b_{1j} 為正值, 表示女生在該題的答對機率比男生高, b_{1j} 為負值, 表示女生在該題的答對機率稍低。 b_{1j} 是我們探求的目標: 若能力相同, 而題目內容沒有偏誤, 則 b_{1j} 應該為零, 或極近於零。但如果某一題的 b_{1j} 不為零, 則有可能是男女之間在該題所測量的能力上有所不同, 或題項的確有偏誤 (DIF)。

5.1 高中職五專

高中職五專學生的入學，是經由聯考的篩選機制而決定；位於城鎮的學校，其學生不一定來自城鎮，反之亦同。而資料庫中的城鄉變數是根據學校位置而定，並非學生的生長受教環境，因此無法以高中職五專的樣本，探討高中職五專的測驗是否因城鄉別而有 DIF。

以下我們探討，若學生在測量目標的能力相同，高中職五專的測驗是否因為學生的性別差異而有不同的答對機會？也就是：能力相同的條件之下，男女生答對題項的機會，有多大的差異？

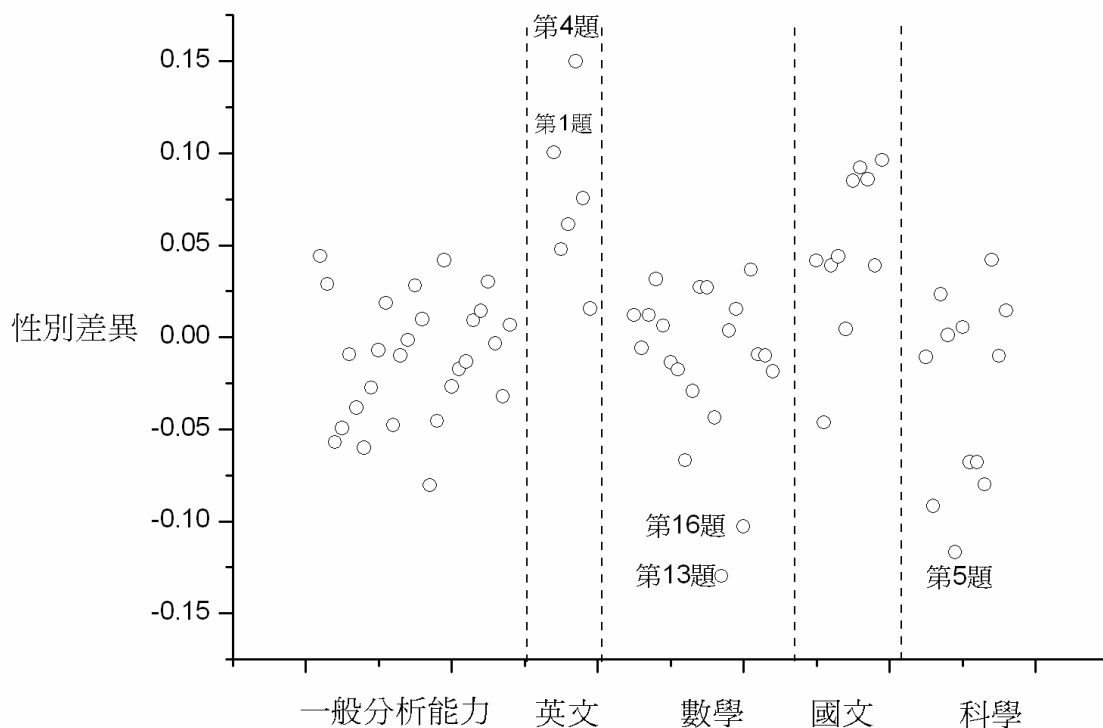


圖 5-1 控制總答對題數後，高中職五專男女生在各題答對機率之差異 (b_{1j})

圖 5-1 是 方程式(5-1)中的 b_{1j} 值。橫軸是各題項，以虛線分隔出不同領域的測驗題；

縱軸則是 b_{1j} 值。圖 5-1 顯示：大多數題項的 b_{1j} 值都落在 -0.05 與 0.05 之間，少部分落在 0.05 與 0.10 之間或 -0.05 與 -0.10 之間，極少一部份則落在這個範圍之外；圖 5-1 的明顯情形是：如果總答對題數相同，男女生答對各題項的機率差異並不大：在一般分析能力方面，男女生答對機率的分佈頗為平均；英文與國文領域方面，女生的答對機率都比男生稍高；數學與科學方面，則男生的答對機率比女生稍微高一些。但這些差異都不大，男女生在大多數題項的答

對機率差異不超過 .10，除了五個題項以外。這五題分別是英文題目中的第一與第四題（在總得分相同的情況下，女生答對該兩題的機率比男生分別高約 .10 與 .15 左右）；數學題目中的第十三題與第十六題（在總得分相同的情況下，女生答對該兩題的機率比男生分別低約 .13 與 .10 左右）；及科學題目中的第 5 題（在總得分相同的情況下，女生答對該題的機率比男生低約 .12 左右）。

經檢查題項內容，英文第一題考的是文法概念，第四題則是閱讀測驗的一部份；數學第十三題的情境是園遊會中轉盤的應用問題，第十六題是以生物的繁殖為情境的應用問題，科學第五題則是物理動力學方面的問題。這些情境都頗為「中性」，也不需特別的知識，看不出有任何因而造成男女生差異的「雜音」。而另一方面，國外做的學生能力測量，發現男女生在文理上的學習成就的確有差異（UNESCO, 2000），因此我們相信這些差異應該是反應男女生實際學習成就的差異，而非題項內容的問題（DIF）。

由於本測驗所用樣本數極大，即使極小的差異都會使顯著檢定（t-檢定）顯得大而顯著，因此 b_{1j} 的 t-檢定可說沒有什麼價值。所以本文不報告 t-檢定。

5.2 國中

我們探討國中樣本裡，性別及城鄉別是否影響學生的答對機率。

國中樣本也同樣以線性機率模式來探討性別是否影響學生答對題項的機率，利用固定效果模式來控制能力及其他因素造成的差異，利用 robust variance 來解決（能力）同組者變異量小而影響標準誤的問題。所使用的模型亦與方程式（5-1）相同。

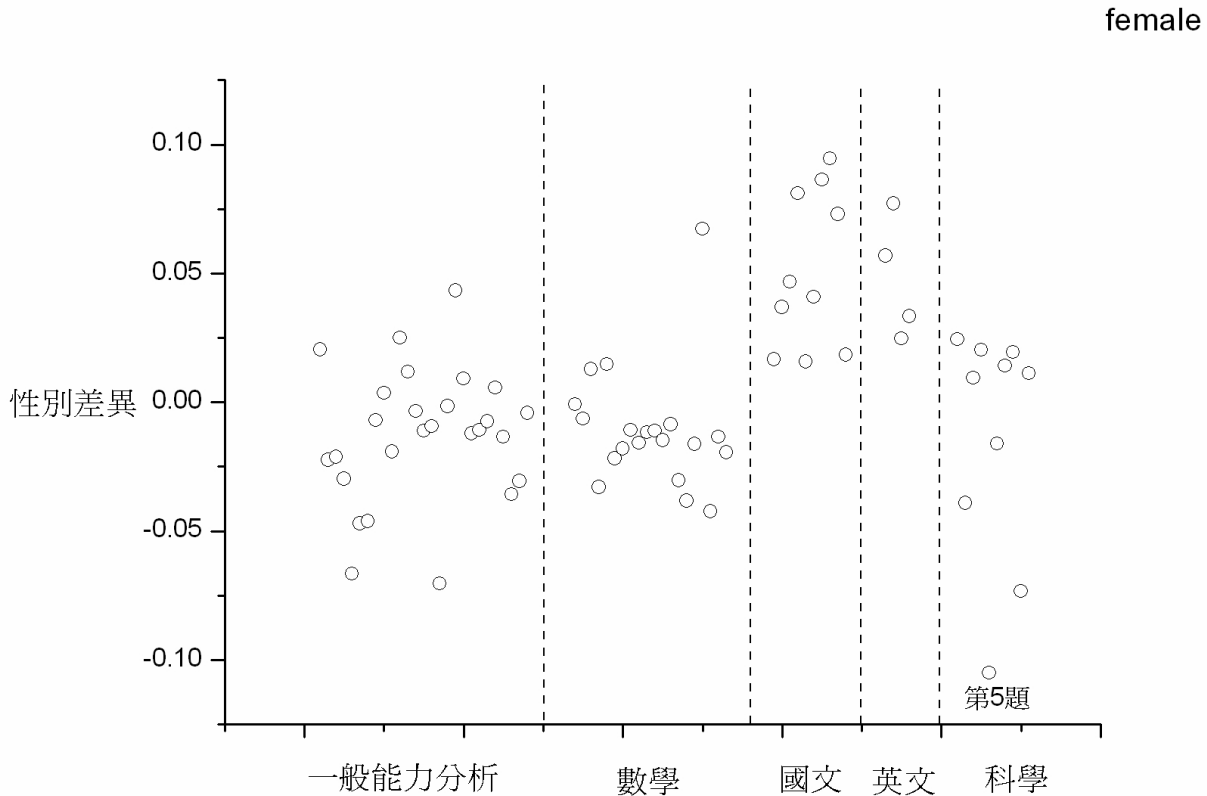


圖 5-2 控制總答對題數後，國中男女生在各題答對機率之差異 (b_{1j})

圖 5-2 與圖 5-1 相同之處在於：控制了總答對題數後，女生的與文科（國文與英文）的答對機率都比男生稍高，數學與科學比男生稍低，而一般分析能力方面的差異較不明顯。因此在科目上的表現，男女生之間的差異，似乎在國中階段就已經出現，而且持續至高中。相較於高中職五專的結果，答對題數相同的國中男女生，在各題答對率上的差異比較小。大部分的差異都在 -0.05 到 0.05 之間，其他則在 -0.05 到 -0.10 之間或 0.05 到 0.10 之間，唯有科學部分的第 5 題，男女差異稍大於 0.10 。

檢查科學部分第 5 題的內容，發現它測量的純粹是關於生物為了要適應環境而如何演化的問題，沒有任何假定的情境。因此這題的差異很可能是因為男女生的興趣不同（男生喜歡科學、探險，女生喜歡人文學科）而造成答對機率的的不同，並非與測量目標不相干的內容影響所致。

此外，由於國中生的入學是以學區為主，學校所處的大環境（都市、城鎮或鄉鎮），也是學生的生活與其他所有學習活動發生的場所。因此我們利用此機會探討：若學生在測量目標的能力相同，他們在各題項的答對機會，是否因都市化程度不同而有差異？

資料庫中，學校所在區域的都市化分三級：都市、城鎮與鄉村。我們探討：

(1) 若答對題數相同，都市或城鎮的學校學生與鄉村的學生相比較，兩者在各題的答對機會是否有差異？

(2) 若答對題數相同，都市學校的學生與鄉村或城鎮的的學生相比較，兩者在各題的答對機會是否有差異？

固定效果模式的方程式與 方程式(5-1) 類似，僅把虛擬變項($girl$)_{*i*} 分別改成 ($city/town$)_{*i*} 及($city$)_{*i*}。新的方程式如下：

$$Y_{ijk} = b_{0j} + b_{1j}(city/town)_i + u_{jk} + e_{ijk} \quad (5-2)$$

$$Y_{ijk} = b_{0j} + b_{1j}(city)_i + u_{jk} + e_{ijk} \quad (5-3)$$

因此 方程式(5-2) 中的 b_{1j} 是當總答對題數相同時，都市或城鎮的學生相較於鄉村學生，在題項 j 答對機會上的差距；方程式(5-3) 中的 b_{1j} 則是當總答對題數相同時，都市學生相較於城鎮或鄉村學生，在題項 j 答對機會的差距。

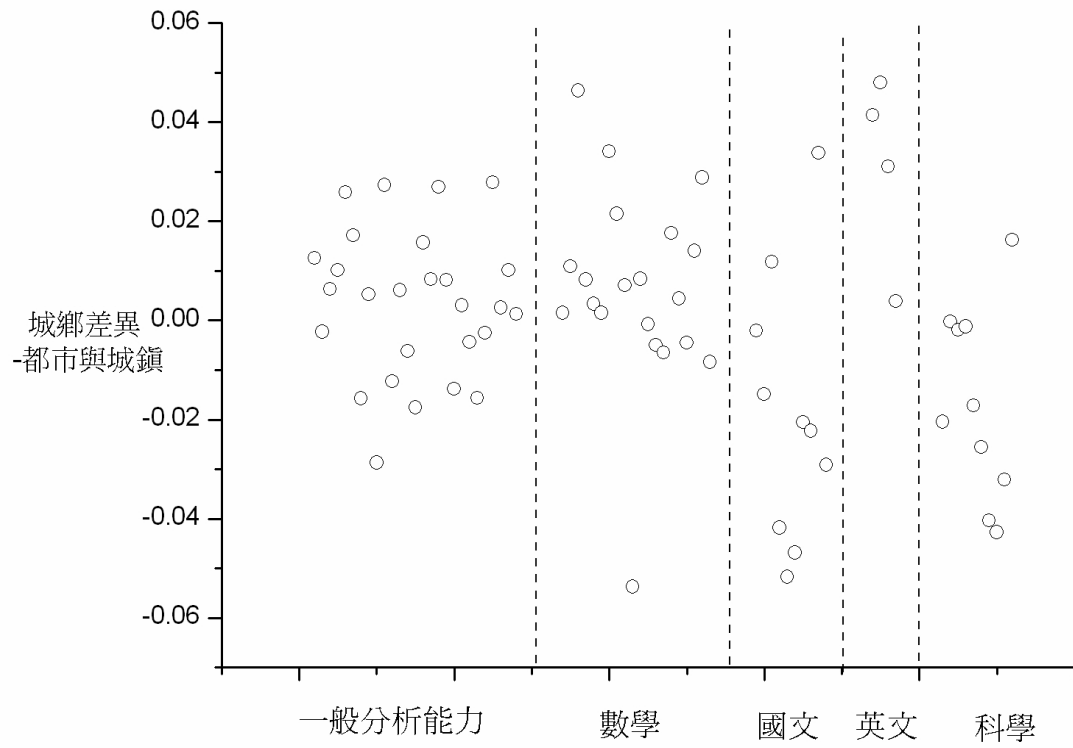


圖 5-3 控制總答對題數後，國中樣本中，都市或城鎮的學生相對於鄉村學生，在各題答對機率上之差異 (b_{1j})。

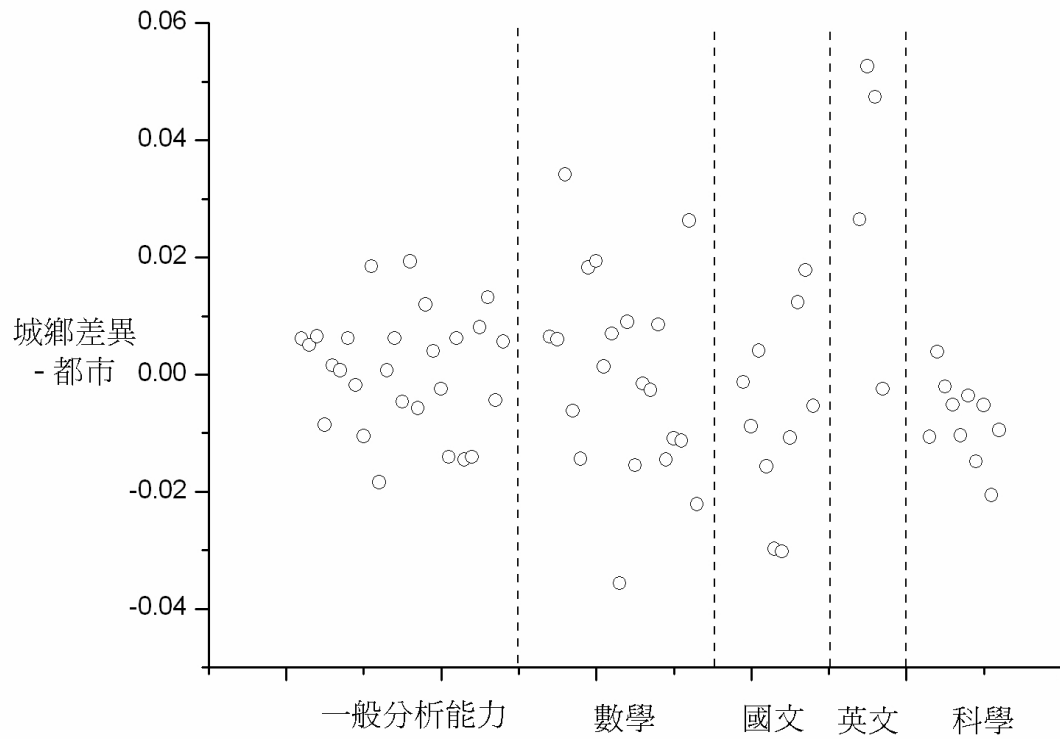


圖 5-4 控制總答對題數後，國中樣本中，都市的學生相對於城鎮或鄉村學生，在各題答對機率上之差異 (b_{1j})。

圖 5-3 與圖 5-4 的型態頗為類似。在答對題數相同的條件下，城市（都市或都市加城鎮）學生答對英文題的機會普遍比鄉鎮學生高，而鄉鎮學生則在科學題的答對機會較高；其他科目的題目則沒有明顯的差異。但這些都只是相對的差異，實際的差距頗為微小：若以都市或城鎮相對於鄉村而言，大部分的差異集中在 -.04 到 .04 之間；若以都市相對於城鎮與鄉村而言，大部分的差異都在 -.02 與 .02 之間。這些差異的範圍都比性別的差異小。

學生的答題意願

學生作答本次調查的測驗題之意願有多高？有多少學生是在不甘願的情況下閉著眼睛瞎猜？雖然可以從學生的答案卡是否呈現明顯的圖案，或者是否每個答案都相同等，各種方式來探索這方面的問題；但限於人力，第一波的資料沒有蒐集這方面的訊息。在這裡，我們以比較粗略的方法來大略瞭解這個問題。

由於每道測驗題都有四個答項，因此若純粹亂答，答對的機會有四分之一。我們就利用答對題數作為答題意願的指標：若答對題數低於或接近 18（高中職五專部份：73/4；國中部份：71/4）題的學生比率偏高，那麼測驗的內部效度可能受到威脅。表 5-1 列出高中職五專答對一半以下題數的學生比率。

表 5-1、高中職五專學生在 73 題中答對題數的累積比率（%）

≤ 18	≤ 22	≤ 26	≤ 30	≤ 36
.92	2.89	6.80	12.67	25.06

表 5-2、國中學生在 71 題中答對題數的累積比率（%）

≤ 18	≤ 22	≤ 26	≤ 30	≤ 36
2.22	5.60	11.25	18.80	33.63

從上表得知，高中職五專學生非常可能亂猜（答對題數 ≤ 18）的學生比率約百分之一，而國中學生的比率約佔百分之二（2.22%）。由於參加測驗的學生樣本很大（高中職五專學生有 19232 人；國中生樣本有 19957 人），我們可以假定亂猜的答對題數呈常態分佈，若假定完全亂猜的答對題數之期望值為 18，則此亂猜的答對題數之常態分佈另一半（亂猜而答對 18 題以上）所佔的比率將與「亂猜而最多答對 18 題」的比率相同；從這個角度而言，高中職五專學生亂猜的比率還算可接受的範圍（.92% + .92% = 1.84%），國中生的比率則稍高（2.22% + 2.22% = 4.44%），但應該也不致對內部效度構成太大的威脅。

參考資料

- 張荳雲（2002）「台灣教育長期追蹤資料庫」之簡介。「調查研究」第 12 期，頁 111-118
- 譚康榮（2002）群體差異的分析：隱性變項之難題。台灣社會學刊，第 28 期，頁 231-259。

UNESCO (2000). *Literacy Skills for the World of Tomorrow — Further Results From PISA 2000: Executive Summary*.

<http://www.pisa.oecd.org/Docs/Download/ExecutiveSummaryPISAplus.pdf>

附錄一 高中職五專 IRT 分析結果圖示

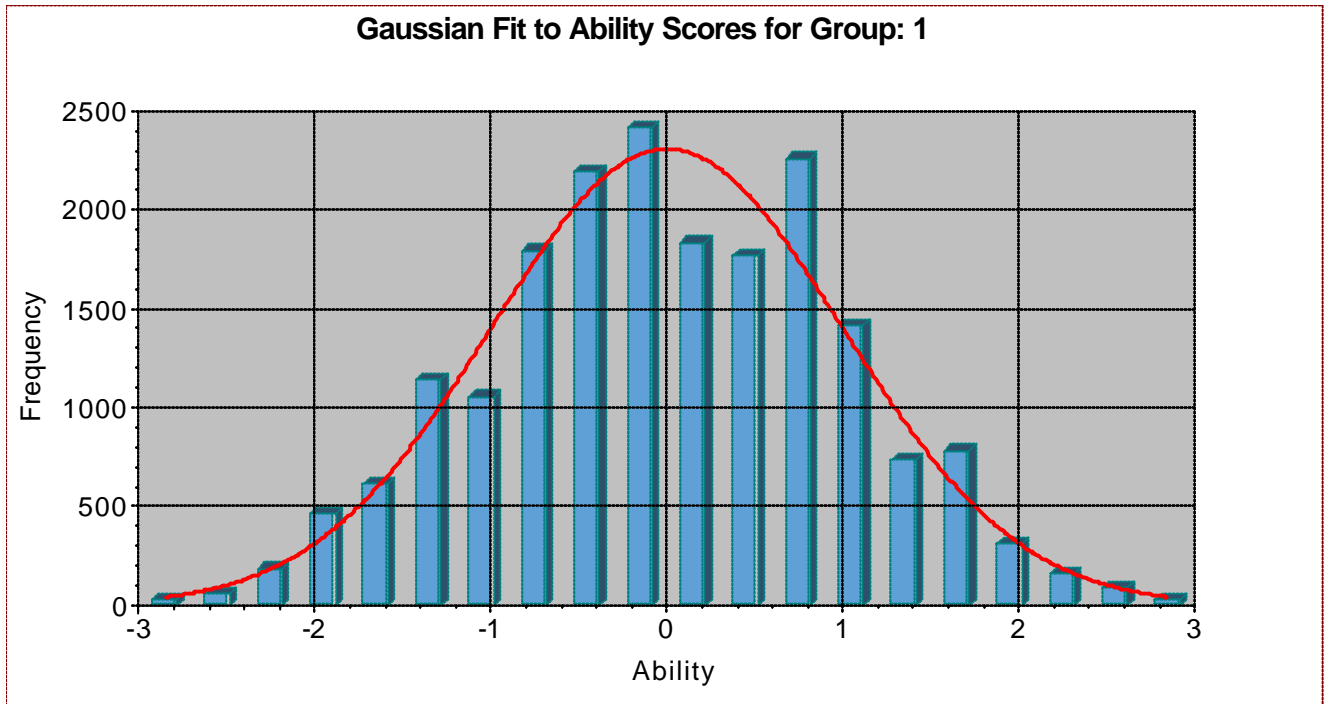
本附錄包括高中職五專樣本中，各分項測驗在 1PL 與 3PL 模式下，學生能力估計值分佈長條圖、測驗訊息 (Test Information) 曲線圖、測驗中各題項的答對機率曲線圖 (Item Characteristic Curve) 及總答對題數與能力估計值的散佈圖 (scatterplot)。

從能力估計值分佈長條圖可以看出，在 3PL 模式下的能力分佈比較趨近於常態分佈，尤其當題目較多 (綜合分析能力題) 時，3PL 的能力估計值長條分佈圖幾近於常態分佈，雖然能力高於平均值 (0) 的學生似乎稍多 (圖 附一-5)；而 1PL 模式下的分佈長條圖則稍有類似雙峰的情形；但當題數不多時，即使 3PL 模式下的分佈也與常態分佈差異頗大。

測驗在某一個能力點提供的訊息愈多，能力估計值等於該點者，該估計值的標準誤 (standard error) 也愈低。一般而言，不論是哪一種分項測驗，1PL 和 3PL 都顯示，本測驗在能力範圍為平均值左右的訊息較為豐富，但 1PL 模式下訊息較豐富的區域較集中於中間偏低的能力，而 3PL 模式下，訊息較豐富的區域則稍為集中於中間偏高的能力範圍。比較特殊的是，一般能力測驗在 3PL 模式下，測驗訊息密集於能力為 1 附近，而其餘部分訊息都極低，使得這些地方的能力估計值之標準誤急遽上升。數學分析能力題的測驗訊息在 3PL 下則呈一個雙峰的型態，一個較低峰是在能力估計值為 -2 左右，另一個較尖而高的峰則是在能力估計值為 0 左右 (圖附一-22)。數學或數字型分析能力題的訊息圖與前述數學分析能力題的訊息圖類似，但可能是因為多了九題數字型的題項，補足了能力在 -.5 左右的不足，使得該處的坡度變得平緩 (圖附一-30)。

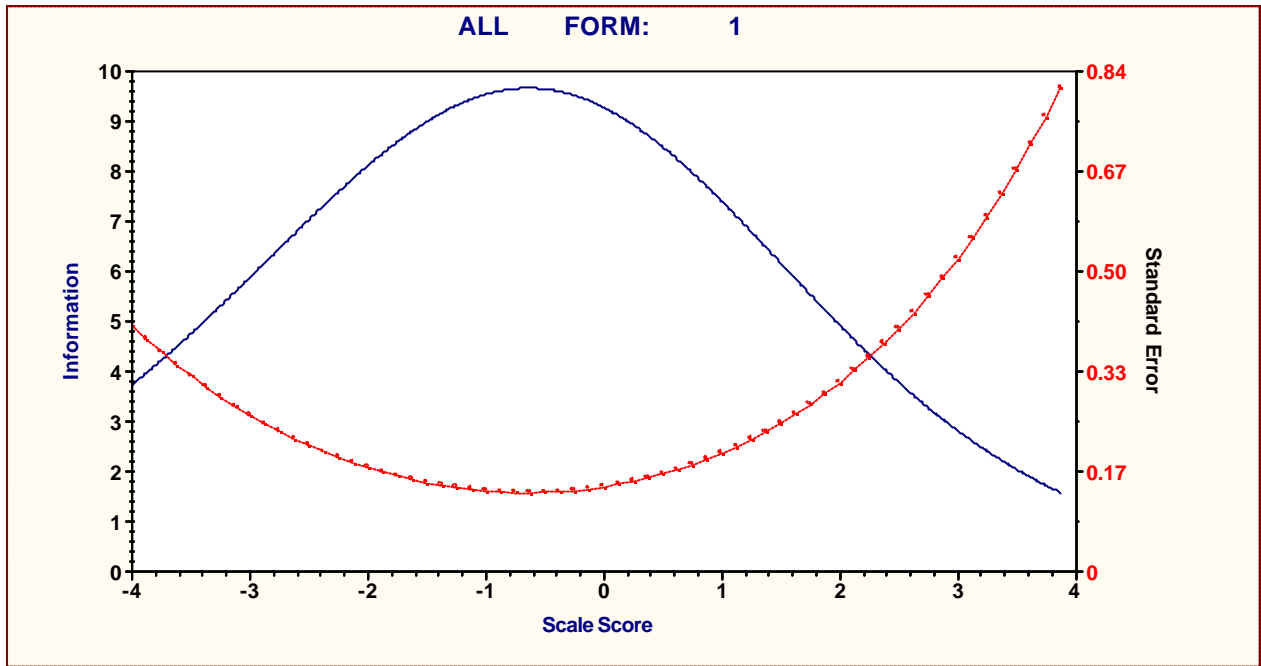
由於分析模型的不同，各題項答對機率曲線圖的形狀差異頗大。在 1PL 模式下，由於鑑別度預設為相同，因此，若答對率不致太高或太低時，同一分項測驗中的答對機率曲線圖形狀頗為相近；然而在 3PL 模式下，由於每個題項有三種參數在變化，使得每一題的答對機率曲線圖形狀差異頗大。也因此，在 1PL 模式下看似類似的題項，3PL 模式的解讀可能很不同。例如在 1PL 模式下 (圖 附一-3)，綜合分析能力題的第 31 題與第 72 題頗類似，但從 3PL 模式的答對曲線圖 (圖附一-7) 而言，第 31 題的猜對機會 (曲線與縱軸相交處) 很低，鑑別度很高 (曲線的斜率大，坡度陡)，難度也高；而第 72 題的猜對機會稍高，但難度極高，鑑別度極低，使得曲線幾乎成為與橫軸平行的直線。

至於答對題數與能力估計值的散佈圖，由於項目反應理論模型中，答對機會和能力及項目的參數之間並非簡單的線性關係，因此散佈圖呈現的，並不是直線而是平滑的曲線 (或帶狀)。而因為在 1PL 模型中，答對題數是能力估計值的充分統計量，只要答對題數相同，能力估計值就相同，答對題數和 1PL 的能力估計值之散佈圖都是一條曲線。3PL 模型則較為複雜，相同的答對題數，可能因為所答對的題項之參數值不同，而使能力估計值不同；因此散佈圖都呈帶狀。儘管如此，兩種模式下的能力估計值都和答對題數有極高的相關。



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附一-1、綜合分析能力題 1PL 模式下，學生能力分佈估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附一-2、 綜合分析能力題 1PL 模式下的測驗訊息圖

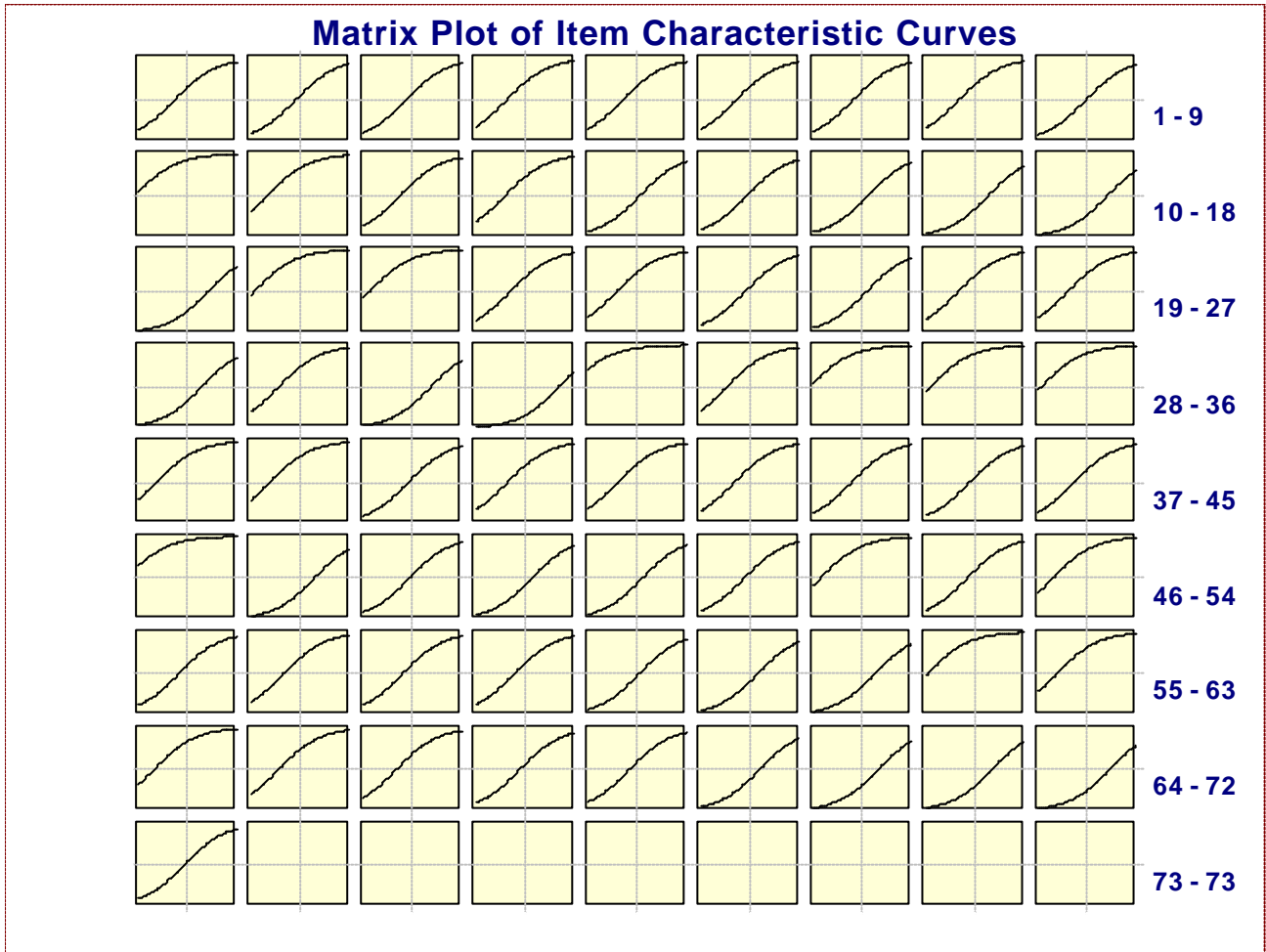


圖 附一-3、 綜合分析能力題 1PL 模式下，各題項的答對機率曲線圖

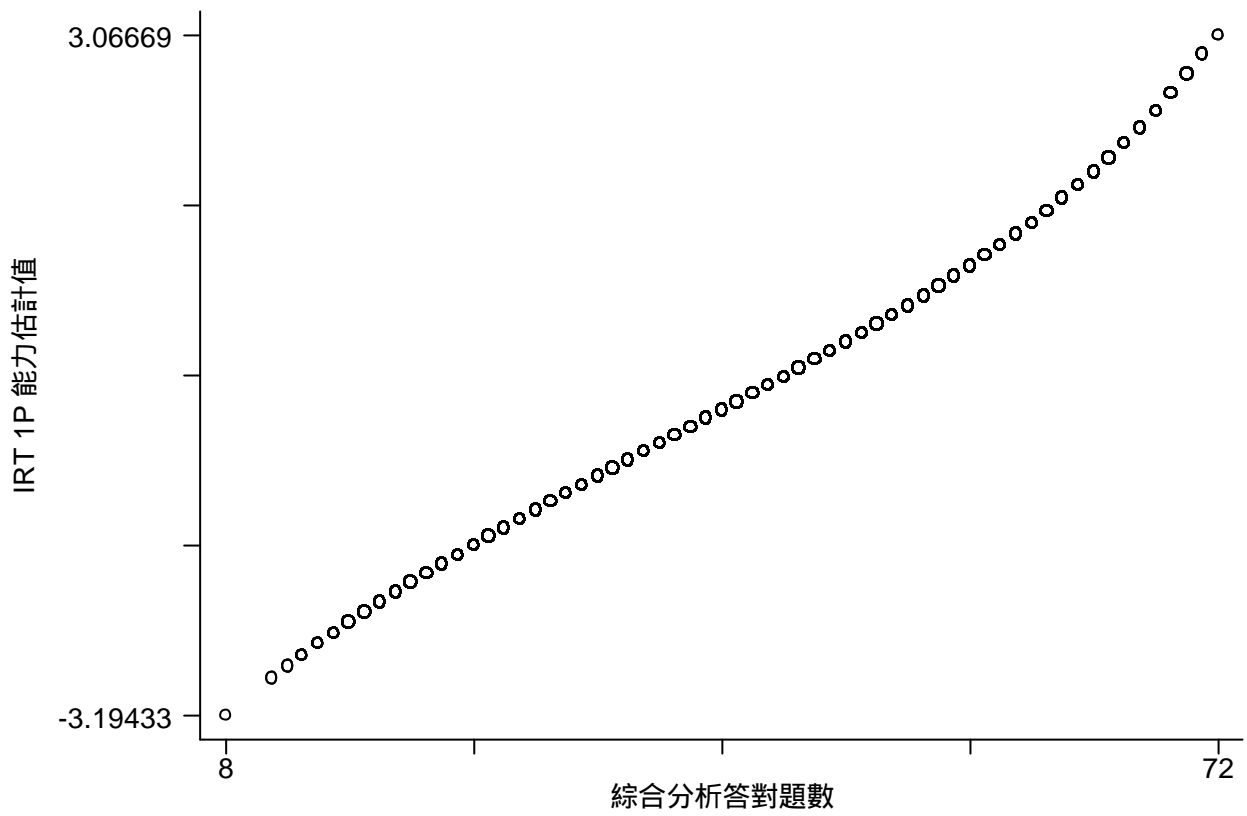
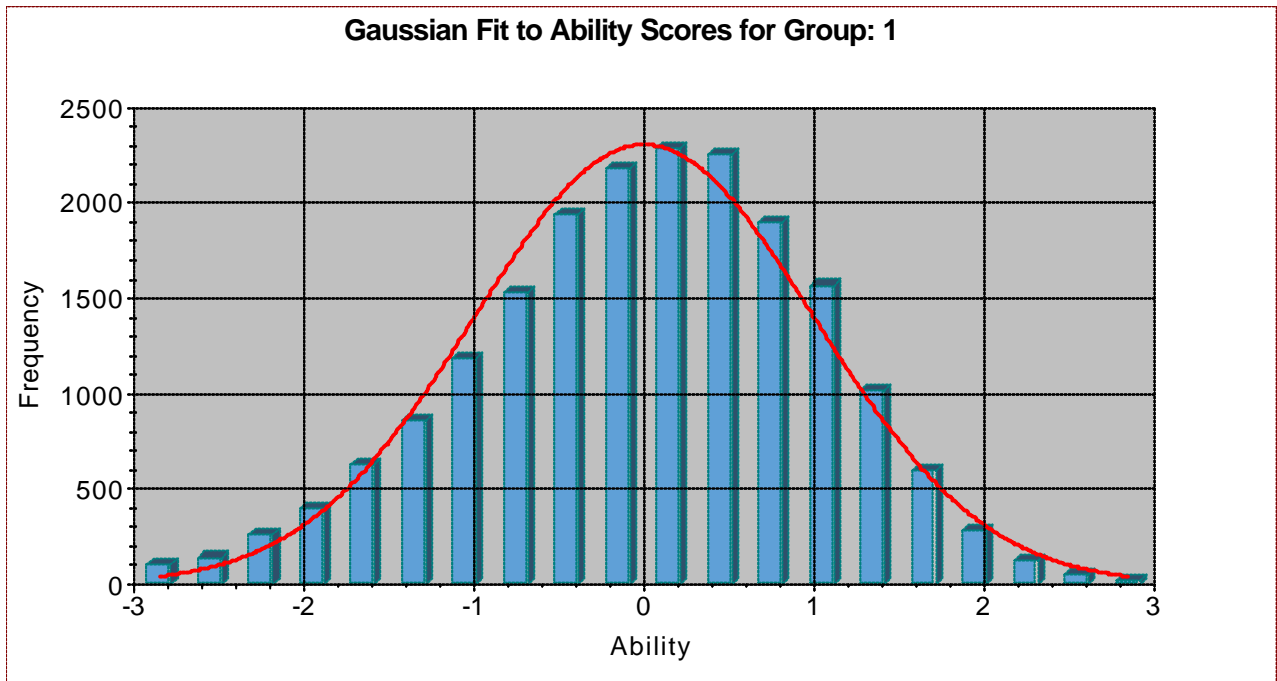
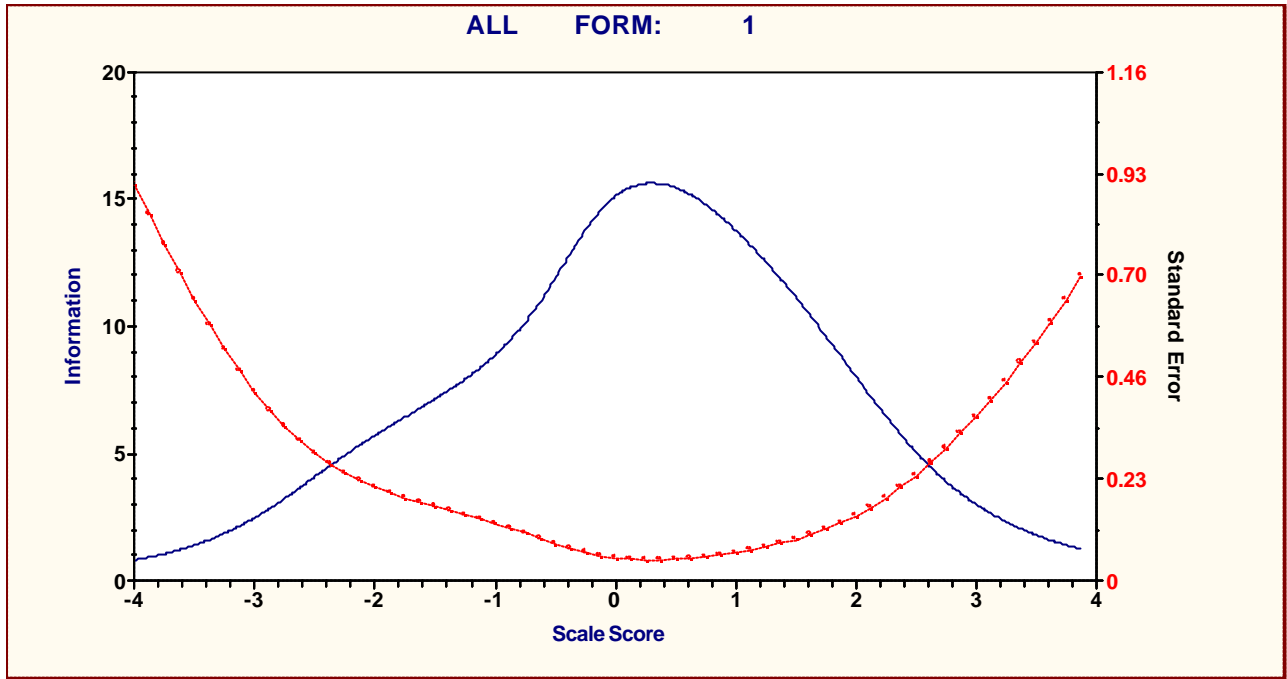


圖 附一-4、 綜合分析能力題 答對題數與 1PL 模式下能力估計值的散佈圖， $r = .997$



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附一 -5、 綜合分析能力題 3PL 模式下，學生能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附一-6、 綜合分析能力題 3PL 模式下的測驗訊息圖

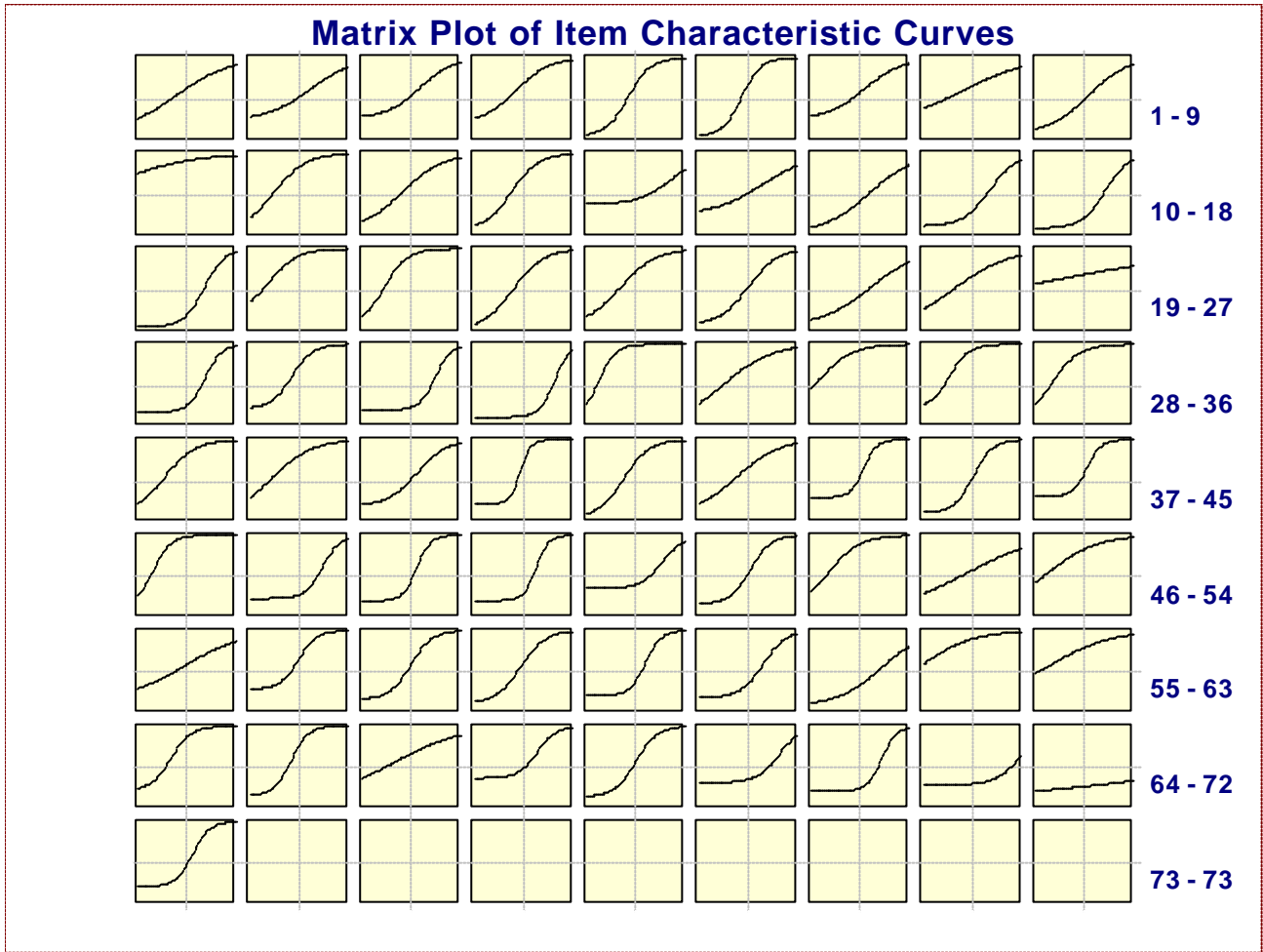


圖 附一-7、 綜合分析能力題 3PL 模式下，各題項的答對機率曲線圖

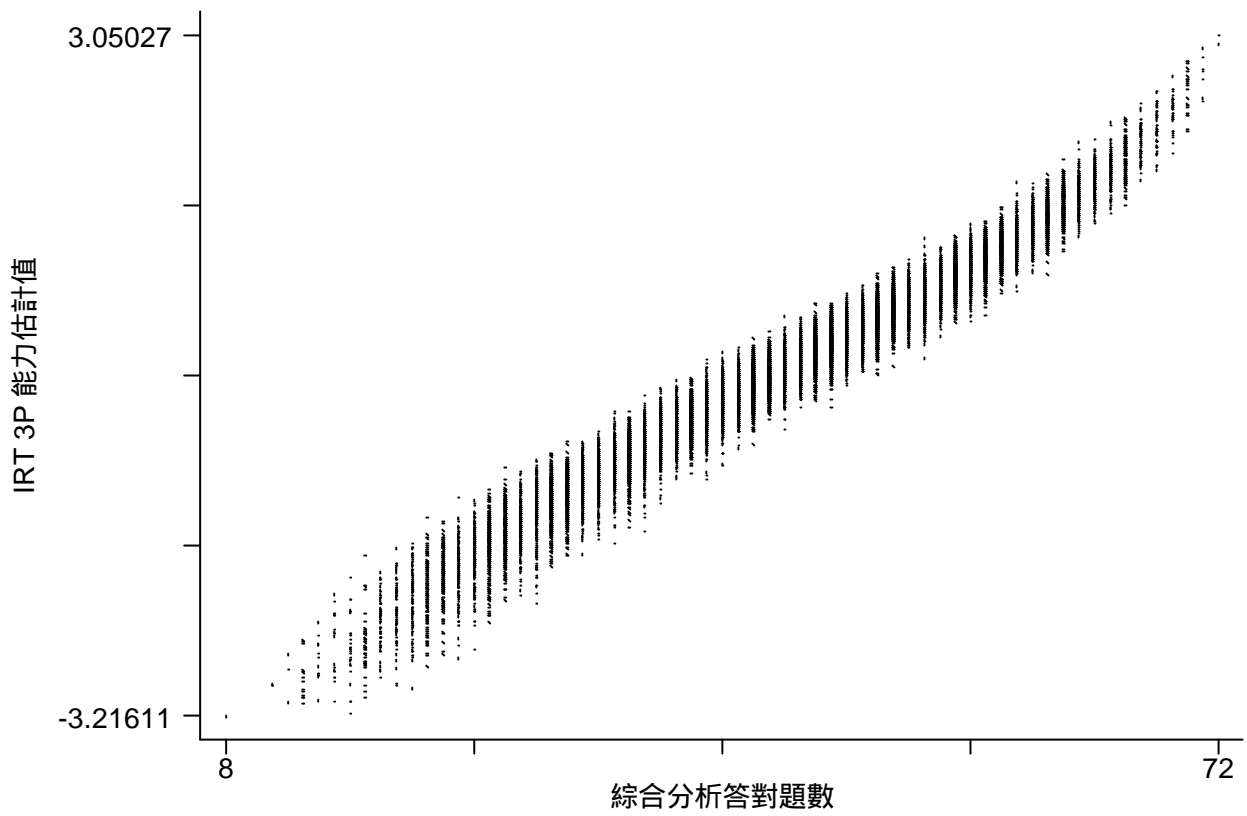
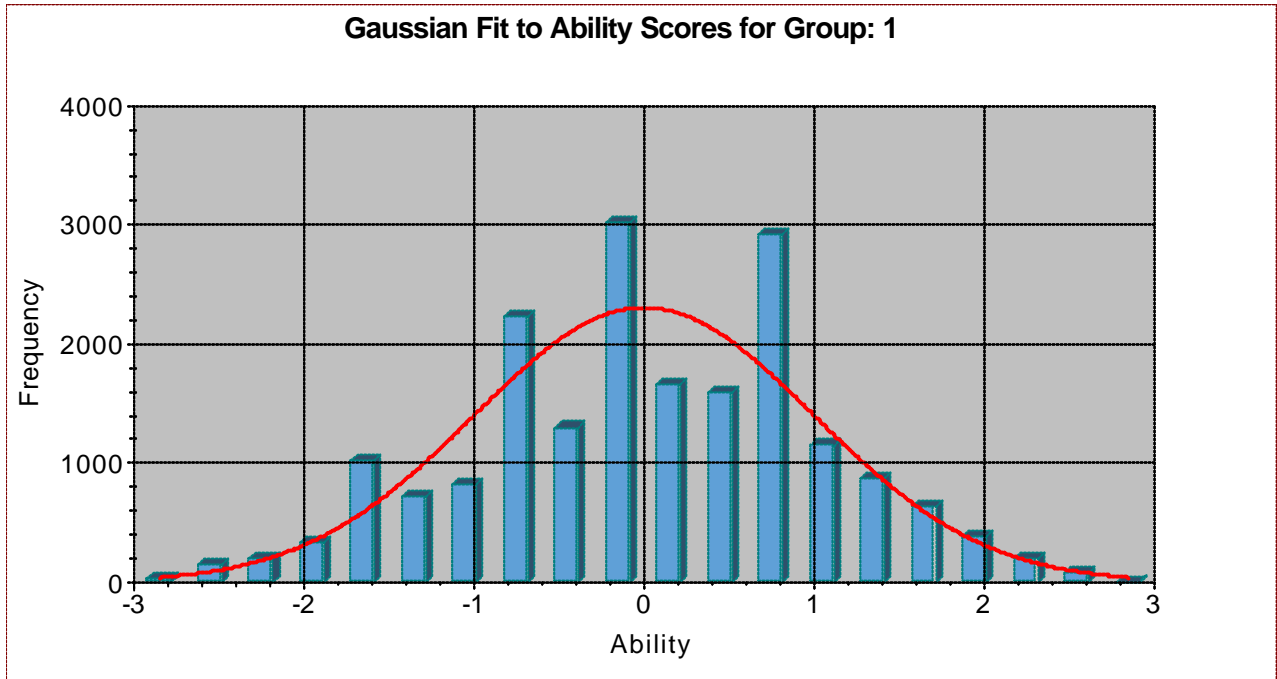
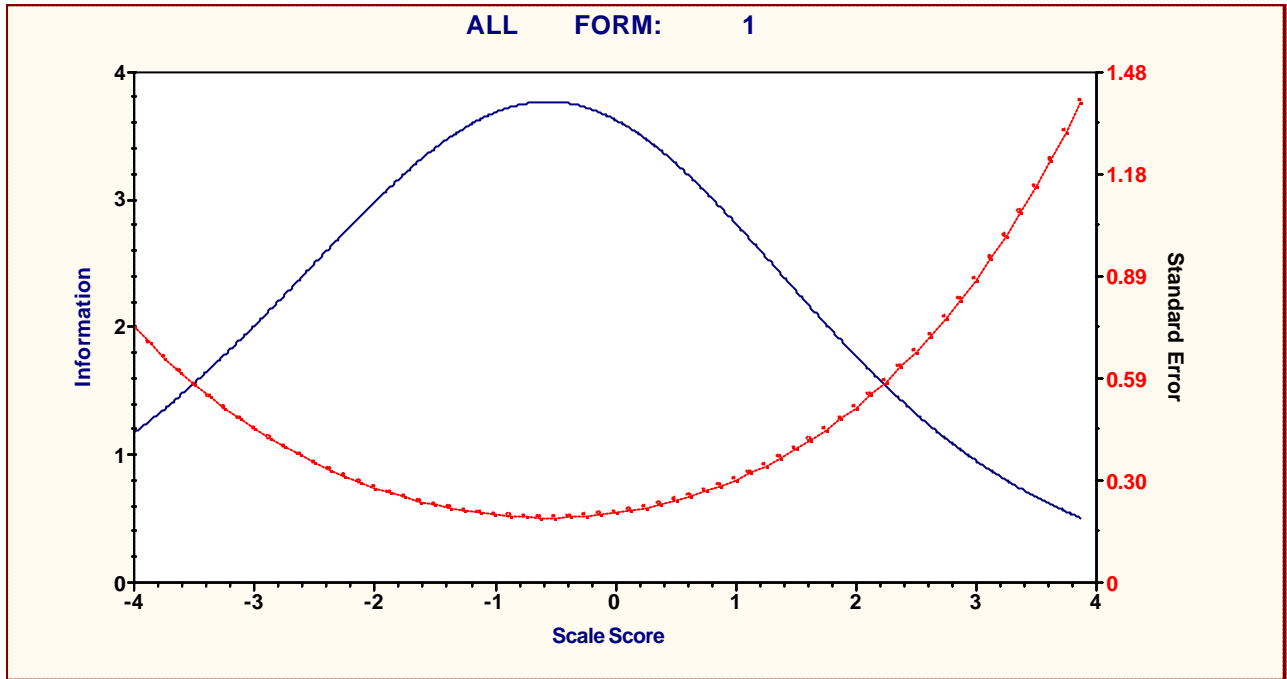


圖 附一-8、 綜合分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $r = .984$



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附一-9、一般分析能力題，1PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附一-10、一般分析能力題 1PL 模式下的測驗訊息圖

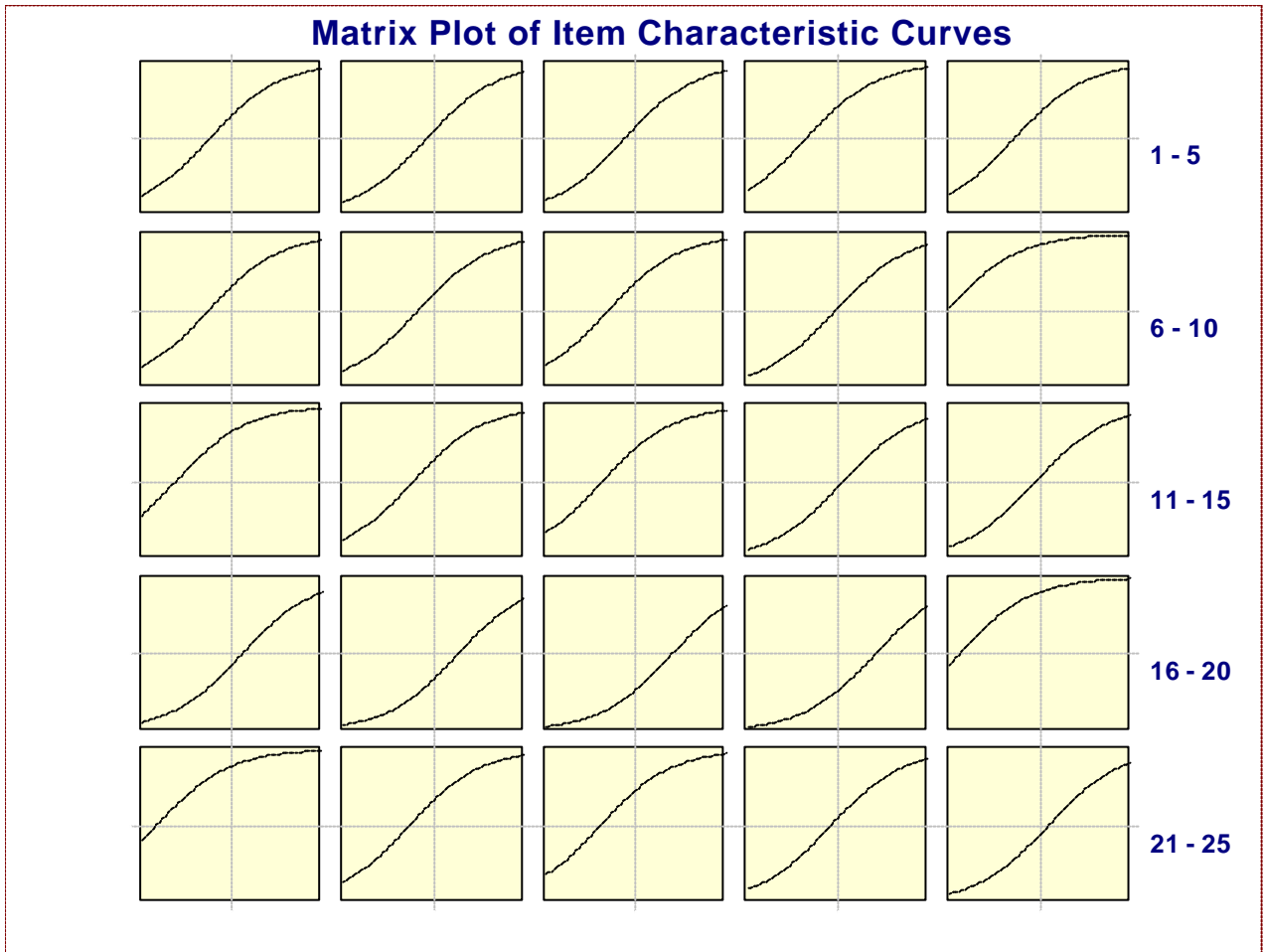


圖 附一-11、一般分析能力題，1PL 模式下各題項的答對機率曲線圖

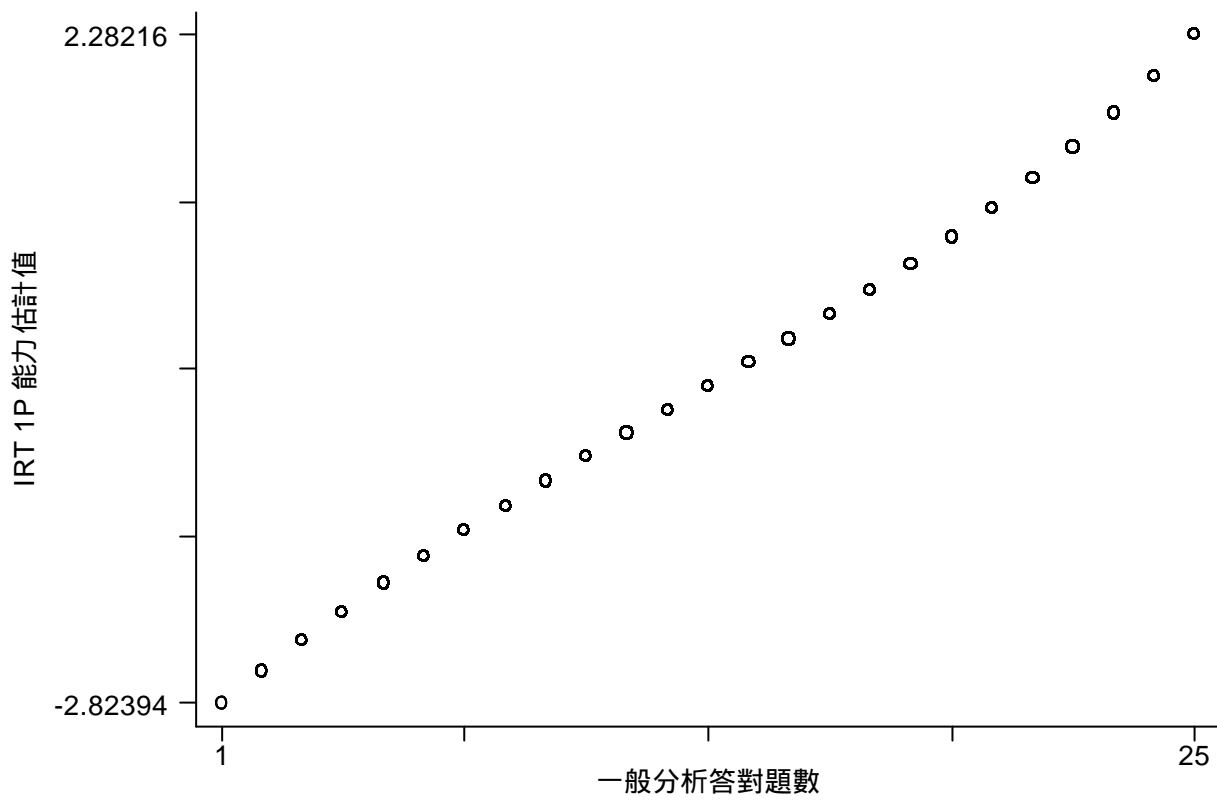
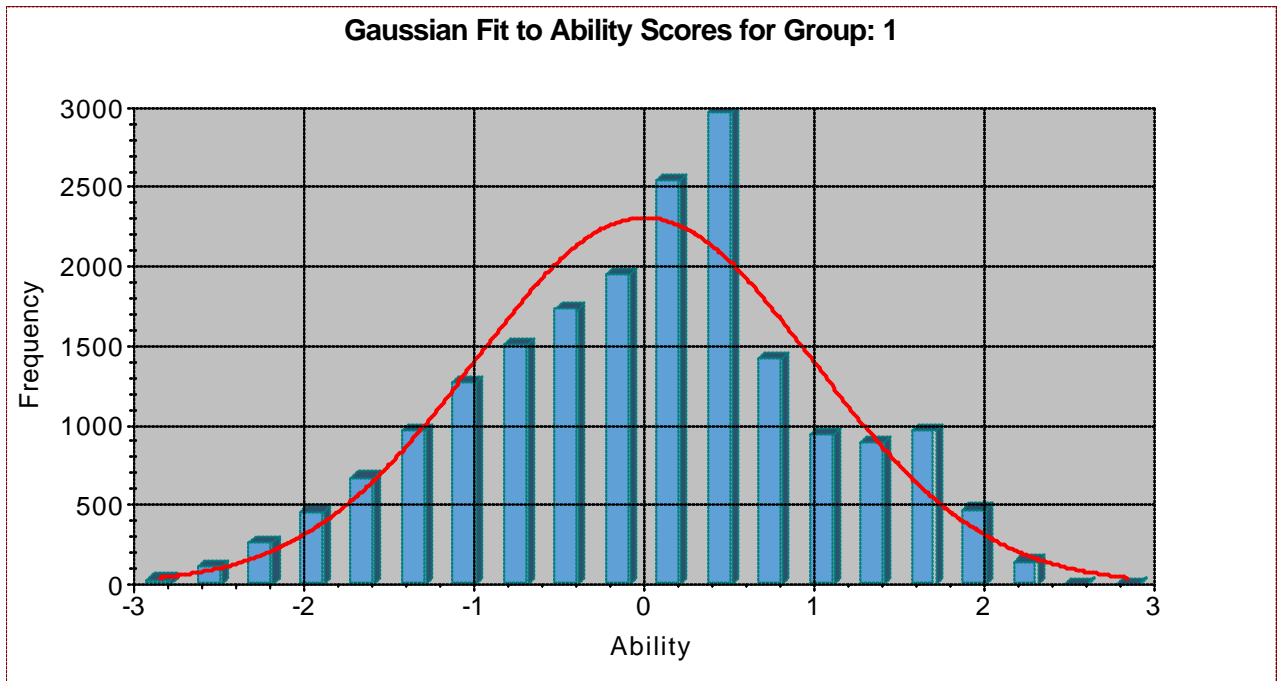
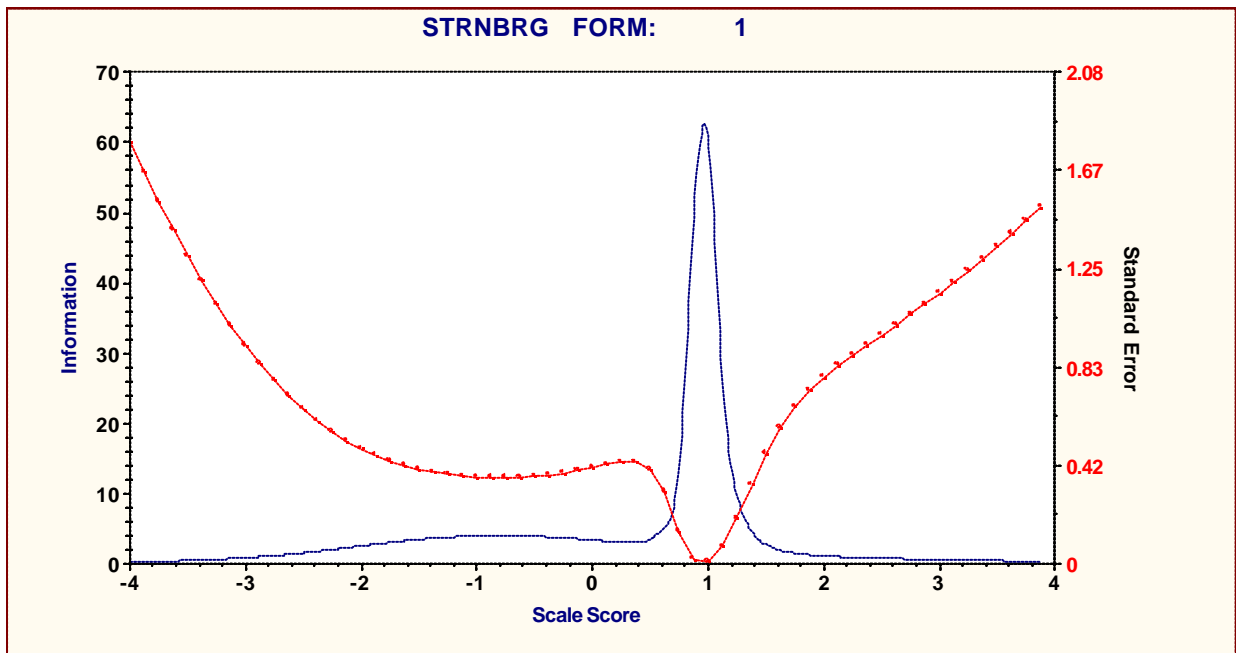


圖 附一-12、一般分析能力題，答對題數與 1PL 模式能力估計值的散佈圖， $r = .998$



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附一-13、一般分析能力題，3PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附一-14、一般分析能力題，3PL 模式下的測驗訊息圖

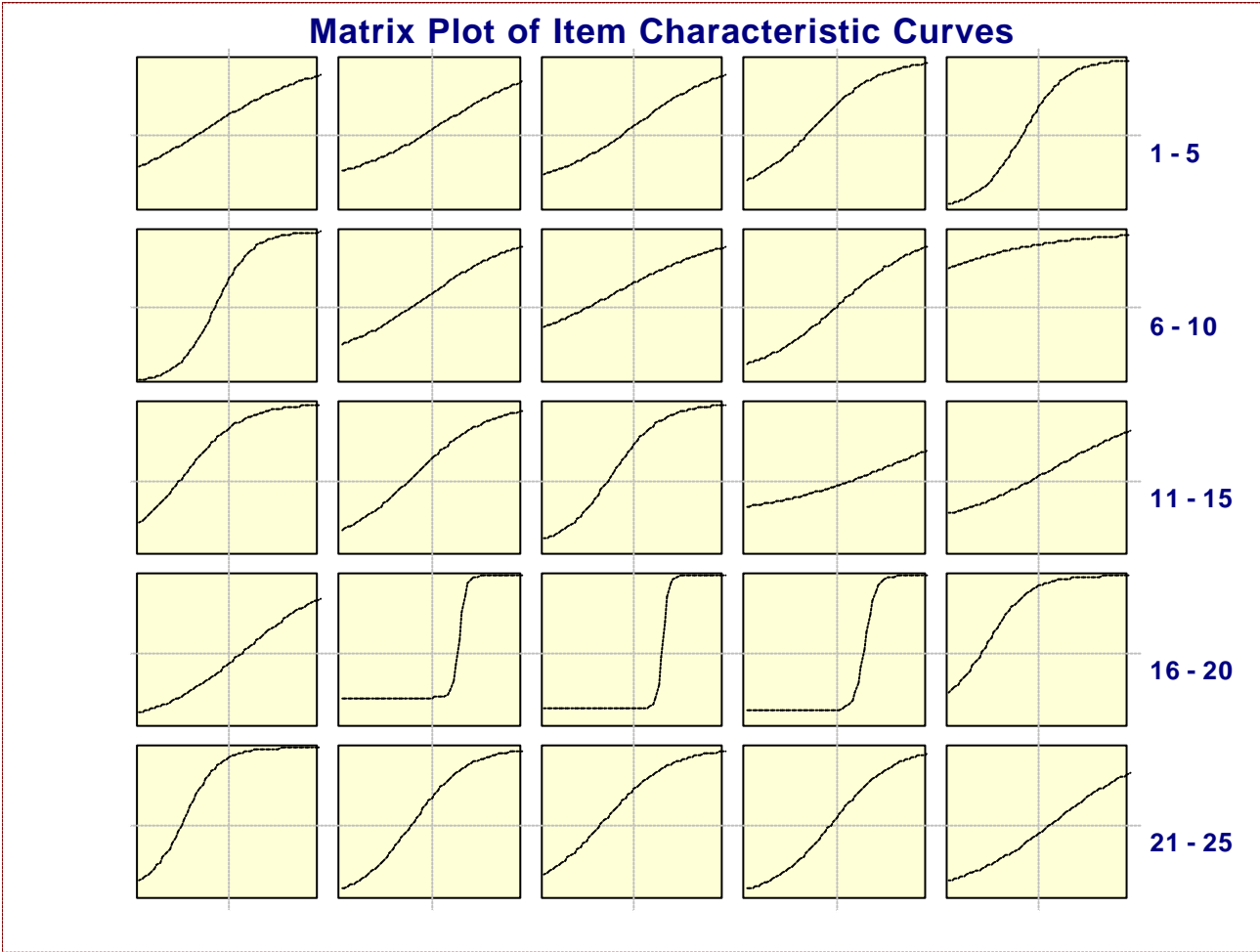


圖 附一-15、一般分析能力題，3PL 模式下各題項的答對機率曲線圖

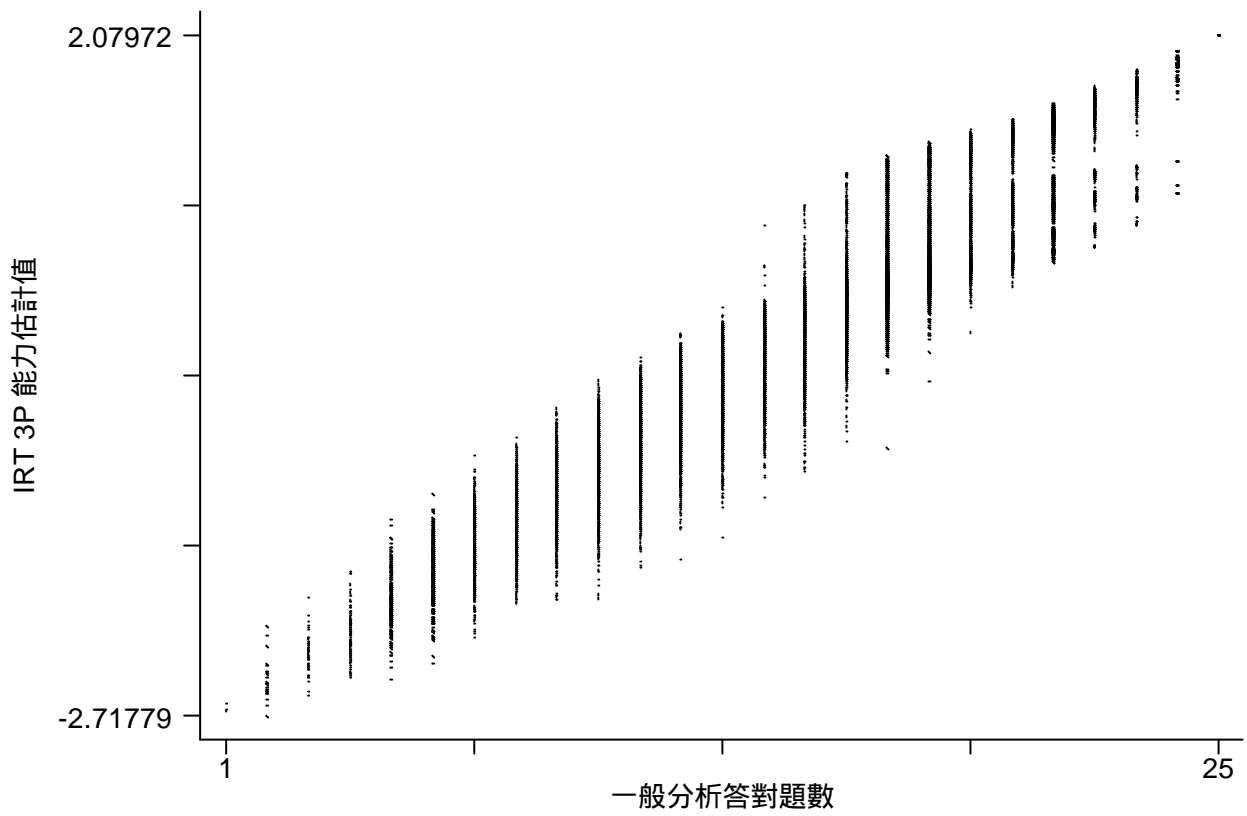
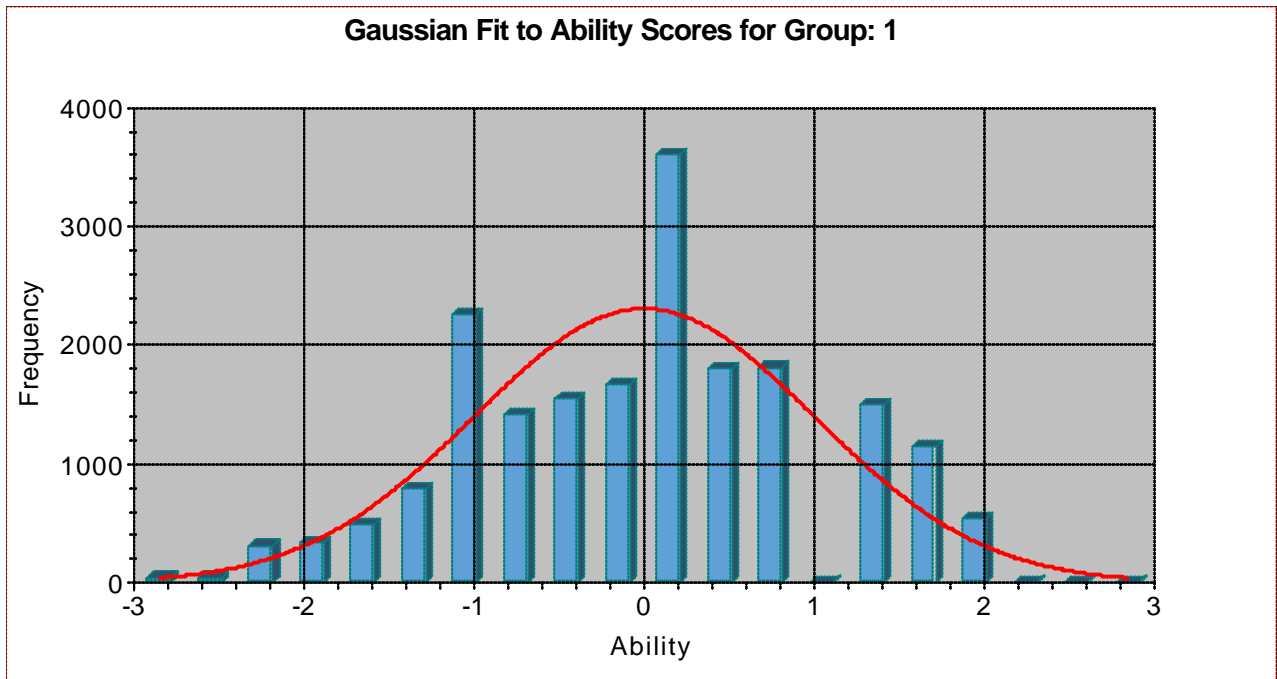
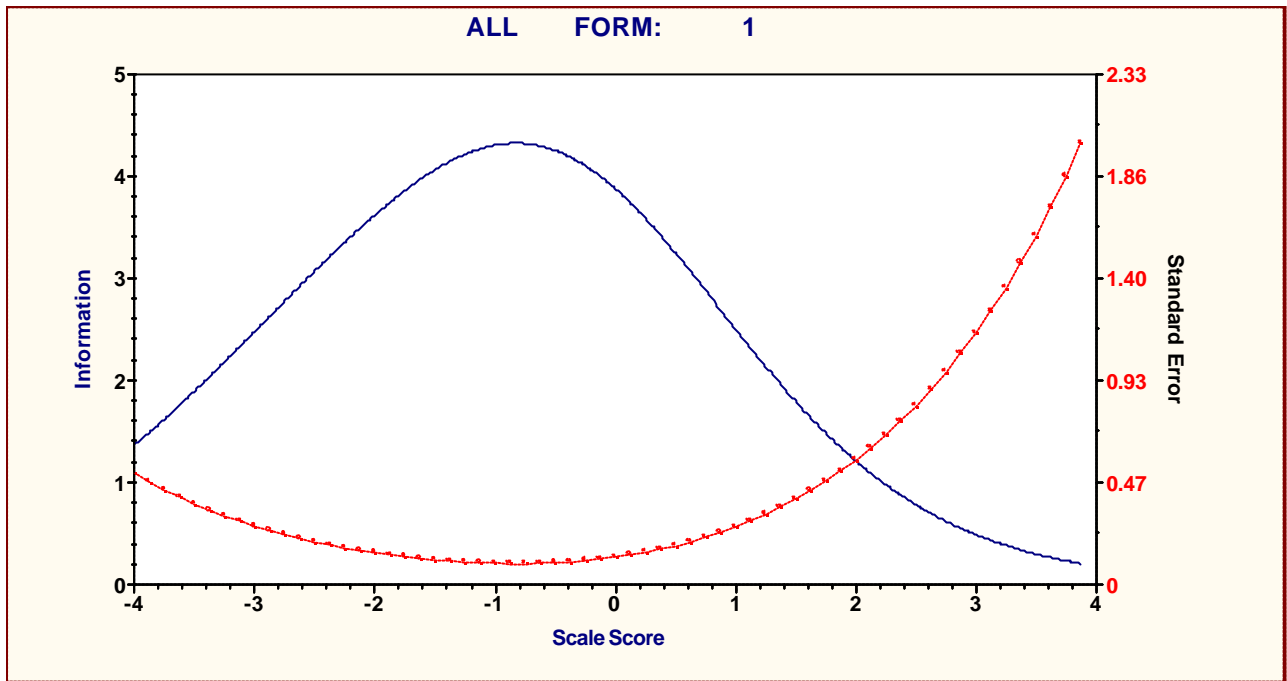


圖 附一-16、一般分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $r = .952$



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附一-17、 數學分析能力題，1PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附一-18、 數學分析能力題，1PL 模式下的測驗訊息圖

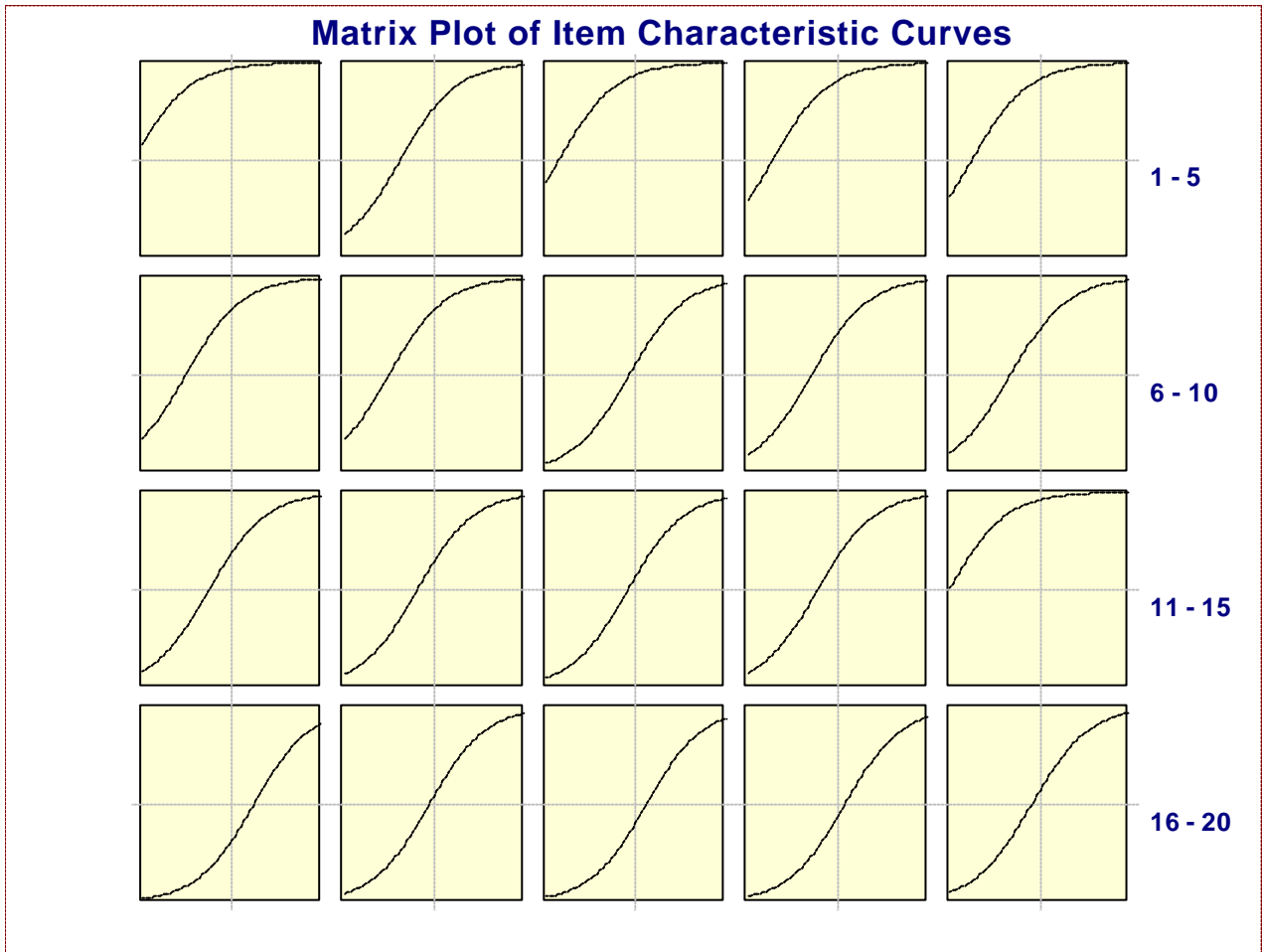


圖 附一-19、 數學分析能力題，1PL 模式下的各題項答對機率圖

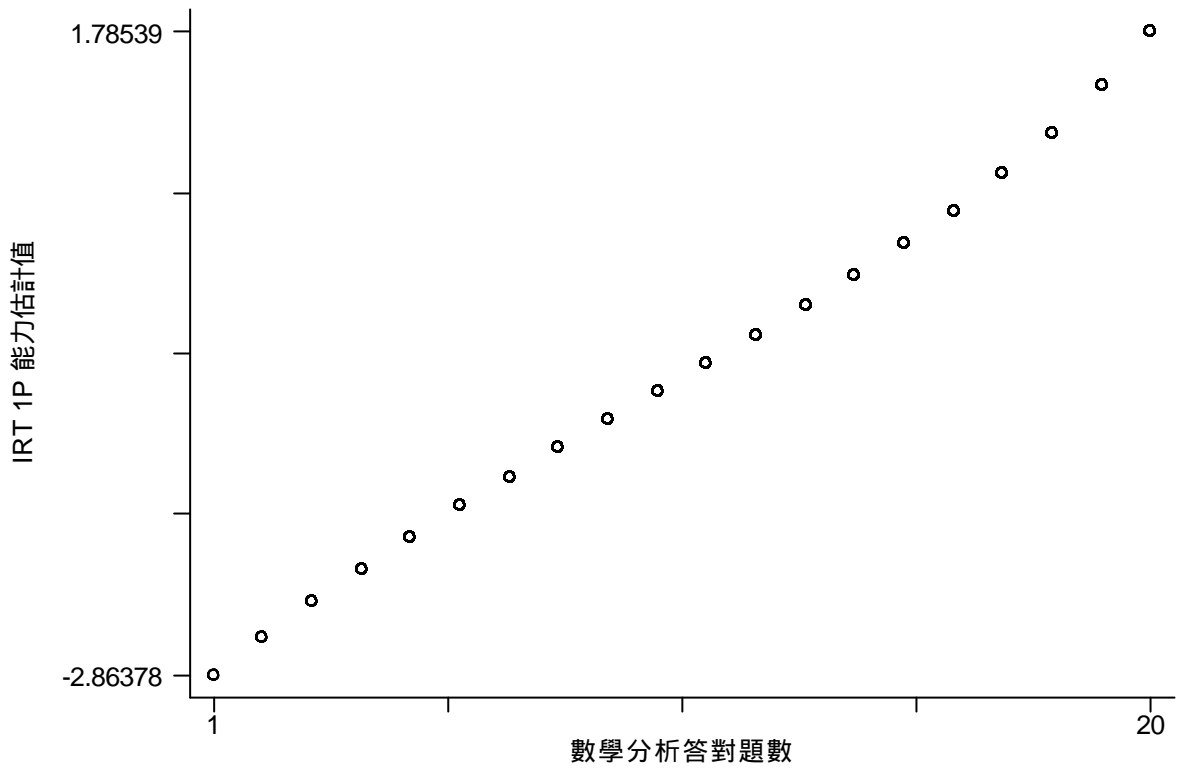
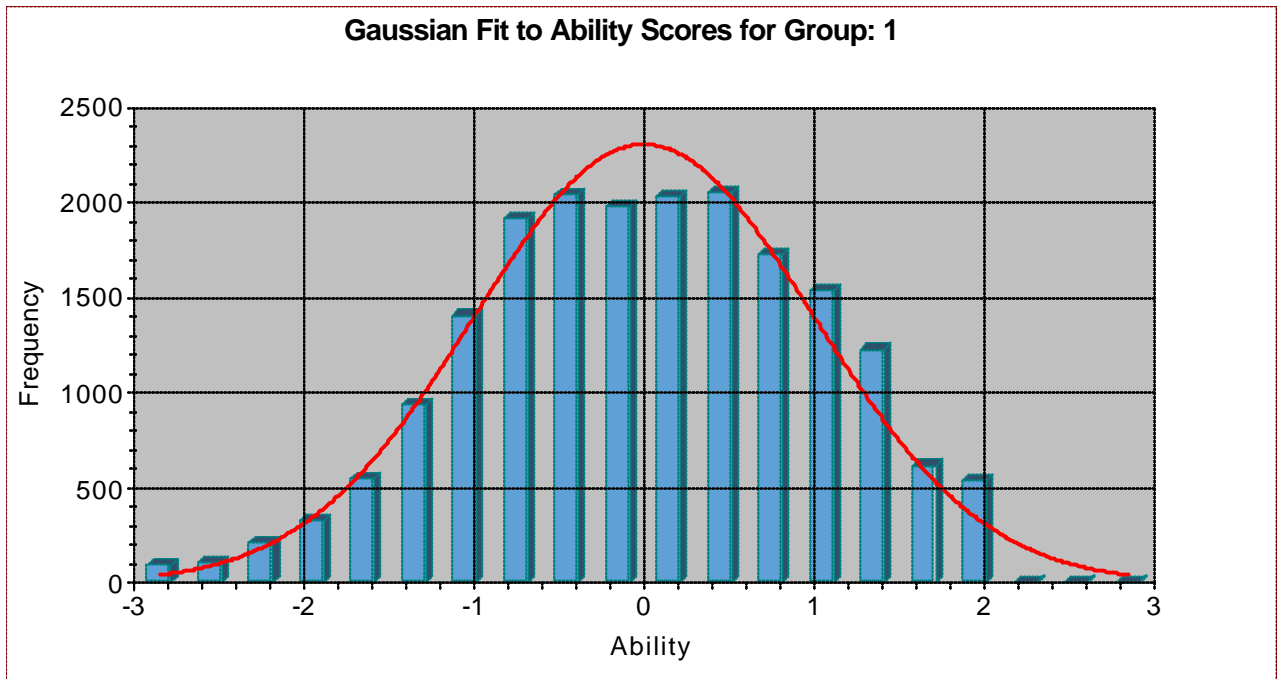
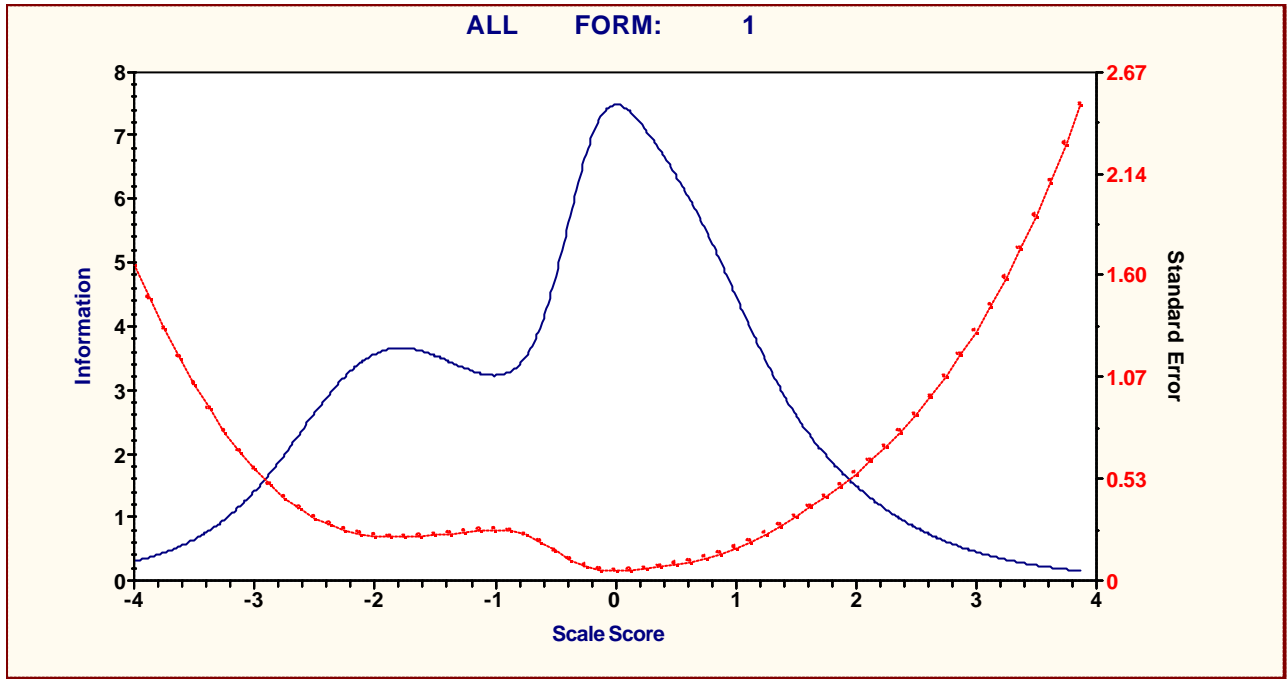


圖 附一-20、數學分析能力題 1PL 模式下，能力估計值與答對比率的迴歸線， $r=.996$



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附一-21、 數學分析能力題，3PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附一-22、 數學分析能力題，3PL 模式下的測驗訊息圖

Matrix Plot of Item Characteristic Curves

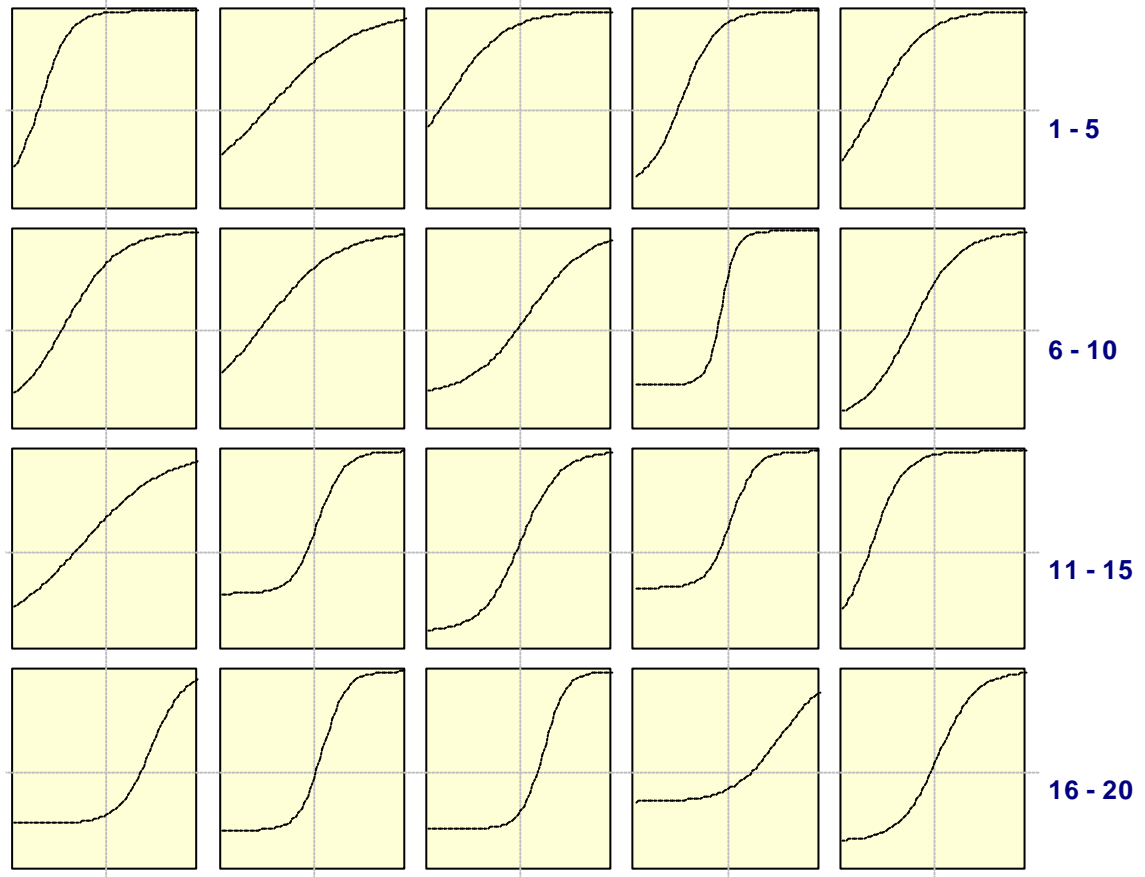


圖 附一-23、數學分析能力題，3PL 模式下的各題項答對機率曲線圖

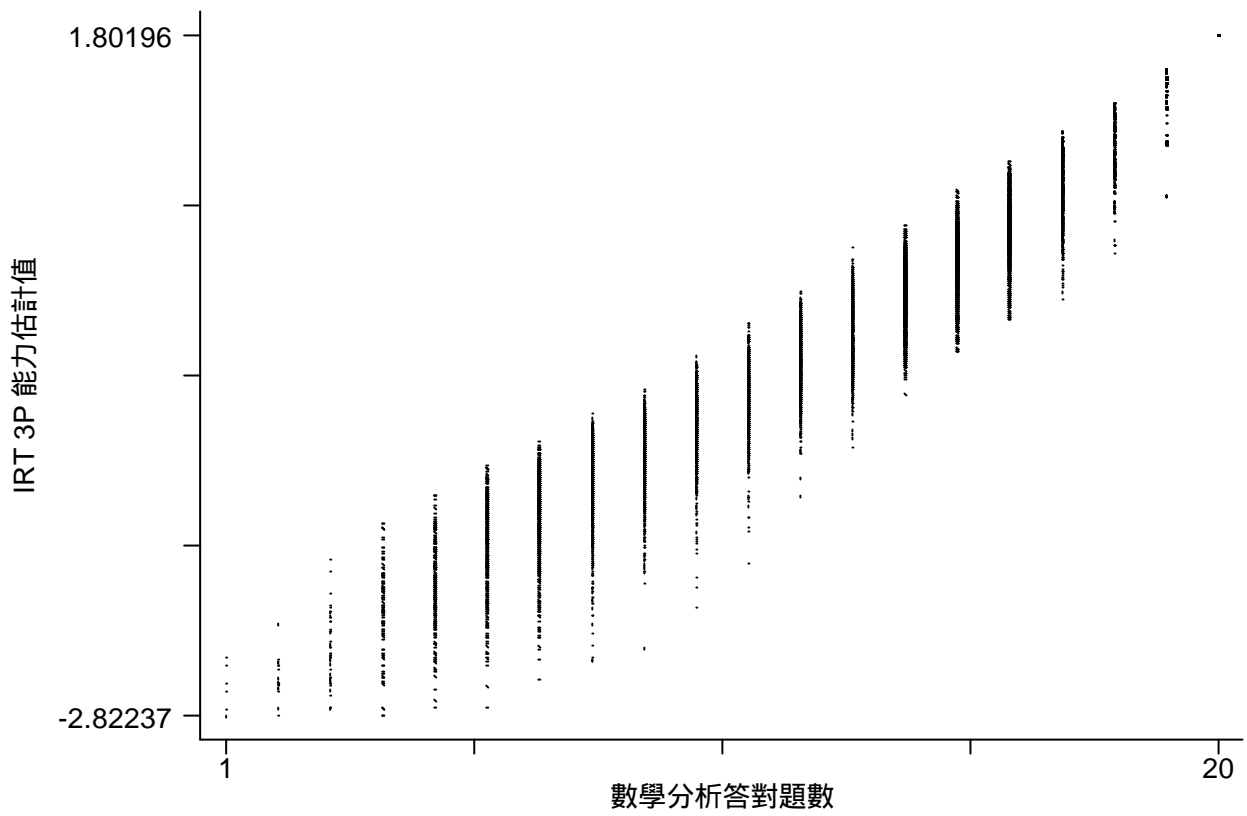
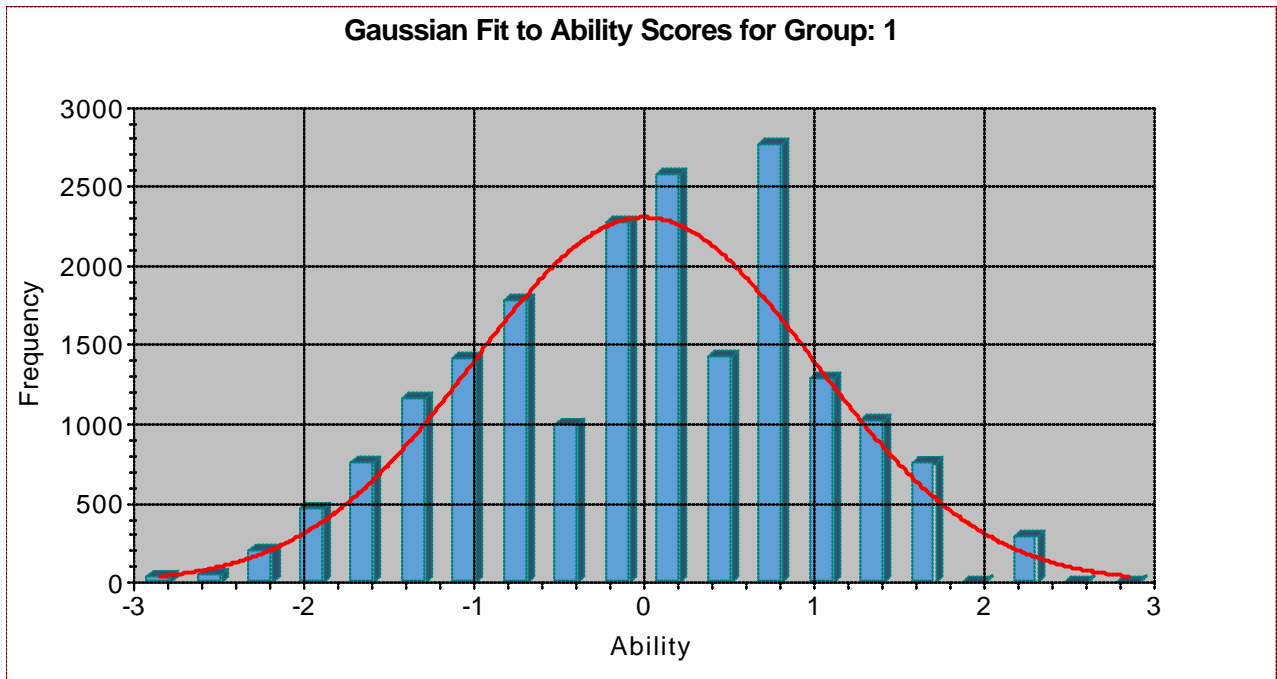
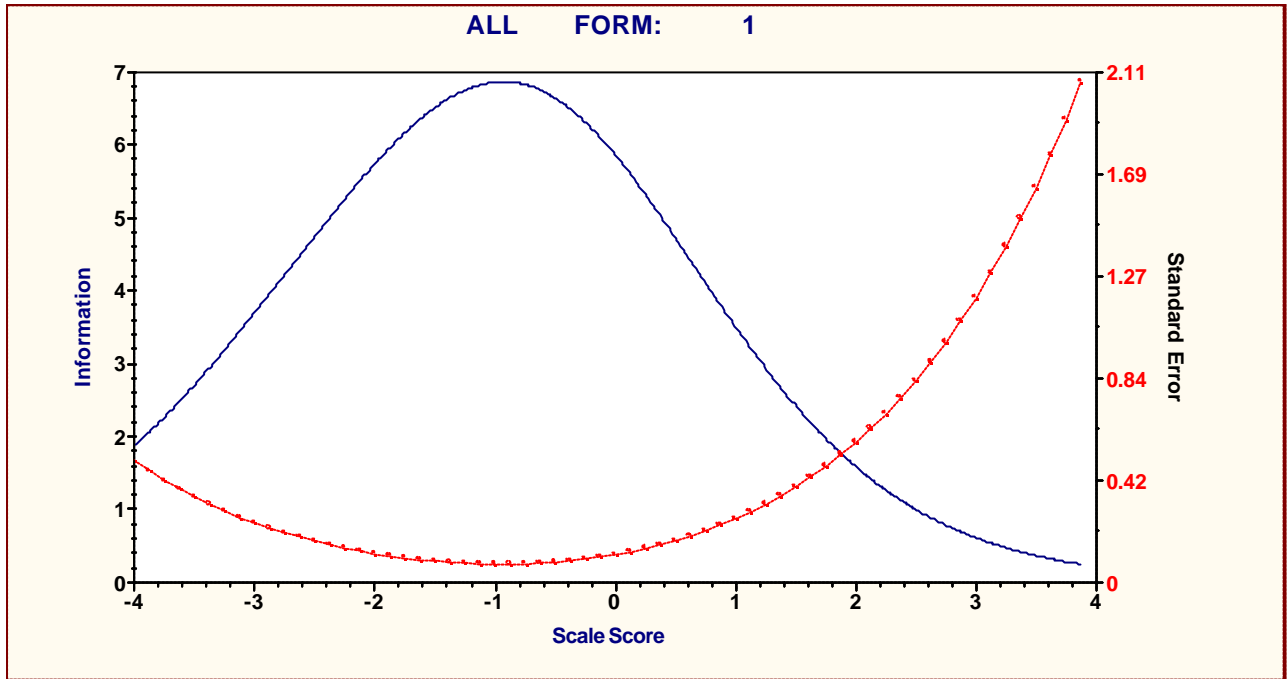


圖 附一-24、 數學分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $r=0.971$



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附一-25、 數學或數字型分析能力題，1PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附一-26、 數學或數字型分析能力題，1PL 模式下的測驗訊息圖

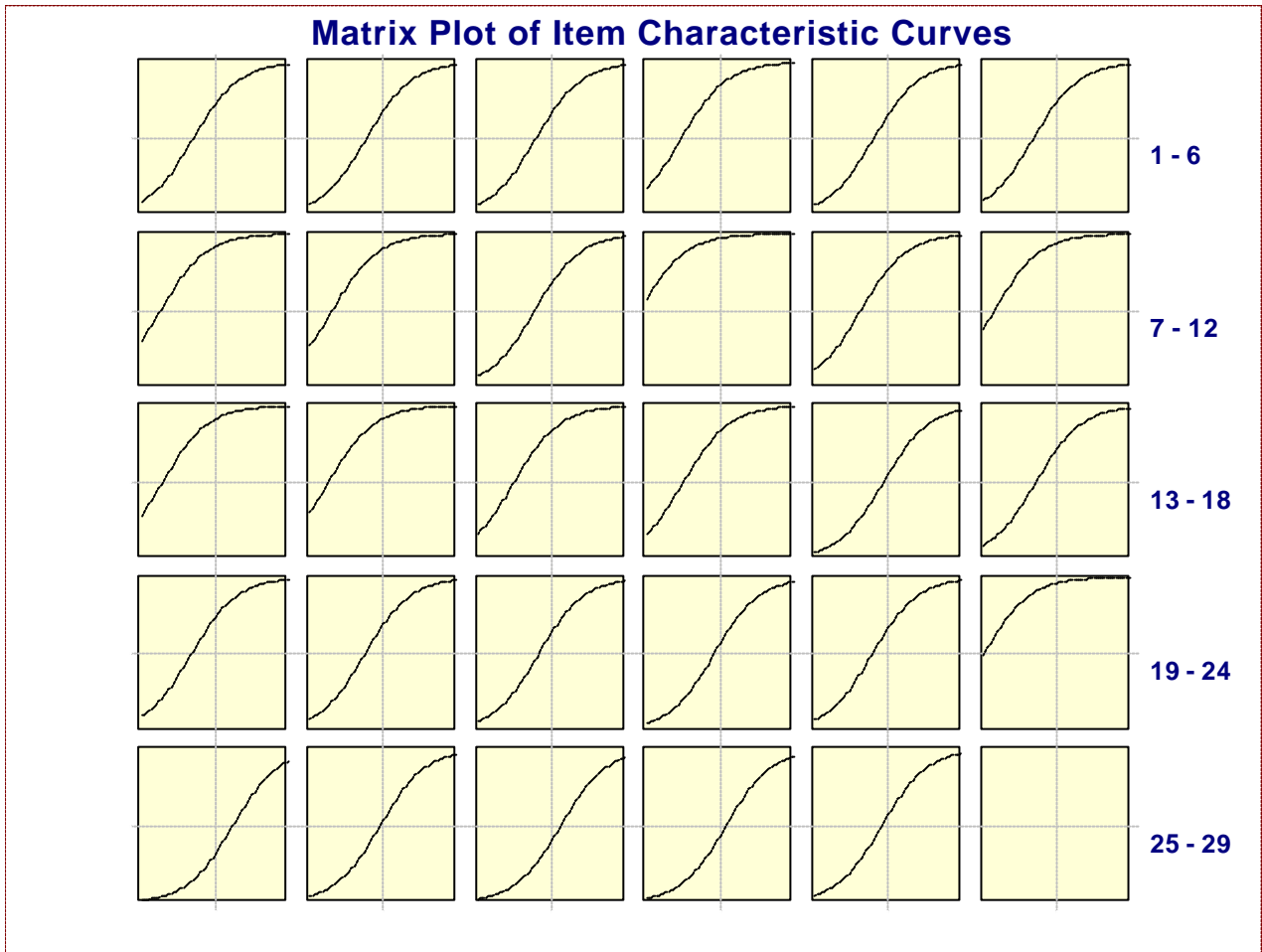


圖 附一-27、 數學或數字型分析能力題，1PL 模式下的各題項答對機率圖

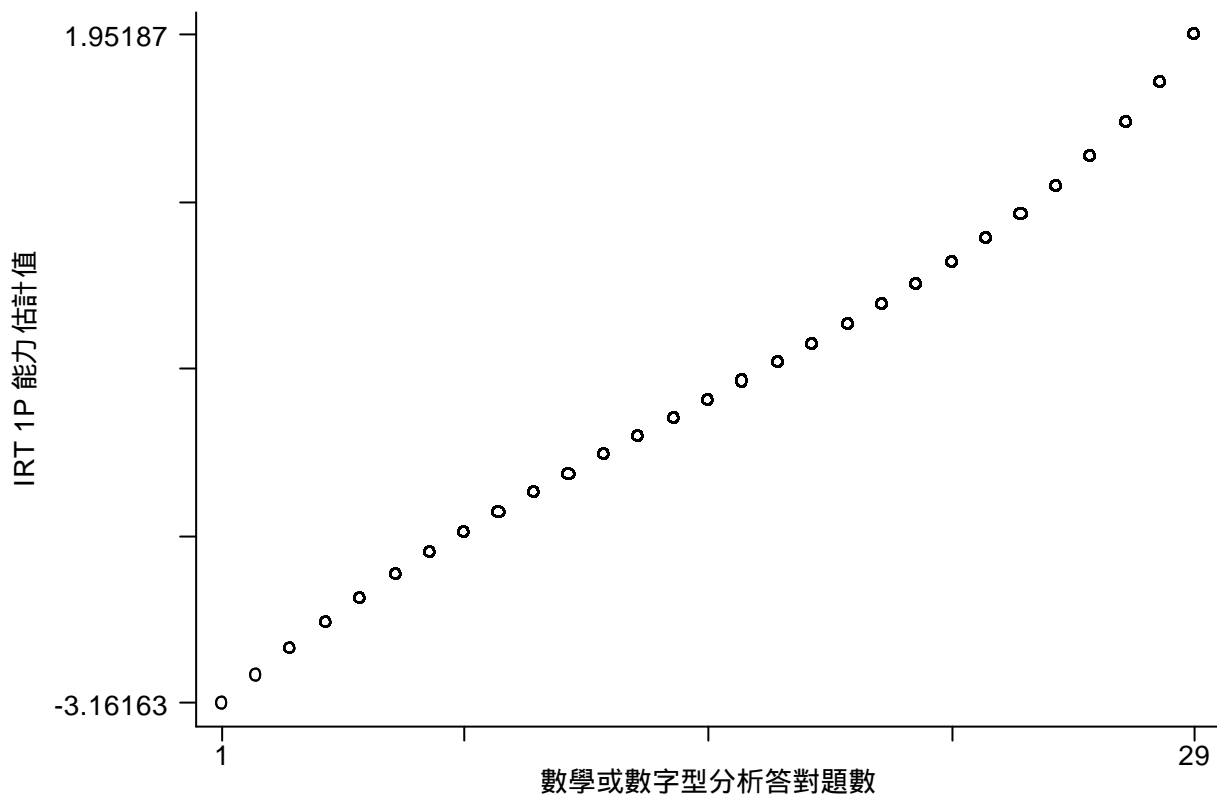
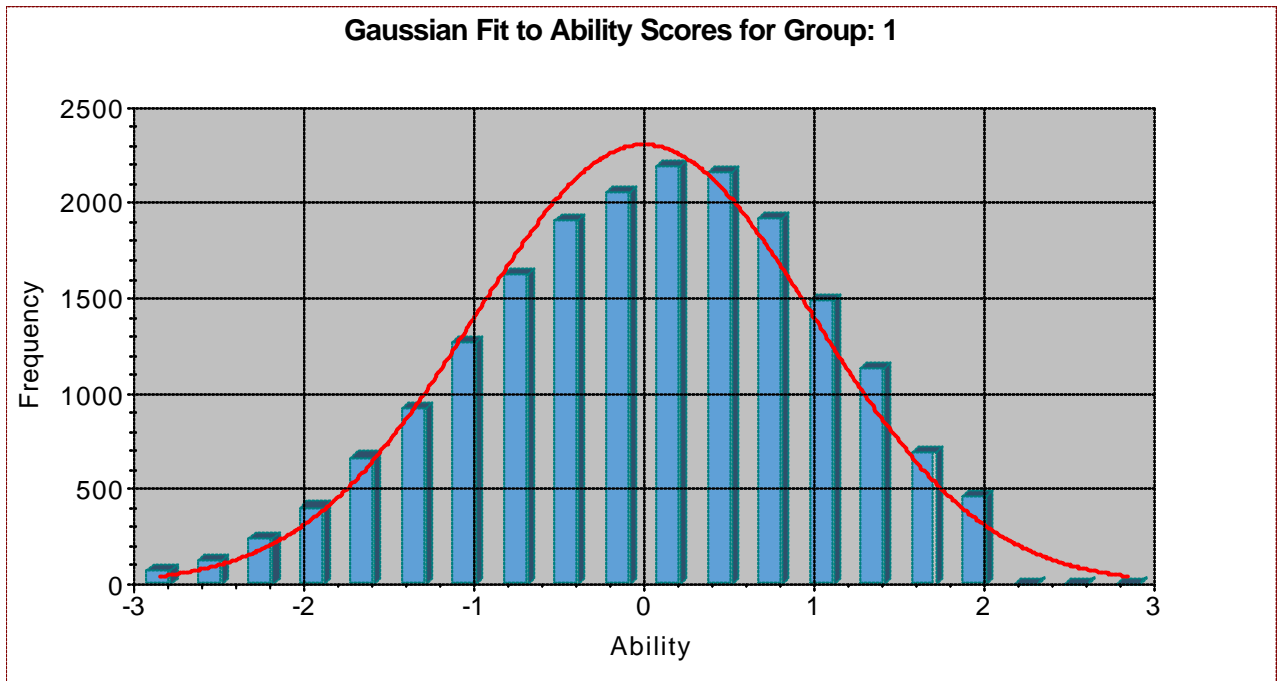
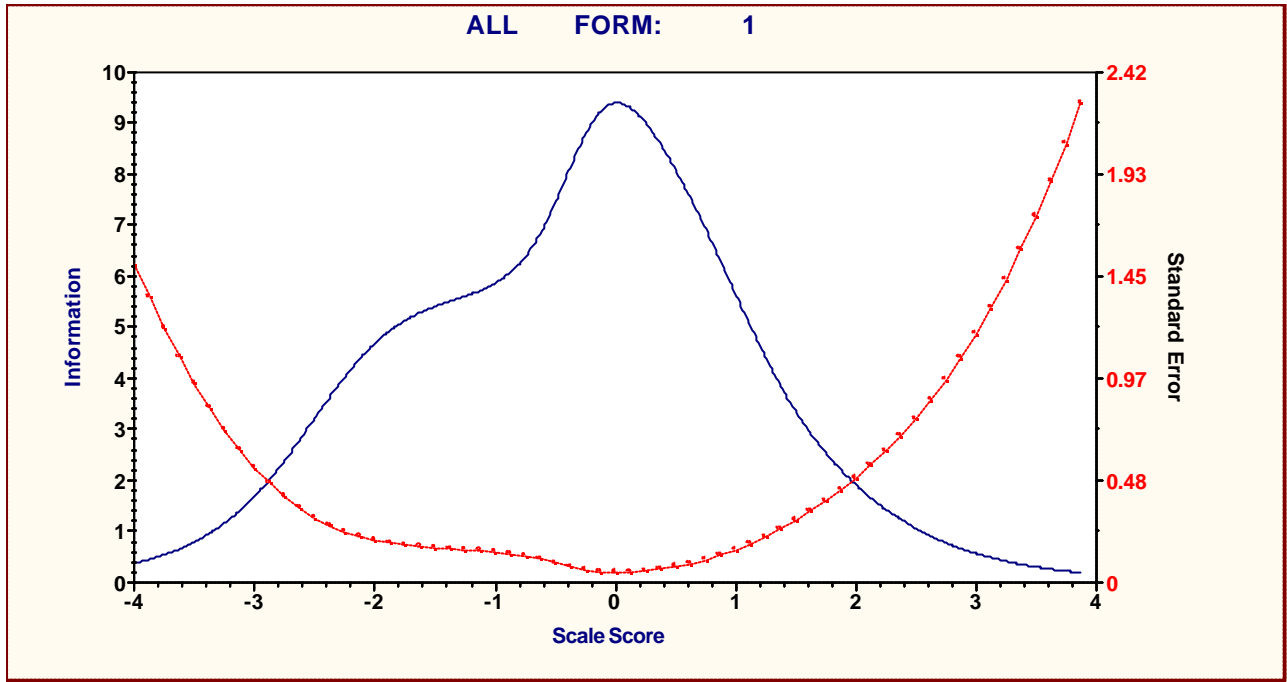


圖 附一-28、 數學或數字型分析能力題，答對題數與 1PL 能力估計值的散佈圖， $r=.992$



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附一-29、 數學或數字型分析能力題，3PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附一-30、 數學或數字型分析能力題，3PL 模式下的測驗訊息圖

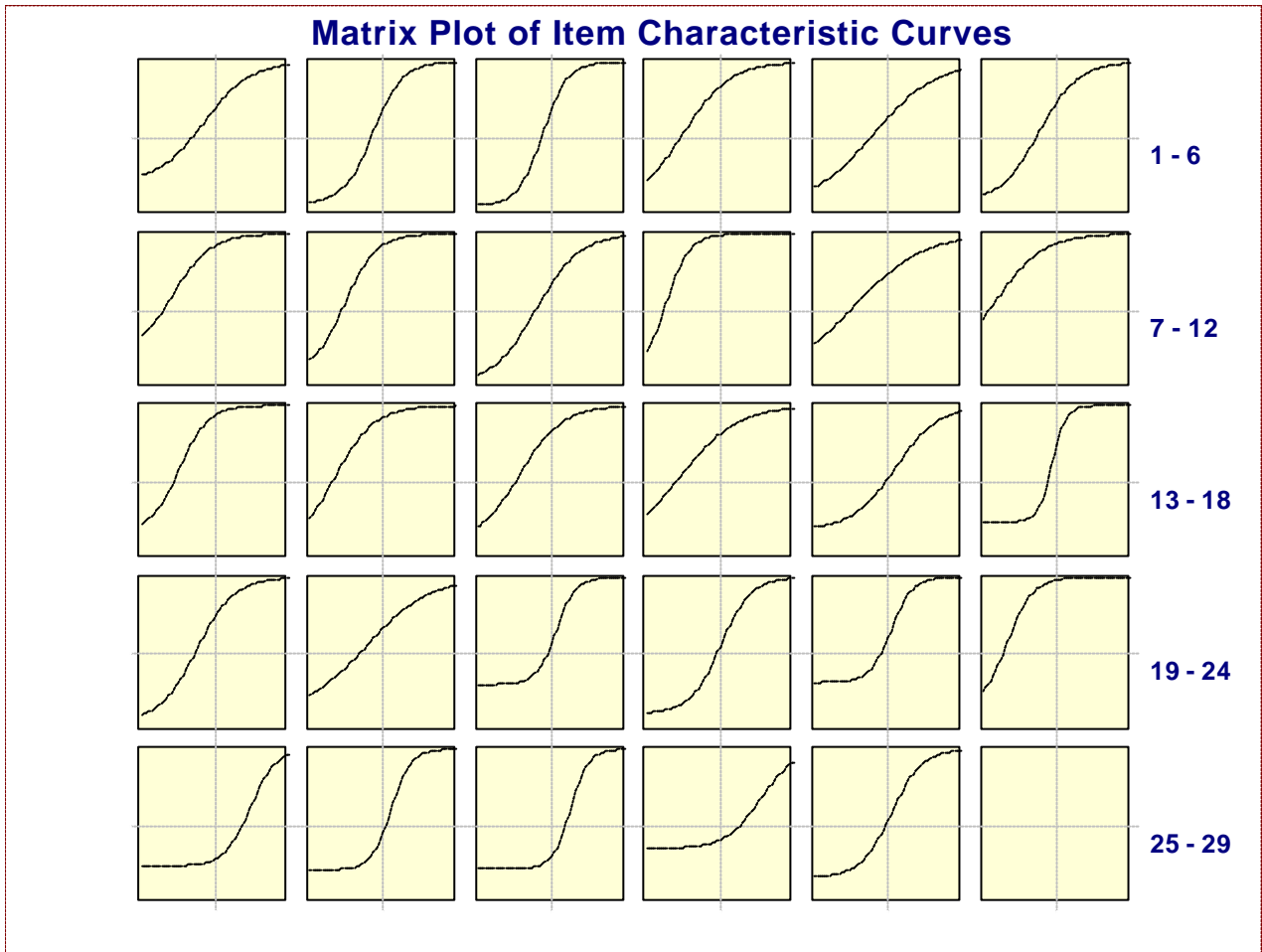


圖 附一-31、 數學或數字型分析能力題，3PL 模式下各題項的答對機率曲線圖

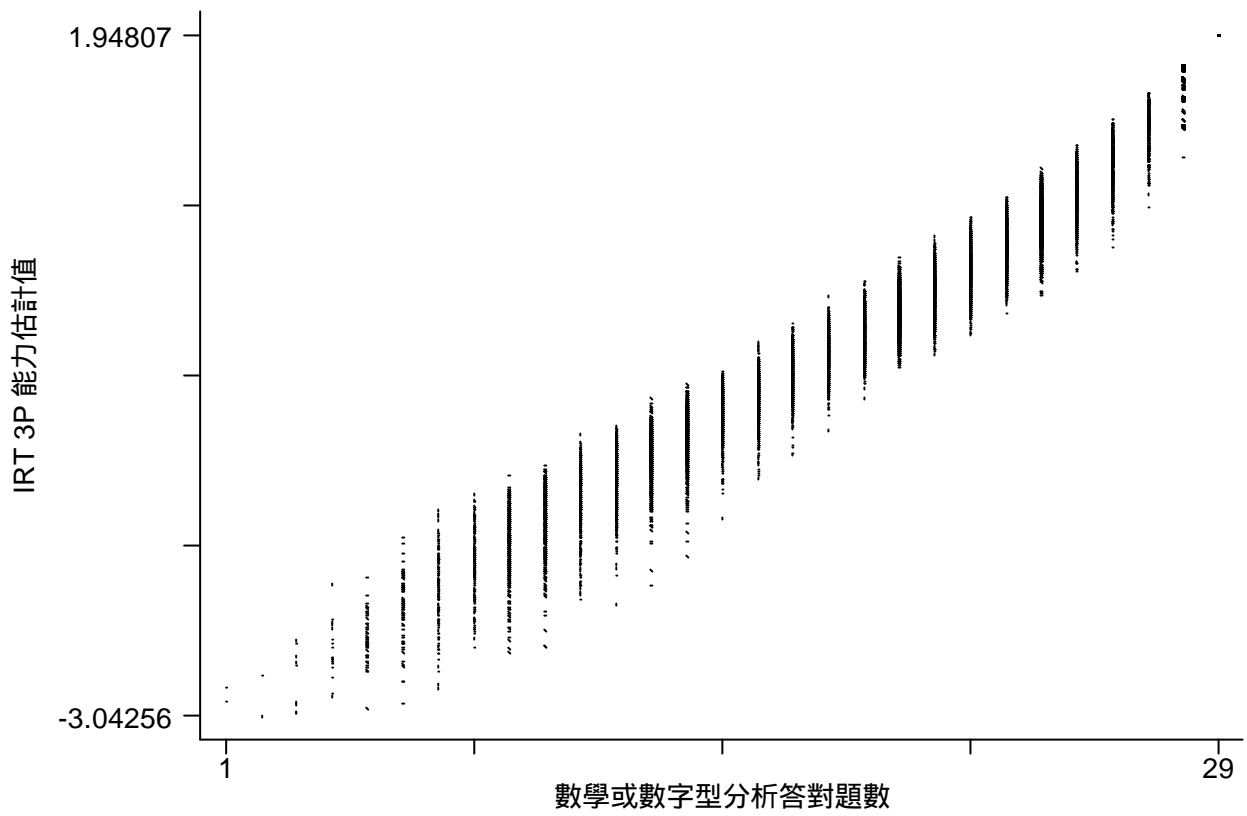


圖 附一-32、 數學或數字型分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $r=.979$

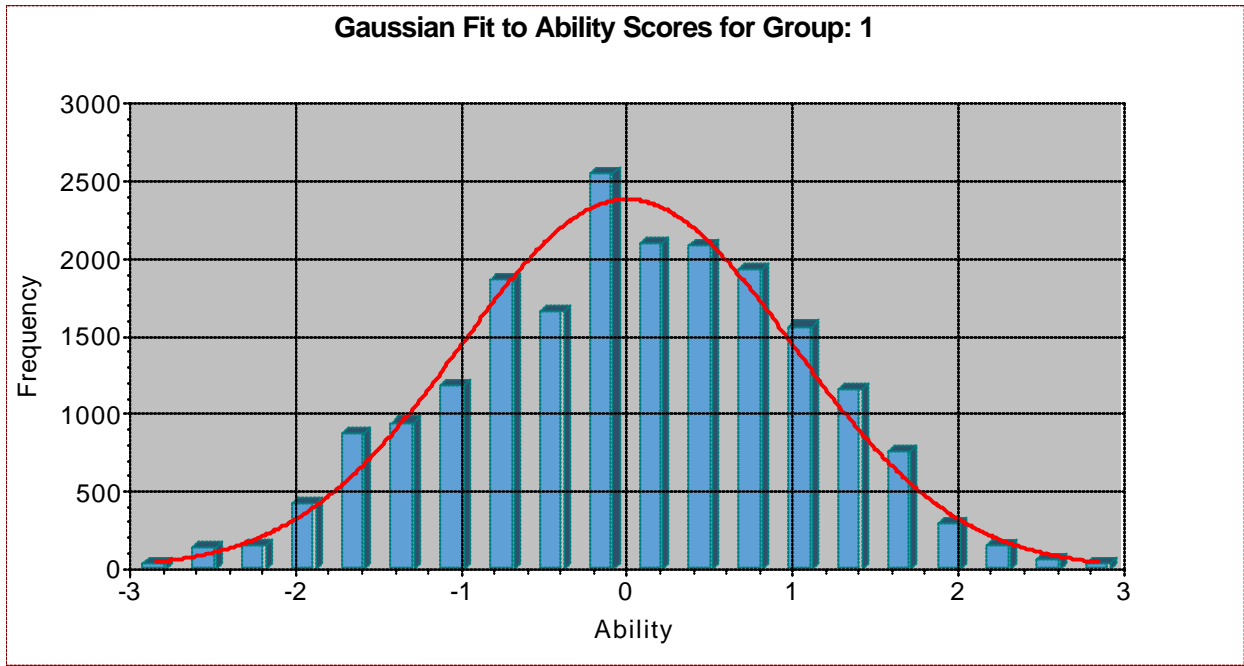
附錄二 國中 IRT 分析結果圖示

本附錄包括國中樣本裡，各分項測驗在 1PL 與 3PL 模式下，學生能力估計值分佈長條圖、測驗訊息 (Test Information) 曲線圖、測驗中各題項的答對機率曲線圖 (Item Characteristic Curve) 及總答對題數與能力估計值的散佈圖 (scatterplot)。

國中樣本的能力估計值分佈長條圖型態，與高中職五專樣本的型態類似，在 3PL 模式下的能力分佈比較趨近於常態分佈，尤其當題目較多 (綜合分析能力題) 時，3PL 的能力估計值長條分佈圖幾近於常態分佈，雖然能力高於平均值 (0) 的學生似乎稍微較多 (圖 附二-5)；而 1PL 模式下的分佈長條圖則稍有類似雙峰的情形；但當題數不多時，即使 3PL 模式下的分佈也與常態分佈差異頗大。

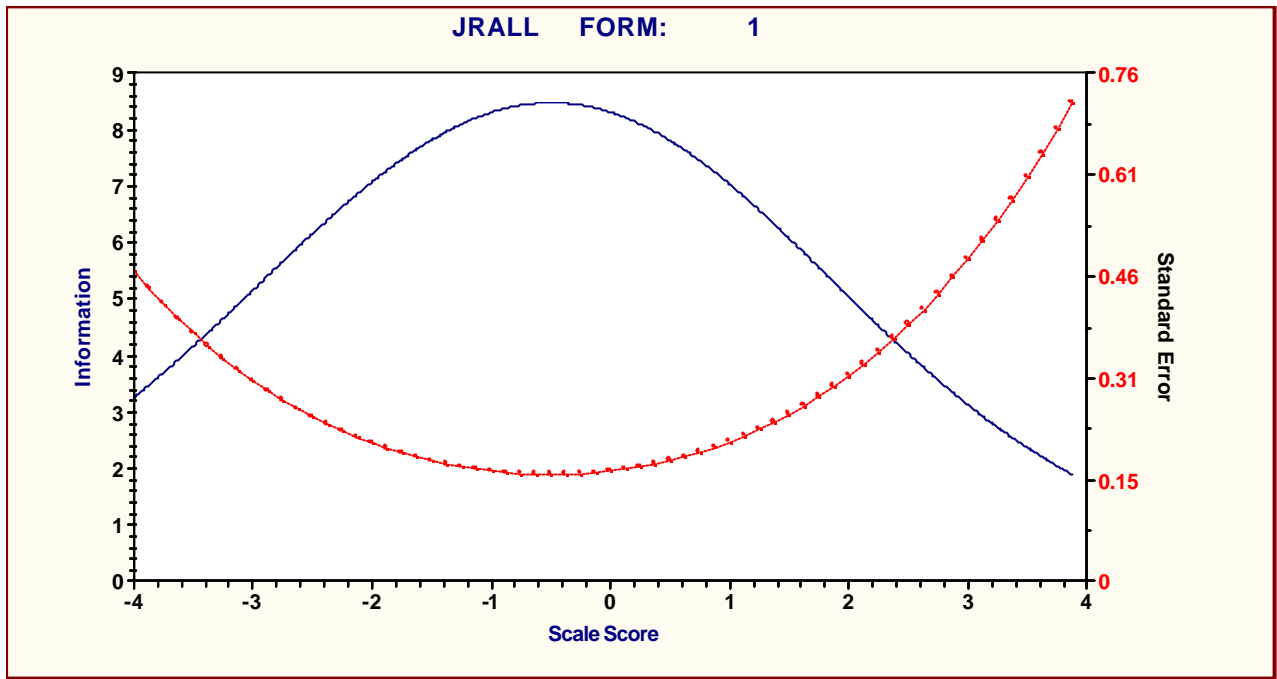
一般而言，不論是哪一種分項測驗，1PL 和 3PL 都顯示，本測驗在能力範圍為平均值左右的訊息較為豐富，但 1PL 模式下訊息較豐富的區域較集中於中間偏低的能力，而 3PL 模式下，訊息較豐富的區域則稍為集中於中間偏高的能力範圍。與高中職五專的測驗不同之處是，國中所使用的測驗似乎能提供訊息之範圍較廣，高訊息處形成一個高原的形狀 (如 圖 附二-6)；不似高中職五專的測驗訊息密集於一處，高訊息處是一個尖峰；尤其兩者的 (圖 附一-14 與 圖 附二-14) 一般分析能力題在 3PL 模式下的測驗訊息圖，是強烈的對比。國中樣本的數學分析能力題之訊息圖也呈雙峰現象 (圖 附二-22)，但與高中職五專樣本不同的是，其高峰是在 -0.5 左右，低峰在 0.8 左右，而且高低峰之間的分隔不明顯；加入了一般分析能力中的九題數字型分析能力題後，兩峰成了一個頗為平緩的高原，使能力在 -1 與 1 之間的訊息都頗為豐富 (圖 附二-30)。

至於答對題數與能力估計值的散佈圖，由於 IRT 模型中，答對機會和能力及項目的參數之間並非簡單的線性關係，因此散佈圖呈現的，並不是直線而是平滑的曲線 (或帶狀)。而因為在 1PL 模型中，答對題數是能力估計值的充分統計量，只要答對題數相同，能力估計值就相同，答對題數和 1PL 的能力估計值之散佈圖都是一條曲線。3PL 模型則較為複雜，相同的答對題數，可能因為所答對的題項之參數值不同，而使能力估計值不同；因此散佈圖都呈帶狀。儘管如此，兩種模式下的能力估計值都和答對題數有極高的相關。



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附二-1、國中綜合分析能力題 1PL 模式下，學生能力分佈估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附二-2、國中綜合分析能力題 1PL 模式下的測驗訊息圖

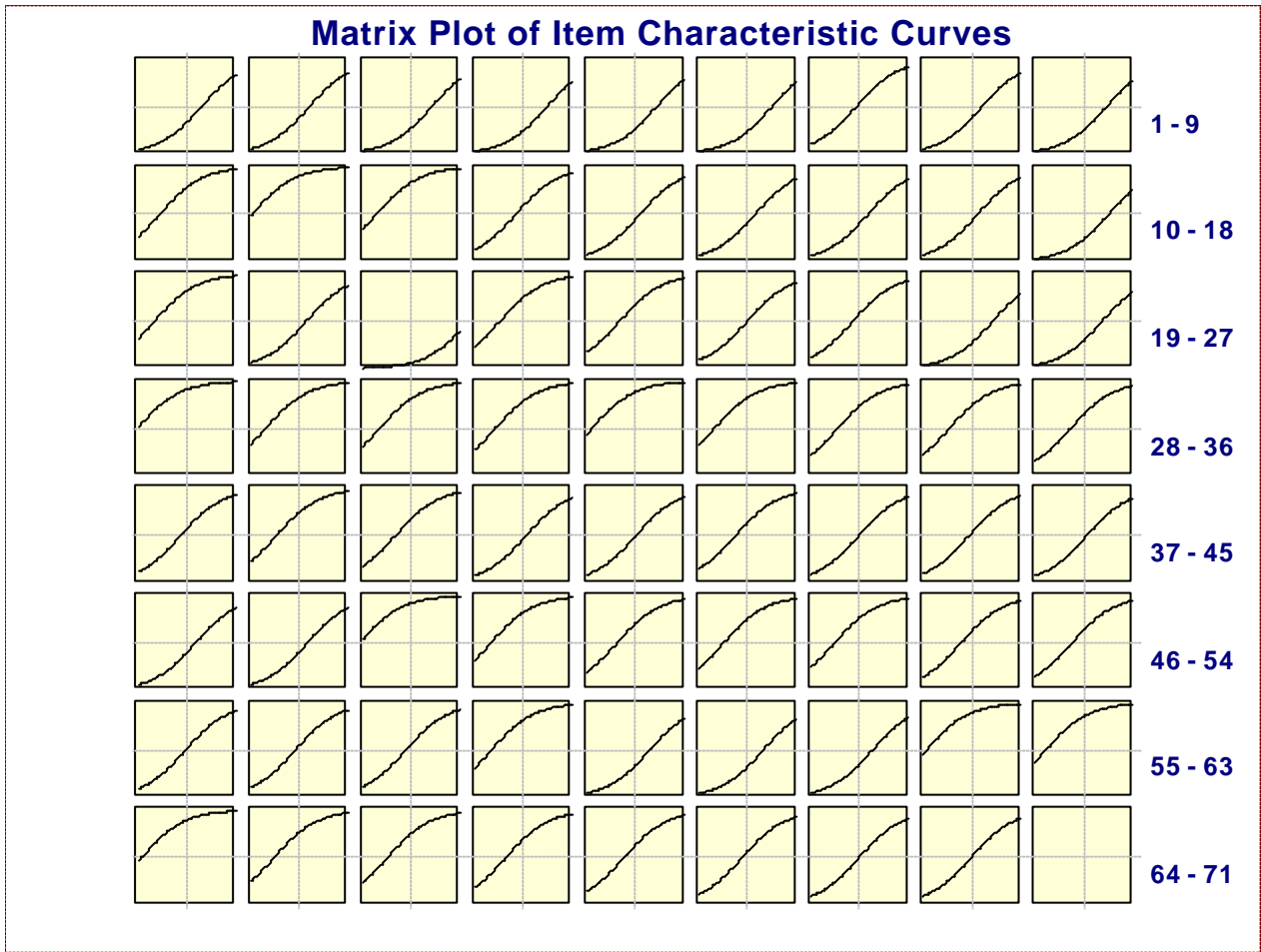


圖 附二-3、國中綜合分析能力題 1PL 模式下，各題項的答對機率曲線圖

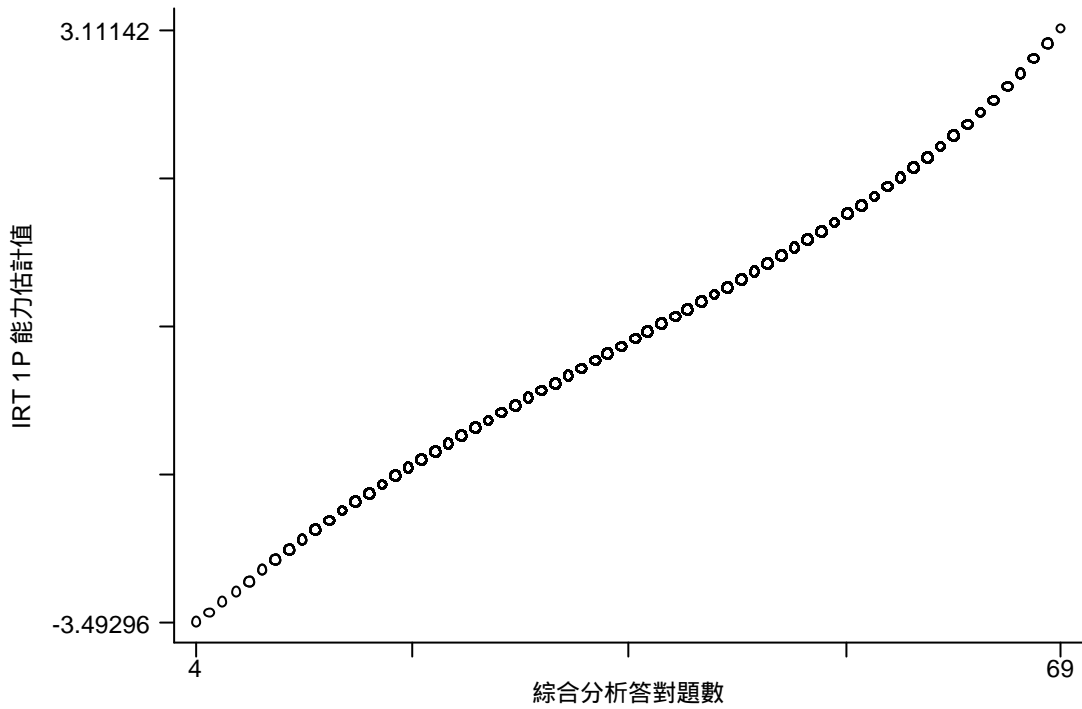
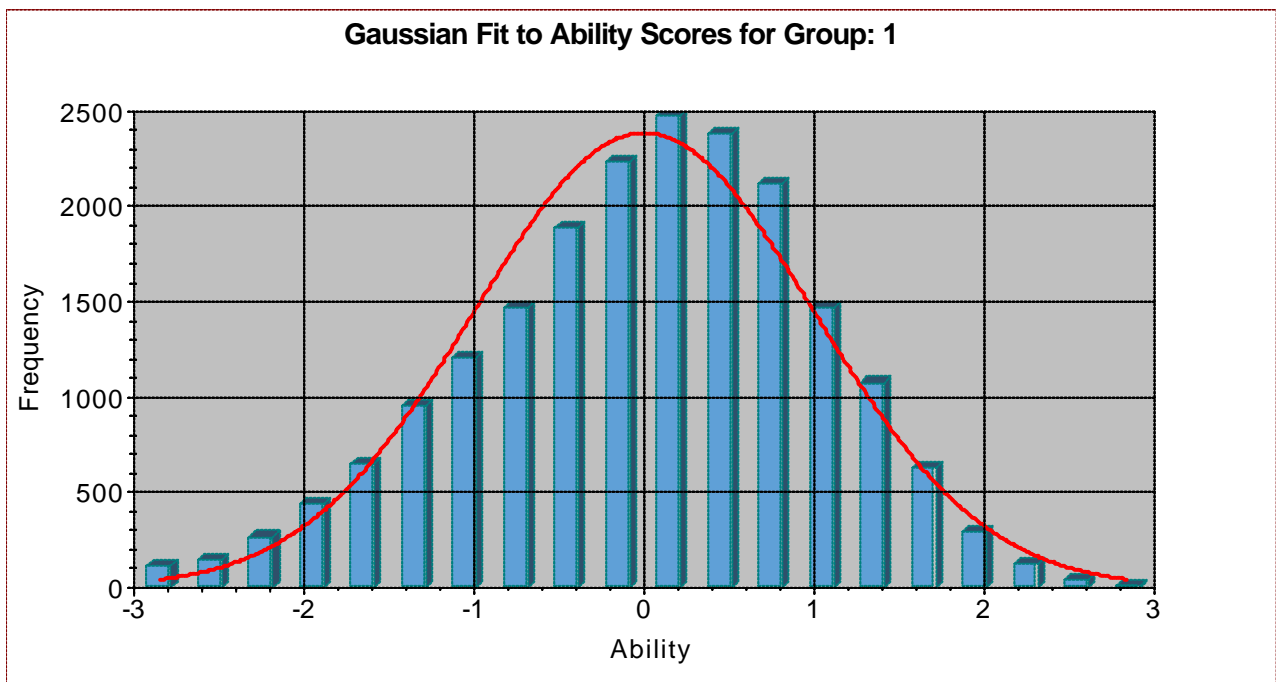
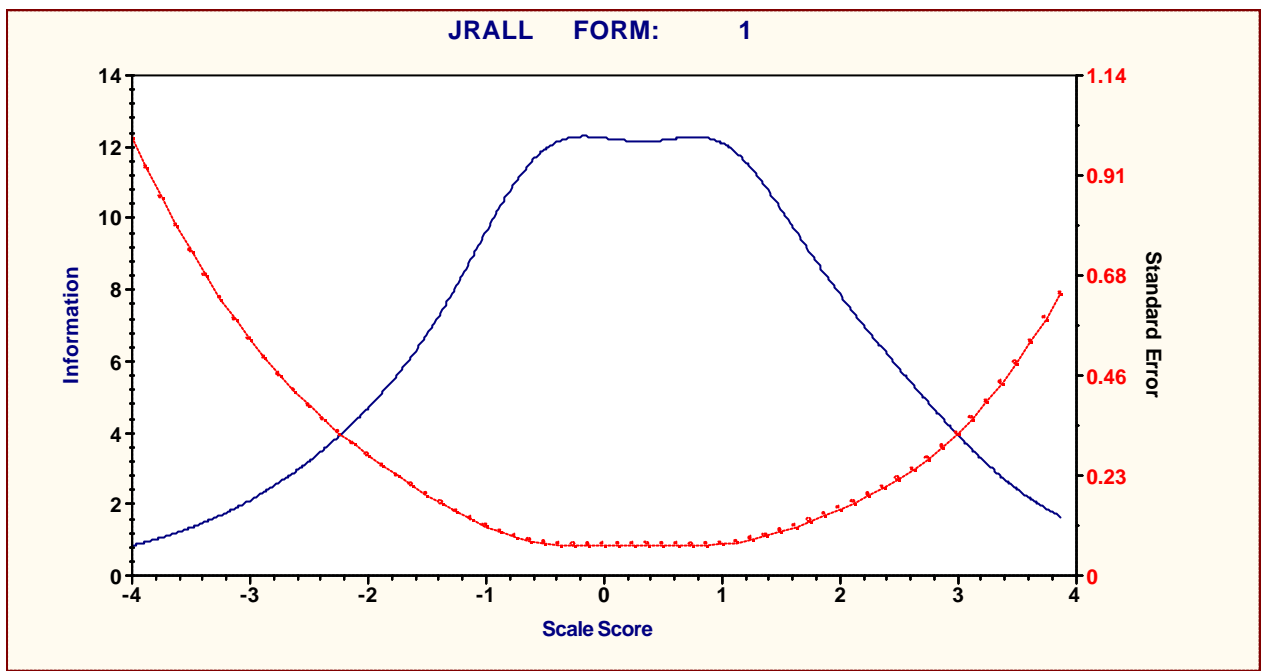


圖 附二-4、國中綜合分析能力題 答對題數與 1PL 模式下能力估計值的散佈圖， $r = .999$



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附二 -5、國中綜合分析能力題 3PL 模式下，學生能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附二-6、國中綜合分析能力題 3PL 模式下的測驗訊息圖

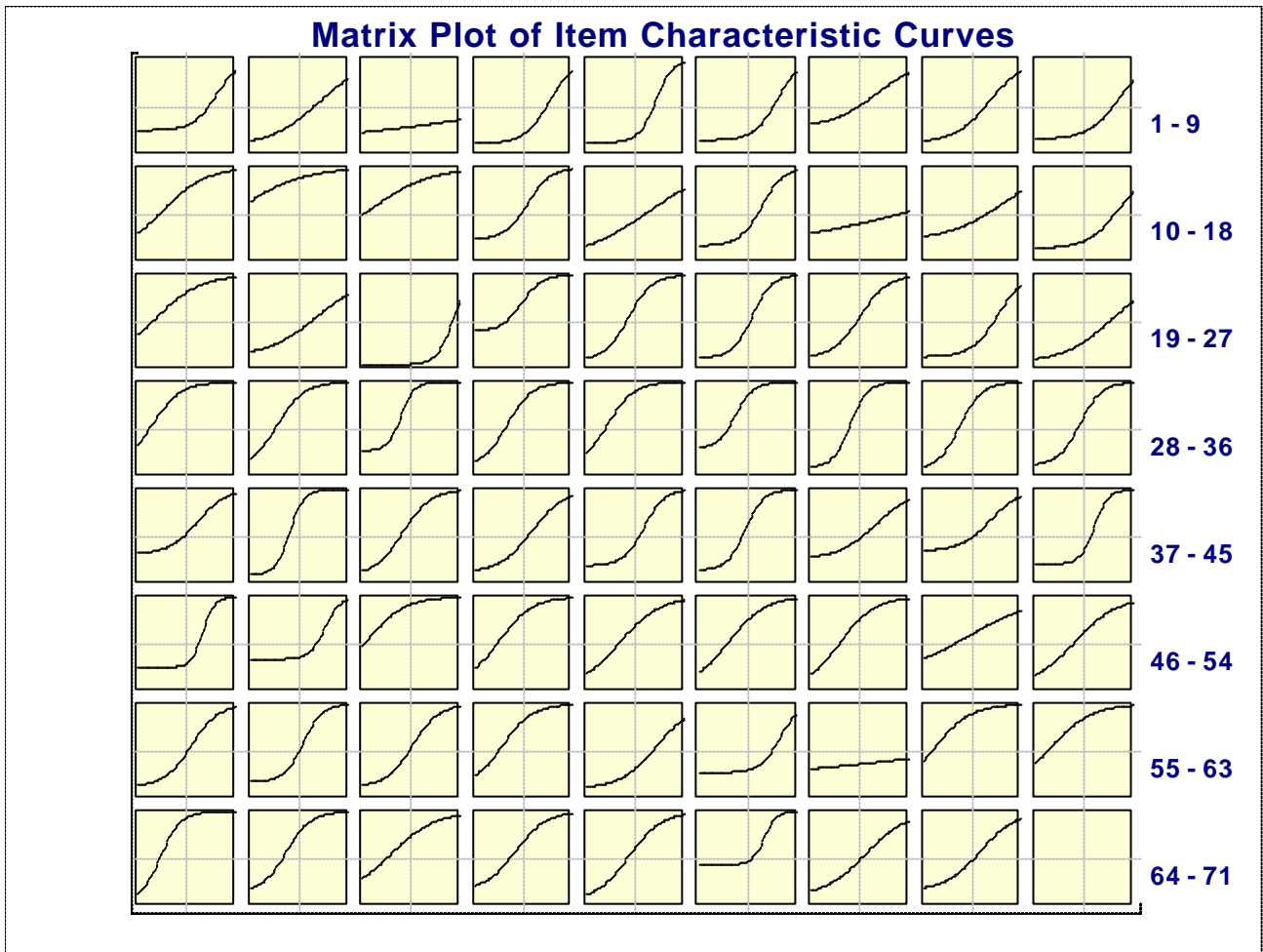


圖 附二-7、國中綜合分析能力題 3PL 模式下，各題項的答對機率曲線圖

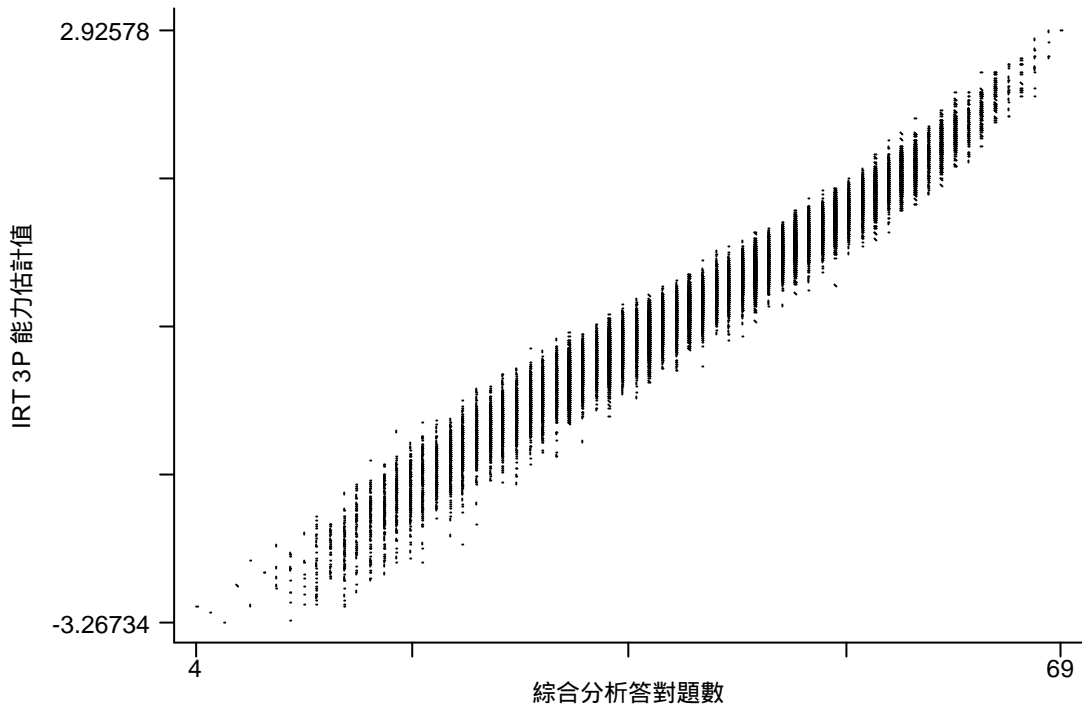
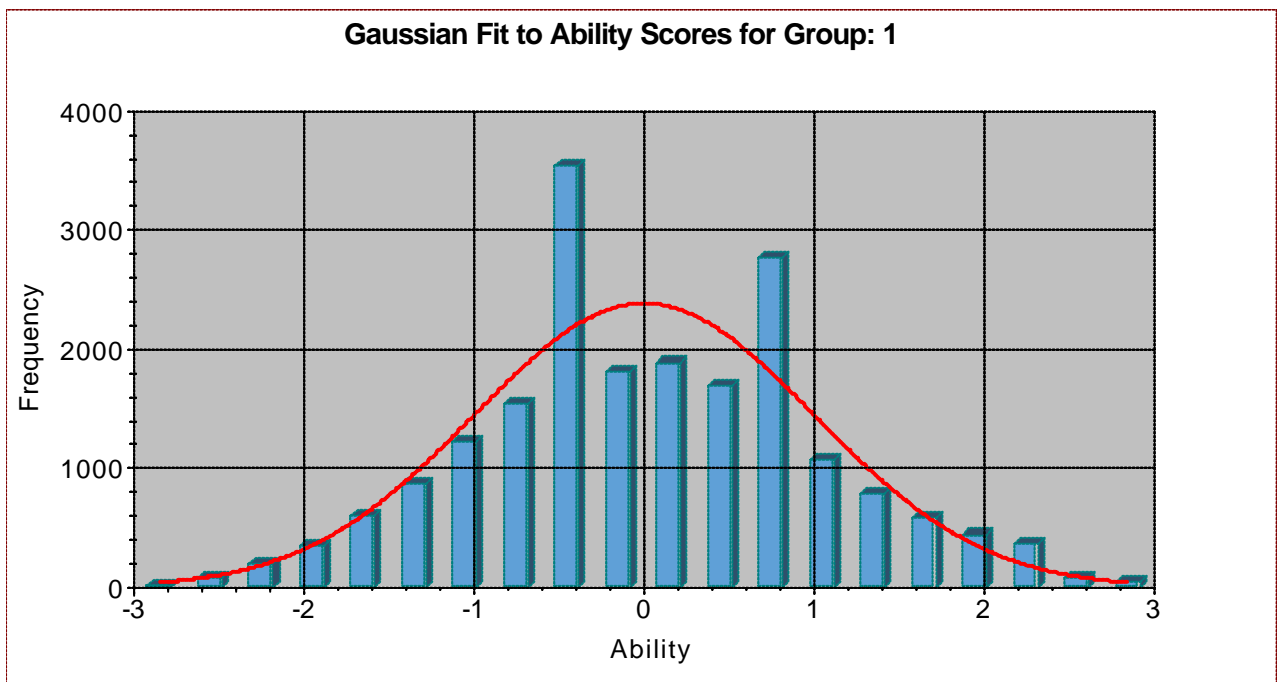
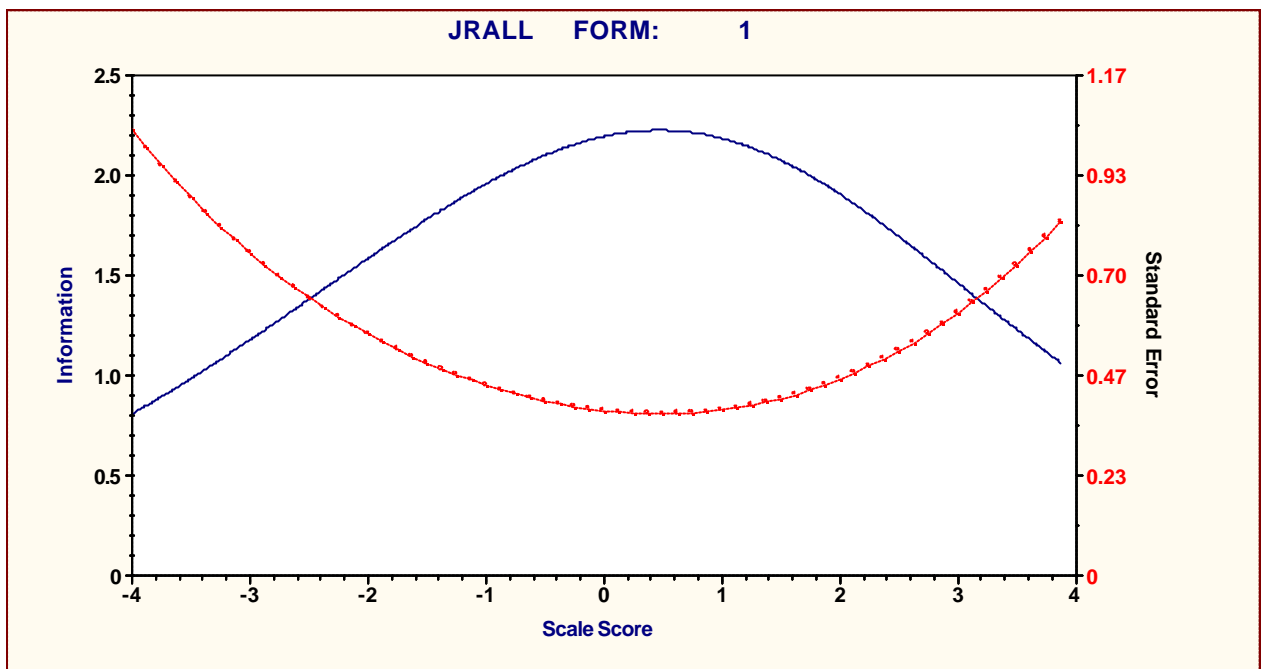


圖 附二-8、國中綜合分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $r = .984$



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附二-9、國中一般分析能力題，1PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附二-10、國中一般分析能力題 1PL 模式下的測驗訊息圖

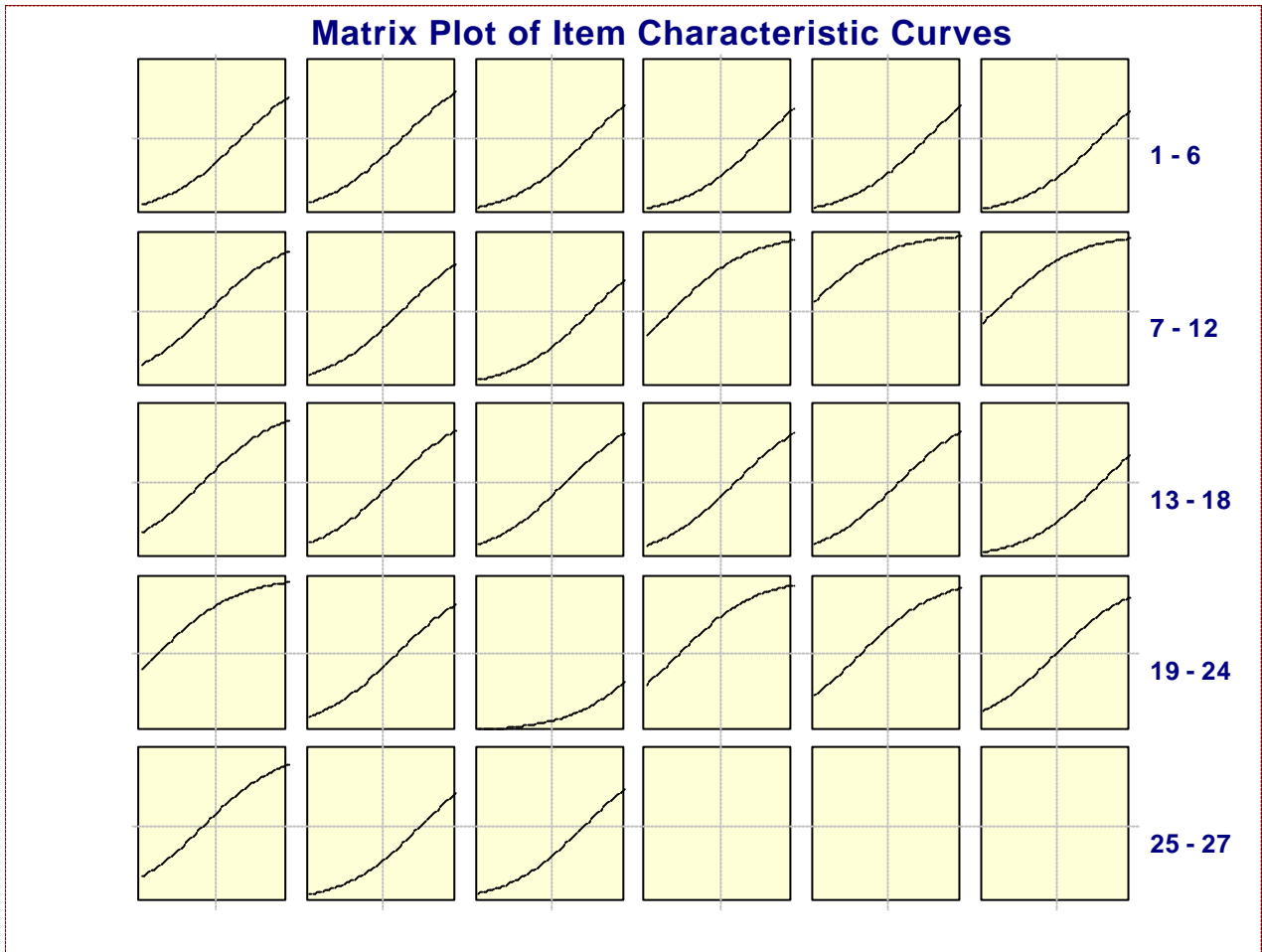


圖 附二-11、國中一般分析能力題，1PL 模式下各題項的答對機率曲線圖

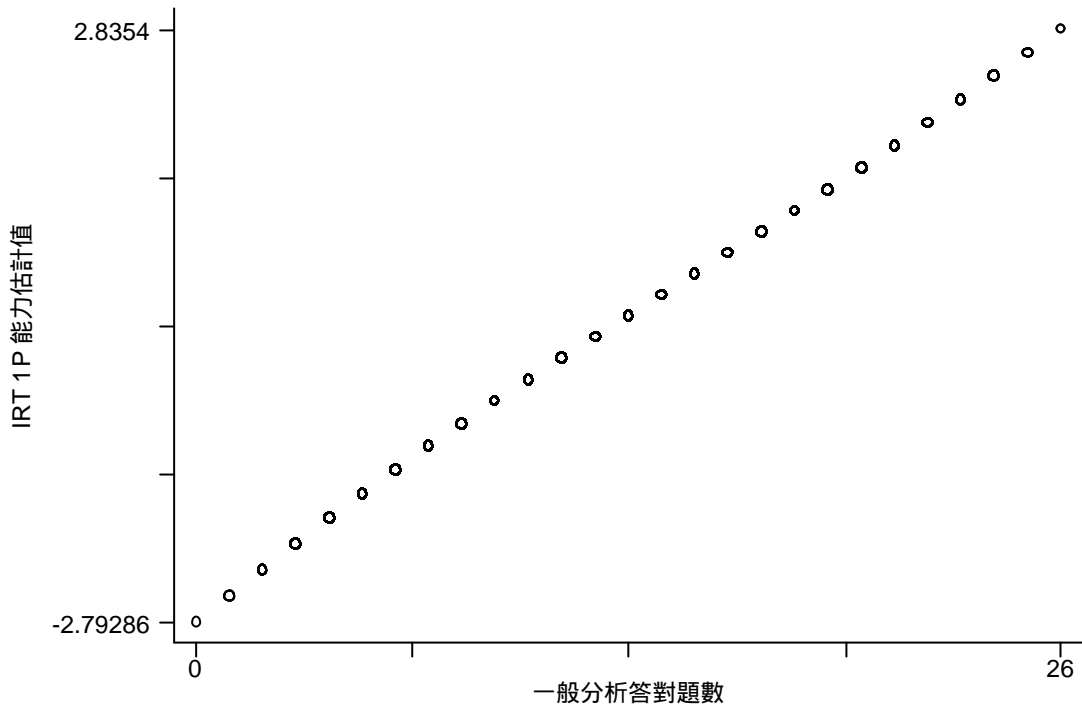
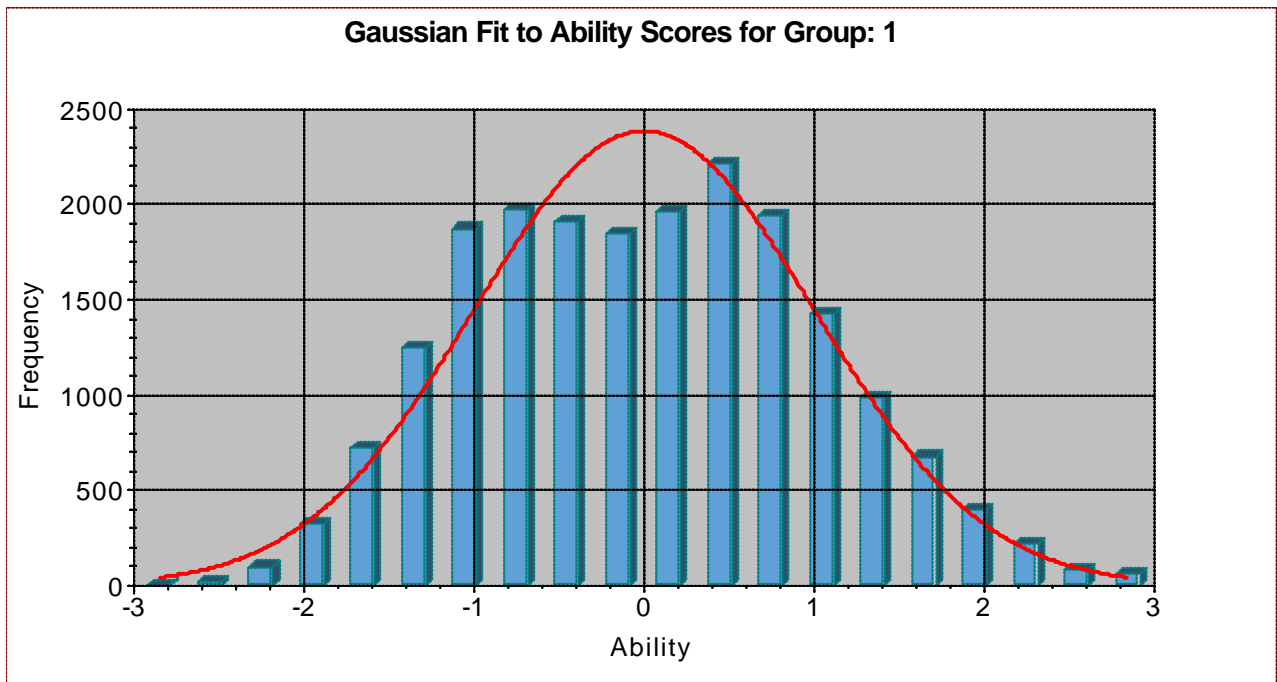
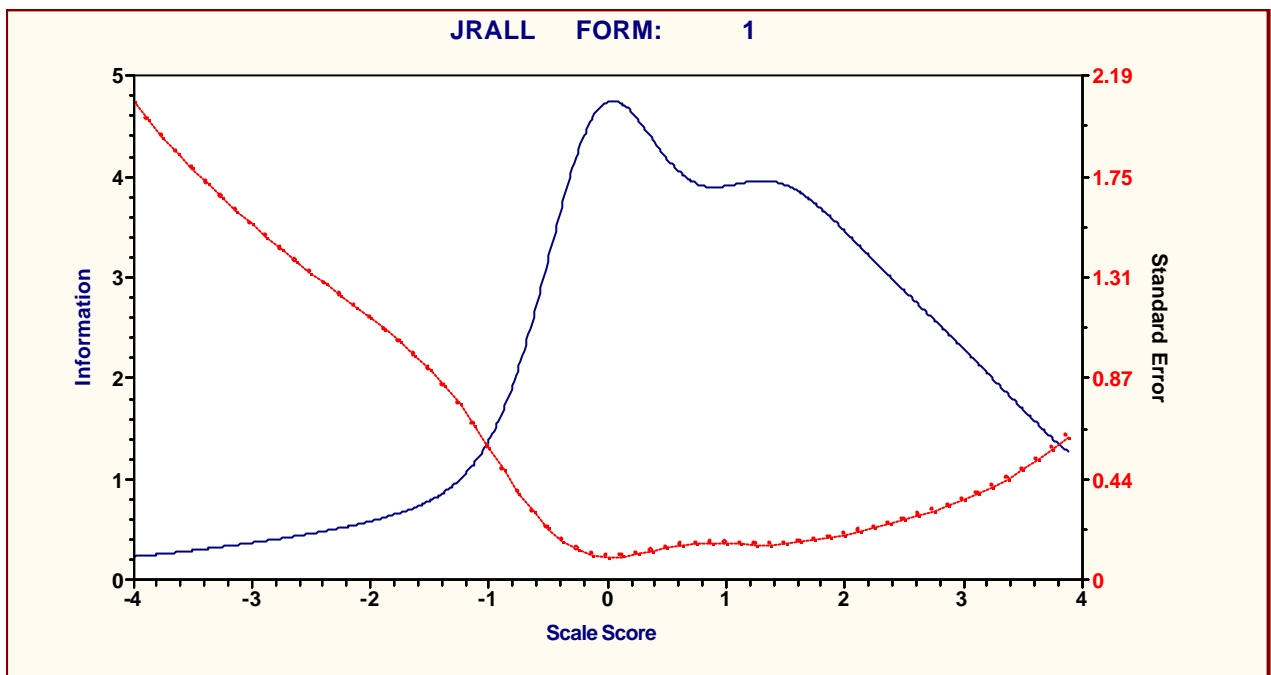


圖 附二-12、國中一般分析能力題，答對題數與 1PL 模式能力估計值的散佈圖， $r = 1.000$



Histogram of abilities
 Ability scores are transformed to have a mean of zero and a standard deviation of 1.
 The area under the bell-shaped curve equals the total area of the histogram.

圖 附二-13、國中一般分析能力題，3PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附二-14、國中一般分析能力題，3PL 模式下的測驗訊息圖

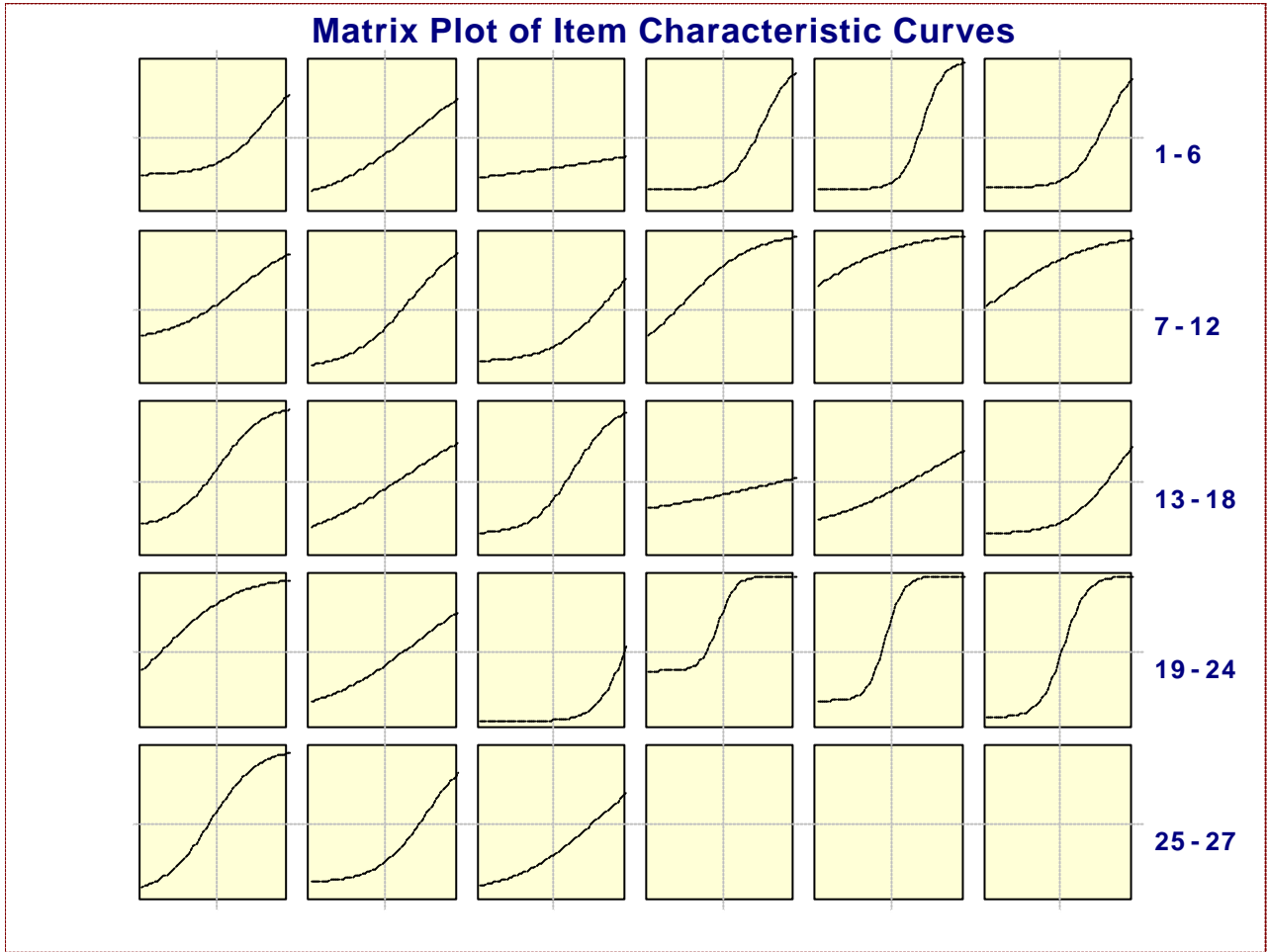


圖 附二-15、國中一般分析能力題，3PL 模式下各題項的答對機率曲線圖

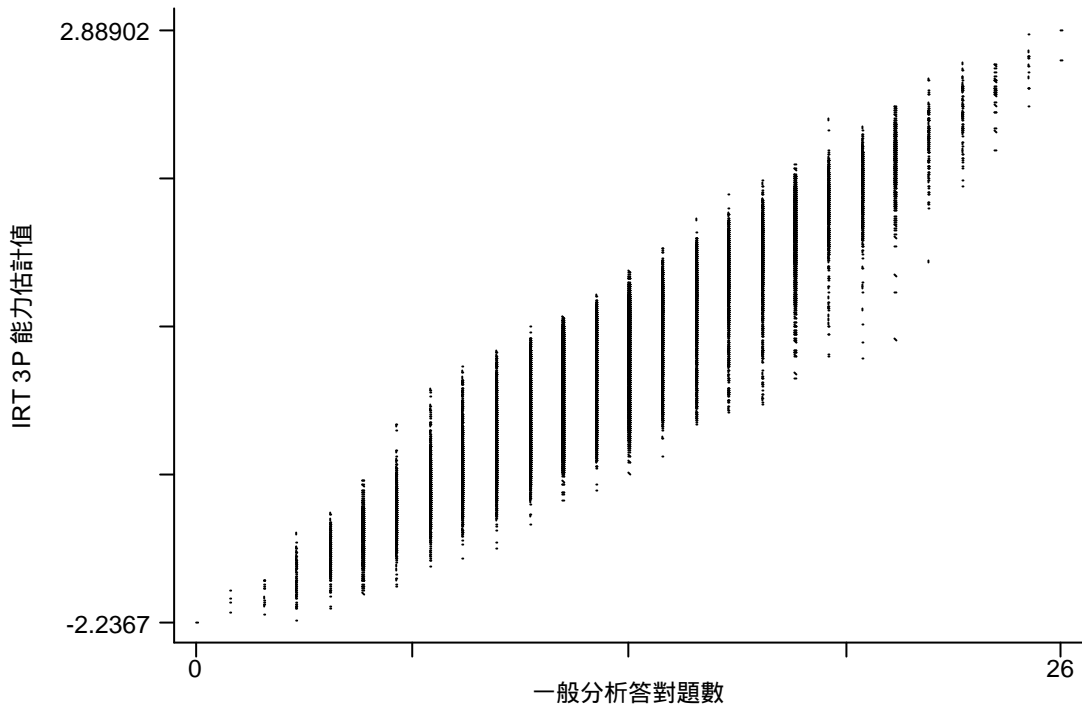


圖 附二-16、國中一般分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $r = .934$

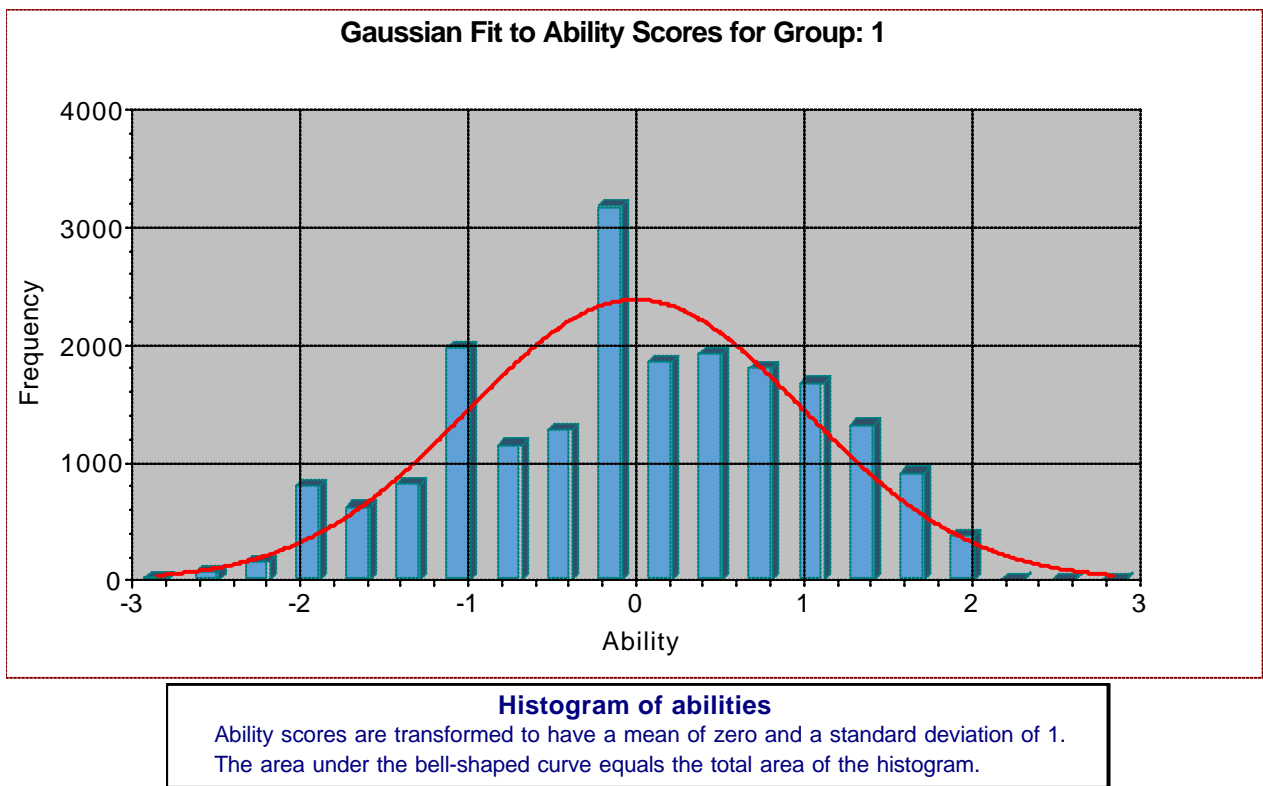
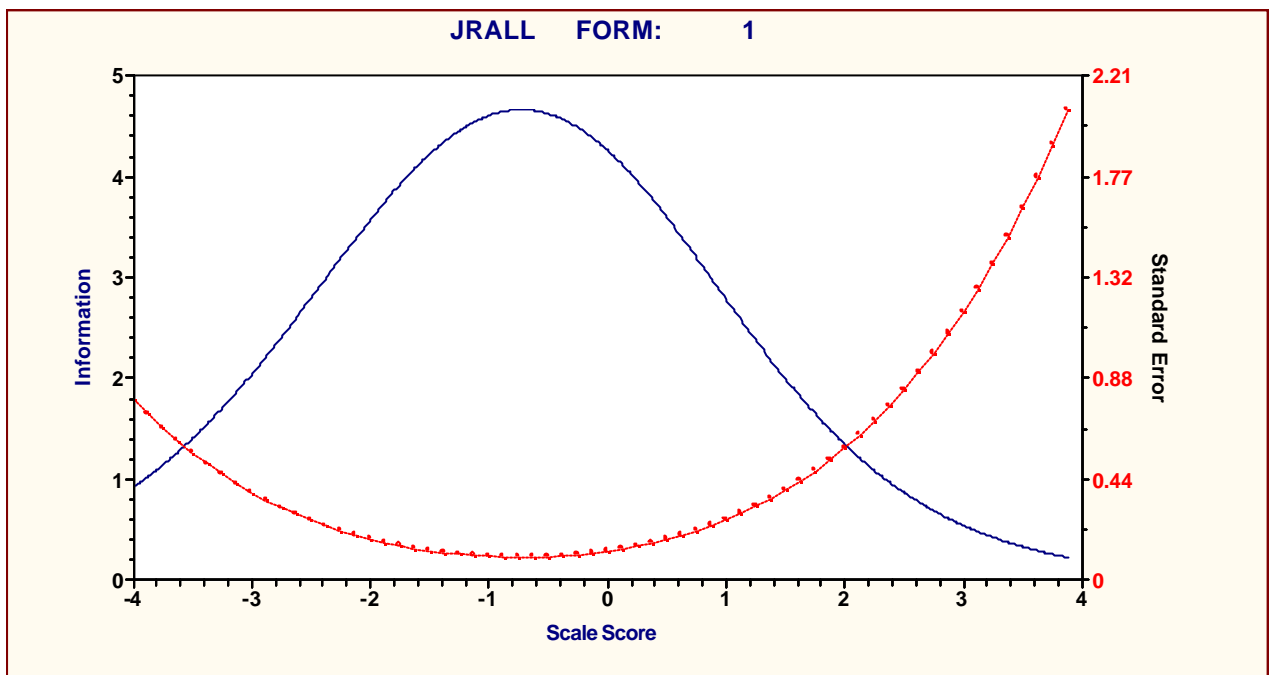


圖 附二-17、國中數學分析能力題，1PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附二-18、國中數學分析能力題，1PL 模式下的測驗訊息圖

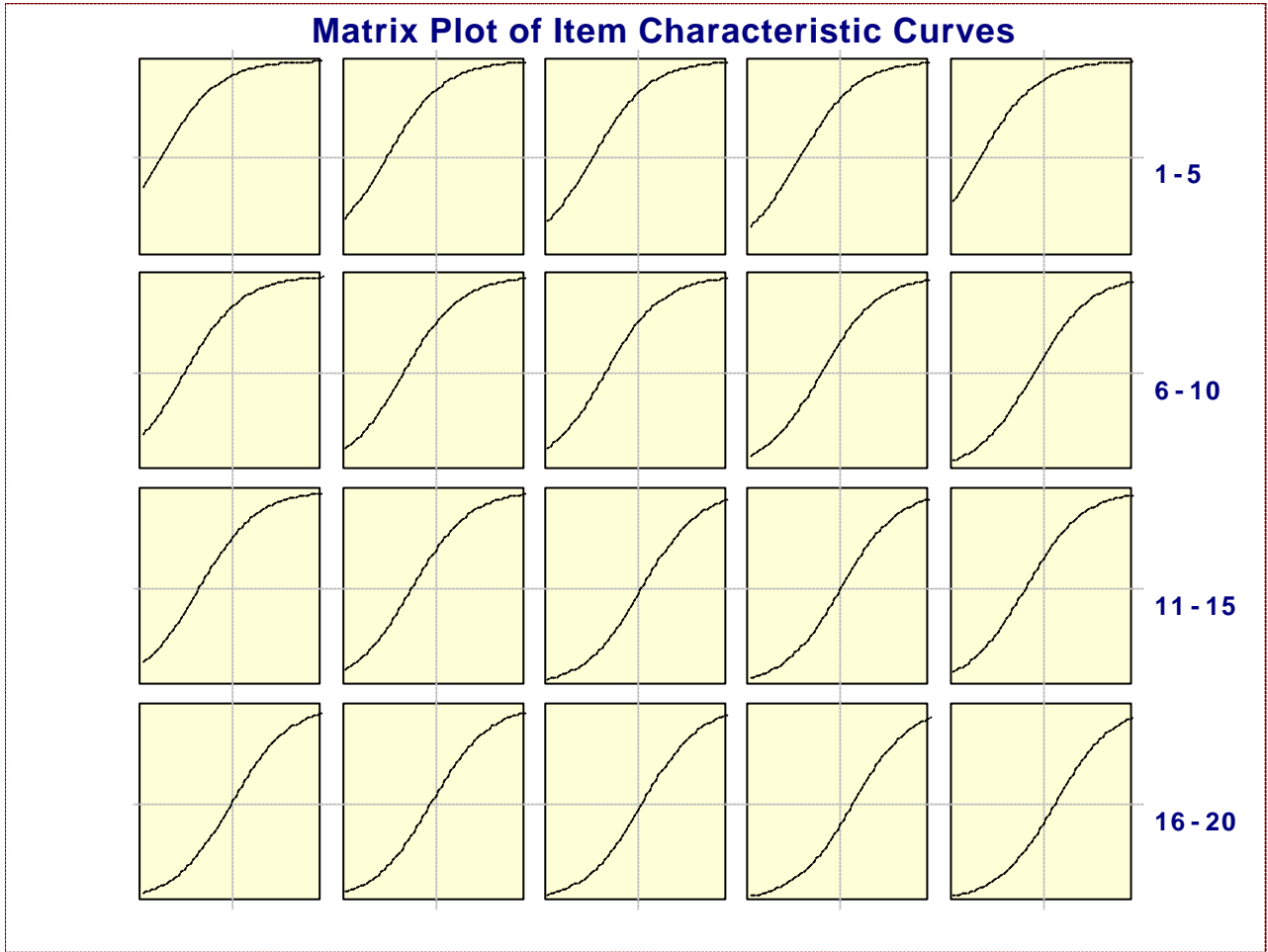


圖 附二-19、國中數學分析能力題，1PL 模式下的各題項答對機率圖

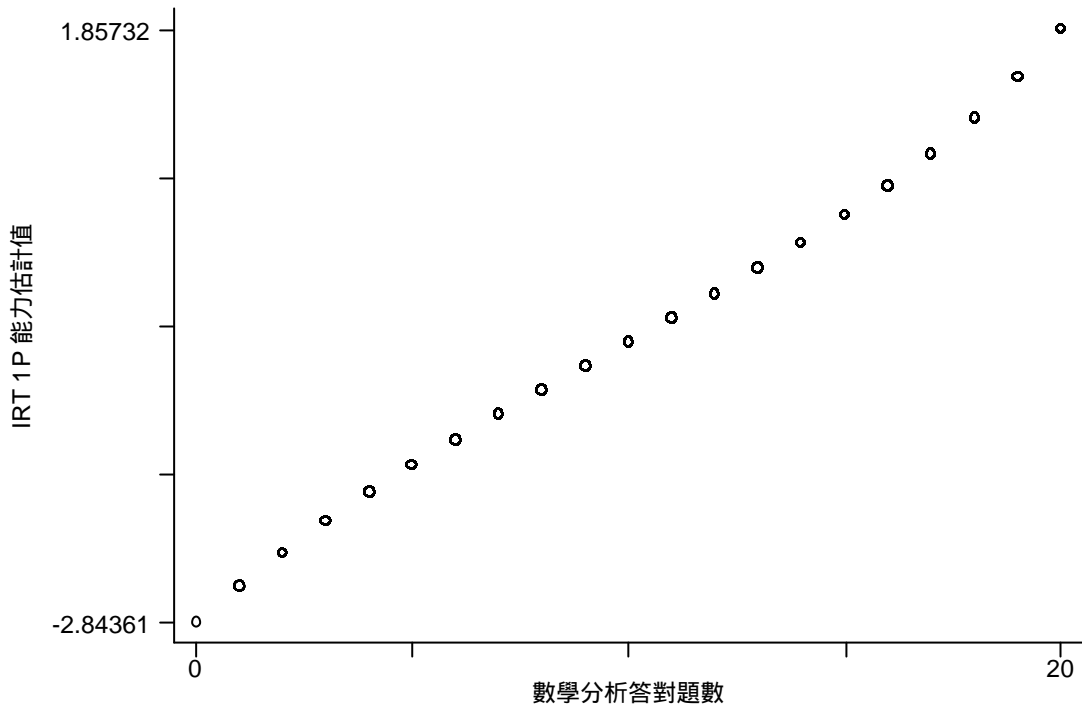


圖 附二-20、國中數學分析能力題 1PL 模式下，能力估計值與答對比率的迴歸線， $r = .997$

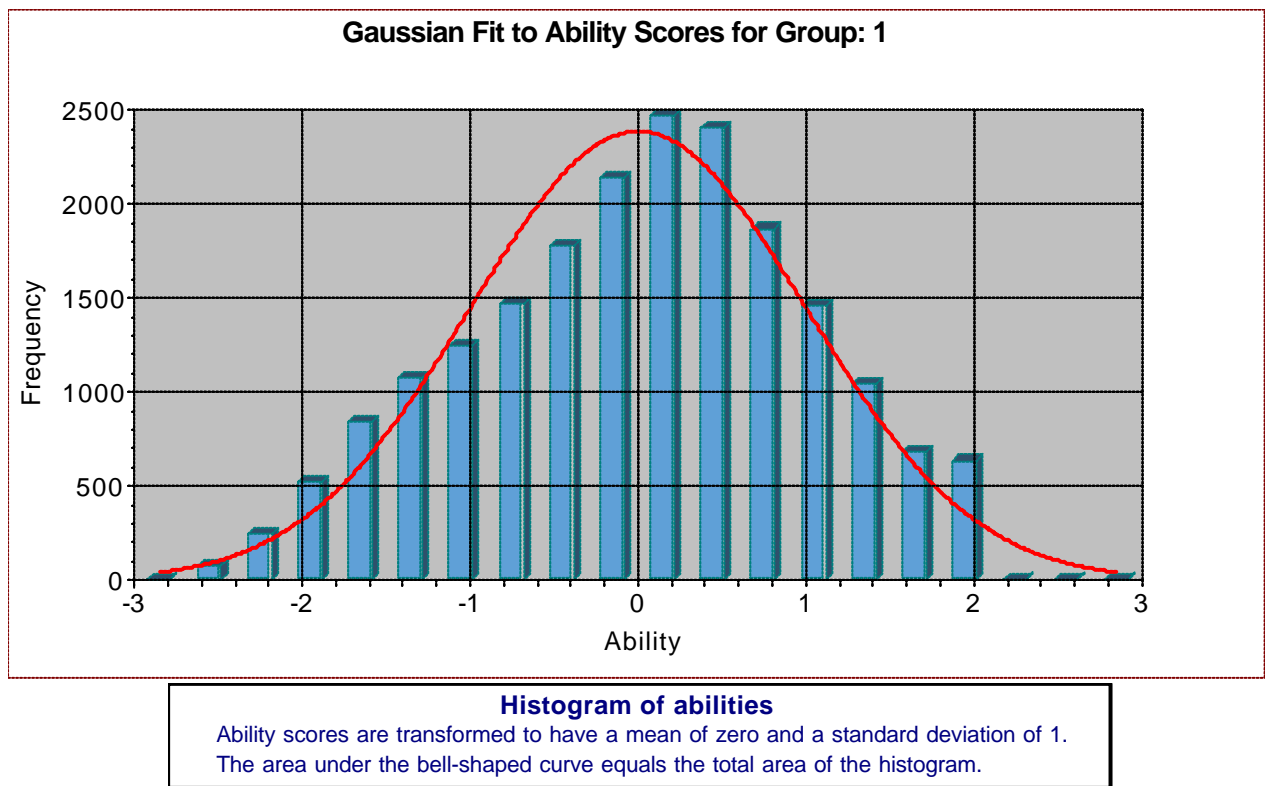
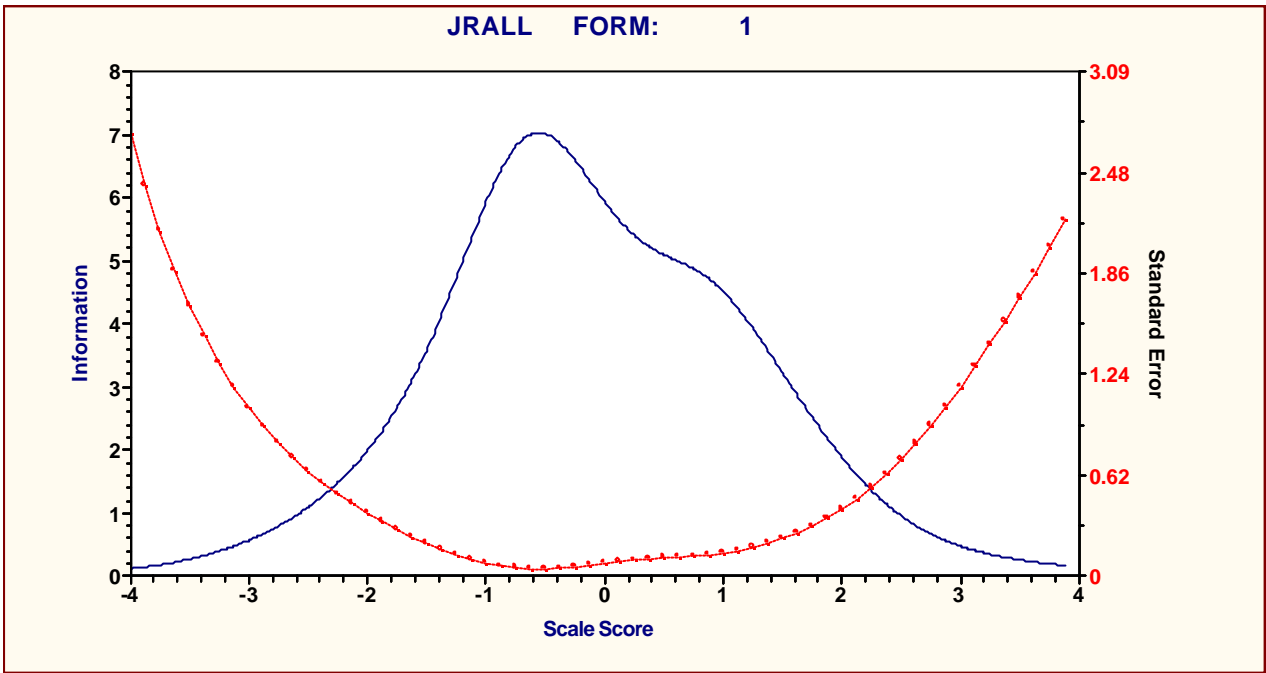


圖 附二-21、國中數學分析能力題，3PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附二-22、國中數學分析能力題，3PL 模式下的測驗訊息圖

Matrix Plot of Item Characteristic Curves

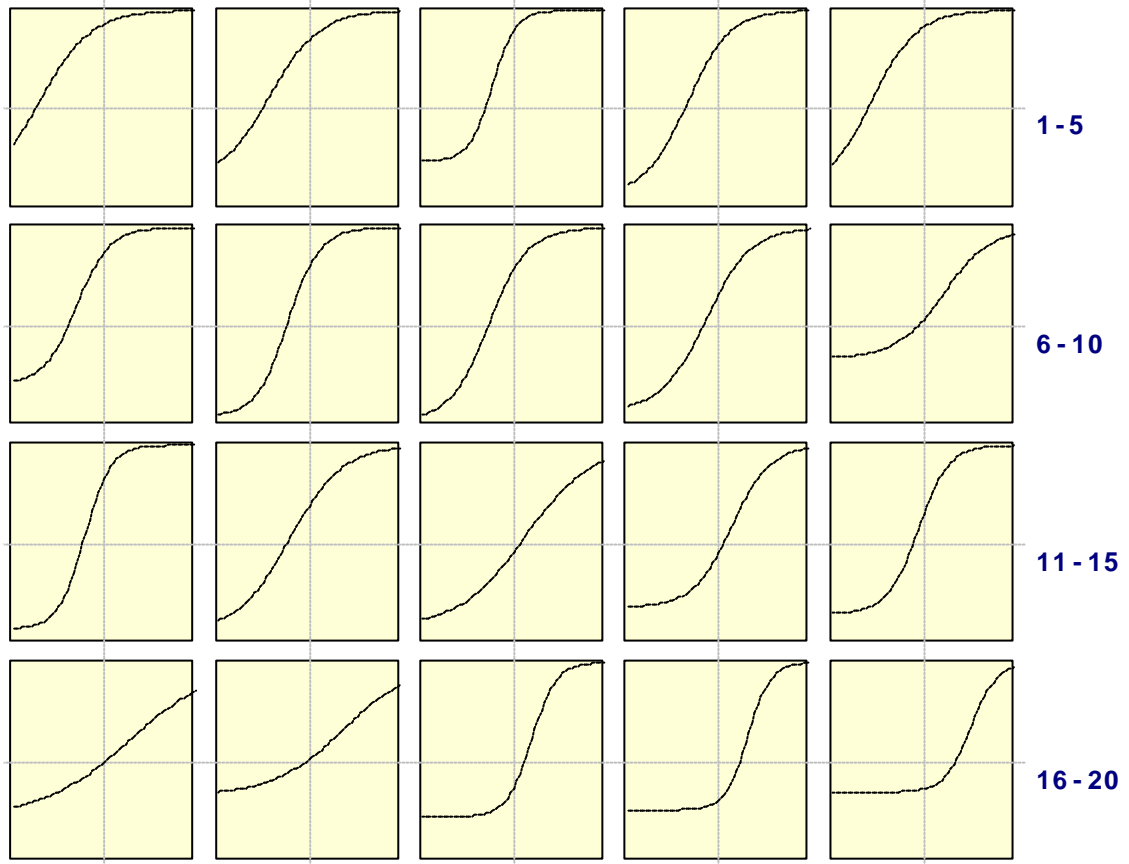


圖 附二-23、國中數學分析能力題，3PL 模式下的各題項答對機率曲線圖

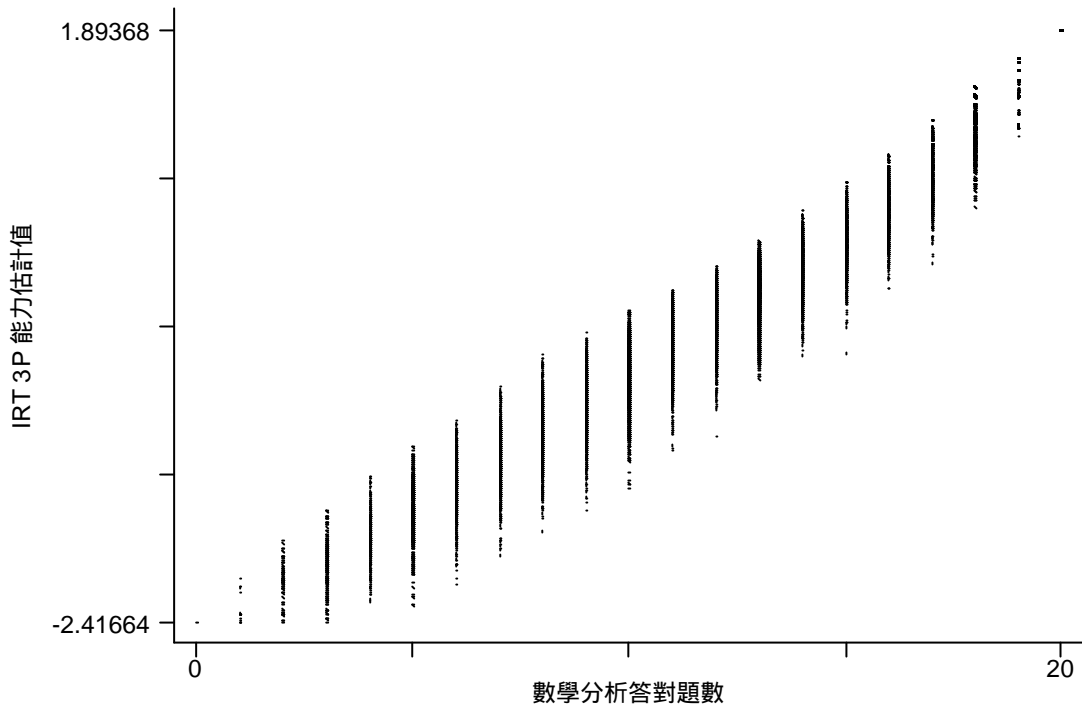


圖 附二-24、國中數學分析能力題，答對題數與 3PL 模式能力估計值的散佈圖， $r = .975$

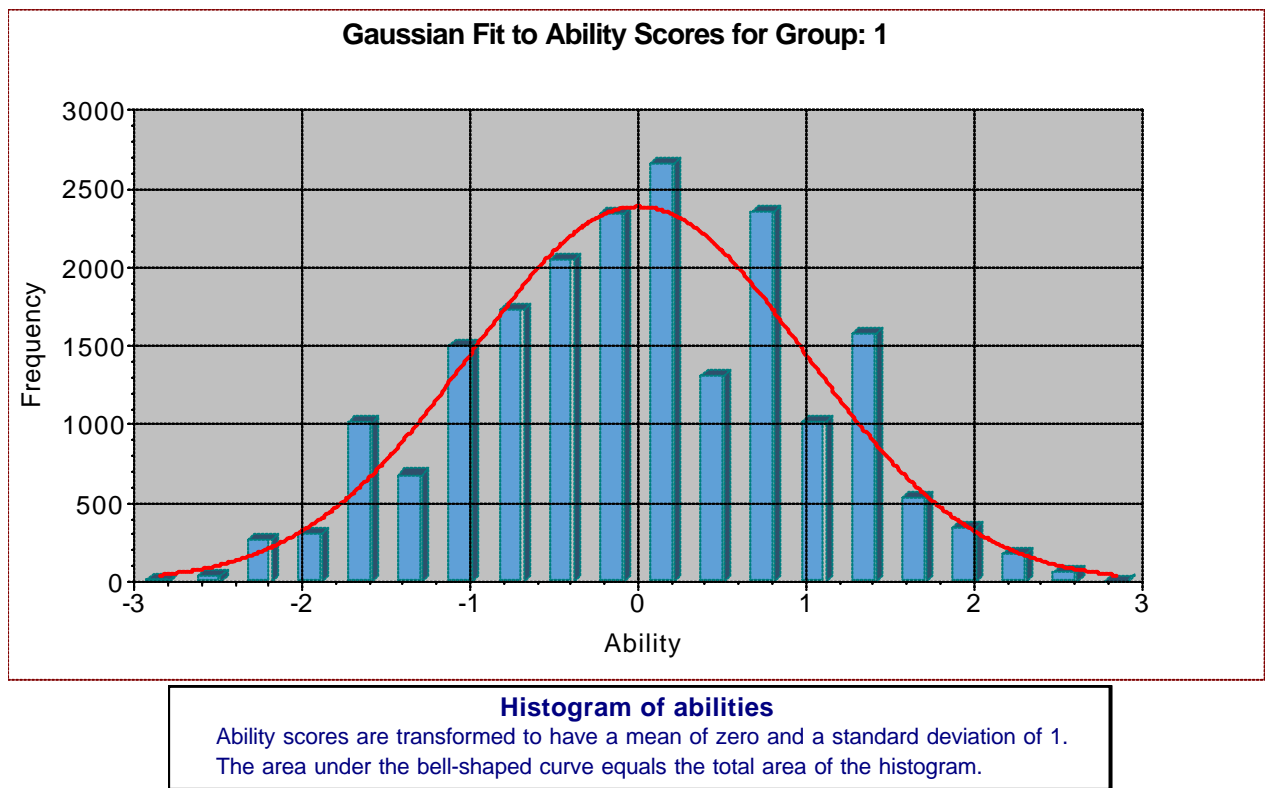
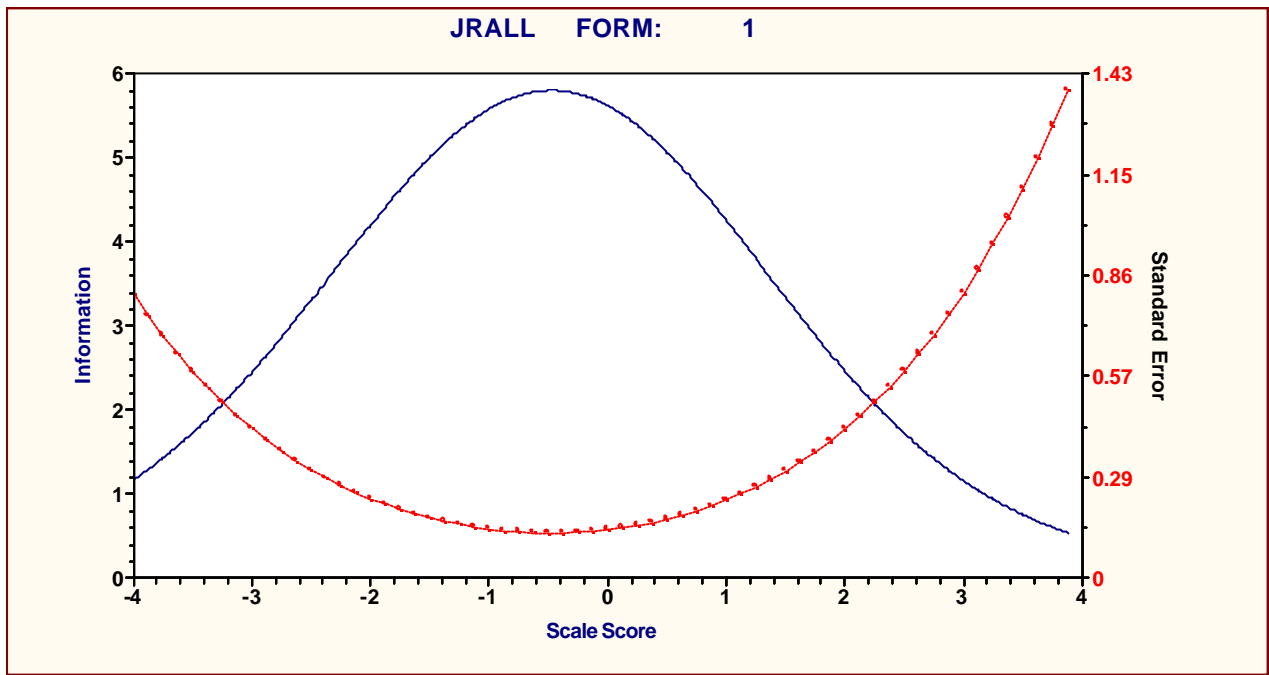


圖 附二-25、國中數學或數字型分析能力題，1PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附二-26、國中數學或數字型分析能力題，1PL 模式下的測驗訊息圖

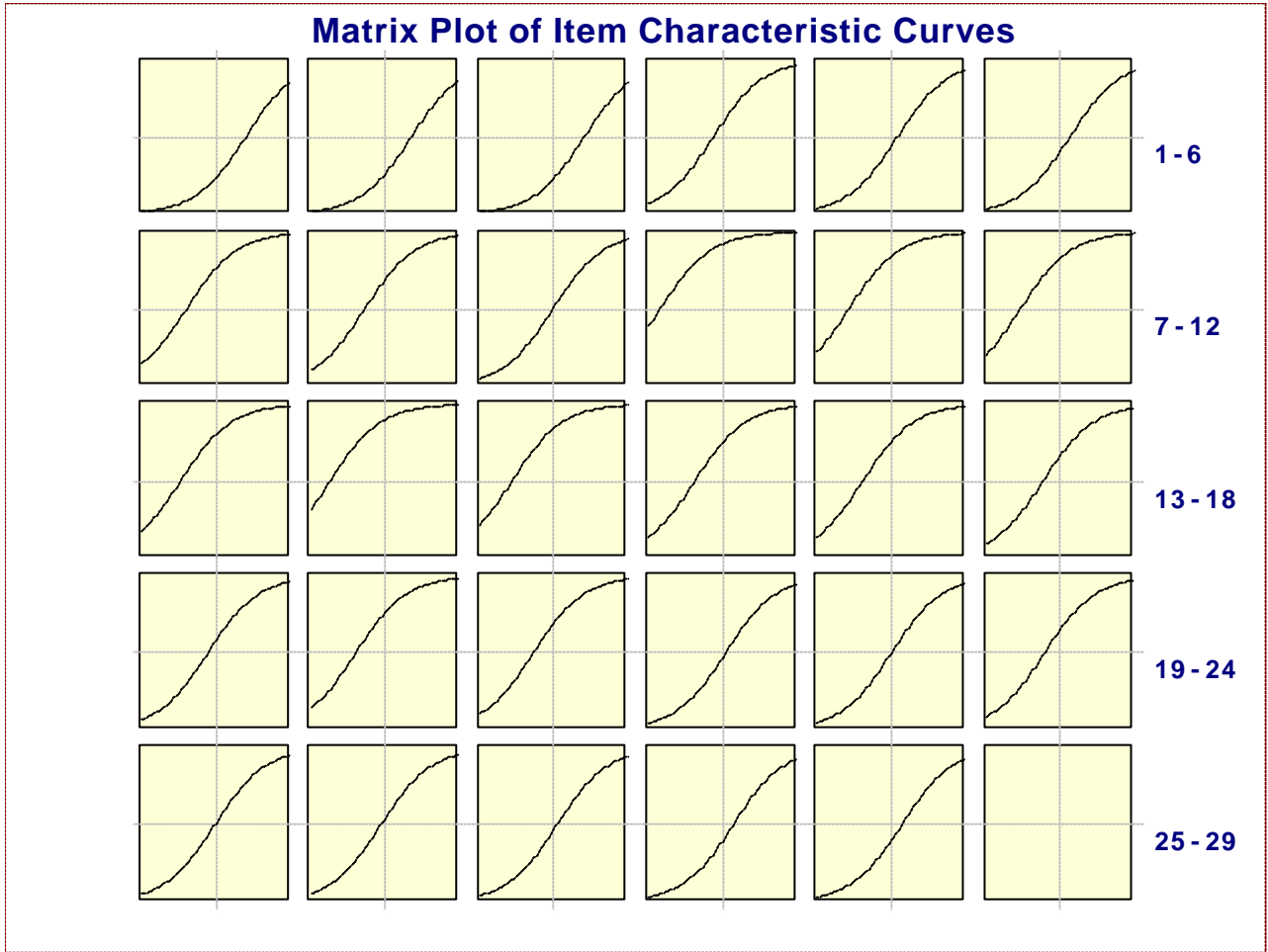


圖 附二-27、國中數學或數字型分析能力題，1PL 模式下的各題項答對機率圖

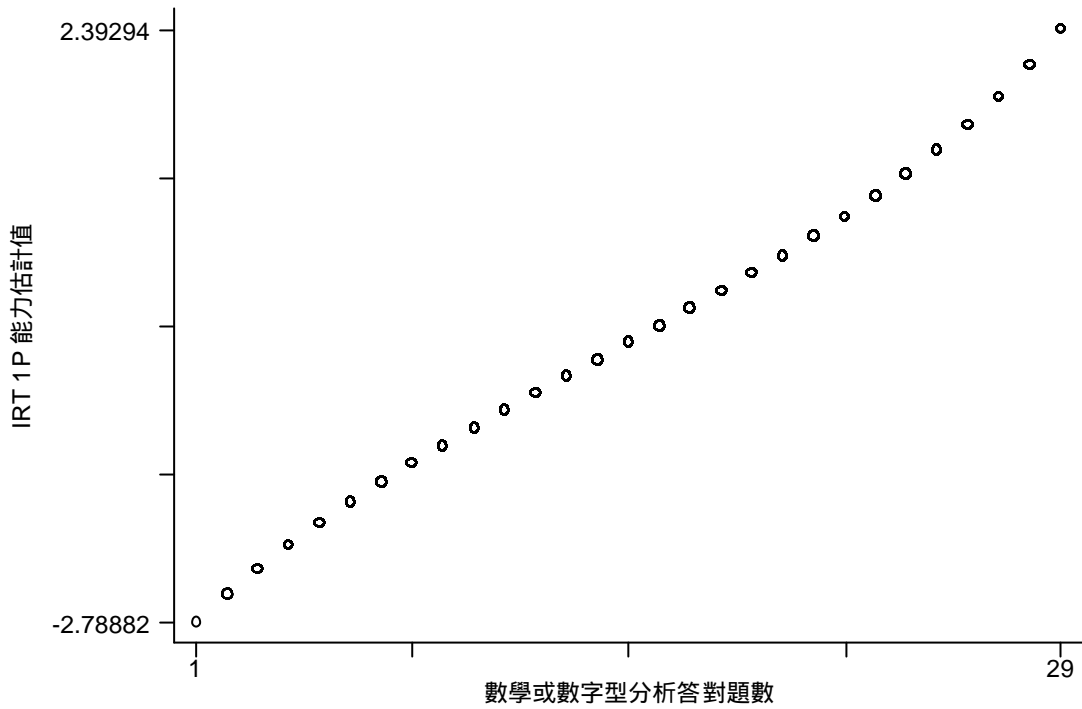


圖 附二-28、 數學或數字型分析能力題，答對題數與 1PL 能力估計值的散佈圖， $r=.997$

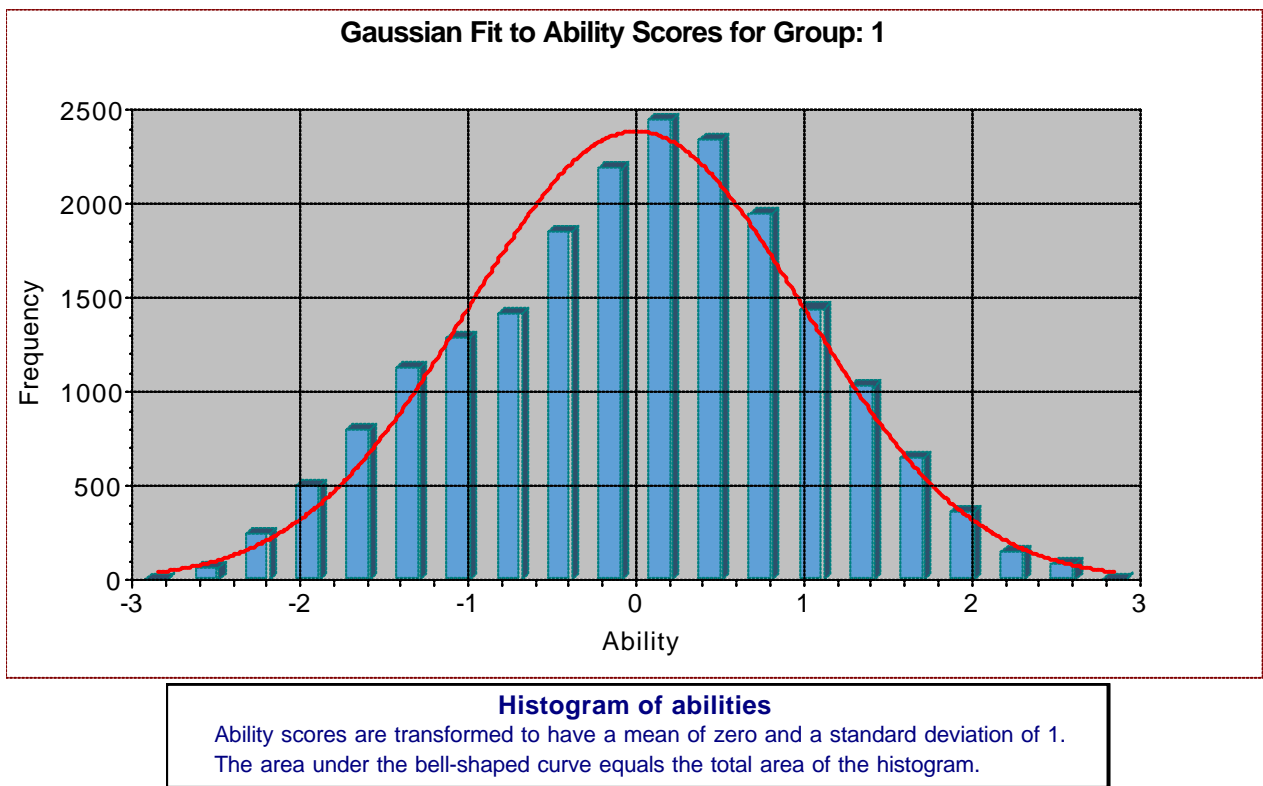
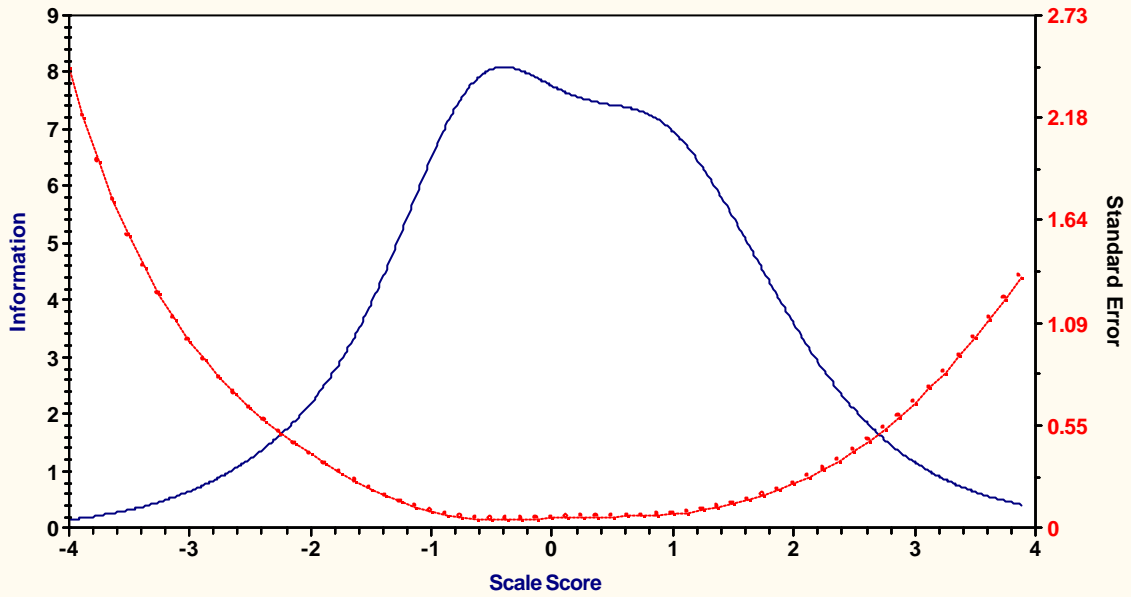


圖 附二-29、 國中數學或數字型分析能力題，3PL 模式下的能力估計值分佈長條圖



Test information curve: solid line **Standard error curve: dotted line**
 The total test information for a specific scale score is read from the left vertical axis.
 The standard error for a specific scale score is read from the right vertical axis.

圖 附二-30、國中數學或數字型分析能力題，3PL 模式下的測驗訊息圖

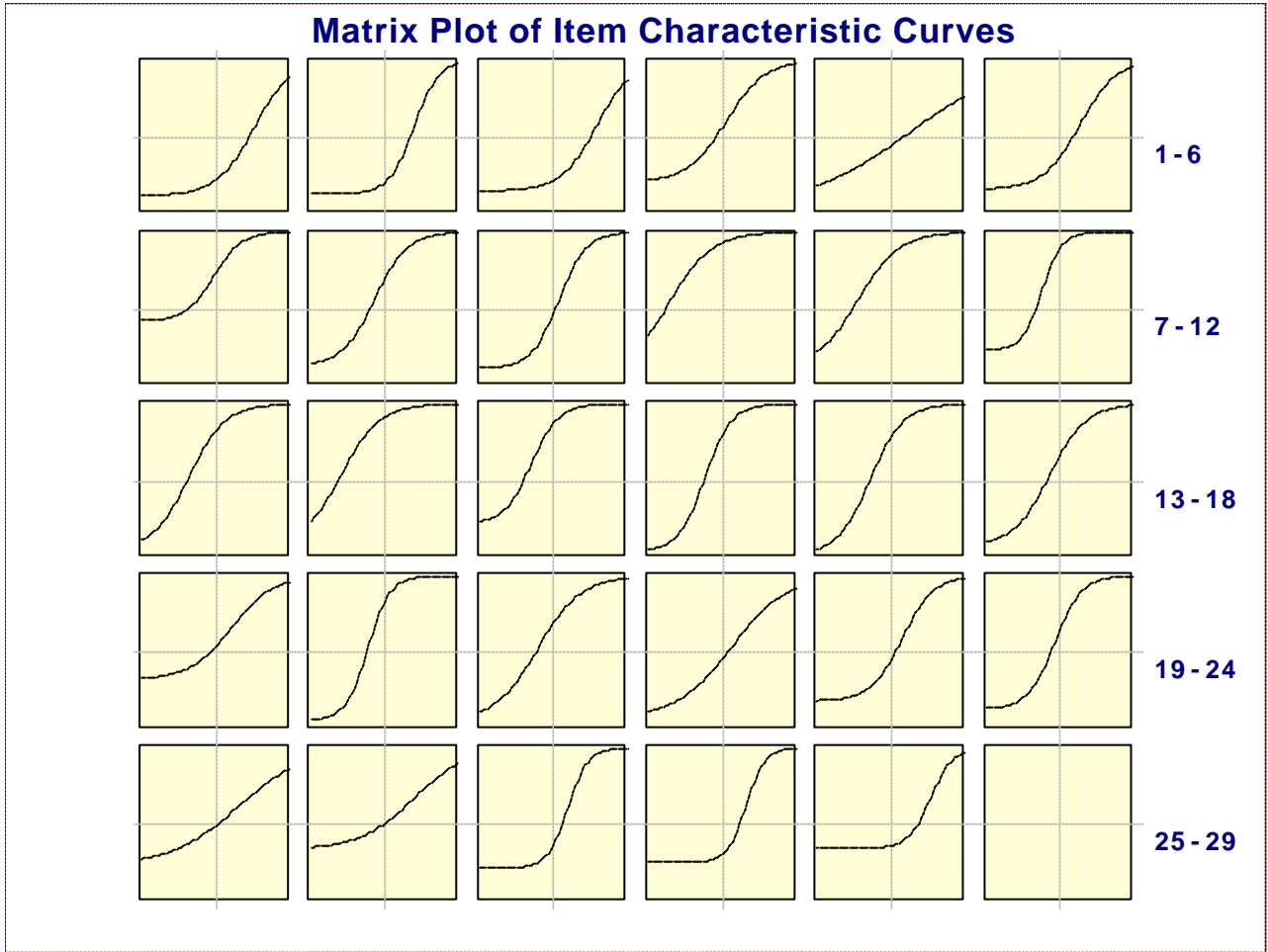


圖 附二-31、 數學或數字型分析能力題，3PL 模式下各題項的答對機率曲線圖

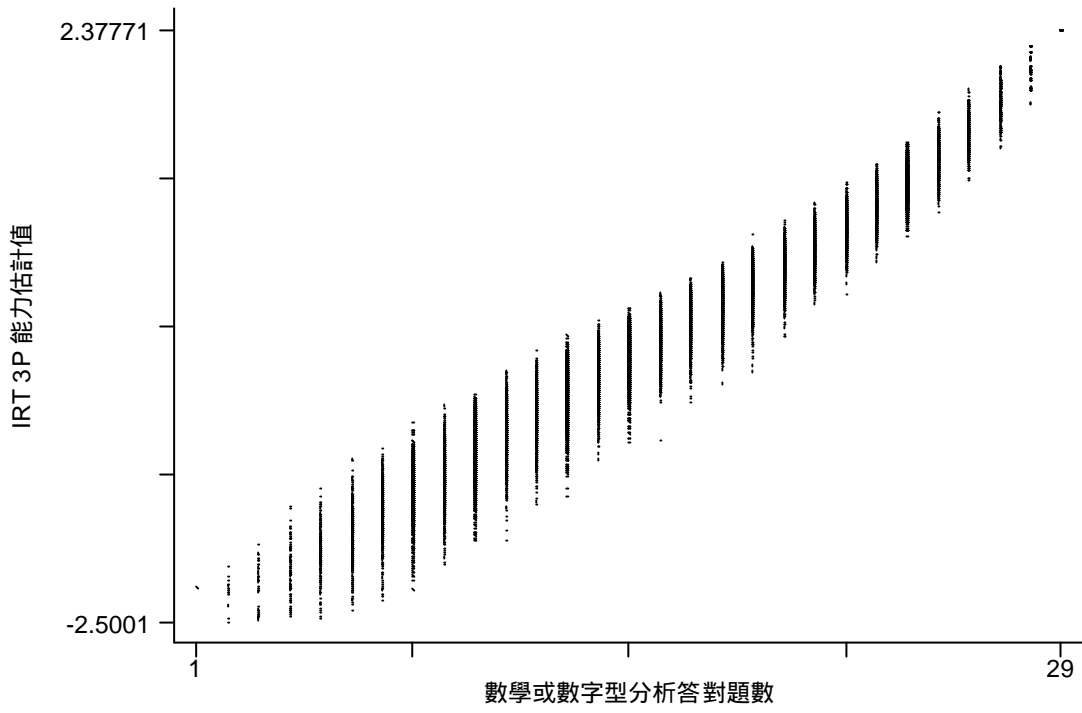


圖 附二-32、國中數學或數字型分析能力題，答對題數與 3PL 模式能力估計值的散佈圖，
 $r = .980$