

受試特徵曲線在能力測量之應用研究

范 德 鑫

本研究旨在探討試題邏輯分析與受試特徵曲線 (person characteristic curve) 在能力測量上可能之應用情形。受試者取自國中一、二年級男、女生前、後段班四二九人, 以實施普通分類測驗語文理解與方塊計算兩類試題的結果作為分析與研究的資料。具體研究項目包括: (1) 利用試題邏輯分析方法, 找出造成方塊計算試題之難度差異的因素。(2) 探究受試特徵曲線在鑑定試題測量單一能力上運用之適當性, (3) 男、女限制羣體 (restricted group) 是否影響到受試特徵曲線之運用。(4) 前後段限制羣體是否影響到受試特徵曲線之運用。(5) 試題鑑別力指數是否影響受試特徵曲線之功能。

研究結果發現: (1) 試題圖形之是否清楚, 和方塊之排列方式影響方塊計算試題的心理計量的難度。(2) 受試特徵曲線具有鑑定測驗是否測量單一能力之功能。(3) 限制羣體不影響受試特徵曲線之運用。(4) 刪除鑑別度 .30 以下之試題後, 所剩優良試題仍能維持該能力測驗之特徵。

筆者的建議: (1) 重視使用各種方法找出造成試題困難差異之因素, 以作為能力界定之基礎。(2) 使用受試特徵曲線以鑑定試題是否測量單一能力¹, 以保障測驗之構念效度 (construct validity)。(3) 測驗編製者應進一步利用受試特徵曲線之功能以改進測驗之實施。

能力 (ability) 就是個人的學習潛能與經由教育或訓練而獲得的知識, 通常包括性向 (aptitude) 和成就 (achievement) 二者。所謂性向是個人進一步訓練獲益之潛能, 而成就係指個人已獲得能力的程度 (Aiken, 1985)。用來測量個人能力的測驗就是能力測驗, 它包括性向測驗和成就測驗, 其中性向測驗主要是用於預測將來的成就; 成就測驗則在評量現有之知識與了解。就測驗題目本身, 實難以判斷它到底是屬於性向測驗或成就測驗; 以區分性向測驗組合中之機械性向測驗槓桿的題目為例, 吾人實難以判定它是測量性向還是成就的。因此, 從受試作業結果觀察界定能力, 性向與成就的差別實無意義 (Carroll, 1985)。

欲對個人的能力進行測量, 應從「能力」的界說開始探究。依據美國傳統字典 (American Heritage Dictionary), 能力是「能做某事的才能; 包括生理的、心理的、財務的或法律的作業能力」(the quality of being able to do something: physical, mental, financial, or legal power to perform)。這本字典先是把能力界定成「能做某事」, 然後又把「能」(able) 界說成有足夠的能力 (having sufficient ability)。這種繞圈子的定義方式對能力一詞的界定並無直接的幫助。在日常生活中, 對能力的解釋大多採事後方式 (posterior manner), 即先知道一個人在某件事的作業情形, 再予以概括性的界說。例如某人鋼琴彈得好, 就說這個人有音樂方面的能力。這種事後方式的說法人人習以為常, 言者、聽者皆能順利溝通。Carroll (1985) 認為從心理與測量的觀點言, 美國傳統字典對能力雖未予以明確的界定, 但界說中所提及的「某事」二字却提供心理測驗編製者對能力測量研究之重要訊息。在心理與教育的情境裏, 定義中的「某事」不必一定是某類的工作 (task); 因為有運動或音樂的能力的人未必運動或音樂上各種工作都能做得很好。當心理學家與教育學者提到心理能力 (mental ability), 也不應指各種心理工作上的作業 (

performance)；因為一個具有音樂欣賞素養的人不一定是能擅於演奏樂器，一個擅於彈琴的人也未必是一個優秀的作曲家，而彈琴、作曲、欣賞却都是音樂能力之表現方式 (Davies, 1975)。由此可見，人類能力是具有相當分化特徵的 (Carroll, 1985)。

過去學者們 (如 Thurstone, 1938; Vernon 1965; Cattell 1971) 曾使用相關與因素分析的方法研究人類的心理能力。他們發現心理能力包含一些彼此不相關或相關甚低的能力，這種研究方法已被學者們證實能提供能力的鑑定和分類資料；指出那些工作或試題可以測量某一因素所代表的能力，那一些測量另一種能力；那些能力可以合併，那些應該分開，也就是能提供需要不同能力的工作分類的訊息。在心理與教育情境中，需要某一特定能力的工作不計其數，而難度是造成這些工作不同的原因之一。到底工作難在那裏，一般說來應該是可以利用一些方法找出來的，例如：距離的遠近就是造成足球員射門的困難度的原因之一，音調高低之相差量即是形成受試者辨別音調之困難的因素。但在心理與教育測量中，通常是以能通過試題人數與總人數之比例來代表這個题目的難度。測驗的試題就是工作受試者做對的工作愈多，分數愈高；分數愈高，則能力愈強。

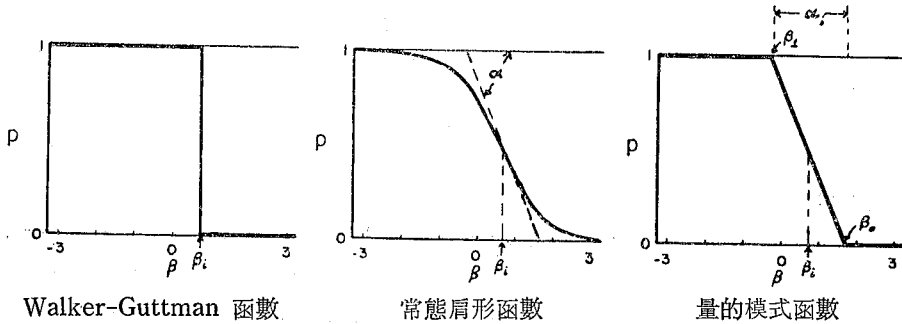
試題是否測量單一能力與一個測驗的構念效度有密切的關係。所謂構念效度就是能夠測量到理論上的構念或特質的程度 (Anastasi 1982)。過去初步評鑑測驗構念效度的方法都是只從測驗中的工作是否屬於同一種類來考慮 (Carroll, 1985)。例如，如果所有題目與語文推理有關，這個測驗就認為語文推理能力的測驗。所有題目若與視覺表徵空間關係操弄有關就視為是空間能力的測驗。由於工作的性質關係，這種直覺的工作分類常無法保證只是測量唯一語文推理能力或空間關係的能力，這種情形已被學者們所證實 (Lohman, 1979)。若有這種情形發生，不但影響能力測驗之構念效度，測驗結果也無法作正確的解釋。Allen 和 Yen (1979) 認為了解受試者對試題的處理過程能提供測驗分數的構念效度之訊息。處理過程則可經由檢查試題的內容、試題間的相關研究、分析與學生之晤談等方法達成。毫無疑問的，Allen 和 Yen 的看法已指出了改進測驗需從受試者對試題處理的過程開始。Bejar (1985) 認為，近年來對認知心理學的強調，影響心理計量學至少有兩方面：一是從認知構念了解測驗作業之可能性；另一是利用認知心理學改進現有的測驗或編製新的測驗。Allen 和 Yen 認為要改進測驗必須了解受試者在試題的反應過程；和 Bejar 強調認知心理學之構念在改進心理測量方面實相一致，都是強調在測驗設計時，構念效度需要進行試題間差異的研究。測驗編製者一旦能夠解釋試題間特徵差異，就能掌握住綜合利用試題特徵的知識 (Egan, 1979)。若能統整心理計量模式與認知科學兩者不但有助於提升心理計量工具之科學地位，而且能夠將進步的科技使用在施測過程上。

測驗試題是否測量單一能力 (single ability) 對測驗的編製與測驗結果分數的解釋兩方面皆極其重要。近年來單一能力之研究一直是傳統測驗理論與項目反應理論 (item response theory) 學者所重視，因為測驗試題是否測量單一能力不但影響到測驗結果分數的解釋，亦影響施測過程的運用。僅從試題的相關研究結果難以判定試題是否具有單一能力的性質，原因是試題反應之雙變量分配 (bivariate distribution) 解釋非常困難 (Carroll, 1985)。過去有很多有關於能力區分的知識，係來自於因素分析的研究；但由於傳統因素分析方法抽取因素無絕對的標準，命名亦相當主觀，以及試題本身的缺陷，很難保證試題是否測量單一能力。近年來，有學者 (Wilson, Wood, & Gibbons, 1982) 曾設計一套有別於傳統因素分析的電腦程式，極有可能解決這個問題。

英國教育心理學家 Walker (1931, 1936, 1940) 曾於一九三〇年代提出「單一」(unig) 和「零亂」(hig, 即 higgledy-piggledy) 兩個名詞，以敘述測驗試題是否測量單一能力或多種能力。如果工作的難度水準和受試者之特徵呈現系統關係的測驗，Walker 稱它為單一能力測驗。所謂系統關係，簡單的說，就是能夠作答較難試題的受試者比無法作答者，有一致較高通過較易題目之可能性；受試者不會作答容易試題，也就無法通過較難的題目。相反地，沒有這種特性的即稱為非單一能

力測驗。

根據 Walker 單一能力的說法，受試特徵函數 (person characteristic function) 應可用作檢驗試題是否測量單一能力之用。受試特徵函數所呈現之圖形稱之為受試特徵曲線，它是表示個人或羣體成功機率與工作難度間潛在關係的圖形。成功機率與工作難度的關係，通常有 Walker-Guttman 函數，常態肩形函數 (normal ogive function) 和量的模式函數 (quantal model function) 三種。其函數所形成之曲線如圖一(a)、(b)和(c)。圖形中 p 值表示成功之機率， β 表示工



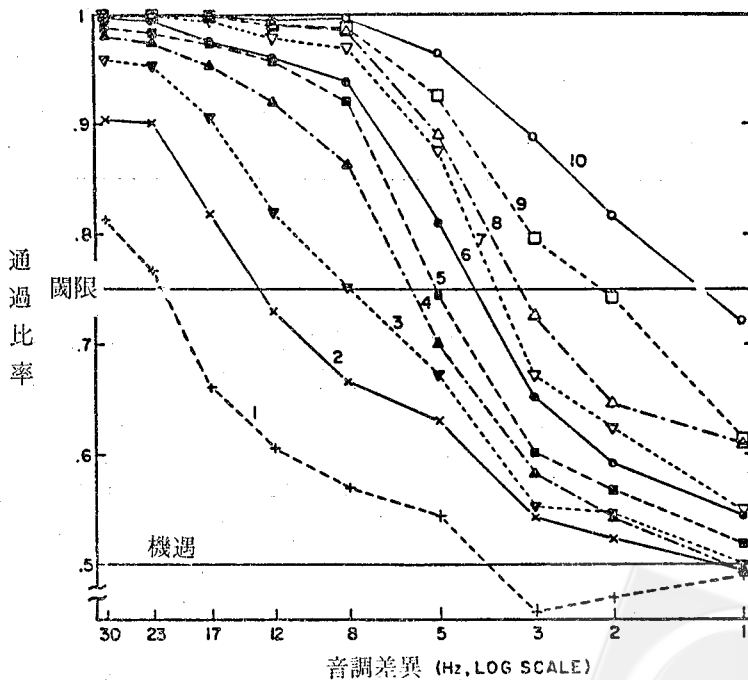
Walker-Guttman 函數

常態肩形函數

量的模式函數

圖一 成功機率與工作難度受試特徵函數之三種可能形式 (改畫自 Carroll, 1983)

作的難度， α 表斜率。在心理與教育測驗情境中，此三種函數，以常態肩形函數之出現率最高。這種函數顯現成功機率隨着工作難度之增加而變小。Carroll (1985) 曾把受試者的能力 (特徵) 區分成高低不同的十組，繪製受試特徵曲線檢驗西沙音調感覺測驗 (Seashore Sense of Pitch Test) 試題。各種能力組受試特徵曲線如圖二。



圖二 西沙音調感覺測驗十組總分不同受試特徵函數 (改畫自 Carroll, 1985)

從圖中看出受試特徵與工作難度之函數顯現系統關係，證實了此類測驗試題所要測量的音調辨別能力的確是存在的。

本研究主要是以普通分類測驗中的語文理解與方塊計算兩類試題為研究材料，以探求試題邏輯分析在能力測量上的重要性和受試特徵曲線在個人能力測量上之應用。具體研究項目包括：

- (1) 利用試題邏輯分析方法，找出造成方塊計算試題之難度差異因素；
- (2) 探究受試特徵曲線在鑑定試題測量單一能力運用之適當性；
- (3) 男女限制羣體是否影響到受試特徵曲線之運用；
- (4) 前後段限制羣體是否影響到受試特徵曲線之運用；
- (5) 試題鑑別度是否影響到受試特徵曲線之鑑定功能。

方 法

一、受 試 者：

本研究受試者取自臺北市立和平國民中學一、二年級學生，該校為配合數學資優班的成立，男、女生各取二~三班較好的學生編為前段班，其餘的學生常態編班。在本研究中稱之為前、後段班。故在取樣時考慮段別，由於前段班級較少，所以前段班各年級只取二班（男、女各一班），後段班則各年級取四班（男、女各兩班），總共十二班，465人。不合作或太多未完成之受試試卷先行剔除，最後僅根據 429 人之資料加以分析研究。至於受試者之年級、段別、性別等統計如表一。

表一 取樣受試人數

年 級	一		二		合 計
	前 段	後 段	前 段	後 段	
男	37	76	42	67	222
女	39	63	43	62	207
男 + 女	76	139	85	129	429

二、研究材料：

本研究所使用的材料係取自於普通分類測驗中之語文理解與方塊計算兩類試題。普通分類測驗係由路君約、黃堅厚(民65)依據美國民間出版之陸軍普通分類測驗 (Army General Classification Test) 修訂而成，全測驗包括 語文理解、算術推理、方塊計算等三類試題，每類試題各有五十題，測驗材料採循環式排列，適合於國中一年級到高中三年級學生。本研究係屬探索性研究 (exploratory study)，故僅使用其中語文理解與方塊計算兩類試題。語文理解類題目主要測量受試者對兩個字彙之理解能力，如「繼續」的意義和什麼最相近？(1)連綿(2)間歇(3)始終(4)動靜。而方塊計算類的題目則在測量受試者某一種空間關係能力，此類題目也是選擇式試題，由受試者就每題的一堆方塊之透視畫去選答案方塊的個數。

三、研究程序：

本研究對象選出之後，以班級為單位實施上述測驗，為使學生對測驗題能作最大表現，筆者商請該校的導師或輔導老師擔任施測工作。施測時，主試者對受試者作指導說明，使其了解作答方法後，

正式做此兩類試題，測驗時間約為25分鐘，但為使學生能有充分時間作答，俟學生答完後再收卷。施測時筆者要求主試者注意觀察那些學生不認真作答，俾便於收回試卷後剔除之。受試者之答案係用2 B鉛筆塗寫在師大教心系印製之多種用途答案紙上，以便於電腦閱卷。

四、資料處理：

測驗之計分及資料統計工作係經由師大教育心理系光學閱讀機及小型電腦處理。根據全體受試之測驗結果，計算出每個試題的難度指數及鑑別力指數。難度指數包括P值（即從最高分部份向下取總人數的27%與從最低分部份向上取27%，答對人數之比率之平均數）和 Δ 值（美國教育測驗服務社創用之以平均數為13，標準差為4之等距量尺分數）。用分數最高的27%與最低的27%答對人數比率之差值當做鑑別力指數。其他的統計工作則由人工處理。

結 果

一、語文理解與方塊計算試題之難度指數

根據全部研究對象 429人在普通分類測驗語文理解與方塊計算兩類試題之結果，算得之各題難度指數如表二、表三。表中難度指數 Δ 值愈大表示題目愈難，反之，則愈簡單。又為便於讀者了解試題內容，表中使用之題號與原測驗題本之題號相同。

表二 語文理解五十題之難度指數

題 號	1	2	3	4	5	6	7	8	9	10
Δ 值	5.9	15.3	9.9	10.1	12.1	12.8	13.6	16.8	9.5	11.5
題 號	31	32	33	34	35	46	47	48	49	50
Δ 值	8.1	8.8	11.0	9.5	12.9	14.9	11.2	10.0	7.3	11.0
題 號	61	62	63	64	65	76	77	78	79	80
Δ 值	11.1	11.5	9.8	18.2	12.6	14.0	9.9	9.0	12.5	9.5
題 號	91	92	93	94	95	106	107	108	109	110
Δ 值	8.6	11.4	13.8	14.5	10.2	8.2	10.9	11.3	11.0	9.8
題 號	121	122	123	124	125	136	137	138	139	140
Δ 值	9.0	9.9	10.3	10.5	15.0	9.5	10.7	9.8	8.5	8.7

由表二資料顯示，普通分類五十個語文理解題目之難度指數（ Δ ）介於 5.9 至 18.2 之間，難度指數在 14.5 以上者計有 2、8、46、64、94、125 等六個題目。



表三 方塊計算五十題之難度指數

題 號	21	22	23	24	25	26	27	28	29	30
△ 值	6.8	6.5	7.8	7.3	8.2	11.8	7.9	8.0	8.0	8.6
題 號	41	42	43	44	45	56	57	58	59	60
△ 值	8.6	7.5	8.7	7.6	13.6	10.1	9.6	8.0	10.1	9.7
題 號	71	72	73	74	75	86	87	88	89	90
△ 值	10.9	10.8	11.4	9.9	10.5	9.4	10.9	11.1	10.9	12.1
題 號	101	102	103	104	105	116	117	118	119	120
△ 值	9.5	12.1	10.7	13.0	12.0	13.2	13.0	14.0	13.5	16
題 號	131	132	133	134	135	146	147	148	149	150
△ 值	14.8	14.0	10.5	14.5	12.0	13.6	12.2	13.3	13.1	14.8

由表三資料得知，方塊計算試題之難度指數 (△) 介於 6.5 至 16.0 之間。難度指數在 14.5 以上計有 120、131、134、150 等四個題目。

如果將兩類試題按其難度指數 (△) 由小到大之順序排列，結果如表四、表五：

表四 語文理解易難順序表

題 號	1	49	31	106	139	91	140	32	78	121
△ 值	5.9	7.3	8.1	8.2	8.5	8.6	8.7	8.8	9.0	9.0
題 號	9	34	80	136	110	63	138	3	77	122
△ 值	9.5	9.5	9.5	9.5	9.8	9.8	9.8	9.9	9.9	9.9
題 號	48	4	95	123	124	137	107	33	50	109
△ 值	10.0	10.1	10.2	10.3	10.5	10.7	10.9	11.0	11.0	11.0
題 號	61	47	108	92	10	62	5	79	65	6
△ 值	11.1	11.2	11.3	11.4	11.5	11.5	12.1	12.5	12.6	12.8
題 號	35	7	93	76	94	46	125	2	8	64
△ 值	12.9	13.6	13.8	14.0	14.5	14.9	15.0	15.3	16.8	18.2

表四資料指出，五十個語文理解題目之易難順序與原測驗之排列順序並不完全吻合，這或許是因本研究之受試者與原修訂者之預試樣本不同所致。但為本研究分析上的需要，後面的統計分析均以上述難度指數及其順序為依據。

表五 方塊計算五十題易難順序表

題 號	22	21	24	42	44	23	27	28	29	58
△ 值	6.5	6.8	7.4	7.5	7.6	7.8	7.9	8.0	8.0	8.0
題 號	25	30	41	43	86	101	57	60	74	56
△ 值	8.2	8.6	8.6	8.7	9.4	9.5	9.6	9.7	9.9	10.1
題 號	59	75	133	103	72	71	87	89	88	73
△ 值	10.1	10.5	10.5	10.7	10.8	10.9	10.9	10.9	11.1	11.4
題 號	26	105	135	90	102	147	104	117	149	116
△ 值	11.8	12.0	12.0	12.1	12.1	12.2	13.0	13.0	13.1	13.2
題 號	148	119	45	146	118	132	134	131	150	120
△ 值	13.3	13.5	13.6	13.6	14.0	14.0	14.5	14.8	14.8	16

表五指出，此五十個方塊計算試題之易難順序與原測驗之排列順序雖較語文推理部分相符合，但仍有部分不一致。如第 26、45 題竟位在易難順序表之後面，而 101、103、133 等題却呈現在易難順序表之前面，其可能的理由同前述語文理解題目者。

二、語文理解與方塊計算兩類試題之易難分組

由於這兩類試題各有五十題之多，若以五十題之難度指數作為試題難度水準，將有兩個缺點：第一，因為題數多，難度水準多，將造成受試特徵曲線不易閱讀之現象。第二，這兩類試題不若前述之西沙音調感覺測驗有造成試題難度之具體因素（音調之差別），相近之難度實難以表示其難度水準有別，這是筆者進行試題分組之理由。

按試題的難度指數，將語文理解、方塊計算兩類試題各分成四組。分類之基準如下：凡△值在 8.9 以下的試題列為第 I 組，在 9.0~10.9 歸為第 II 組，在 11.0~12.9 者列為第 III 組，在 13.0 以上則列入第 IV 組。至於題號，實際難度範圍，題數及平均難度等詳見表六。語文理解第 I、II、III、IV 組試題之平均難度分別為 8.01, 9.88, 11.70 和 15.12；方塊計算四組試題之平均難度依次為 7.83, 10.25, 11.83 和 13.89，上述各組試題之平均難度將作為以下繪製受試特徵曲線試題難度水準之依據。



表六 語文理解與方塊計算試題易難分組資料

組別	語文理解					方塊計算				
	題號	難度範圍	題數	平均難度(Δ)	題號	難度範圍	題數	平均難度(Δ)		
I	1 31 32 49	5.9 }	8	8.01	21 22 23 24 25 27	6.5 }	14	7.83		
	91 106 139 140	8.8			28 29 30 41 42 43 44 58				8.7	
II	3 4 9 34 48	9.0 }	19	9.88	56 57 59 60 71	9.4 }	14	10.25		
	63 77 78 80 95				72 74 75 86 87					
	107 110 121 122 123	10.9			89 101 103 133					
	124 136 137 138									
III	5 6 10 33 35	11 }	14	11.7	26 73 88 90 102	11.1 }	8	11.83		
	47 50 61 62 65	12.9			105 135 147	12.2				
	79 92 108 109									
IV	2 7 8 46 64	13.6 }	9	15.12	45 104 116 117 118	13.0 }	14	13.89		
	76 93 94 125	18.2			119 120 131 132 134 146 148 149 150				16.0	

三、普通分類測驗語文理解部分單一能力之研究：

繪製受試特徵曲線，了解試題所測量之能力是否存在，除了必須有試題難度水準外，尚須將受試按其分數高低分成若干組來表示受試者的特徵。測驗分數的高低可視為受試在此測驗所測量之能力的高低。

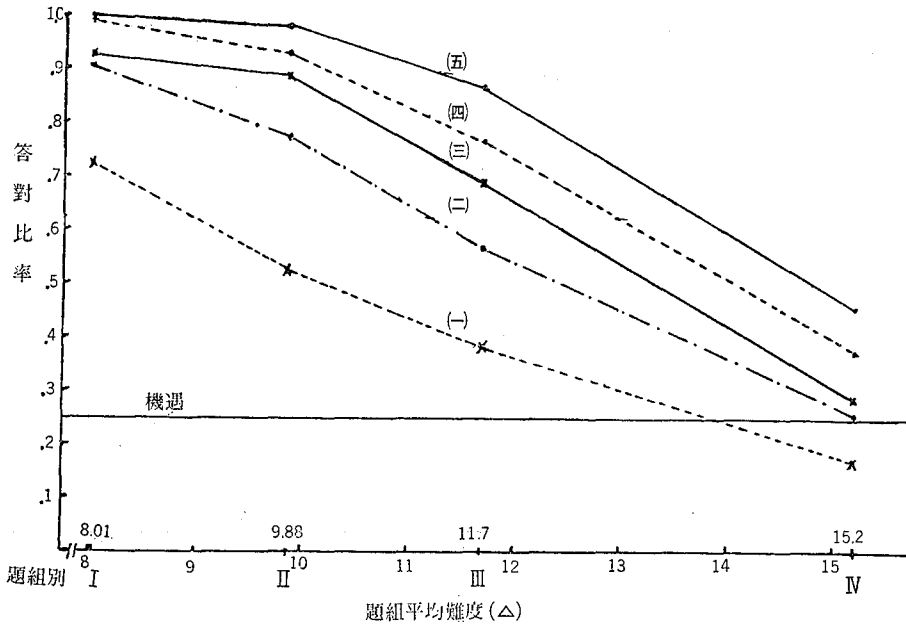
(一) 語文理解類全體受試高低能力分組及其受試特徵曲線

根據全體受試者 429人之語文理解分數將學生區分成五組。第一組是代表語文理解分數較低的一組，而第五組表示分數最高的一組。由於很多受試者之分數一樣，如等分幾組，將造成各組之能力無差別之現象，因此，本研究把相同分數之受試者歸入同一能力組，但遵照每組約佔百分之廿左右之受試為原則。詳細資料如表七：

表七 語文理解類高低能力分組資料 (N=429)

組別	分數範圍	人數	累積百分比
五	41 ~ 47	84	100
四	39 ~ 40	71	80.4
三	35 ~ 38	103	63.9
二	30 ~ 34	79	39.9
一	10 ~ 29	92	21.4

而各組受試特徵曲線如圖三：



圖三 全體受試之語文理解題組受試特徵曲線

由於筆者目前尚未找到造成語文理解試題困難之客觀標準，本研究暫使用美國教育服務社所創用之試題難度指數作為基線之量尺，縱軸則為答對比率（通過比率），由於語文理解試題之選項有四，就理論上說其猜對之機率为四分之一，因此在縱軸 0.25 處有一條與基線平行之機遇線，基線上之四個標點：8.01，9.88，11.70，15.20 分別是第 I、II、III、IV 題組之平均難度值。

由圖三可知，資料呈現極有系統的關係。在四個題組中，高分組答對比率始終高於低分組，最高能力的受試者在較簡單的題目（第 I 組、第 II 組試題）答對比率幾近於百分之百，稍難的第 III 組試題答對比率仍高達百分之八十七左右，到最難的一組題目才有困難；而最低能力的受試者在第 II 題組與第 III 題組試題開始感到困難，至最難之第 IV 組試題就更加感覺困難了，答對比率降至機遇之下。

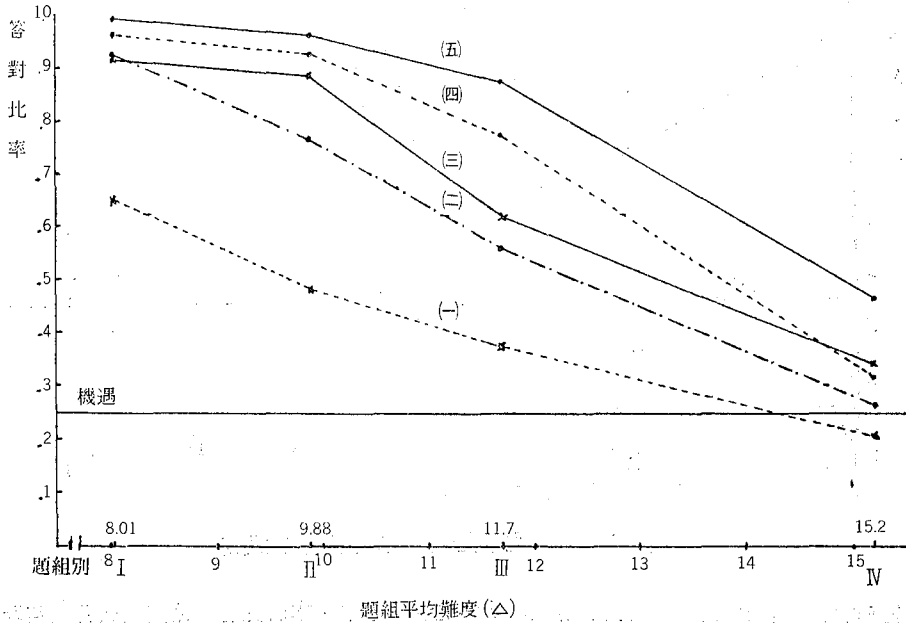
(二) 語文理解研究受試男生、女生高低能力分組及其特徵曲線

表八 語文理解男、女生能力分組資料

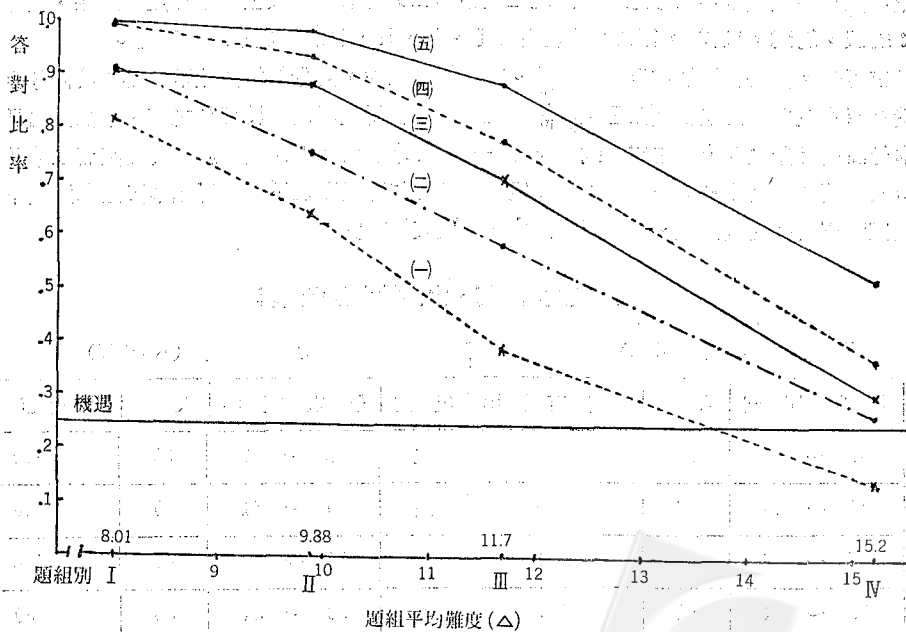
男 生 (N=222)				女 生 (N=207)			
組 別	分 數 範 圍	人 數	累 積 百 分 比	組 別	分 數 範 圍	人 數	累 積 百 分 比
五	41 ~ 46	38	100	五	42 ~ 47	30	100
四	38 ~ 40	55	82.9	四	39 ~ 41	49	85
三	35 ~ 37	40	58.1	三	35 ~ 38	46	61.8
二	30 ~ 34	41	40.1	二	30 ~ 34	38	39
一	10 ~ 29	48	21.6	一	17 ~ 29	44	21.3

依據男生 222 人，女生 207 人語文理解測驗上的分數，各分成五個能力不同之組別，各組分數範圍、累積百分比、人數等資料如表八。

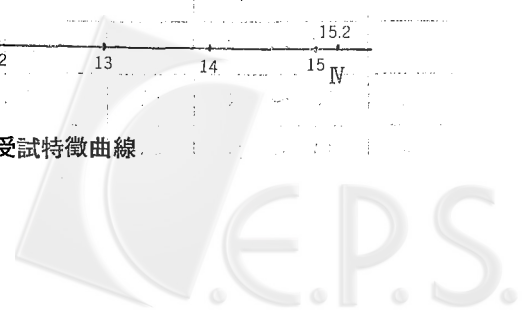
兩性高低能力分組之方式與全體相同，其各組分數範圍與表七全體受試分組之範圍極相近似。男性、女性受試者特徵曲線分別見圖四與圖五：



圖四 男生之語文理解題組受試特徵曲線



圖五 女生之語文理解題組受試特徵曲線

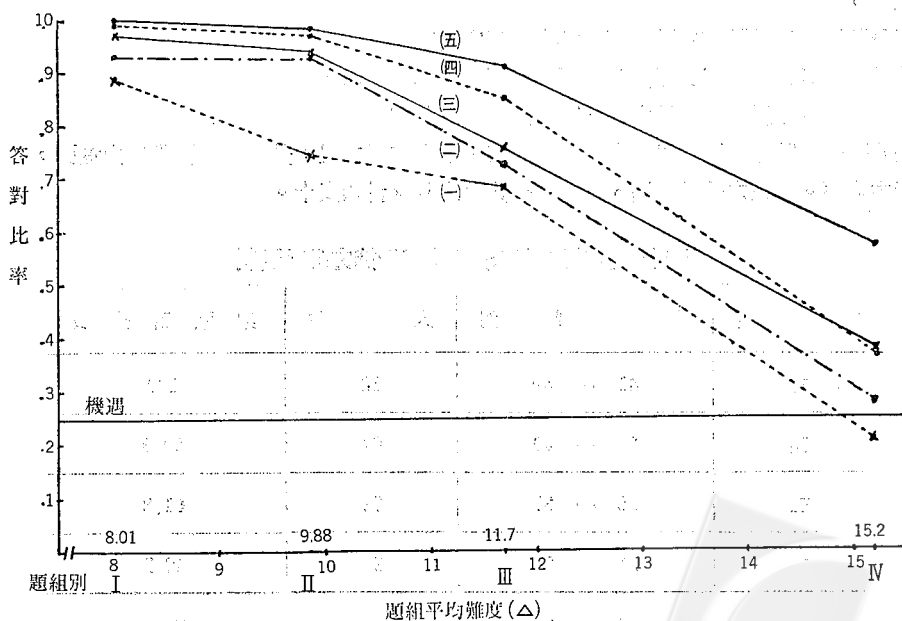


由圖四、圖五觀之，雖然使用限制的 (restricted) 受試，資料仍皆呈現相當的系統性，高分組的受試，大部分容易的題目都能答對，少數困難的題目 (如第IV類題組) 才會有困難。低分組的受試，即使很簡單的問題 (如第I類題組) 作答也感困難，對困難的問題就更少有機會通過了。男、女生能力最差的一組在第IV題組之答對比率都落在機遇線以下。進一步的比較第四、五兩圖不難發現：女生部分之受試特徵曲線斜率幾近相同，呈現有系統的排列，只有第二、第三能力組在第一試題組上答對比率有輕微的紊亂現象，男生組中第二與第三能力組在第一試題組上答對比率有顛倒現象，在第三、第四能力組在第IV題組亦有反常之現象，這些顛倒現象之發生有兩個共同點：(1)發生於相鄰二組，(2)答對比率相差很小；因此，這種現象似可忽略，換言之，系統性的趨勢仍然存在的。

(三) 語文理解前後段受試能力分組及其特徵曲線

表九 語文理解前後段受試能力分組資料

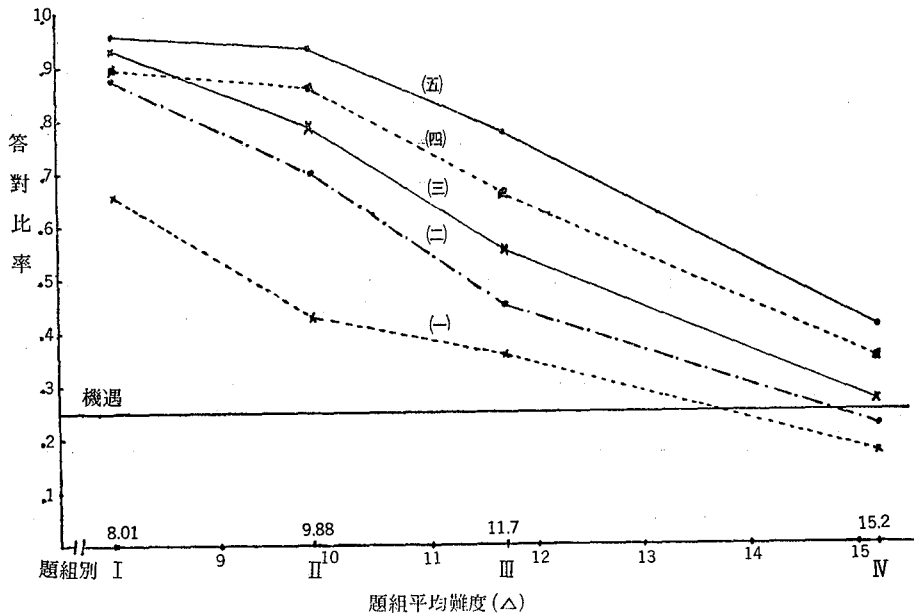
前 段 (N=161)				後 段 (N=268)			
組 別	分 數 範 圍	人 數	累 積 百 分 比	組 別	分 數 範 圍	人 數	累 積 百 分 比
五	43 ~ 47	28	100	五	38 ~ 44	54	100
四	41 ~ 42	43	82.6	四	35 ~ 37	56	79.9
三	39 ~ 40	41	55.9	三	31 ~ 34	56	59.0
二	37 ~ 38	25	30.4	二	27 ~ 30	51	38.1
一	28 ~ 36	24	14.9	一	10 ~ 26	51	19.0



圖六 前段班之語文理解題組受試特徵曲線



按照前後段生之語文理解上的分數，將其分成不同的五個能力組，各組分數範圍、累積百分比、人數分配等資料如表九。前後段生各組分數範圍相差較大，前後段受試特徵曲線如圖六、圖七。



圖七 後段班之語文理解題組受試特徵曲線

從圖六、圖七得知，大體上高分組的受試者在各題組上答對比率皆比低分組為高，除了最低分組以外，其他四組之斜率十分相近，前段班受試特徵曲線第三、第四能力組在第IV題組答對比率高低有顛倒現象；而圖七之第三能力組在第I題組上的答對比率亦高於第四能力組。不過，整個圖形仍呈相當的系統性。

四、普通分類測驗方塊計算類試題單一能力測量之研究

方塊計算類試題受試高低能力分組原則與語文理解類相同。

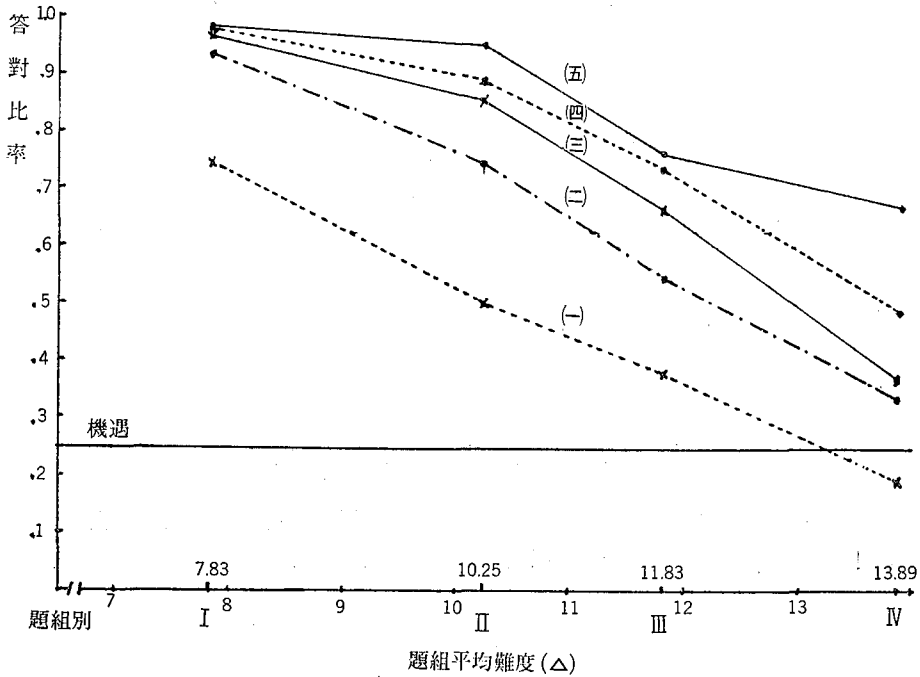
(一) 方塊計算全體受試高低能力分組及其受試特徵曲線

依據 429 位受試方塊計算試題之總分分成五個不同的能力組，分數愈高視為能力愈強，分數愈低則視為能力愈弱。各組之分數範圍、累積百分比及人數資料如表十。

表十 全體受試方塊計算高低能力分組資料

組別	分數範圍	人數	累積百分比
五	41 ~ 49	82	100
四	38 ~ 40	84	80.9
三	35 ~ 37	79	61.3
二	31 ~ 34	90	42.9
一	6 ~ 30	94	21.9

全體受試在方塊計算類試題之特徵曲線則如圖八，圖中顯示各能力組曲線之斜率除最高能力組稍有不同外，其他大致相同。各能力組受試之平均答對比率隨工作難度之增加而減少。較高能力組在各種工作難度水準之答對比率始終比低能力組高。能力最高一組之曲線，在前三種工作難度水準呈現逐漸下降之趨勢，這與其他四組極為一致。但在難度為13.89之工作難度時，其曲線走向與其他四組稍有不同，比預期答對之比率稍微高些。大體來說，受試答對機率與系列難度工作間仍呈現相當單調的(monotonic)密切關係。



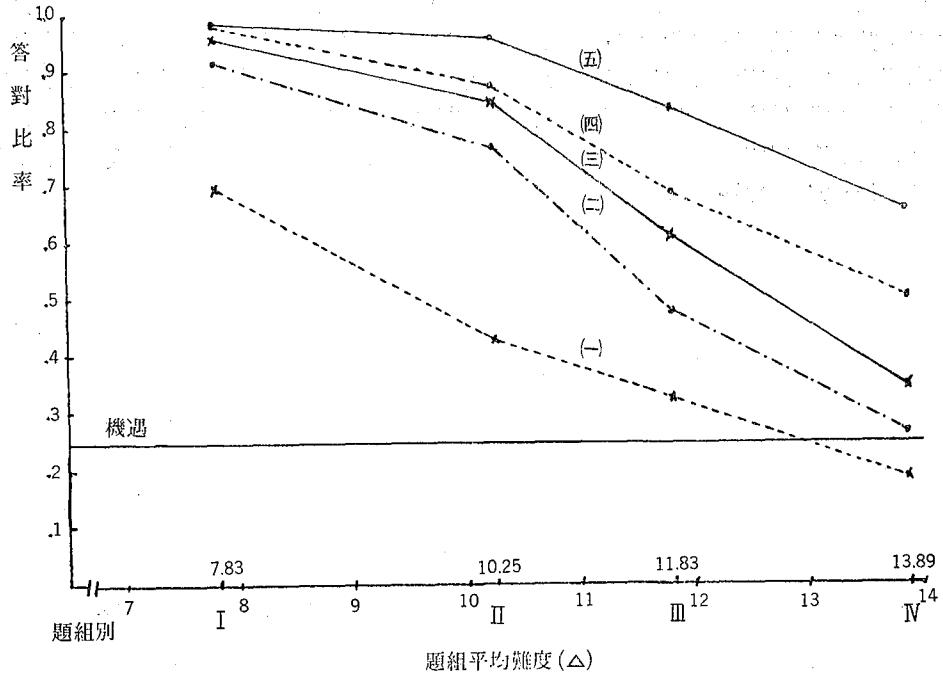
圖八 全體受試之方塊計算題組受試特徵曲線

(二) 方塊計算男、女受試高低能力分組及其特徵曲線

男、女受試在方塊計算類試題之能力分組詳細資料如表十一，資料顯示除第一組分數範圍差異較大外，其他各組之分數範圍，男、女受試十分相近。男、女受試在這類試題之特徵曲線如圖九、圖十。

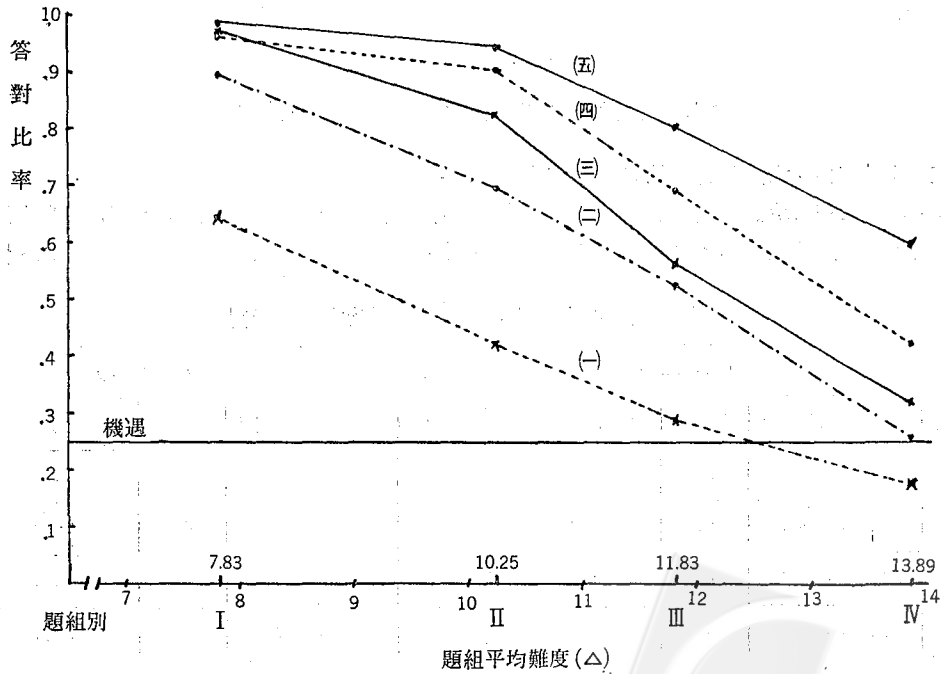
表十一 方塊計算男、女受試高低能力分組資料

男 生 (N=222)				女 生 (N=207)			
組 別	分 數 範 圍	人 數	累 積 百 分 比	組 別	分 數 範 圍	人 數	累 積 百 分 比
五	41 ~ 47	39	100	五	41 ~ 49	45	100
四	37 ~ 40	50	82.4	四	38 ~ 40	44	79.2
三	34 ~ 36	48	59.9	三	35 ~ 37	38	58.0
二	29 ~ 33	42	38.3	二	31 ~ 34	45	39.6
一	6 ~ 28	43	19.4	一	13 ~ 30	37	17.9



圖九 男生之方塊計算題組受試特徵曲線

方塊計算類試題，男、女受試特徵曲線皆呈現明顯系統之關係。兩性特徵曲線除了第四能力組以外，其他三個能力組之曲線走勢幾乎一致。高能力組在簡單的題目之答對比率幾近於百分之百，即使



圖十 女生之方塊計算題組受試特徵曲線



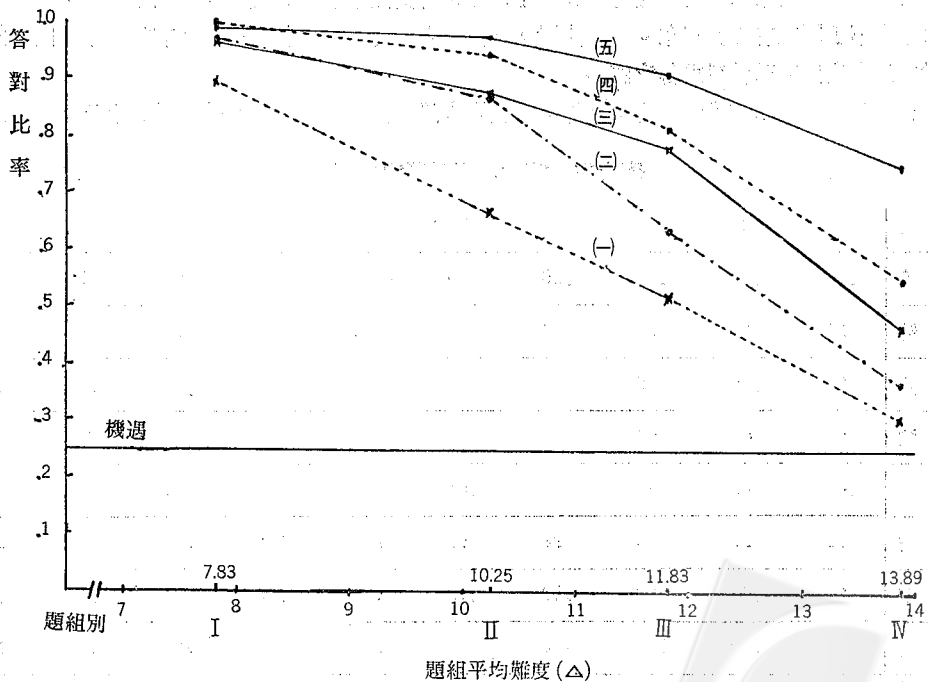
是難的題目組平均答對比率在 .70 左右。能力差的組就是極簡單的題目答對比率也不高，到了第 II 或第 III 難度題組就更感困難，而到第四題組時更是困難重重了。另外，女生第一能力組在各個試題難度水準之答對比率皆顯著高於男生同一能力組，其他三個能力組則無太大之差別。

(三) 方塊計算前後段班高低能力分組及其特徵曲線

表十二 前後段受試方塊計算能力分組資料

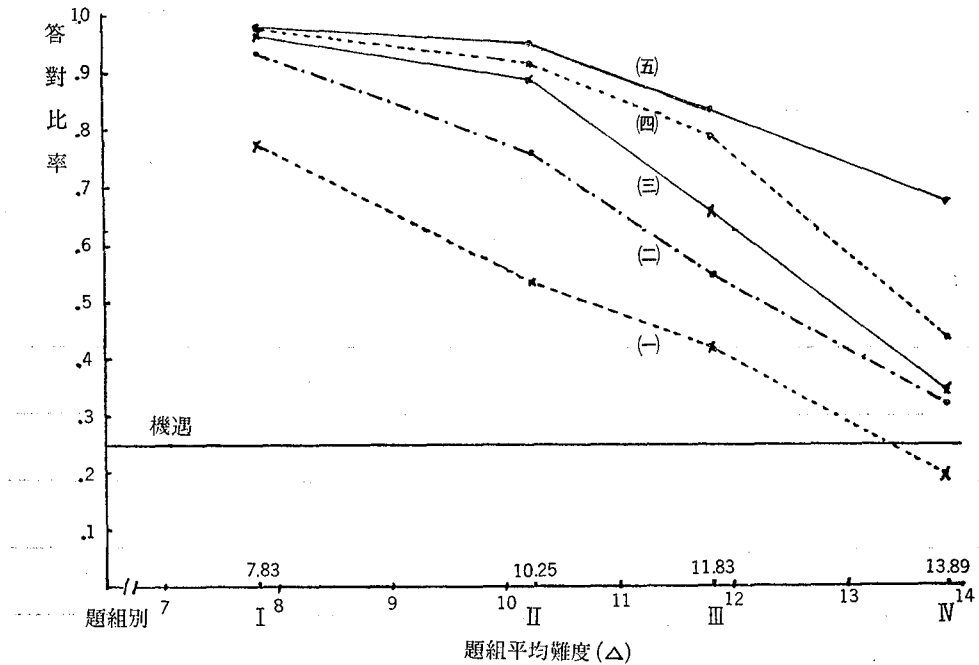
前 段 (N=161)				後 段 (N=268)			
組 別	分 數 範 圍	人 數	累 積 百 分 比	組 別	分 數 範 圍	人 數	累 積 百 分 比
五	44 ~ 49	26	100	五	40 ~ 45	39	100
四	40 ~ 43	37	83.9	四	36 ~ 39	60	85.4
三	38 ~ 39	31	60.9	三	32 ~ 35	49	63.1
二	35 ~ 37	38	41.6	二	27 ~ 31	71	44.8
一	13 ~ 34	29	18.0	一	6 ~ 26	49	18.3

根據前後段班方塊計算類試題總分分成不同的五個能力組，詳細資料如表十二。由表中得知前段受試各組之分數範圍均顯著大於後段對應組之受試分數範圍。其特徵曲線如圖十一、圖十二。



圖十一 前段受試之方塊計算題組受試特徵曲線





圖十二 後段受試之方塊計算題組受試特徵曲線

從圖十一、十二可以看出，前段各能力組受試在各個題組難度水準之答對比率顯著高於後段組，前後段各能力組受試，除了第一能力組外，其他四組在較易的題組答對比率相差甚微。在圖十一中，第四與第五能力組及第二、第三能力組在第 I 題組之答對比率有交錯現象。在圖十二中，第三、第四能力組亦在第 I 題組有顛倒現象。整體來說，前後段各能力組受試特徵曲線具有相當系統關係的。

五、試題鑑別度與受試特徵曲線之關係

(一) 普通分類測驗語文推理與方塊計算兩類試題之鑑別力指數 (D 值) 分別見表十三、表十四。

表十三 語文理解試題之鑑別指數

題 號	1	2	3	4	5	6	7	8	9	10
D 值	.04	.24	.33	.36	.10	-.11	.14	.07	.18	.21
題 號	31	32	33	34	35	46	47	48	49	50
D 值	.21	.16	.55	.30	.50	.26	.56	.45	.10	.59
題 號	61	62	63	64	65	76	77	78	79	80
D 值	.41	.43	.35	.09	.57	.45	.36	.28	.53	.36
題 號	91	92	93	94	95	106	107	108	109	110
D 值	.22	.57	.38	.29	.42	.23	.57	.50	.49	.28
題 號	121	122	123	124	125	136	137	138	139	140
D 值	.27	.39	.38	.43	.32	.38	.47	.39	.26	.28

從表十三知道，語文理解試題鑑別力指數呈負值的僅有第六題一題。根據 Ebel (1979) 試題鑑別力評鑑標準，鑑別指數在 .30 以上者可視為優良題目。在這種標準下，語文理解五十個試題中，屬優良题目的共有29題。由表十四知，在五十個方塊計算試題中，有28個試題屬優良試題。

表十四 方塊計算試題之鑑別指數

題 號	21	22	23	24	25	26	27	28	29	30
D 值	.09	.10	.16	.14	.13	.07	.20	.16	.17	.25
題 號	41	42	43	44	45	56	57	58	59	60
D 值	.27	.12	.27	.18	.49	.30	.33	.22	.34	.34
題 號	71	72	73	74	75	86	87	88	89	90
D 值	.45	.36	.40	.38	.48	.26	.50	.49	.48	.42
題 號	101	102	103	104	105	116	117	118	119	120
D 值	.22	.58	.49	.50	.53	.26	.43	.31	.56	.19
題 號	131	132	133	134	135	146	147	148	149	150
D 值	.26	.45	.47	.40	.27	.30	.48	.48	.39	.23

(二) 語文理解與方塊計算鑑別力指數達 .30 以上之試題易難分組

從語文理解與方塊計算兩類試題中選出鑑別力指數 .30 以上之優良試題，再依其難度指數大小分

表十五 語文理解與方塊計算鑑別指數 .30 以上之試題易難分組資料

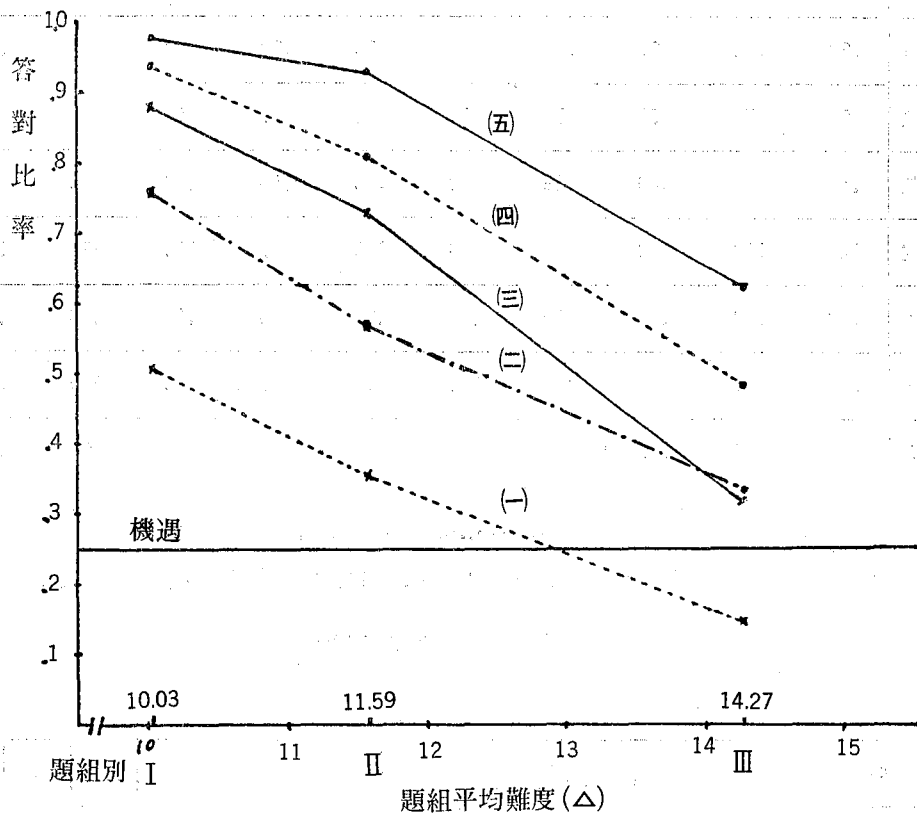
組 別	語 文 理 解							方 塊 計 算								
	題 號					難度範圍	題數	平均難度(度△)	題 號					難度範圍	題數	平均難度(△)
I	3	4	34	48	63	9.0 } 10.9	15	10.03	56	57	59	60	71	9.0 } 10.9	12	10.38
	77	80	95	107	122				72	74	75	87	89			
	123	124	136	137	138				103	133						
II	33	35	47	50	61	11.0 } 12.9	11	11.59	73	88	90	102	11.0 } 12.9	6	11.82	
	62	65	79	92	108				105	147						
	109															
III	76	93	125			13.6 } 18.2	3	14.27	45	104	117	118	119	13.6 } 18.2	10	13.56
									132	134	146	148	149			

組。在汰除品質不良的試題過程中，已經把原歸為第 I 題組之題目剔除（鑑別力均未達 .30，見表四、五、六），故在此分組之基準調整如下：凡 Δ 值在 .90~10.9者歸入第 I 組，在 11.0~12.9者列入第 II 組，13.0以上者歸入第 III 組，詳細分組資料見表十五。語文理解第 I、II、III 題組之平均難度分別為 10.03、11.59 和 14.27；方塊計算三組試題之平均難度依次為 10.38、11.82 與 13.56。上述平均難度將在繪製受試特徵曲線時作為試題組難度水準之依據。

表十五與表六比較發現，表十五第 I、II、III 組之平均難度指數與表六之 I、II、III 組之平均難度指數相近。

(三) 語文理解與方塊計算鑑別力指數 .30 以上題組受試特徵曲線

兩類試題經保留優良題目之後，其受試特徵曲線如圖十三與圖十四。



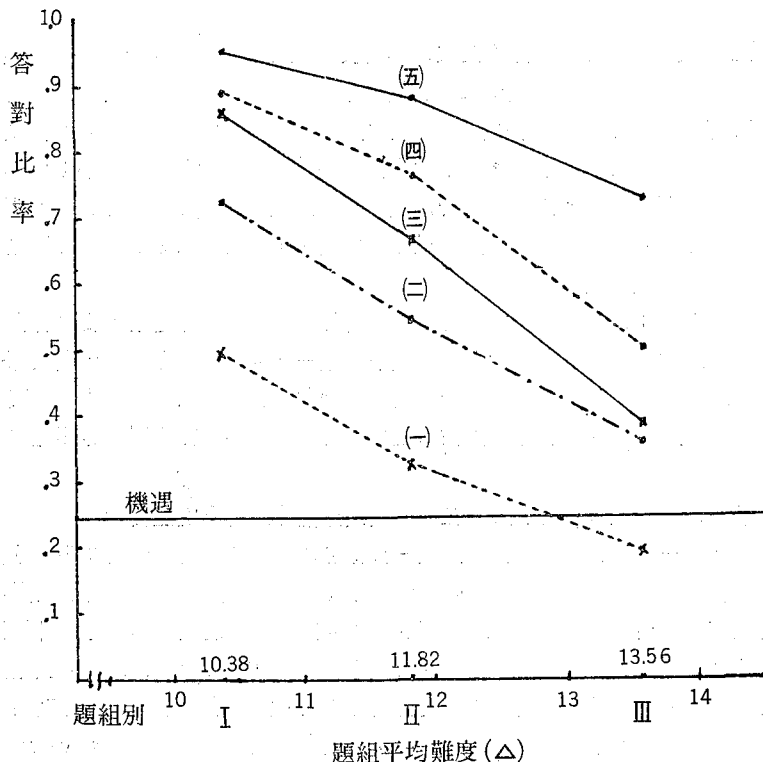
圖十三 全體受試之語文理解鑑別力指數 .30 以上題組受試特徵曲線

雖然語文理解試題第二能力組在第 III 題組比第三能力組答對比率要高些，但整體而言，語文理解與方塊計算兩類試題之受試特徵與試題難度水準呈現相當有系統的關係，若將圖十三與圖三之後三組，圖十四與圖八之後三組相比較，可發現函數關係極為一致。

六、方塊計算試題難度水準之因素

分析造成試題困難之因素，對能力界說、測驗是否測量單一能力、試題編排與實施等皆極為重要。將 429 位受試方塊計算試題之難度指數依序排列。由於試題過多，故僅選出逢 5 的倍數順序之題目共 10 題進行試題之邏輯分析，題號及難度指數如表十六。

分析表中 10 個試題內容，發現前兩個極易的題目係屬對稱題型，而其他 8 個皆屬不對稱的題型。



圖十四 全體受試之方塊計算鑑別力指數 .30 以上題組受試特徵曲線

不對稱題型中的最難三個試題共同的特徵是題目圖形不夠清楚；與其他幾個不對稱圖形相較，方塊之形狀也較小。

表十六 難度因素分析之方塊計算試題題號與難度指數*

題 號	44	58	86	56	72	73	102	116	118	120
Δ 值	7.6	8.0	9.4	10.1	10.8	11.4	12.1	13.2	14.0	16.0

*由於測驗內容係屬機密，故不附表中10個題目之內容，若必須參考時，請向師大教心系洽借。

討 論

一、受試特徵曲線在能力測量運用之可能性

能力測驗的效度與其施測的策略是目前測驗學者們重視的兩個問題。測驗編製者所宣稱擬測量之能力與能力測驗的構念效度有密切的關係。若按照前面Walker對單一能力的說法及受試特徵曲線之性質，受試特徵曲線應適用於勘定能力之存在與否；只要對某一受試，或某一能力相近之受試，正確作業之機率和一系列難度不同之工作間呈現一種有系統的、單調的密切關係，而且高能力組較低能力組在這些不同工作難度上之答對比率為高時，這個工作所欲測量能力就可獲得證實 (Mosier, 1941)。本研究中，是使用被學者們(如 Carroll 1982, 1985)認為屬單一能力的語文理解與方塊計算試題為材料，使用受試特徵曲線加以驗證，結果證實了兩類測驗題係測量單一能力的。另外，受試特徵曲線不但用在未限制的全體受試具有運用的價值，就是有限制的受試(男女或前後段受試)亦可

使用。

從上面結果各個受試特徵曲線知道：語文理解所要測量的字彙知識；方塊計算所欲測量的空間視覺能力的確是存在的；因為從圖三到圖十四皆呈現常態肩形系統關係。各個特徵曲線也有極相似的斜率，此斜率應可各別視為字彙知識或空間視覺能力之特徵（Carroll, 1985）。從受試特徵曲線斜率之高低能說明這個測驗信度受限制情形，斜率愈大表示測驗的信度受到限制愈小；斜率愈低則表示這個能力測驗之信度受到限制愈大。按照傳統的測驗理論，適度增加良好品質之試題，測驗信度將會提高，測驗編製者若能了解某類能力測驗，其特徵曲線之斜率，將有助於了解提高信度之限制。

二、編製測驗應重視品質優良的試題

從圖十三、十四知道：語文理解與方塊計算兩類淘汰約一半的試題之後，受試特徵曲線與原來未汰除鑑別力指數在 .30 以下題目之受試特徵曲線幾乎完全一致。這說明了少數優良試題一樣能證實欲測量的能力之存在。同時，從汰除不良試題後，受試特徵曲線斜率看來，仍可表現出與未淘汰前能力特徵及信度受限的特性。因此，置放鑑別力指數低的題目於測驗中，非但對能力測量無助益更耗費人力與物力。

三、試題邏輯分析，試題解決過程的了解，在編製與改進能力測驗上之重要性

測驗很少是純粹某一心理特質之測量，因此一教育工作者如欲適當解釋受試測驗分數，則必須先找出影響作業之因素（Cronbach, 1984）。Sternberg (1981) 也認為試題解決過程之清晰的了解有助於能力特質與其評估（assessment）新的領悟。同樣地，語文理解或方塊計算測驗有構念效度則須先證實語文理解或方塊計算概念具有科學或概念的意義（Green, 1981）。試題的邏輯分析，試題解決過程的訊息皆有助於了解造成試題難度差異的因素。能力也應該從這個角度來界說，如果能力的界定是根據造成試題難易之因素，則這種能力測驗之構念效度將獲得保障。Vernon (1962) 曾對閱讀理解測驗作過邏輯分析，就發現除了閱讀理解之外，還有許多解釋受試測驗結果之因素。就以方塊計算類試題為例，如果仔細分析這類試題內容，會發現這類試題表面上有兩個影響試題難度差異之因素——看不見的方塊個數之多少和方塊的排列是否對稱。看不見的方塊個數愈多試題愈難，但這個因素與對稱因素有交互的作用（Carroll, 1985），換言之，對稱與不能立見的方塊數是造成方塊計算試題之難度水準的因素。但根據筆者對10個各種難度题目的分析結果，對稱性和不能立見之方塊的個數並不是決定題目困難的因素，因為大部分對稱性題目儘管不能立見的方塊數很多，其難度仍不高，題目之清晰性與方塊之排列方式才是最重要的。方塊計算試題理論上應該是測量空間視覺的能力（Spatial visualization ability）而這種能力與空間關係（Spatial relations）不同，前者屬於較為複雜，且難度較高的空間能力，後者屬於較簡單、偏於速度性的能力（Pellegrino, 1982）。由試題分析結果，提醒吾人將來編製方塊計算試題以測量空間視覺能力時，應盡可能將題目印刷清楚，以免影響受試者視不能立見方塊的能力。

至於語文理解試題型態極為單純，試題難度因素不易從試題分析或訪問受試者來了解。就筆者所知，國內尚無對字彙使用次數之研究報告，因此無法像方塊計算試題容易找出試題難度之具體因素。在日常生活中或課本中次數出現愈多的字彙，受試較易了解它的意思，因而能找出其意思相近之字彙。字彙在日常生活中出現的頻率之多寡，以為此類試題難度之一個向度當無問題。是否有其他之參考準據尚待學者們的研究。

四、試題難度分析與受試特徵曲線有助於能力測驗之實施

如何使用最短時間、最少題目找出學生的能力是傳統測驗學者與項目反應理論學者同感興趣的問題。透過試題內容之邏輯分析或是經由與受試者晤談而了解受試對試題之解題過程皆有可能找出造成試題難度差別的因素。有些試題或工作較易找出客觀的因素，有些較為複雜的工作則不易分析。不易分析的工作本身即可能潛藏著測量不同能力的線索，同時也不易對所要測量之能力給予明確的名稱，

測驗的編製者應注意這個問題。

根據一系列之難易題目所繪製之受試特徵曲線的另一個運用是：對於今後能力測驗之實施方式提供了訊息，如果能找到試題難易的客觀標準，將來可以從那一種難度水準的題目推測受試可能正確作答之機率，進一步運用在測驗之實施時將可節省人力物力。

結論與檢討

利用各種方法進行試題分析，以找出造成試題難易之客觀標準，是能力界說所不可缺的。同時也是鑑定試題是否測量單一能力之受試特徵曲線基線之根據。單一能力之證實又有助於測驗分數正確解釋，因此一個良好的能力測驗應該考慮進行試題分析，驗證試題是否測量單一能力。本研究以被學者們認為較單純的語文理解與方塊計算兩類試題進行試題分析與使用受試特徵曲線驗證其是否測量單一能力。結果發現此兩類試題的確測量單一能力；而且單一能力之驗證不因受試段別及性別的影響。同時也發現汰除鑑別力指數在 .30 以下之題目，並不影響到受試特徵曲線之斜率，也就是不影響到測驗所要測量的能力特徵，這也說明了汰除不良題目，不但不會影響到能力的測量，同時還可節省人力與物力。

由於本研究係屬初步探究，選取之材料是以學者們認為較可能測量單一能力的兩類試題，試題分析也是以較具客觀標準之方塊計算為例，以說明受試者特徵曲線在能力測量方面之用途。今後如能借助認知心理學的知識從事試題或工作分析，找出受試作答之工作差異特質，以作為能力界說的基礎，則對個人能力能作更正確而合理的測量。

參考書目

- 黃堅厚、路君約 (民68) 普通分類測驗實施手冊。見師大教育心理系 心理測驗實施手冊 (第一册)
- 路君約、黃堅厚 (民65) 普通分類測驗。師大教育心理系印製
- Aiken, L. R. (1985). *Psychological testing and measurement* (5th ed.). Boston: Allyn and Bacon.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, Ca: Brooks/Cole.
- Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.
- Bejar, I. I. (1985). Speculations on the future of the test design. In S. E. Embretson (Ed.), *Test design* (pp. 279-294). Orlando: Academic press.
- Carroll, J. B. (1983). The difficulty of a test and its factor composition revisited. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement* (pp. 257-282). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Carroll, J. B. (1985, April). *New perspectives in the analysis of abilities*. Paper presented at the Buros-Nebraska symposium on Measurement and Testing, University of Nebraska, Lincon, Nebr.
- Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton Mifflin.
- Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.
- Davies, J. B. (1975). *The psychology of music*. Ca: Stanford University Press.

- Ebel, R. L. (1979). *Essentials of educational measurement* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall, 1979.
- Egan, D. E. (1979). Testing based on understanding: Implication from studies & spatial ability. *Intelligence*, **3**, 1-15.
- Green, B. F. (1981). A primer of Testing. *American Psychologist*, **36**, 1001-1011.
- Lohman, D. F. (1979). *Spatial ability: Individual differences in speed and level*. Stanford, Ca: Aptitude Research Project, School of Education, Stanford University.
- Mosier, C. I. (1941). Psychophysics and mental test theory, II, The constant process. *Psychological Review*, **48**, 235-249.
- Pellegrino J. W. (1982). Process analysis of spatial aptitude. In R. J. Sternberg (Ed.), *Advances on the psychology of human intelligence* (vol 1). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sternberg, R. S. (1981). Testing and cognitive psychology. *American Psychologist*, **36**, 1181-1189.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Vernon, P. E. (1962). The determinants of reading comprehension. *Educational and Psychological Measurement*, **22**, 269-286.
- Vernon, P. E. (1965). Ability factors and environmental influences. *American Psychologist*, **20**, 723-733.
- Walker, D. A. (1931, 1936, 1940). Answer pattern and score scatter in tests and examinations. *British Journal of Psychology*, **22**, 73-86; **26**, 301-308; **30**, 248-260.
- Wilson, D., Wood, R., & Giddons, R. D. (1984). *TESTFACT* [computer program]. Mooresville, IN: Scientific Software.



Bulletin of Educational Psychology, 1987, 20, 107—130
Taiwan Normal University, Taipei, Taiwan, China.

ASTUDY ON THE APPLICATION OF THE PERSON CHARACTERISTIC CURVE TO ABILITY MEASUREMENTS

DER-HSIN FAN

ABSTRACT

The present study was conducted to investigate the capacity of the application of the item content logic analysis and the person characteristic curve to ability measurements. A total of 429 seventh and eighth graders served as examinees. The verbal comprehension test and the block counting test, two subtests of the Army General Classification Test in the Chinese revision, were administered to all subjects.

The major findings are as follows:

- (1) The degree of difficulty in the block counting test item depends upon the way in which the figures of blocks are arranged or printed.
- (2) The person characteristic curve gives information whether a test measures a single ability.
- (3) The restricted group has no influences on the application of the person characteristic curve.
- (4) A few items with high indices of discrimination (larger than .30) still maintain the characteristics of a test.

The author suggests that test users and writers should:

- (1) Adopt a variety of methods to find those factors that cause difficulty in test items with a view to defining exactly the ability which a test is claimed to measure.
- (2) Make use of the person characteristic curve to determine whether a test measures a single ability for the sake of construct validity.
- (3) Employ the person characteristic curve to further improve the administration of a test.

