國立中山大學 應用數學研究所

碩士論文

高級中學教師自編測驗評量方式之探討

研究生：江仲翔撰

指導教授：羅夢娜博士

中華民國 九十二 年 六月

中央研究院 eThesys
學位論文

摘　　要

本研究在探討如何編製一份高級中學數學科測驗試題，及如何從測驗結果，了解學生在各項概念中所可能產生之解題錯誤與原因。本研究以高中三年級數學甲第一章 “機率與統計(二)” 作為測驗評量試題之範圍。以高雄市兩所高級中學，共七個班級 249 位高三學生作為本次研究的測驗對象。

我們針對學生進行重測複本測驗，對樣本所得資料進行題目之試題分析、複本信度與效度分析，並建立各組測驗之常模。對於學生測驗成績進行學生問題表分析、迴歸分析等統計分析。分析結果，利用重測複本方式運算未檢討教學班級之信度值為 0.6，利用 Z 分數轉換所得信度值為 0.54，各校標效度值介於 0.29 到 0.6 之間，表面效度為 0.83。整份試題對於此群體之學生屬於偏難之試題。就試題部分分析，約有 25%試題難度低於 0.3(難度愈低則試題愈難)，87.5%鑑別度適當，試題為優良與尚可之試題。就學生部分分析，10%學生學習穩定與良好，約有 32%之學生學習力不足，學習不夠充分。38%學習尚稱穩定但仍需用功。利用回歸模型檢定，發現性別與分數並無顯著的相關；而複習與未複習之學生，則具有相當顯著的成績差異。

試卷成績分析表示此份試卷屬於還算優良之試題卷。教師可利用本研究所提之方法，在自編相關測驗時了解其所編製測驗之試題好壞，進而加以修改，並可建立各教師專屬之題庫。並可利用本研究所提之方法，讓教師能了解各學生之作答狀況與觀念不清之所在，進而對症下藥，提早補救，期能藉此提高學生之學習動機增進學習的信心與興趣。

目　　錄

摘　要…… 1

目　錄…… 2

表　次…… 4

圖　次…… 5

附錄次…… 6

第一章　緒論…… 7

第一節　研究動機…… 7

第二節　研究目的…… 8

第三節　測驗評量的發展史…… 9

第二章　理論與相關文獻探討…… 11

第一節　試題分析理論…… 11

第二節　信度、效度分析理論…… 15

第三節　學生問題表分析理論…… 23

第四節　數學能力與性別差異…… 28

第三章　測驗設計與實施…… 29

第一節　測驗對象…… 29

第二節　測驗方法與編制過程…… 31

第三節　研究之限制…… 34

第四章　結果與討論…… 35
第一節　測驗題項…… 35
第二節　信度、校度與常模之研究…… 37
第三節　學生問題表分析…… 41
第四節　各類統計檢定…… 43
第五章　結論與建議…… 48
第一節　研究結論…… 48
第二節　研究建議…… 50
參考文獻…… 52
附錄…… 55

表　　次

表 2-1　鑑別度評鑑標準表…… 15

表 2-2　$D_B(M)$ 查表……25

表 3-1　研究樣本分配表…… 29

表 3-2　複本信度樣本分配表…… 30

表 3-3　效度樣本明細表…… 30

表 3-4　雙向細目表…… 32

表 4-1　試題題數分配表…… 35

表 4-2　試題項目分析表…… 36

表 4-3　重測複本信度與測量標準誤…… 38

表 4-4　本測驗之內部一致性信度…… 39

表 4-5　本測驗之效標關聯效度…… 40

表 4-6　各組與全體之平均數與標準差…… 40

表 4-7　各組差異係數…… 41

表 4-8　前測與後側平均數差異之顯著性檢定…… 44

表 4-9　各組平均數差異之顯著性檢定…… 45

表 4-10　逐步迴歸變異數分析表…… 45

表 4-11　模型係數顯著性定…… 46

圖　　次

圖 4-1　後測試題屬性分佈圖…… 42

圖 4-2　試題屬性診斷分佈圖…… 42

圖 4-3　後側學生屬性分佈圖…… 43

圖 4-4　學生屬性診斷分佈圖…… 43

圖 4-5　兩校前後測驗成績分佈圖…… 47

圖 4-6　全體前後測驗成績分佈圖…… 47

附　錄　次

附錄一　*S-P* 表製作方式…… 55

附錄二　試題項目分析…… 57

附錄三　常態分配 Q-Q 圖…… 61

附錄四　注意係數分析表…… 64

附錄五　S-P 表分析程式碼…… 71

附錄六　試題編製流程圖……... 74

第一章　緒論

第一節　　研究動機

教育的目的在於培養學生健全的人格，對於教師角色而言，無論受教對象的年齡層是由小學的小朋友，到研究所博士班的成年人，教師的任務都離不開「教學(teaching)」與「評量(assessment)」這兩件事。任何的教學計劃常需依照學生的能力、性向等資料來加以設計。而每位教師在進行教學時，皆有屬於自己的教學方式，如要去比較哪種方式為優良之教學方式，在無法明確量化的討論上常會因人而異，屬於很主觀之觀點。

提供學生適當有效的教學，常有賴於教師對學生的了解。而利用評量方法，通常是了解學生狀況的最主要方法。有完備的評量方式，教師才能有較明確的教學策略與適時的改變自身的教學方法，畢竟，不同類型的學生都應有不同的教學方法來教導，教師應自許勿只使用一種方式在進行教學，如教師都是使用相同的方式教學，這樣不管是學生或者是教師本身都是無法進步的。

要評量學生的各項能力，常常需使用各種的評量工具，當接觸的情境有所不同時，評量的方式與工具也就會因此不同。雖然現今提倡許多多元評量方式(例如：實作評量(Linn, 1994)、檔案評量(LeMahieu, et al., 1995))，但在大多數為了要求客觀、公平的測驗評量上，還是常常會使用可以量化的評量方式來進行。有鑑於此，我們希望能利用這些量化數值，來加以討論教師在製作測驗評量時，應多加注意的地方與問題。

根據資料(教育資料文摘，民 87)調查，國中、小學生最不喜歡的科目中，以數學佔全部科目的第一位，且總是居高不下。而教師們也常因此需面對許多普遍缺乏興趣上數學課的學生，再加上學生之間的數學程度常具有很大個別差異，使得教師在教學上更是倍感困擾。

有感於現今的教育狀況，我們認為不只在加強自身的教學技巧之外，仍要利用一些客觀方法去了解學生的能力、興趣和是否有進步的情形。而要去蒐集到學生個別資料的方式有非常多種，例如：晤談法、觀察法、同儕互評等。而這些獲得資料的方法，有許多都是由觀察者個別判斷的，有時會過於主觀且常常會花費許多時間與精神。而利用「測驗法」的好處，為能在最短時間就能獲得大量的資料，且屬於很客觀之評量方式，不容易受特定的主觀事物所影響，因此，測驗法往往是教師必須在短時間內蒐集到學生能力與學習狀況的最便捷方法(Brueckner ＆ Band,1955)。

而編製一份測驗評量試題，對於教師來說應該都已是駕輕就熟之

事，但往往教師不了解所編的試卷是否真的適合於這一群的學生，常只能依照以往的教學經驗，年復一年的使用這些未經過任何分析的評量試卷。本研究中為了探討教學的過程中，能了解教師所編製或所使用的測驗試題是否優良，進而研究探討這些試卷中所隱藏的測驗理論與分析出來的數據意義。利用這些量化的數據來提供教師了解那些試題與試卷為適合學生的試卷，那些試題是需要再加以修改。並能利用這些資料所呈現的現象來了解整體學生的學習狀況是否有困難，或者教師本身的教學方式是否不適合這類型之學生。並且可利用一些評量分析的方式來發現各個學生的問題所在，與觀念不清的地方，進而加以對症下藥。有此可知，對於每一位教師，如果皆能了解評量分析的方法與其所代表之意義，絕對有助於各教師的教學進行，並且能因此幫助到更多的學生，使他們產生興趣與學習動機，進而提高學生的自信。

第二節　研究目的與方法

針對前節所述，可發現學生之學習成效，常可由評量中發現，教師如果能從這些評量的結果中，發現學生們的問題所在，並適時的提出補救方法，那必可以幫助學生更進一步的學習，為了使教師能夠學習如何分析這些試卷評量與了解分析後所帶來的意義，本研究針對了下列之目的加以說明：

一、利用試題分析理論判別教師自編測驗之試題是否屬於優良之試題。

(一)針對各試題題目進行微觀分析(micro-analysis)：

分析各試題之難度(difficulty)、鑑別度(discrimination)與誘答力選項(distraction) (Haladyna, 1994)是否屬於適當之範圍。

(二)對整份試卷進行鉅觀分析(macro-analysis) (郭生玉 民 85)：

1.信度(reliability)研究：

以重測複本法、折半法(split-half correlation)、庫李法(Kuder-Richardson method)分析測驗的穩定性與一致性，並利用組別 T 分數轉換檢驗複本法的相關信度值。

2.效度(validity)研究：

從兩方面進行測驗之效度研究：

(1)內容效度(content validity)：

擬請有關數學、測驗專家、與高級中學教學豐富經驗教師，依據課程標準，與教學目標，編製雙向細目表，分析本測驗之效度。

(2)校標關聯效度(criterion validity)：

以九十一年全國學科能力測驗作為預測效標。學生學校模擬考成績、月考成績、平時考成績，與試卷相同範圍成績作為同時效標，分別求取各種效度。

(三)建立各班、各類組與各校之常模。

二、針對學生的作答反應組型(response pattern) (Sato, 1985)了解試題之出題狀況與學生之學習情形

(一)說明如何製作 *S-P* 表，與分析反應組型之差異係數(disparity coefficient)(Sato, 1985)。

(二)分析試題注意係數(item caution coefficient)以判斷試題之適當性與分析學生注意係數(student caution coefficient)讓教師能了解學生學習狀況。

三、進行各項統計分析

(一)檢定性別、類組(自然組與社會組)、校系差異之顯著性。

(二)建立回歸模型，檢驗經過複習教學後之學生與未經過複習教學後學生之差異。

第三節　測驗評量的發展史

近百年來，有許多的學者或教科書常會將心理計量學的發展以某某學派來作區隔與分類的依據。Suen 於 1990 提出大概可以將其依下列最為分類與區隔：

1.(Gulliksen, 1950)提出古典測驗理論(classical test theory，簡稱 CTT)。
2.(Cronbach, et al., 1972)提出推論力理論(generalizability theory，簡稱 GT)。
3.(Lord, 1980)提出試題反應理論(item response theory，簡稱 IRT)。

而上述各學派的理論發展，絕對不是互相區隔而沒有交集的。其實，有許多理論到現在仍然是並存且持續的發展著。

要說起測驗評量的發展，我們首先介紹一套被心理計量學界喻為百科全書，記載著近百來測驗發展的重要著作。此套書籍是由美國教育委員會(American Council on Education)所發行出版的，其書名為教育測量(Educational Measurement)。教育測量一書將測驗發展從一八五零年至今分為三個不同階段，大致可分為開創期(1850~1950)、盛行期(1951~1970)與擴展期(1970~1988)。

教育測量第一版(Linquist et. al., 1951)主要在探討一百年來教育測驗發展的重要回顧與整理。Binet ＆ Simon(1905)首先提出了第一份適用於評量人類智力的測驗：比西量表(Binet-Simon Intelligence Scale)。隨著比西量表的發展，許多有名的心理測驗也隨之發表誕生，為心理計量學揭開發展的序幕。開創期階段強調再測信度的重要性，重視內容效度與效標關聯效度的取得，並提倡常模參照測驗，利用個人在團體中相對得分位置來加以比較解釋測驗分數的意義。在此階段亦實行試題分析，利用統計方法來進行分析，以提昇試題與測驗品質。

教育測量第二版(Thorndike, et al., 1973)發行，主要記載盛行期心理與教育測驗的重要回顧與發展。在此階段由於電腦科技在軟體方面均有顯著進步，所以使得評量工具蓬勃發展，也因資訊技術的輔助使得測驗應用的範圍不斷地擴大，導致了許多測驗的誤用與濫用，因此在此期也出現了許多檢討的聲浪，如美國心理學會(American Psychology Association)制定的「心理學家的倫理標準」。在盛行期階段重視複本信度概念，擴展了測驗的應用層面，而效度方面，建構效度概念的建立，補充了前期效度概念的不足，並提倡效標參照測驗，以事先既定的精熟標準來促進學生的學習成就。在此階段由於電腦的快速運算與分析，產生了許多大樣本式的客觀測驗，因此因應個別化測驗的需求而產生了題庫的建立。

第三版教育測量(Linn, et. al.,1989)。由於電腦科技的突飛猛進與測驗理論的發展日趨成熟，使得在擴展時期的測驗發展變得更多樣性。試題反應理論的誕生，改進了古典測驗理論的缺失，藉著電腦快速精確的估算，分析受事者的潛在特質。而在此期的測驗分析不像過去在分析或解釋測驗時，總是凌亂不完整。注重整合分析，降低了測量的變異數，使得研究結果更能接近事實。而在擴展期更強調受試者個差異的需求，編製適合各特性之試題、施測方式與記分方式，已達到因材施測的理想。而在本期也強調測驗的倫理規範，強調測驗使用者資格、結果詮釋皆應符合倫理與道德的規範，確保受試者的隱私與人權。

由 1989 年至今，測驗理論的發展仍然在持續進行中。隨著多元智力理論的提出(Gardner, 1983)，與標準化測驗的誤用，以致於測驗評量的轉型發展，已有逐漸重視實作評量或真實評量的趨勢，甚至新式的檔案評量方式也被使用為評量多元心智能力的主要架構，但可惜的是，這些評量方式的信度、效度或者測驗的品質與客觀公平性等，皆未獲得合理的解決，還有許多需要努力的地方。

基于第一節研究動機所提，本次研究對於實作評量、檔案評量等多元評量部分，在此不加以進一步探討。本研究只針對開創期與盛行期中能進行量化部分的古典測驗理論進行探討與分析，並也會針對試題反應理論中的學生作答反應組型加以進行說明與分析。

第二章　理論與相關文獻探討

在教師生涯過程中，編擬試題以評量學生的學習成就，是一件例行工作。或許有許多教師會自認為自己所編擬的試題一定是優良試題。即使沒有修習過諸如測驗編製等教育專業課程，多年的教學經驗也會幫助教師體驗出編擬試題的法則；這種看法，其實不然。教師如果只注重教學，而疏忽評量的重要性，其教學成效通常很難從評量中獲得具體改進，即使有，則成效亦不會顯著。因此，學會編擬優良試題與適當的分析試題以評量學生學習成就，是改進教學的管道之一。本研究就試題分析部分，提供教師在自編測驗時，如何判別所編試題即為優良試題，以幫助教師增進教學評量的品質，進而改進教學。

第一節　　試題分析理論

在教師自編測驗的過程中，判定教師所編製出來的測驗試題是否為優良品質試題，常可由兩方面來著手：

1. 對試題內容進行品質分析(qualitative analysis)，可由試題的內容審查，及一些有效的命題原則與教學目標等評鑑工具來進行(Linn, & Gronlund, 2000)。
2. 針對試題進行量化分析(quantitative analysis)，主要在分析每個試題所具備的三大統計特徵：難度(difficulty)、鑑別度(discrimination)、誘答力(distraction)此為本文所需討論之重點之一。

經過試題分析後，教師除了可以得知自編測驗的品質好壞之外，也可以間接提昇測驗的信度與效度，將在下節之中再做詳細的介紹。除了提昇試題的品質，教師更可以透過試題分析來挑選優良試題，建立起自己的題庫，以供日後使用。接者，我們針對試題量化分析的三大統計特徵再進一步加以說明。首先，先來探討試題的難度指標。

一、難度指標分析

(一)、答對百分比與難度指標

難度分析的主要目的在確定每一個試題的難度為何？最簡單且最常見的分析方法是直接計算全體學生中，每題答對人數占總人數的百分比值，所以可得答對百分比值公式為：

$$P_i = \frac{R_i}{N} \times 100\% \qquad \mathrm{i} = 1,2,\cdots\cdots,\mathrm{n} \tag{2-1}$$

其中，P_i 為第 i 題試題的答對百分比指標，R_i 為第 i 試題答對人數，N 為

總人數。

雖然答對百分比指標運算簡單且觀念顯而易見，但卻無法顯示出當兩個試題的答對百分比相同時，兩試題的難度是否真的相同。是高分組的同學答對較多？還是低分組的同學答對較多？

基於上述疑慮，提出了另外一種分析試題難度的運算方式，此方法利用選取高、低分組兩個極端族群，來計算其難度指標，而非只利用答對百分比指標來表示。而此難度指標之公式如下：

$$P_i = \frac{P_{iH} + P_{iL}}{2} \qquad \mathrm{i} = 1,2,\cdots\cdots,\mathrm{n} \tag{2-2}$$

P_i 為第 i 題試題的難度指標，R_{iH} 為高分組答隊第 i 試題人數(依測驗原始得分高低順序排序後，選取最高 25%~33%之人數)，R_{iL} 為低分組答隊第 i 試題人數(依測驗原始得分高低順序排序後，選取最低 25%~33%之人數)。

由公式可以發現試題難度指標，是以高分組和低分組學生答對人數百分比之平均數來表示。固當「P 值越大則代表試題難度越容易；反之，P 值越小則越困難」。且難度指標之值域將會介於 0 到 1 之間，可參考附錄二中本測驗各試題難度指標運算結果。

(二)、等距量尺分析

不過，利用上列方法所表示之難度指標，只能用來表示試題難易的相對位置，為一種順序量尺(ordinal scale)，無法指出各難度間差異大小的數學涵義。為了克服這項缺失，若當每個試題所測驗出的特質數值為常態分配，可將試題之難度指標，依照常態分配表所分佈的機率大小，轉換成具有相等單位的等距量尺(interval scale)指標。

常用等距量尺為美國教育測驗機構(Educational Testing Service，簡稱 ETS)所使用之公式，以轉換過的分數來表示試題難度值。其轉換公式如下：

$$\Delta = 13 + 4Z \tag{2-3}$$

Δ為試題難度指標，Z 為標準化常態分配的標準分數。

例：P=.9987　Z=-3　Δ=13+4(-3)=1　；　P=.0013　Z=3　Δ=13+4(3)=25

P=.8413　Z=-2　Δ=13+4(-2)=5　；　P=.1587　Z=2　Δ=13+4(2)=21

P=.5000　Z=0　Δ=13+4(0)=13

Δ值介於 1 與 25 之間，平均難度值為 13；Δ越大，表試題越困難；反之則越容易。Fan (1952) 已將其整理轉換製成表格，教師們可從「范氏試題分析表」(Fan’s item analysis table)直接查得(郭生玉 民 85,附錄二)。

（三）、難度指標分析

而要如何選取適當的難度指標呢？在常模參照測驗中，常以難度指標為 0.50 左右為最理想，此時試題最能夠區別學生之能力高低，而針對試題作答方式的不同，應該就會有不同的難度指標值。Ahmananm 與 Glock (1981) 就針對選擇題試題，主張應以 0.4 到 0.7 之間為難度值的範圍。但在測驗的目的上，大都希望檢驗學生是否達到教學預設的精熟程度，因此，教師都會期望學生在試題表現上，應該都接近百分之百答對。所以，教學品質與學習的優劣才是影響難度指標最直接的因素。理論上，在做測驗試題的難度指標時，應與教學前預先設定通過標準一致。因此，教師可以不必專門挑選難度指標為 0.50 左右之試題，而是應該依據教學情況來選取適當之難度指標即可。

接者，我們在來討論試題量化分析的第二項特徵—鑑別度指標。

二、鑑別度指標分析

鑑別度分析的目的在於確定試題是否具有區分學生能力高低的作用。鑑別度高，則表示試題能區別不同能力學的的功能越強；反之，若試題區別不同能力學生的功能很弱，則其試題鑑別度一定較低。但若鑑別度低，則不一定代表試題就不是優良之試題。例如當有些教師希望試題難度較容易或較艱深時，常會造成鑑別度過低之現象(如本研究附錄二中前測第 5 題)。如果經過仔細檢查這些鑑別度低的題目後，發現它們不但沒有命題的缺點，而且是測驗重要的學習結果，則應保留之。

鑑別度依分析時所依據的標準，大約可以歸納為兩類：一是內部一致性(internal consistency)分析法，另一是外在效度(external validity)分析法。

（一）內部一致性分析

主要在瞭解各個試題的功能是否與整個測驗的功能相符合一致。也就是說，我們在做測驗時多半希望高能力學生在每個試題上的答對百分比值皆能比低能力學生的答對百分比值要大上很多，而如果每個試題皆有這樣的特性，則說各試題與整個測驗的期望相符合。基於此，所以採用下列之公式來做計算：

$$D_i = P_{iH} - P_{iL} \qquad (2\text{-}4)$$

其中，D_i 為第 i 試題的鑑別度指標，P_{iH} 為第 i 試題高分組答對百分比，P_{iL} 為第 i 試題低分組答對百分比。

Kelly(1939)指出當測驗分數為常態分配時，以 27%分組可以獲得試題最可靠的試題鑑別力。當低於 27%時，結果的可靠性較低，而百分比太大時，會影響題目的作用。所以對教師而言，合理的分組百分比為 25%~33%之間。

鑑別度指標(item discrimination index)通常以小數表示，其值介於 -1.00 到 1.00 之間。當指標數愈高，表示鑑別力愈大，反之指標數愈小，則鑑別力愈小。當鑑別度過低時，其可能原因為：

(1)題目太容易或太艱難，使的大部分的人皆答對或答錯。

(2)題目題義不清。

如有試題其鑑別度為負時，則為一種很不尋常之狀態，教師應將試題刪除或者修改。

（二）外在效度分析

目的在於檢驗試題是否具有預定的某種鑑別作用，學生在每個試題上的反應與在校標上的表現所具有的相關情形為何，其分析過程與內部一致性分析方法大之相同，唯一不同處在於外在效度分析是依據外在效標的分數來分為高、低分兩組。

本次測驗只使用內部一致性進行分析，而不進行鑑別度的外在效度分析的主要原因，在於沒有適當之外在效標來作為區分高分組與低分組的分類指標。本測驗之各鑑別度可參考附錄二之結果。

除了上列分析鑑別度指標方式外，另外還有一些分析各試題鑑別度之方法，如林清山(民 84)文中所提到之二系列相關γ (biserial correlation)來進行分析，此分析方法主要都是利用試題作答結果的對或錯，與測驗總分間的相關係數，作為試題的鑑別度指標。可利用范氏試題分析表直接查得二系列相關γ (如附錄二)。

當二系列相關係數值愈高，則代表試題的作用與測驗總分的作用之間愈一致，其鑑別功能愈強。以下再加以說明如何利用鑑別度指標篩選適當的試題題目。

(三)、鑑別度指標分析

評鑑試題性能的優劣，並沒有一致的標準，常因編製測驗的目的與性質而有所不同。常用的評鑑原則是先選出鑑別力較高的試題，再從中選出難度較適合的題目。而鑑別度大部分可接受的標準為 0.2 以上，低於此標準之下則視為鑑別力不佳試題。Ebel 與 Frisbie (1979, 1991) 提出評鑑鑑別度的標準如下：

表 2-1　鑑別度評鑑標準表

鑑別度指標	試題評鑑
0.40 以上	非常優良
0.30~0.39	優良，但可能需要修改
0.20~0.29	尚可，但應該再做修改
0.19 以下	劣，需要刪除或修改

而要如何選取適當的鑑別度分析方法，在此提出一些使用建議；就測驗的使用目的而言，如果教師自編測驗的目的在於想了解每個試題在整份測驗中的相對好壞，則建議使用內部一致性分析或二系列相關係數。如目的是想了解學生未來在外在效標上的可能表現則才使用外在效度分析(效標分數必須來自自編測驗以外，且具有某項標準的效標分數)。

三、選項誘答力分析

Haladyna (1944) 指出客觀測驗(例如：選擇題)之所以要進行選項的誘答力分析(distraction analysis)，乃因可藉此來提供教師近一步的試題分析指標，協助教師改進編製試題的技巧與初步瞭解整體學生的作答情況。而要分析試題選項是否具有誘答功能，我們仍然只要分析高、低分組學生在各試題選項中的選答次數，再加以判斷即可。主要的判斷基本原則為：

1.每個不正確之選項，都至少有一位低分組的學生來選。
2.低分組選擇不正確之答案要比高分組選擇不正確之答案的人來的多。

若發現有違反上述兩個參考原則之一者，則表示該試題不正確的選項需要加以修改或者是刪除，以保持選擇型試題的基本誘答功能。

本節所討論的部分，主要是針對各試題的選項部分進行微觀分析(micro -analysis)，分析各試題之難度、鑑別度與誘答力。而下節之中將主要針對整份測驗卷的信度與效度來進行試卷的鉅觀分析(macro-analysis)。

第二節　信度、效度分析理論

信度(reliability)與效度(validity)是測驗上的兩大特徵(Gronlund, 1976)，教師在編製每一份測驗評量時，都希望測驗具有較高的信度值與效度值。也為了使測驗能成為一份公正、客觀、且優良的教學評量工具，分析試題的信度與效度更是必要的步驟。也唯有利用這些數值的分析，教師才能知道自己所編製之測驗是否有使用之價值，與是否適合所教導

之學生。首先，我們先針對信度的意義與原理來做討論。

一、信度分析理論

（一）信度的意義與原理

Anastasi (1988) 指出信度的意義可從兩個觀點來討論，從測量的一致性來看，信度就是指相同的個人在不同的時間，以相同的測驗或以複本測驗測量，所得結果的一致性。如果兩次或多次測驗的結果一致，表示測驗分數具有穩定性與可靠性。

若從測量誤差來看，信度就是指測驗分數，反映出真實量數(true measure)的程度。簡單來說，信度是在估計測量的誤差有多少。亦即在測驗分數中，有多少比率是由於特質的真實差異所造成，以及有多少比率是由於測量誤差所造成。如果測驗中包含有很多的誤差成份，信度就會低。當信度指數為 1 時，則為完全是真實的分數，根據這個觀點來看，也可以發現信度高表示測驗具有精確性與可靠性。

在傳統測驗理論中，實得分數(observed score)的公式表示為：

$$X = T + E \tag{2-5}$$

其中，X 為實得分數，T 為真實分數，E 為誤差分數。

由於任何測驗都有誤差存在，所以在所得分數中一定會包含許多若干程度的誤差在內。Payne (1974) 將測量誤差大概可分為非系統誤差(unsystematic error)和系統誤差(systematic error)兩大類。非系統誤差主要是來自受試者本身的身心狀況，屬於隨機不可預測所以又可稱為隨機誤差(random error)。系統誤差則是屬於測驗本身或測驗環境所產生之誤差，是一種一致的方式下高估或低估分數(例如：將測驗時間統一延長)。所以又稱為常誤(constant error)。

真實分數(true score)又稱為普遍性分數(universe score)，它是無法正確地被測量到。但真實分數的求得可由同一組受試者在標準的測驗情境下，接受相同的測驗(或複本測驗)相當多次(理論是無限多次)所得測驗分數的平均數估計而得。

在傳統測理論下，可將公式(2-5)之等號兩邊各求其變異數，則實得分數變異量則為真實分數變異量與測量誤差變異量之和(其實，公式(2-6)等號右邊應在加上兩倍的真實分數與測量誤差的共變數，但由於傳統測驗理論的假設之一為兩分數的相關係數為 0)(Allen ＆ Yen, 1979)。所以可得公式如下：

$$S^2_x = S^2_t + S^2_e \tag{2-6}$$

S^2_x 為實得分數總變異量，S^2_t 為真實分數變異量，S^2_e 為誤差分數變異量。

可加以推演則可推得信度係數公式如下：

$$r_{xx'} = \frac{S_t^2}{S_x^2} = \frac{S_x^2}{S_x^2} - \frac{S_e^2}{S_x^2} = 1 - \frac{S_e^2}{S_x^2} \qquad (2\text{-}7)$$

（二）信度的種類

由上列敘述可知，決定信度最理想的方法是在相同的情境下去獲得兩組測驗分數，比較兩組分數的差異。但是，事實上我們根本不可能在完全相同的情形下去實施兩個測驗，因此，有許多的替代方式來估計測驗之信度。估計信度的方法有非常多種，在不同的情況下就會有不相同的信度估計方式，以下列出本文會使用到之相關估計方法，教師如想了解其餘不同信度信度之計算方式可參考其它各相關資料。

1.重測複本法：以等值的兩份測驗，在不同的時間來測量同一群學生。重測複本法不但可以反映出測驗內容的誤差，也可以反映出學生狀況所造成的誤差，此種信度又稱為穩定與等值係數(coefficient of stability and equivalence)。因此為估計測驗信度最嚴謹的良好方法之一(Gronlund, 1993)。雖然如此，但仍有以下缺點。

(1)測驗試題的編製不易，不但費時而且所要花費之成本較高。

(2)仍然會受到學生的練習與記憶性質而有所影響。

2.內部一致性法：上述估計信度方法，為需使用兩次測驗才能估計，學生的合作意願或動機都會影響到測驗的實施。因此只根據一次測驗結果就可以估計到信度的方法，則會受到大多數教師的歡迎，本文將介紹幾種在本測驗中曾使用到之方式讓各位教師作為參考。

（1）折半方法

折半法是憑著一次測驗的結果，求兩半分數的相關係數，這種相關係數即稱為折半相關(split-half correlation)。而折半的方式最常用的方法有兩種：一是依題目隨機分為兩半，另一種則依奇數題與偶數題分為兩半。至於哪種方式較好，目前尚無一個標準的參考原則。本文使用後者方法分析，因試題難度大約是依據順序排列。本測驗之折半信度如表 4-4 所示。

而折半信度其實只是使用了半分測驗來估計信度，所以一定會低估原來試題長度的測驗信度。因此，為了能夠估計原來試題長度測驗之信度，Spearman (1910) 與 Brown (1910) 兩位學者便提出了斯布校正公式(Spear- man-Brown formula)，將折半信度加以還原，不過與重測方式估計之信度仍然會有些差距。斯布校正公式如下：

$$r_{SB} = \frac{nr}{1 + (n-1)r} \qquad (2\text{-}8)$$

r_{SB} 為估計信度，r 為原測驗信度，n 為測驗加長或縮短之倍數。

此時因使用折半之方式，固 n 值取值為 2，$r_{hh'}$ 為折半信度。

$$r_{xx'} = \frac{2\,r_{hh'}}{1 + r_{hh'}} \tag{2-9}$$

雖然斯布校正的折半信度估計方式最為普遍流行，但是，斯布公式的基本假設是在兩半測驗的「變異數相等」時才可使用，如違反假設，則會導致信度的高估(Hopkins, Stanley, & Hopkins, 1990)。於是 Flanagan(1937)與 Rulon (1939)則各提出另外兩種等價的校正公式，公式分別如下：

$$r_{Fl} = 2(1 - \frac{S_0^2 + S_e^2}{S_x^2}) \tag{2-10}$$

$$r_{Ru} = 1 - \frac{S_d^2}{S_x^2} \tag{2-11}$$

其中，r_{Fl} 為 Flanagan 信度，r_{Ru} 為 Rulon 信度，S_0^2 為奇數題分數的變異數，S_e^2 為偶數題分數的變異數，S_x^2 為測驗總分的變異數，S_d^2 為兩半測驗分數之差的變異數。

上述三種校正方式的算法，都是在計算「真實分數變異數」占「實得分數總變異數」之比率。這些估計方式皆符合上述之信度理論，且可證明公式之間是可互相推導的。

以下將依據內容之異質性介紹另外一種估計信度的方式。

（2）庫李方法

庫李方式(Kuder & Richardson, 1937)也是只需要實施一次測驗即可，不過，不必將測驗分成兩半。主要想法是依據學生對所有試題的反應，分析試題間的一致性(inter-item consistency)已確定題目是否測量相同特質。主要受到兩種誤差影響：第一為內容取樣，另一則是取樣的內容異質性，而所謂的內容異質性，指的是測驗中的題目所要測量的特質超過兩種以上(Anastasi, 1988)。

庫李公式對於測驗試題的基本假設：

(1)試題記分方式使用對或錯的二元化記分。
(2)試題不受作答速度之影響。
(3)試題大都為同質性試題。

滿足這些假設時，則所估計信度將與折半信度相近。本文使用庫李 20 來做估計，其公式如下：

$$r_{KR20} = \frac{n}{n-1} \times \left(1 - \frac{\sum_{i=1}^{n} p_i q_i}{S_x^2} \right) \tag{2-12}$$

r_{KR20}為測驗信度，n 為測驗題數，p_i為答對第 i 試題的百分比值，q_i為答錯第 i 試題的百分比值，S_x^2為測驗總分之變異數。以下將針對公式部分加以分析解釋。

因為試題使用二元化記分，所以可知每個試題的變異數為 $p_i q_i$，所以 $\sum_{i=1}^{n} p_i q_i$ 則為個別試題的變異數總和，當 $S_x^2 > \sum_{i=1}^{n} p_i q_i$ 時則表示試題間，具有共變數存在，表示試題間具有相互關聯，所以當共變數越大時代表試題的同質性越高，所以，所測量出來之信度相對的也將愈大。

由上述可知，測量信度的方法有許多種，在不同的測驗環境下常常都會使用不同的估計方式，當你使用了錯誤的估計方式時，常常會導致信度高估或者低估，所以教師在選擇估計方式時，還是需依照自己班級之狀況與所使用之測驗方式來選擇所需之估計信度方法。

二、效度分析理論

在測驗所應具備的重要條件中，效度為一最重要的特性(Ahmanan ＆ Glock, 1981)。它可以告訴教師一項測驗是否能真實且正確地測量到所預測的能力。因此，如果缺乏效度，測驗的其它結果就算在好，也是沒有應用的價值。

（一）效度的意義與原理

效度(validity)是指測驗分數的正確性。簡單來說，就是指一個測驗能夠測量到它所想要測量的特質的程度，或測驗能夠達到其編製目的的程度。所以當效度愈高，測驗的分數則愈正確。而在了解效度的意義時，應避免將它當作成一個籠統的名詞看待，誤認為一個測驗僅有一個效度而已。事實上，一個測驗因其使用的目的和不同的效度估計方法，就會有許多種不同的效度。

為了釐清大家對於效度的觀念 Gronlund (1976) 將教師在做測驗時應該了解的效度特性加以說明：

(1)效度是無法直接測量，但可從其他資料推論，所以測驗教師必須審慎判斷效度的證據是否適當或令人滿意。
(2)效度的判斷是依據測驗分數的使用目的或結果來解釋。因此，效度是屬於測驗的結果，而非測驗工具本身。

(3)效度是程度上的差別，所以不能說一個測驗完全沒有效度或效度為1。因此，選擇測驗時，應考慮測驗是否是適合使用的需要？如果是，則應選擇多高的效度？

(4)效度在使用的目的和情境具有特殊性，故不宜視為普遍性的特質。即應依據使用的目的去選擇具有適當效度資料的測驗。

Kerlinger (1986) 依據效度的原理可以經由對測驗分數變異量的分析加以說明。一個測驗的總變數可包含三個部分：共同因素的變異數(common factor variance)、獨特的變異數(specific variance)和誤差變異數(error variance)，公式表示如下：

$$S_x^2 = S_{co}^2 + S_{sp}^2 + S_e^2 \qquad (2\text{-}13)$$

其中，S_x^2為總變異數，S_{co}^2為共同因素變異數，S_{sp}^2為獨特變異數，S_e^2為誤差變異數。

將上式兩邊各以總變異數(S^2_x)除之，則可得三部份變異來源在總變異數中所佔的比率值，結果如下：

$$\frac{S_x^2}{S_x^2} = \frac{S_{co}^2}{S_x^2} + \frac{S_{sp}^2}{S_x^2} + \frac{S_e^2}{S_x^2} \qquad (2\text{-}14)$$

其中，共同因素的變異數佔總變異數的比率值，即定義為「效度」，並以符號 r_v 表示，亦即：

$$r_v = \frac{S_{co}^2}{S_x^2} = \frac{S_x^2}{S_x^2} - \frac{S_{sp}^2}{S_x^2} - \frac{S_e^2}{S_x^2} = 1 - \frac{S_{sp}^2}{S_x^2} - \frac{S_e^2}{S_x^2} \qquad (2\text{-}15)$$

由此可知，效度即是指某個測驗與其他測驗(大部分指外在效標)所「共同分享之變異數」部分佔該測驗總變異數的比率值，即是指兩個或兩個以上測驗所共同擁有的部分。而所謂的「獨特因素」，就是指某測驗本身所單獨具有的系統變異數，不與其他測驗所分享的部分。有關這兩個概念，教師可參考多變量分析(multivariate analysis)等統計專書(林清山 民 69, Hair, et al., 1995)。

（二）效度的種類

測驗的效度既然要依測驗達到其使用目的的程度而定，評估一個測驗的效度就不能只看效度的高低，尚需考慮效度的資料是否適合其原來的使用目的，再據以採用不同的方法收集各種效度資料。根據美國心理學會(American Psychological Association)、美國教育研究學會(American

Education Research Association)和國立教育測驗委員會(National Council on Measurement in Education)等三個教育專業團體於 1985 年出版的「教育與心理測驗標準」(Standards for Educational and Psychological Testing)中，針對測驗使用目的的不同，而將其分類為下列三種：

1.內容效度(content validity)：即確定學生的表現與測驗內容樣本之間的關聯性。例如，給學生做一份數學成就測驗，教師可以確定學生目前的數學成就水準。

2.效標關聯效度(criterion validity)：即預測學生未來的表現，或估計學生在目前在某些效標表現上的未知狀況。例如，給學生做一份學業性向測驗，可以預測他們未來的學業成績表現。

3.建構效度(construct validity)：即推論學生是否具有某些特質的程度。例如，給學生做一分記憶力測驗，用以推論學生的學業性向。

根據上列分類，主要將效度分為三種類型。以下將本文所使用到之效度估計方法列出，如各教師需要知道其餘更詳盡之效度估計方法可參考其他各類教育與心理測驗書籍(如：郭生玉　民 85)。

（1）內容效度

一般而言，測驗試題若能涵蓋所有的教學目標與教學內容，並且是根據雙向細目表來命題，且具有充分的代表性，即能夠確定該測驗具有適當的內容效度。因此，確立教學目標與教材內容是內容效度的兩個重要層面。

建立內容效度的方法有許多種，本文主要利用邏輯的分析方法，仔細的判別每一個試題是否符合教材內容的與教學目標。此項做法為邀請相關教師，針對測驗編製之雙向細目表，仔細判斷這份測驗試題是否與教學內容、目標符合。如果判定結果顯示真是如此，則表示該測驗具有良好的內容信度。由於這種分析方式是屬於邏輯的分析與合理的判斷，故又稱為「合理或邏輯的效度」(rational or logical validity)。

當然，除了利用專家的判斷外，尚可利用複份的編製(duplicate-construction)來分析內容效度(Cronbach, 1971)。此方法是根據編製前所確立的測驗內容之選擇原則，分別編擬兩套試題，此法類似於前面討論的複本信度，之前信度所關心的是：我們有多少信心能夠從測驗上的分數推論到其他相等的測驗分數。同樣的，在內容效度中，我們關心從一個獨特的分數能夠推論到較大教材內容的適切性。因此，依同樣的內容編兩份測驗所得之相關，不僅可以提供複本信度還可作為內容效度之用。

另外有一些效度，如表面效度(face validity)常易與內容效度產生混淆的效度概念。表面效度缺乏系統的邏輯分析，它只是指測量給人的第一個印象好像是在測量某種特質的指標。表面效度的存在對於一個測驗來

說仍然是重要的，因為具有表面效度，有時可以激勵學生認真作答，可讓學生感到有親切感，並表現接受測驗的合作意願。

（2）效標關聯效度

效標關聯效度為另一種估計效度的方式，也是大家普遍使用的一種方法，它是以經驗性的方式研究測驗分數與一些外在效標間的關聯性，故又稱為「經驗效度或統計效度」(empirical or statistical validity)(Thorndike ＆ Hagen 1977)。可將其分為同時效度與預測效度兩大類。

其中，同時效度(concurrent validity)指測驗分數與實施測驗相近時間所取得的效標之間的相關係數。本文利用學生平時成績、與試卷相同範圍考試成績、月考成績和模擬考成績作為同時效度之效標依據，其結果可見表 4-4。

而預測效度(predictive validity)指測驗分數與實施測驗後一段時間所取得的效標之間的相關係數，主要在使用測驗分數預測個人在效標方面的未來表現。因要選取適當的效標是不容易的事情，需要滿足可靠性、客觀性與可用性等，本文選取九十一年大學學科能力測驗各學生數學科成績作為此次的預測效標，運算結果如表 4-4 所示。

（3）建構效度

建構是心理學上的一種理論構想或特質，它本身是觀察不到，並且也無法直接測量到，但卻被理論假設是存在的，以便能夠用來解釋或預測個人或團體的行為表現。Anastasi (1988) 指出建構效度即是指測驗能夠測量到理論上的建構或特質程度，亦即根據理論建構，測驗分數能否達到它的測驗目的所做的分析和解釋。其實，建構效度所涵蓋的範圍包括內容效度和效標效度兩者，所以，所有有關內容效度和效標關聯效度所建立的方法與結果皆可作為分析建構效度的基本證據，但在本文將不使用。

前兩節所提到的理論皆屬於古典理論之範疇，所採用之公式簡單明瞭，適用於大多數的教育測驗驗情境。不過，古典測驗理論卻忽視了學生其作答反應組型所代表意義。對於在原始得分上相同的學生或正確反應總和相同的試題，即將學生看成具有相同能力或試題具有相同的難度估計，其實，這是一種非常不正確的看法。當只要作答反應組型為不同時，其所求出的學生能力與試題參數估計，應該就會有所不同，而非都是一樣。在下一節中將依據各個學生的作答反應組型，來進行各試題的優缺分析，與了解各個學生的作答狀況。

第三節　學生問題表分析理論

一份測驗的試題不能僅在探討其是否屬於優良的試題而已，應也需提供教師如何利用試題的分析來了解學生的問題所在，本節所要探討「學生問題表」的分析技術，是由日本學者佐藤隆博(Takahiro Sato)於一九七０年所創(Sato, 1969)，不僅可幫助教師做好試題分析，改進命題技巧外，也可幫助教師了解學生再學習上的個別差異，及觀念不清之試題為何。以便讓教師能提出各種的補救教學或輔導措施，以便提高學生之學習動機，增加學習的信心與興趣。

一、S-P 表的理論概要與編製

（一）理論概要

教師在進行試題分析時，若仔細去留意學生的作答資料，可能會發現有部分學生的答對總分一樣，但是他們的作答反應組型(response pattern)(Guttman, 1944)卻是有明顯的不同。此所謂的作答反應組型即是指學生在某份試卷上，針對每道試題作答的原始資料所組合成而的一個向量。這種反應組型不僅反應出學生的作答結果，同時也反應出學生的作答思考模式，因此反應組型不同所代表的意義就不會相同。

S-P 表所關心的課題即是學生在測驗試題上的作答反應組型是否有不尋常的一種測驗分析方法。這種分析方式適用於以班級為單位，人數不會很龐大的資料分析，是一種不對母群體特性設定任何假設值的無母數統計方法(nonparametric method)。*S-P* 表分析所使用的指標係數，有以下幾種：差異係數(disparity coefficient)、試題注意係數(item caution coefficient)、以及學生注意係數(student caution coefficient)等，會在之後再做進一步的介紹。

而所謂的 *S-P* 表分析，即在分析每位學生與每個試題的作答反應組型所產生的注意係數，與整份測驗試卷的差異係數。這些指標都是用來協助教師診斷學生表現、試題的品質以及教學的成果的重要資訊之一，也可將這些指標作為改進教學、命題與輔導學生之參考。

（二）編製過程

製作 *S-P* 表過程的概念非常簡單。首先當教師從任教班級收集到一筆 n 名學生在 m 個試題上的反應資料，經過評分之後，將會得到一個未經任何處理的 n ˣ m 的原始得分矩陣資料，此稱為「*S-P* 原始資料表」，簡稱「*S-P* 原表」，如附錄一(ａ)。

接下來，按照每位學生得分總分高低，將整個反應組型與總分由上(總分最高排最上面)往下依序排列，如有總分相同時，為了簡便起見，亦可

照座號之小大順序排列，如附錄一(ｂ)。接者，按照每道試題答對人數多寡，將整個反應組型與其答對人數由左(答對人數最多之試題排在最左邊)到右依序排列，遇到相同得分之試題，亦可依題號之小大順序排列，如附錄一(ｃ)。

最後，依據每位學生所得總分，從左往右數出與總分相同之試題個數，並在其右邊畫上一條分界線，由高分往低分別畫出每位學生總分所對應的分界線，再將這些分界線的下方利用直線連接，則會行成一階梯狀之曲線，此曲線即稱為「S 曲線」，如附錄一(ｄ)。同理，依據每道試題的答對人數，從上往下數出與答對人數相同之學生個數，並在其下邊畫上一條分界線，由左端往右端分別畫出每道試題之答對人數所對應的分界線，再將這些分界線的右方利用直線連接，則會行成一階梯狀之曲線，此曲線即稱為「P 曲線」，如附錄一(ｄ)。

S 曲線是指學生得分的累加分佈曲線，是用來區分學生答對與答錯的分界線。而 P 曲線是指試題答對人數的累加分佈曲線，是用來區分試題答對與答錯人數的分界線。排列在 *S-P* 表左上方者，代表能力較好之學生與較簡單之試題，大多數是被期望答對的試題，所以這個區域應該出現大多數為 1。相反的，在 *S-P* 表右下方者，應該出現大多數為 0。

當 S 曲線以左或 P 曲線以上部分都出現為 1，我們成這種情況為「完美量尺」(perfect scale)的反應組型，而且此時可以發現 S 曲線與 P 曲線將會重疊在一起。但實際上這種完美量尺是不太可能出現，大部分之反應組型皆會出現不規則或不尋常的情形。為了了解這種不尋常或異常的嚴重性，我們則利用了一些量化指標的分析來加以解釋，其中，差異係數便是辨別異常嚴重性的一種方式。

二、差異係數

常見的反應組型皆是屬於不完美量尺，其 S 曲線與 P 曲線會呈現分開的狀況，分離程度太大或太小都不適宜，而其分離程度的標準值常常都是需要靠經驗來加以判斷，其標準值會隨著測驗的內容、種類與形式都會有所不同。而測量兩曲線的分離程度我們可以使用一個量化指標來表示，即所謂的「差異係數」(disparity index)(Sato, 1985)。差異係數是指實際 S 和 P 曲線所圍面積部分，占隨機情況下 S-P 曲線所圍成之面積期望值的比值，其公式表示如下：

$$D^* = \frac{S(n,m,\bar{p})}{S_B(n,m,\bar{p})} \qquad (2\text{-}16)$$

$S(n,m,\bar{p})$ 表示在人數為 n，試題數為m，平均答對率為 $\bar{p}$ 的條件下，實際 S-P 表中兩曲線所包圍部分的面積；$S_B(n,m,\bar{p})$ 表示在人數為 n，試題數為m，平均答對率為 $\bar{p}$ 的條件下，兩曲線呈累加二項分配曲線所包圍部分

之面積。

由於 $S_B(n,m,\bar{p})$ 之公式運算複雜且費時，*S-P* 表之發明者佐藤博士(Sato,1985)提出下列的近似公式來取代：

$$D^* = \frac{C}{4nm\bar{p}(1-\bar{p})D_B(M)} \qquad \text{M} = [\sqrt{nm} + 0.5]\text{表示取高斯整數值} \qquad (2\text{-}17)$$

C 指 S 與 P 曲線所包含之「1」和「0」的個數和，n 指學生人數，m 為試題數目，$D_B(M)$ 的值可由上述計算之 M 值代入表 2-2 中查得。

當為完美量尺時差異係數會等於 0，而當實際得到的 S-P 表純粹是由隨機造成的，則差異係數將會等於 1，所以在正常的情況下，差異係數會介於 0 到 1 之間。在大多的實際例子中，以 D^*=0.5 左右為標準值；當 D^*>0.65 或 D^*<0.4 時，表示測驗含有異質因素，應對學生的反應組型予以注意，對試題加以檢討，並做適當的修改(Sato, 1985)。

表 2-2　$D_B(M)$ 查表

M	$D_B(M)$	M	$D_B(M)$	M	$D_B(M)$	M	$D_B(M)$	M	$D_B(M)$
11	.278	29	.355	47	.384	65	.402	83	.413
12	.285	30	.358	48	.385	66	.403	84	.413
13	.291	31	.360	49	.386	67	.404	85	.414
14	.296	32	.362	50	.387	68	.404	86	.414
15	.302	33	.364	51	.388	69	.405	87	.415
16	.307	34	.366	52	.389	70	.405	88	.415
17	.312	35	.367	53	.390	71	.406	89	.416
18	.317	36	.369	54	.391	72	.407	90	.416
19	.321	37	.370	55	.392	73	.408	91	.417
20	.326	38	.372	56	.393	74	.408	92	.417
21	.330	39	.373	57	.394	75	.409	93	.418
22	.334	40	.375	58	.395	76	.409	94	.418
23	.337	41	.377	59	.396	77	.410	95	.419
24	.341	42	.378	60	.397	78	.410	96	.419
25	.344	43	.380	61	.398	79	.411	97	.419
26	.347	44	.381	62	.399	80	.411	98	.420
27	.350	45	.382	63	.400	81	.412	99	.420
28	.353	46	.383	64	.401	82	.412	100	.420

三、注意係數

注意係數(caution index)是 *S-P* 表針對「個別」學生與試題所使用的另一類係數，可分為學生注意係數(caution index for students，簡稱 CS)與問題注意係數(caution index for items，簡稱 CP)兩種。主要是用來作為判斷學生或試題在反應組型中是否有異常現象的指標，教師可利用這些指標

數來了解學生或試題的狀況與問題所在。

所謂的注意係數指的是 *S-P* 表中實際反應組型與完美反應組型間的差異，占完美反應組型之最大差異的一種比值(Sato, 1975)，其公式表示如下：

$$\text{注意係數} = \frac{\text{實際反應組型與完美反應組型間的差異}}{\text{完美反應組型的最大差異}} \tag{2-18}$$

所謂的最大差異，旨在給定此組型之平均值時，最為不規則之狀態，這種定義方法以 *S-P* 表之邊際次數(即各個學生的 總分或試題之答對人數)為基準變量來表示，因此上述公式可表示如下：

$$\text{注意係數} = \frac{\begin{bmatrix}\text{完美反應組型與基}\\ \text{準變量之共變數}\end{bmatrix} - \begin{bmatrix}\text{實際反應組型與基}\\ \text{準變量之共變數}\end{bmatrix}}{\begin{bmatrix}\text{完美反應組型與基}\\ \text{準變量之共變數}\end{bmatrix} - \begin{bmatrix}\text{隨機反應組型與基}\\ \text{準變量之共變數}\end{bmatrix}} \tag{2-19}$$

上述隨機反應組型與基準變量之共變數一項，再計算時會有差異產生，因此，大都以期望值來表示較為適當，而在此種情況下其期望值為 0。所以公式又可表示如下：

$$\text{注意係數} = 1 - \frac{\text{實際反應組型與基準變量之共變數}}{\text{完美反應組型與基準變量之共變數}} \tag{2-20}$$

因此我們可以推算出學生與試題注意係數之計算公式分別如下：

※問題注意係數：

$$\begin{aligned} CP_j &= 1 - \frac{\sum_{i=1}^{n}(y_{ij} - \bar{y}_{.j})(y_{i.} - \bar{y}_{..})}{\sum_{i=1}^{n}(y'_{ij} - \bar{y}_{.j})(y_{i.} - \bar{y}_{..})} = 1 - \frac{\sum_{i=1}^{n}(y_{ij})(y_{i.}) - (y_{.j})(\bar{y}_{..})}{\sum_{i=1}^{y_{.j}}(y_{i.}) - (y_{.j})(\bar{y}_{..})} \\ &= \frac{\sum_{i=1}^{y_{.j}}(1 - y_{ij})(y_{i.}) - \sum_{i=y_{.j}+1}^{n}(y_{ij})(y_{i.})}{\sum_{i=1}^{y_{.j}}(y_{i.}) - (y_{.j})(\bar{y}_{..})} \\ &= \frac{\begin{bmatrix}\text{試題 j對應於 } P\text{曲線上方}\\ \text{答[0]的學生總分之和}\end{bmatrix} - \begin{bmatrix}\text{試題 j對應於 } P\text{曲線下方}\\ \text{答[1]的學生總分之和}\end{bmatrix}}{\begin{bmatrix}\text{試題 j對應於 } P\text{曲線上}\\ \text{方各學生總分之和}\end{bmatrix} - \begin{bmatrix}\text{試題 j之答}\\ \text{對人數}\end{bmatrix} \times \begin{bmatrix}\text{學生之平}\\ \text{均得分}\end{bmatrix}} \end{aligned} \tag{2-21}$$

※學生注意係數：

$$CS_j = 1 - \frac{\sum_{j=1}^{m}(y_{ij} - \bar{y}_{i.})(y_{.j} - \bar{y}'_{..})}{\sum_{j=1}^{m}(y'_{ij} - \bar{y}_{i.})(y_{.j} - \bar{y}'_{..})} = 1 - \frac{\sum_{j=1}^{m}(y_{ij})(y_{.j}) - (y_{i.})(\bar{y}'_{..})}{\sum_{j=1}^{y_{i.}}(y_{.j}) - (y_{i.})(\bar{y}'_{..})}$$

$$= \frac{\sum_{j=1}^{y_{i.}}(1 - y_{ij})(y_{.j}) - \sum_{i=y_{i.}+1}^{m}(y_{ij})(y_{.j})}{\sum_{j=1}^{y_{i.}}(y_{.j}) - (y_{i.})(\bar{y}'_{..})} \qquad (2\text{-}22)$$

$$= \frac{\begin{bmatrix}\text{學生 i對應於 } S\text{曲線左方答} \\ [0]\text{的試題之答對人數和}\end{bmatrix} - \begin{bmatrix}\text{學生 i對應於 } S\text{曲線右方答} \\ [1]\text{的試題之答對人數和}\end{bmatrix}}{\begin{bmatrix}\text{學生 i在 } S\text{曲線左方各} \\ \text{試題之答對人數和}\end{bmatrix} - \begin{bmatrix}\text{學生 i} \\ \text{之總分}\end{bmatrix} \times \begin{bmatrix}\text{試題之平均} \\ \text{答對人數}\end{bmatrix}}$$

其中，y_{ij} 表實際反應組型學生 i 在第 j 試題的作答狀況，y'_{ij} 表完美反應組型時學生 i 在第 j 試題的作答狀況，$y_{i.}$ 為學生 i 的總分，$\bar{y}_{..}$ 為學生之平均得分，$y_{.j}$ 試題 j 的答對人數，$\bar{y}'_{..}$ 試題之平均答對人數。

由此可知注意係數為非負值的量數，當為完美反應組型時，注意係數為 0；當為隨機反應時則會接近 1。有時注意係數之值也有可能大於 1，此時，其反應組型呈現逆轉現象，在 *S-P* 表中未及異常之反應組型。因此，當注意係數越大時，表示反應組型為不尋常或有異常之情況較嚴重；反之，注意係數越小，則表示不尋常之現象較不嚴重，在容許的誤差範圍之內。

當注意係數(CP 或 CS)介於 0~0.5 之間時，表示該試題或學生之反應組型的不尋常情況並不嚴重，屬於正常程度。當注意係數在 0.5~0.75 之間，則表示不尋常之情況已是嚴重狀況，教師應該加以注意。而當注意係數大於 0.75 時，則表示不尋常之情況已是非常之嚴重，教師應更加特別注意。本次測驗各學生注意係數可見附錄四。

上述判斷標準，只是經驗上所推薦之標準，並非為一並之使用標準。0.5 之標準值，剛好是隨機反應組型下之注意係數值得一半。但在多因子的 S-P 表分析中，其差異系數值均較大，所以注意係數之判斷標準也應該放大。

第四節　數學能力與性別差異

數學能力的表現與性別的差異是否會有顯著的不同，根據各學者的研究結果顯示並不一致。有許多研究顯示男女在數學能力上的表現有明顯的差異(Benbow ＆Stanley, 1983；Maccoby ＆ Jacklin, 1974；Liu, 1998)。但亦有許多研究顯示性別在數學能力的表現上並沒有明顯的差異存在(Janes, 1992；Paulsen ＆Johnson,1983；Philippa ＆Norma, 1984)。美國教育當局針對全國進行教育發展評量(National Assessment of Education Progress)，發現九歲與十三歲的學生，其性別在數學成就上的表現並沒有顯著的不同，但針對整體學生而言，到了十七歲以後，男生的數學成就則較女生的數學成就來的優異，而且愈難問題，性別差異愈顯著(Fennema et. al, 1991)。

Jacklin(1992)針對男生與女生在數學能力發展的表現上作了以下分析：

1.女生的數學空間能力較好
2.在十五歲之前，性別間的差異在數學能力的表現上並沒有顯著的差別。
3.而同性之間的差異，在十一歲之前的差異並不顯著，但隨著年齡的增長，男生之間的能力差異會逐漸增大，且比女生間的差異程度更為顯著。

此外，Paulsen 以高社經地區(high socioeconomic area)的四年級、八年級與十一年級的學生做性別與數學能力差異的研究，發現在高社經地區的各年級學生，其男女在數學能力上的表現並沒有顯著的不同。

第三章　研究設計與實施

本研究旨在探討如何編製一份較好的測驗試題，以提供教師在自編測驗時能有方法的了解所編製之測驗為優良試卷之可信度為何。為了使各位教師能夠更了解各種分析數值所代表的意義與運算方式，研究者依重測複本方式製作了兩份試卷，並實際進行測驗，由測驗所得之資料進行各類分析，讓教師在自行分析時，能有實際數據來加以對照。

第一節　測驗對象

本研究的測驗對象為高雄市立兩所高級中學(分別以甲高中與乙高中代表)---甲高中自然組三個班級 104 為學生，社會組二個班級 79 名學生與乙高中自然組兩個班級 66 名學生，總共合計七個班級 249 位學生。其人數分佈如下：

表 3-1　　研究樣本分配表

校名與班級	類組	男	女	合計
甲高中 3-01	自然組	18	20	38
甲高中 3-04	自然組	31	4	35
甲高中 3-05	自然組	27	4	31
合計		76	28	104
甲高中 3-09	社會組	12	26	38
甲高中 3-11	社會組	12	29	41
合計		24	55	79
乙高中 3-12	自然組	28	3	31
乙高中 3-16	自然組	23	12	35
合計		51	15	66
總和		141	98	249

一、複本信度樣本

在進行第二次測驗時，因有許多學生皆在參加申請入學或者為已經甄試進入大學，所以所取得的樣本資料較第一次測驗來的減少許多，其人數分配如下：

表 3-2　　複本信度樣本分配表

校名與班級	類組	男	女	合計
甲高中 3-01	自然組	16	15	31
甲高中 3-04	自然組	25	4	29
甲高中 3-05	自然組	22	4	26
合計		63	23	86
甲高中 3-09	社會組	7	17	24
甲高中 3-11	社會組	10	27	37
合計		17	44	61
乙高中 3-12	自然組	26	3	29
乙高中 3-16	自然組	19	8	27
合計		45	11	56
總和		125	78	203

二、效度樣本

效度樣本包括利用本次試題所測驗之得分與學科能力測驗成績求取相關 249 人，與最近學校平時考數學成績求取相關 66 人，與此份試卷相同範圍考試成績求取相關 66 人，與最近學校月考數學成績求取相關(自然組 104 人、社會組 79 人)，與學校模擬考試成績求取相關 183 人。由於測驗樣本，屬不相同之學校，且有屬於不同之類組，因此所收集到之效標成績也將有所不同，所以需要分別加以計算，而不應混雜的進行運算。詳細分類情形可見表 3-3 所示。

表 3-3　　效度樣本明細表

校名與班級	類組	男	女	學科能力測驗成績	學校平時成績	相同範圍成績	學校月考成績	學校模考成績
甲高中 3-01	自然組	18	20	38			38	38
甲高中 3-04	自然組	31	4	35			35	35
甲高中 3-05	自然組	27	4	31			31	31
甲高中 3-09	社會組	12	26	38			38	38
甲高中 3-11	社會組	12	29	41			41	41
乙高中 3-12	自然組	28	3	31	31	31		
乙高中 3-16	自然組	23	12	35	35	35		
總和		141	98	249	66	66	183	183

三、試題分析與 S-P 表分析樣本

利用兩次測驗成績作為樣本，分別分析兩次的試題難度、鑑別度(可見附錄二)、誘答力與 S-P 表等，並比較各校或各組間分數的差異(如表 4-9 所示)、兩次測驗分數之間的差異(見表 4-8)，其樣本資料分別如表 3-1 與表 3-2。

第二節　測驗方法與編製過程

本研究之目的，在於幫助自編測驗教師如何瞭解所編製之試題為一份適合學生之試題，並利用多次的測驗與分析後，統整為適合學生測驗之題庫。兩次測驗所利用之測驗方法皆屬傳統的紙筆測驗方式。其編製過程如下。

一、彙集文獻資料

本研究旨在判別所編製之試題是否為適合學生測驗之試題，故一方面探討國內外學者對於測驗理論的分析論點，另一方面也蒐集有關測驗評量分析的各種方法，進而討論這些評量方式的使用時機，與使用的理論原因為何。除此之外，研究者也因應電腦化的試題分析，增對一些的試題分析程式加以解釋與應用，可因此減少了許多不必要的時間花費。

二、試題收集與編製

本次試題測驗參照我國現行高級中學教科用書：數學甲上冊第一章 ”機率與統計(二)”為命題時的範圍依據。收集大學歷屆聯考試題與大學學科能力測驗試題，並參考各版本之高中數學指定用書與大學基礎機率統計概念等相關書籍，作為本次試卷出題之參考資料。

首先根據機率統計(二)之範圍，編製以單元為經，以理解、應用、分析、綜合之性質為緯的雙向細目表，作為命題題數之依據，如表 3-4。然後根據雙向細目表和試題編製基本原則，編製本測驗之試題，而本次測驗編製之原則如下：

(一)每一重要單元，其理解與應用試題皆至少命一試題，易混淆之重要概念以二至三題加以分類澄清。

(二)題目皆以概念理解為主，複雜及龐大的計算之試題盡量避免。

(三)避免使用否定語氣之命題，試題文字與題意務求明確。

(四)盡量增列實用性與常見問題題目。

(五)試題文句盡量避免直接抄襲課本或原來之教材，導致學生在作答時有記憶性作答之疑慮。

(六)排除引起爭論之命題與答案。

(七)編排答案時，注意答案與題目間的邏輯性與似真性，且盡量避免使用「以上皆是」或「以上皆非」之答案。

(八)正確答案出現位置能按隨機排列原則，其次數大致相同。

根據雙向細目表與試題編製原則，選取或自行出題編製試題。試題編製完成後，分別請高中數學教師、數學系教授與研究生數人做試卷審定，綜合各專家意見，再進行測驗試題之修改與編製。本次測驗試題之出題方式皆採用四選一之選擇題方式出題，兩次測驗各為 20 題試題，共 40 題。

表 3-4　試題雙向細目表

單元	理解	應用	分析	綜合
條件機率	1	1	1	
獨立事件	1	2		1
變異係數	1	3	0~1	1
相關係數	1	1	1~2	
整合題型	2	1		1
合計	6	8	3	3

三、進行測驗

第一次測驗(前測)於民國九十一年十一月間進行，選取之測驗樣本為以學習過條件機率、獨立事件、變異係數與相關係數之高三學生為對象。選取高雄市各高級中學中，學生平均程度為中等之學校兩所，甲高中自然組班級 3 班社會組班級 2 班與乙高中自然組班級 2 班，共 7 個班級。

第二次測驗(後側)於民國九十二年三月間進行，測驗樣本與第一次測驗樣本相同。兩次測驗時間相距三個多月之久，以防止學生有強烈之作答記憶來影響整份試卷之信度值。而兩次測驗之施測時間皆以 50 分鐘為限。

施測人員由研究者、甲高中與乙高中數學科教師數人負責，施測前我們先以電子郵件方式煩請各校教師協助，並告知各校教師在試題施測時，應注意之事宜。且告知各教師煩請於試題測驗完畢之後能對學生加以解說檢討，以釐清學生不懂之觀念。

四、資料處理與蒐集資料分析

測驗後之資料，將各學生在各題中之選項輸入電腦；利用電腦程式將學生之得分編排為百分等級，並選取得分最高的 25~30%與最低的 25~30%，分別代表高分組與低分組之成績，以作為項目分析之資料。分別計算出兩次測驗中高分組與低分組在各選項作答的答對百分比，以進一步分析各試題之難度與鑑別度，並利用分析所得之結果，根據「范氏試題分析表」查出常態化等距難度指標Δ與二系列相關γ(如附錄二)。

我們分多次蒐集資料分析時有關之資料；有蒐集複本信度驗證之資料，在效度驗證資料方面蒐集了學科能力測驗成績、學生最近於學校平時考數學成績、與此份試卷相同範圍考試成績、最近學校月考數學成績和學校模擬考試成績等五項資料，作為此次效度數值之分析，並彙集兩次測驗成績建立各組與全體學生之常模。資料蒐集於民國九十一年十一月開始蒐集，於九十二年四月完成全部資料之蒐集。

資料處理包括登錄、記分、及輸入電腦為三項主要事項，因所收集之資料量非常的繁多，所以在進行資料輸入時應需做多次的檢查，以防止有資料上的輸入錯誤。登入之編號方式依照各校各班學生學號作為編號之代碼，例如甲高中三年一班 1 號則使用 301-01 代碼表示。

資料分析包括分析其信度、效度、常模樣本資料、製作學生 S-P 分析表，將所收集到各類樣本資料均輸入專用電腦，並利用試題分析軟體(TESTER for Windows)、文書處理軟體(Excel)、統計套裝軟體(SPSS、R)等相關軟體進行試卷成績的各類分析，並利用 S-plus 程式語言，撰寫製作 S-P 之相關程式，其程式內容可參考附錄五之程式碼。而本研究所進行之各類分析如下：

(一)利用試題分析理論計算出各試題之難度、鑑別度與誘答力選項。

(二)計算重測覆本信度、折半信度、庫李信度與相關效度。

(三)製作學生問題表(*S-P* 表)，分析反應組型之變異係數與試題、學生注意係數。

(四)用平均數、標準差等建立各組常模。

(五)利用 T 檢定比較性別、各組平均數差異與前後測成績差異之顯著性。

(六)建立回歸模型，檢驗在經過教學後之學生與未經過教學之學生間是否有顯著差異，並再次驗證性別對分數影響的顯著性。

第三節　研究之限制

(一)本測驗主要在測驗高級中學三年級學生之機率統計概念，但因現階段高級中學各校在招收學生時皆已有將學生的程度差異作劃分，故如將相同試題給予不同學校進行測驗，可能將會得到成績上相當大的差異，故本次測驗之常模只適用於甲、乙兩所高中，如要建立全國或全高雄市之常模，則需針對全國或全高雄市之各高級中學進行隨機抽樣，再加以進行分析建立。

(二)本研究因為是針對高雄市學生平均程度屬於中等之學校進行試題編製，所以無法達到隨機取樣之客觀要求，而在進行測驗之時並非於各班教學後馬上進行，使得測驗成績有可能受到各教師是否進行複習而有所影響。

本研究盡量採取質的研究方法，難免會有主觀因素介入，在研究過程，我們盡將主觀因素降至最低。最後，仍需再次強調，本研究樣本屬非隨機取樣，所以在於研究結果之推廣上，只能提供相似程度之學校作為製作試題評量之參考資料或提供教師在對學生進行試題分析時之參考資料。

第四章　結果與討論

為了使自編測驗教師能夠有方法的更了解自己所編排之測驗試題是否為優良試題，與利用試題評量方式了解學生之學習狀況。依自行編製之試題來進行分析，讓教師們在自行做分析時能有參考的資料最為對照。茲將編製過程中有關測驗題項、信度、效度、學生問題表與各類統計分析之實證結果，分別加以說明如下：

第一節　測驗題項

本測驗之試題是根據高級中學數學甲上－機率與統計(二)為標準來加以製作，分為條件機率、獨立事件、變異係數與相關係數等單元，其試題分配如下表所示：

表 4-1　試題題數分配表

試題單元		題數	試題題號		題數	試題題號	合計
條件機率	前測	3	1,10,16	後測	3	1.10.16	6
獨立事件		4	2,3,6,9		4	2.3.6.9	8
變異係數		6	7,8,13,17,18,20		5	7.8.13.17.18	11
相關係數		3	11,15,19		4	11.15.19.20	7
整合題型		4	4,5,12,14		4	4.5.12.14	8
總和		20			20		40

兩次測驗試題其試題分配相關程度為 0.866，試題內容相似程度為 0.93，大致屬於兩份同質性試題。各試題編製完成後，皆送請高級中學教師審核，經修改與刪除部分題項後，再進行測試。將試題分別以兩次測驗對象，做項目分析，利用本文第二章所提之難度、鑑別度公式，分別求出各題項之難度 P 值與鑑別度 D 值；將所得結果再依據范氏項目分析表查出鑑別度 γ 與常態化等距難度指數 Δ，如附錄二。

表 4-2　試題項目分析表

	統計量	第一次測驗				第二次測驗				兩次測驗			
		P	Δ	D	γ	P	Δ	D	γ	P	Δ	D	γ
全體學生	最高值	0.85	17.8	0.58	0.58	0.81	17.9	0.64	0.69	0.85	17.9	0.64	0.69
	最低值	0.11	8.7	0.03	0.06	0.12	9.3	0.12	0.14	0.11	8.7	0.03	0.06
	全距	0.74	9.1	0.55	0.52	0.69	8.6	0.52	0.55	0.74	9.2	0.61	0.63
	平均值	0.45	13.4	0.34	0.39	0.45	13.6	0.41	0.47	0.45	13.5	0.37	0.43

P 表難度指標，Δ表常態化難度指標，D 為鑑別度指標，γ為二系列相關。

一、難度指標

兩次測驗之測驗難度分佈皆非常廣泛，前測介於 0.11~0.85 之間，後側介於 0.12~0.81 之間。就平均難度來看兩次皆約為 0.45，難度指標通常代表試題受試者的通過百分比，所以數值越大代表試題越容易，相反的，越小則越難，如從數據顯示此份試卷對於此次受測學生而言，為較偏難之試題。

二、常態化難度指標(Δ)

常態化難度指標為一以 13 為平均數，4 為標準差的標準分數(Anastasi, 1982)，Δ值越大，則難度越低；越小，則難度越高。兩次測驗大致介於 8.7~17.9 之間，平均值約為 13.5。

三、鑑別度指標

鑑別度指標「D」在前測介於 0.03~0.58 之間，，後測介於 0.12~ 0.64，兩次測驗之平均 D 值分別為 0.34 與 0.41。在前測時，第九題與第十題之鑑別度指標有偏低的傾向，可見附錄二，經研究者與教師專家們討論，發現試題並無疑問，屬於學生觀念不清所造成，經過解釋複習後。於後側可發現其鑑別度皆有明顯提高。

四、二系列相關

測驗中的γ為二系列相關(biserial correlation) ，γ值表示該試題答對者與答錯者與總分之間的相關情形。前測的γ值介於 0.06~0.58 之間，有些題目的觀念是大家都不熟悉的，所以造成一些試題的鑑別率偏低，後測時，有些班級應已經過教學，使得γ值都介於 0.14~0.69 之間，所以除了少數容易的題目外，各試題跟總分之間皆有顯著的關係。

五、誘答力選項

誘答力選項方面，各試題選項大致都滿足誘答力的基本原則(可參考第二章第一節誘答力分析)，除了前測試卷中的第三題選項 3、第十題選項 1、第十五題選項 2、後側試卷中第十二題選項 2 與第十三題選項 1 等五題違反了基本原則，其餘試題之誘答力都屬非常優良，而研究者觀察此五題的選項，發現為學生觀念不清，導致掉進研究者出題的陷阱之中，應將學生此部分之觀念再次釐清。

此兩次測驗試題在研究者與專家們討論，與由上列各項數據顯示發現，整份試卷試題屬於中上難度之試題，試題內容部分，避免許多需要複雜計算之試題，大多以觀念試題為主，所以對於觀念不清楚之學生，將可能造成分數較為低弱之現象。

第二節　信度、效度與常模之研究

本研究利用兩次測驗，七個班級之學生作為測驗信度、效度與常模分析之樣本，以下分別針對各研究進行分析：

一、信度研究

本研究為了了解測驗試題本身的一致性與與穩定性，採用了重測複本方式進行測驗，利用重測複本信度分析、折半法、庫李法驗證測驗之信度值並利用重測複本信度計算出測量標準誤。

(一)、重測複本信度

本測驗的重測複本信度，兩次測驗時間具有三個月之間隔，前後測驗所得之分數利用 Pearson correlation 求取兩者相關，代表測驗之重測複本信度，如表 4-3。但因在前測結束後並非所有班級皆有再進行試題的檢討與複習，所以造成後測時，有經過檢討之班級成績接明顯之提高，而未檢討仍然維持較低分數之狀態(甚至有降低之趨向，而降低之原因應為學生皆已學測結束，有鬆懈之狀態)。有於此種原因造成了整個信度值之降低。研究者為了解決這個問題，決定將各校學生之成績分別依照各校之變異數與平均數皆轉換為 Z 分數，再利用所轉換過後之 Z 分數求出兩測驗之 Pearson 相關，整理資料如表 4-3 所示。本次測驗因為為選擇題試題，學生容易有猜題現象，所以信度值會有較低之趨勢，本測驗經過經過 Z 分數轉換後所求得之重測複本性度值為 0.536。

(二)、折半信度

本測驗根據前、後測之結果，將各測驗試題以奇數試題與偶數試題區分為相等的兩半，並分別計算學生在這兩半測驗的總得分，然後以 Pearson correlation 求取兩半測驗的相關係數，最後再將所求之相關係數以斯布公

式加以校正，所得結果即為本次測驗之折半信度，如表 4-3。

但因兩半測驗之變異數為非相等之變異，所以最好仍使用第二章所提到之 Rulon 公式加以計算，所得資料如表 4-4 所示。本測驗若使用折半信度與 Rulon 信度皆不是非常之適當，因使用折半信度之條件為兩半試題屬於非常相近之試題，簡單來說就是希望奇數題與偶數題是類似試題且難度相差不遠，但本次測驗只有讓兩試題之難度相似而非為相似試題，所以在此不適用折半與 Rulon 信度估計測驗之信度，此時進行運算只為了讓教師們作為參考之用。

(三)、庫李信度

庫李信度主要在探討試題間是否大都為同質性試題。本測驗使用庫李 21 號公式計算出測量的信度，其運算公式如公式(2-12)，計算結果如表 4-4 所示。由庫李信度值可發現屬於偏低之情況，主要原因在於本次測驗所測量之概念有四個單元，所以有存在非完全同值性之試題，且雖屬同值性試題但其難度皆有差異，如依單元去做分析又因試題過少，庫李信度在進行之校正項會有較大之影響。如要解決庫李信度過小問題，可在下次編製測驗之時將各單元分類出來再分別進行測驗，這樣將可提高庫李信度之值。

(四)、測量標準誤

本測驗根據 z 分數轉換所得之複本信度係數與前測各標準差計算出各個測驗之測量標準誤(Linn, 2000)。其公式為：

$$SE_{meas.} = S_X \sqrt{1-r} \qquad (4\text{-}1)$$

$SE_{meas.}$ 為測量標準誤， S_X 為測驗標準差， r 為測驗之信度

若測驗之信度越高，則測量之標準誤將會越小，因此測量標準誤的大小，可以推之測驗的信度，及解釋個別分數與誤差的關係。本測驗之測量標準誤可見表 4-3。

表 4-3　重測複本信度與測量標準誤

樣本群體	整體學生	甲高中	乙高中
原始分數之重測複本信度	0.329	0.595	0.391
依照各組進行Z分數轉換之複本信度	0.536	0.557	0.391
測量標準誤	1.794	1.787	1.770

表 4-4　本測驗之內部一致性信度

分析對象	折半信度	Rulon 信度	K-S 信度
前測	0.457461	0.457388	0.44
後測	0.6297	0.6294	0.6417
甲高中後測	0.414636	0.414565	0.41
乙高中後測	0.48868	0.488644	0.592

二、效度研究

本測驗針對內容效度、效標關聯效度的同時效度(concurrent validity)與預測效度(predictive validity)分別進行效度之運算，以說明本測驗效度考驗之結果。

(一)、內容效度

本測驗敦請資深高級中學數學科教師、數學系教授與研究生數人針對本測驗之試題，配合雙向細目表如表 3-4，進行試題內容效度之考驗。研究者經各專家之指正與接受各專家之寶貴意見，進行多次的修改，經認為內容效度頗佳，方才進行測驗。在早期，許多專家學者認為內容效度是無法用一個精確的數量來表示其大小，其實不然，有些學者及建議利用內容信度取代內容效度(Mehrens ＆Lehmann, 1991)，以獲得內容效度的量化分析。Aiken (1985) 提出了內容係數的計算公式，教師可自行查閱參考。

本次測驗也敦請六位高級中學教師針對整份試卷進行表面效度評估，以 0 至 100(0 表示整份試卷完全無法測驗出學生之能力，100 表示整份試卷可以 100％的測驗出學生的能力)作為評分基準，經過整體平均所得到之值為 82.5。表示整份試卷大致可測驗出學生之能力。

(二)、效標關聯效度

1、同時效度

利用本測驗之結果，分別與最近學校平時考數學成績、與此份試卷相同範圍考試成績、最近學校月考數學成績、及學校模擬考試成績求取相關，以驗證測驗之同時效度。其各同時效度值如表 4-3 所示。有表 4-3 可以發現與試卷相同範圍之信度值為最高，為 0.596，最近的平時考試成績為其次，其餘大範圍考試的信度值皆只介於 0.3~0.47

之間。依照實際觀點來看得卻非常之符合，唯讀信度值有點較低之情形，其主要原因在於試題屬選擇題試題，或測驗之時間與實際教學已有一段距離。

2、預測效度

利用本測驗之結果，與民國九十一年之學生學科能力測驗成績求取相關，作為本次測驗之預測效度，如表 4-5。不過此次預測效度屬於偏低，主要原因在於此次測驗試題只屬於學科能力測驗範圍之其中一章，故無法真正了解學生在大範圍時候之表現，如要真正的求取預測效度，可將測驗出題範圍擴大成與學科能力測驗範圍相同，再進行分析。

表 4-5 本測驗之效標關聯效度

校名	類組	學科能力測驗成績	最近學校平時考試成績	相同範圍成績	最近學校月考成績	學校模考成績
全體學生		0.375				
甲高中	自然組	0.312			0.463	0.307
甲高中	社會組	0.305			0.466	0.293
乙高中	自然組	0.413	0.509	0.596		

三、建立各組常模

本測驗在建立常模時，測驗時間皆以 50 分鐘為限制，分別對各組與整體分數檢定是否為常態分佈，發現各分數大致滿足常態分佈之假設，可見附錄三。再運算出各平均數與標準差，建立各組與全體學生之常模，見表 4-5。

表 4-6　各組與全體學生之平均數與標準差

學校	組別	統計量	前測	後測
全體學生		平均數	9.20	8.64
		標準差	2.63	3.24
甲高中	自然組	平均數	9.92	8.03
		標準差	2.88	2.59
甲高中	社會組	平均數	8.67	6.67
		標準差	2.21	1.94
乙高中	自然組	平均數	8.68	11.7
		標準差	2.41	3.09

第三節　學生問題表分析

學生問題表分析主要在探討如何使用一份好的試卷題目，在經過學生測驗作答之後，利用學生們的作答反應組型來讓教師了解各個學生的學習狀況，與各試題的出題情況。利用第二章第四節所提之編製過程(可參考附錄一)，針對全體與各組分別製作其學生問題表(*S-P* 表)，或利用試題分析軟體-TESTER for Windows(余民寧,民 91)加以製作與分析。

一、差異係數

利用公式(2-17)進行兩次測驗各組與全體之差異係數，計算結果如表 4-7 所示。由表可發現差異係數皆介於 0.50~0.63 之間，表示測驗沒有包含太多種異質，整份試卷試題仍屬於穩定狀態。

表 4-7　各組與全體之差異係數

	全體	甲自然組	甲社會組	乙自然組
前測	0.55	0.54	0.57	0.63
後測	0.51	0.50	0.56	0.60

二、注意係數

利用 TESTER 程式分別運算出各試題之問題注意係數與各學生之學生注意係數，如附錄四所示。當注意係數小於 0.5 時，表示該試題或學生之反應組型的不尋常情況較不嚴重，屬於可以接納之範圍，若大於 0.5 則表示不尋常情況已屬於嚴重之狀態，教師應該加以注意。以下針對試題注意係數與學生注意係數分析再加以討論：

(一)、問題注意係數

根據問題注意係數與各試題的答對人數百分比，可以將每個試題的屬性分類為四個區域，圖 4-1 為後側各試題的屬性分部，可與圖 4-2 進行對照，即可了解各試題的狀況為何。

試題注意係數取之前所提之 0.5 作為診斷試題功能之分界，試題答對百

分比方面，因為此次試題有偏難之傾向，所以，以 40%作為試題的分界。其實不管是百分比或者是注意係數的分界方面接沒有一定的規定，教師可以依照自行編製試題的難度，與學生作答反應的情況加以調整，並非一定需要依照上列圖表所示。

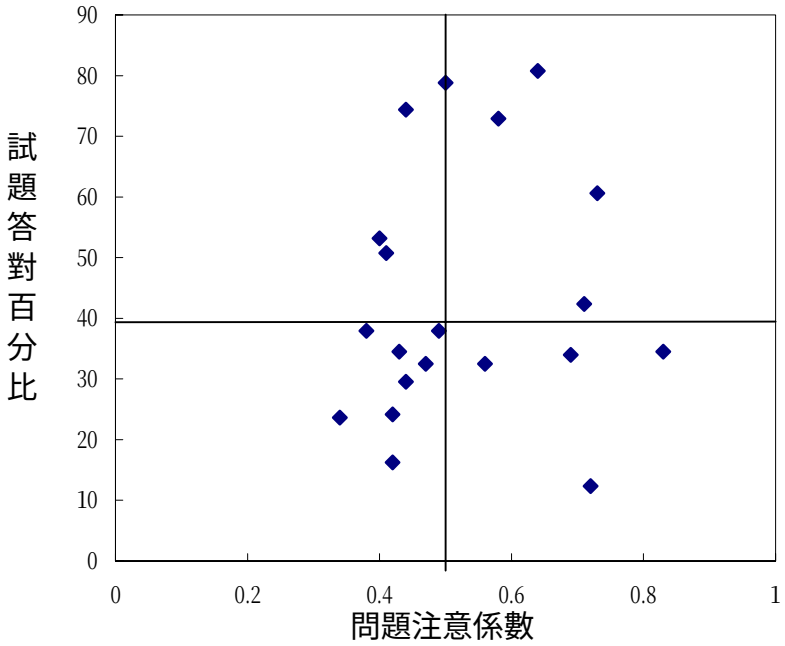


圖 4-1　後側試題屬性分佈圖

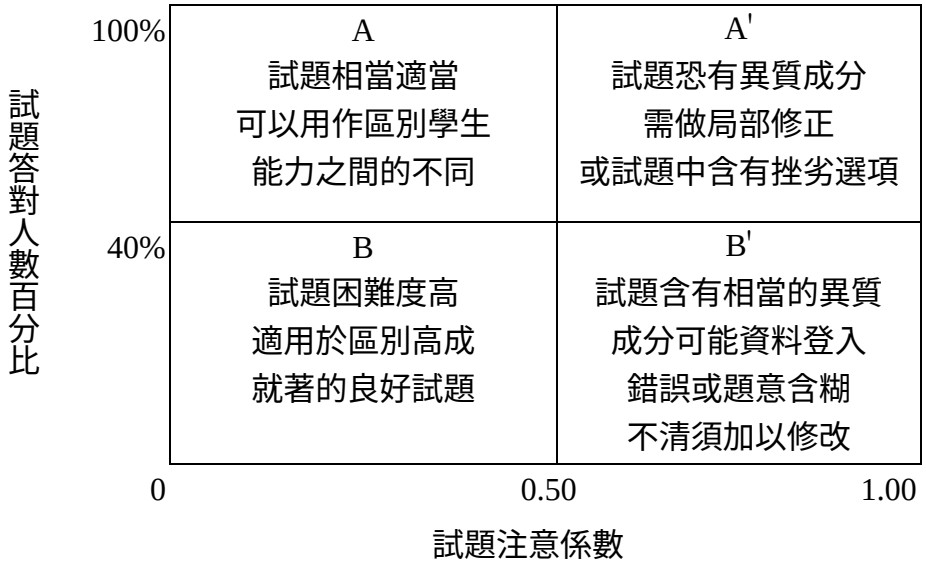


圖 4-2　試題屬性診斷分析圖

(二)、學生注意係數

利用學生後側注意係數與學生得分百分比，繪畫出學生的診斷分析圖，如圖 4-3。可將學生分為六種類型其對照類型可參考圖 4-4 所示。

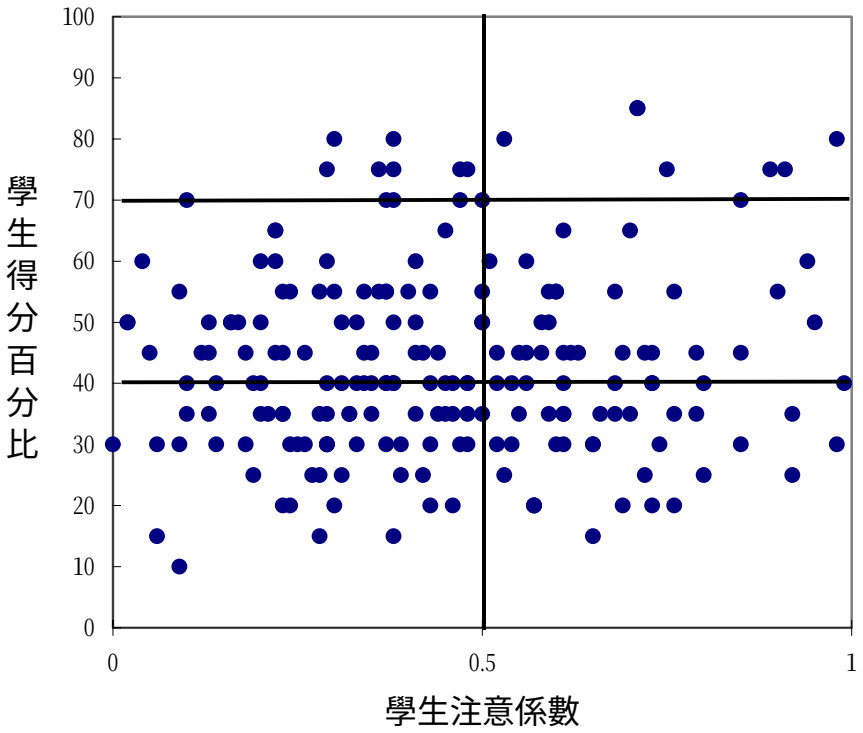


圖 4-3　後側學生屬性診斷分佈圖

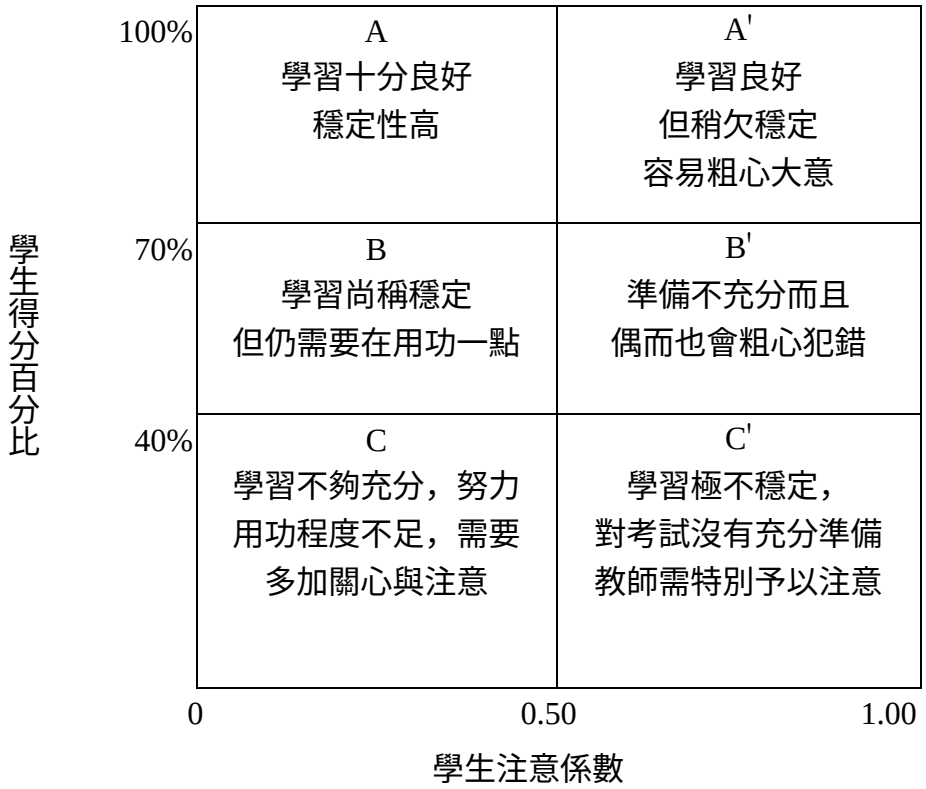


圖 4-4　學生屬性診斷分析圖

學生注意係數仍然以 0.5 作為診斷學生學習類型之分界，學生得分百分比方

面，則以 70%與 40%作為學生類型的分界。與前述相同，教師仍可依照編製試題的難度，與學生作答狀況加以調整，上圖只作為一種分類參考之用。

第四節　各類統計檢定

一、性別檢定

為了進一步了解高級中學學生之數學能力是否會因性別不同而有所差別，研究者將各學生性別另為另一變數 X_{sex}，若為男生則變數值為 0，為女生則變數值為 1，利用 SPSS 程式中之 Pearson 相關進行運算為 0.004。且再利用研究之樣本為分析對象，利用 T 檢定，檢定性別之間的差異。分析結果 *t* 值等於-0.063，表示性別對於本測驗並無顯著之差異。

二、類組與校系差異檢定

利用平均數檢定，分別檢定各組在前、後測時的成績差異如表 4-8，並檢定兩次測驗中各組之間否有顯著差異，其結果如表 4-9。研究者建議請勿使用 SPSS 程式中之比較平均數來作檢定，主要原因在於使用 SPSS 之檢定時，需樣本個數相同時才能進行檢定。在本測驗之樣本數皆不相同，教師可利用 S-plus 程式或者參考 Montgomery (2001 ,P47)所著實驗設計專書中的檢定方式直接進行運算。

由表 4-8 可發現前、後兩次測驗皆有明顯之差異，乙高中經過複習檢討後有明顯之進步，但未檢討之甲高中反而呈現分數降低之現象，所以可說明經過複習檢討事一件很重要的事，之後會利用模型再次說明其之間的差異。由表 4-9 可知，在前測中，甲高中自然組學生程度較其他兩組來的高，但經過一段時間未練習則可發現成績有明顯的降低，而乙學校恰好相反，在前測中與甲高中社會組的程度相似，但經過複習後反而成績皆大幅提昇。

表 4-8　前測與後測平均數差異之顯著性檢定

學校組別	統計量	平均數	標準差	人數	t 值
甲高中自然組	前測	9.92	2.88	104	4.759
	後測	8.03	2.59	86	
甲高中社會組	前測	8.67	2.21	79	5.691
	後測	6.67	1.94	61	
乙高中自然組	前測	8.68	2.41	66	-5.940
	後測	11.7	3.09	56	

表 4-9　各組平均數差異之顯著性考驗

T 檢定	甲高中自然組		甲高中社會組		乙高中自然組	
	前測	後側	前測	後側	前測	後側
甲高中自然組			3.322	3.639	3.027	-7.362
甲高中社會組	3.322	3.639			-0.026*	-10.439
乙高中自然組	3.027	-7.362	-0.026*	-10.439		

三、建立模型，檢定複習教學與未複習教學之差異

此次測驗在前測結束之後，乙學校及有馬上進行檢討與複習，並於進行後測之前一個月內也曾複習到相同範圍之內容，但甲學校不管在前測結束後或後測之前皆未加以檢討複習，導致後側分數在兩校之間呈現明顯之差異。觀察圖 4-5 與圖 4-6，本研究利用迴歸方式建立模型，希望能藉此加以說明複習過後的顯著差異。迴歸模型如下：

$$E(Y)=\beta_0+\beta_1 X+\beta_2 U+\beta_3 XU+\beta_4 W \qquad (4\text{-}1)$$

$Y=(y_1,y_2,\cdots,y_i,\cdots,y_n)'$， $X=(x_1,x_2,\cdots,x_i,\cdots,x_n)'$， $U=(u_1,u_2,\cdots,u_i,\cdots,u_n)'$， $XU=(x_1u_1,x_2u_2,\cdots,x_iu_i,\cdots,x_nu_n)'$， $W=(w_1,w_2,\cdots,w_i,\cdots,w_n)'$， y_i 表各學生後側成績，x_i 表各學生前測成績，$u_i=\begin{cases}0\text{,為甲高中之學生}\\1\text{,為乙高中之學生}\end{cases}$， $w_i=\begin{cases}1\text{,為男學生}\\0\text{,為女學生}\end{cases}$。

利用逐步回歸方式檢定最適模型，逐步過程如下：

表 4-10　逐步迴歸變異數分析表

變數	F 值	P-value
一、選擇 XU 變數		
常數, X	24.454	<0.005
常數, U	103.698	<0.005
常數, XU	122.392	<0.005
常數, w	3.455	0.065
二、加入 X 變數		
常數, XU , X	96.935	<0.005
常數, XU , U	60.935	<0.005
常數, XU , W	60.901	<0.005
三、加入 U 變數		
常數, XU , X , U	72.809	<0.005
常數, XU , X , W	64.382	<0.005

四、剔除 XU 加入 W		
常數, X , U , W	73.261	<0.005
五、剔除 W		
常數, X , U	109.738	<0.005

表 4-11　迴歸係數顯著性檢定

變數	β 估計值	T 值	顯著性
一、			
常數	7.462	35.699	4.09×10^{-20}
XU	0.506	11.063	4.09×10^{-20}
二、			
常數	3.518	5.683	4.62×10^{-8}
XU	0.509	12.294	4.18×10^{-20}
X	0.434	6.694	2.14×10^{-10}
三、			
常數	2.229	3.186	0.002
XU	-0.023	-0.150	**0.881**
X	0.563	7.770	3.768×10^{-13}
U	4.925	3.600	0.0004
四、			
常數	2.392	3.753	0.0002
X	0.562	8.788	4.487×10^{-16}
U	4.801	12.974	4.274×10^{-20}
W	-0.275	-0.819	**0.414**
五、			
常數	2.277	3.665	0.0003
X	0.558	8.759	5.393×10^{-16}
U	4.727	13.190	4.181×10^{-20}

1.由表 4-10(一)可知加入 XU 變數時，其 F 值為最大，所以首先我們選取 XU 變數進入，利用 T 檢定，如表 4-11(一)，檢定變數係數並無需要省略之變數。

2.隨後分別加入 X、U 與 W 變數，由表 4-10(二)可發現選取 X 變數加入為優，再利用 T 檢定檢定是否有需剔除之變數 4-11(二)，發現皆須保留。

3.將 U 與 W 變數加入並檢定個變數係數如表 4-10(三)，選取 U 變數加入較適當，再由表 4-11(三)發現 RX 變數是可將其剔除。

4.將 W 變數在加入並利用 4-11(四)檢定，發現是可以剔除的變數。

5.最後，檢定各變數係數，如 4-11(五)，發現並沒有需要再刪除之變數。

可得最適模型為：$E(Y)=\beta_0+\beta_1X+\beta_2U$　　(4-2)

由上述可發現 U 變數為一項重要之變數，若剔除後會影響到整個模型，而由 U 的係數可發現之值為正，代表乙高中之學生成績真的有明顯

的分數題高，表示經過複習教學的確是有助於學生的學習。

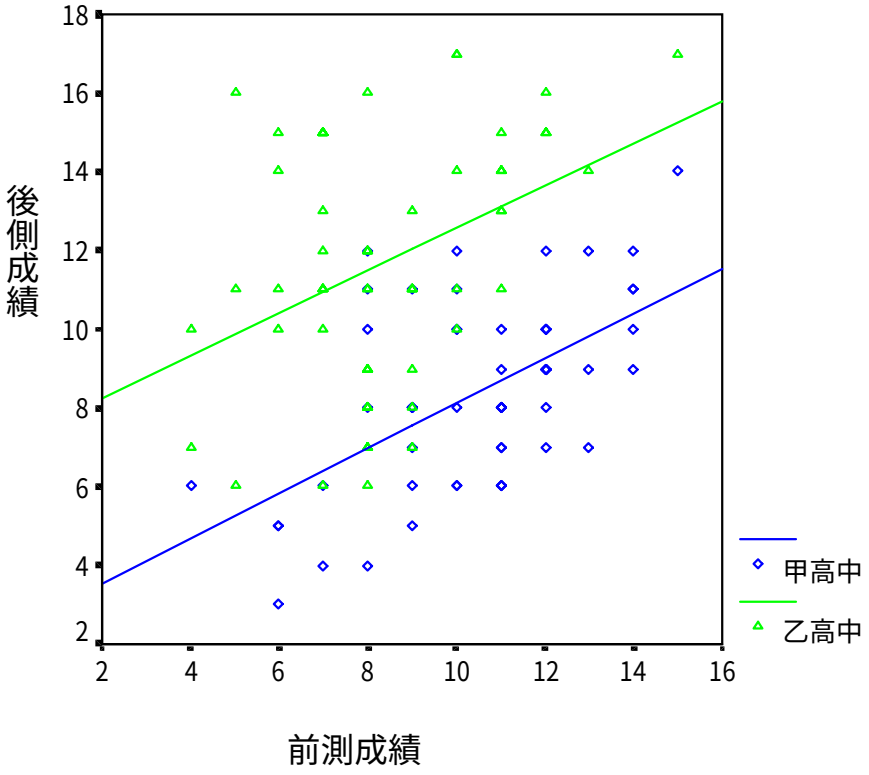


圖 4-5　兩校前後測驗成績分佈圖

我們也可以利用模型來檢定成績與性別是否有顯著關係，由逐步迴歸發現 W 變數皆會被剔除，屬於不影響成績之變數。也就是說，性別的差異不會因此影響到成績的高低。

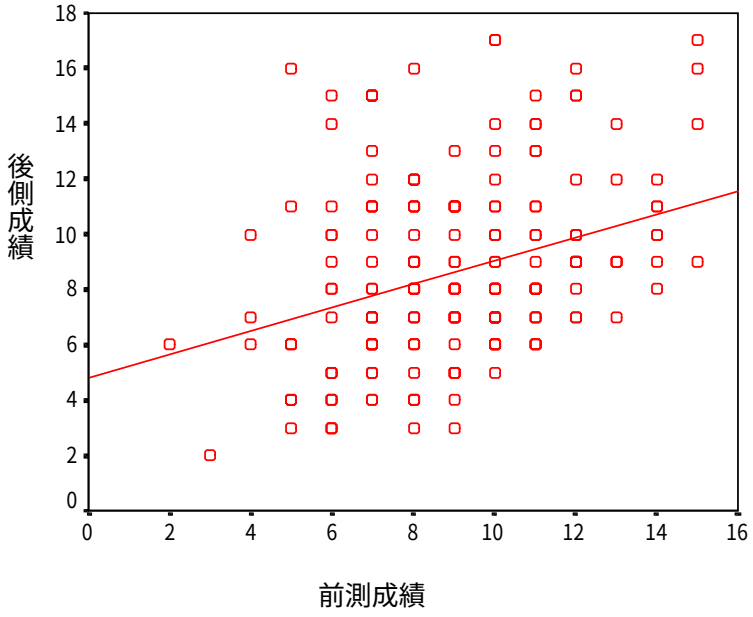


圖 4-6　全體前後測驗成績分佈圖

第五章　結論與建議

要提供學生適當的教育，必須先瞭解學生能力的優缺點，而要瞭解學生能力的最便捷方法莫過於利用測驗的方式。而在測驗的各種方式裡又以紙筆的試題卷測驗最常被教師們所使用，所以學習如何去編製一份適當的試題，常是每位教師的一項重要的任務。數學科是最有系統的學科，具有結構性與次序性，必須按一定的步驟學習，若能提早發現學生的學習問題所在，必將影響其後期的學習效果。當然，不管任何科目，若都能利用測驗方式來了解到學生的問題，那一定就能幫助教師改進教學方式，以提昇整體的教學效果。

本研究針對高中三年級學生學習機率統計概念相關能力，製作測驗試題卷，並作相關資料收集，以分析試題的各項指標，作為判定測驗適當之依據。本章根據第四章結果與討論，可以得到以下結論，並根據結論提出一些建議。

第一節　研究結論

一、研究樣本與編製依據

本研究選取在高雄市中，學生平均程度屬於中等之學校兩所，共七個班級 249 名學生，作為本次研究的樣本。針對樣本資料進行試題分析、信度分析、效度分析、學生問題表分析與各類統計檢定。

以高級中學數學甲上－機率統計(二)做為本次測驗的編製依據，針對內容中的條件機率、獨立事件、變異係數與相關係數四個部分，加以研究編製。本次測驗編製目標，對於學生方面，希望受測學生能利用此次測驗了解許多機率統計的基礎原理，並幫忙釐清學生的錯誤概念；在教師方面，能利用此次測驗了解學生之間的差異，知道學生的問題所在，進而改進自身教學方式。也期許教師能因本研究而對教育測驗的分析方式有進一步的了解。

二、結果摘要

(一)測驗題項

1.難度：兩次試題難度介於 0.11~0.85 之間(見附錄二)，約有 15%之試題難度超過 0.7，屬於較容易試題；有 25%之難度低於 0.3，屬於較困難之試題；其餘試題皆屬於難度適當之試題，由表 4-2 平均難度約 0.45 可知整份試卷屬於較偏難之試題。教師可在下次測驗時將一些較難之試題稍作修正或者在教導時能更釐清其中觀念。

2.鑑別度指標：兩次測驗之試題鑑別度節介於 0.03 到 0.64 之間。平均鑑別度為 0.37。鑑別度指標求得方式是由高分組答對百分比減去低分組的答對百分比所得到，由數據分析可發現兩次試卷除了少部分試題要

在修改外，其餘大都屬於鑑別度良好的試題。

(二)測驗信度

1.重測複本信度：如利用整體原始分數進行運算其值為 0.329，屬於過低之信度值，主要原因在於兩所高中在後側實施之前，已有複習與未經過複習之差異，導致信度值的偏低。若只依甲高中來說明，因為前測與後側皆未經過複習所以可發現其信度值則明顯較高，為 0.595。對於選擇試題屬於試題信度值可接納之範圍。本次研究為了進行分析整體信度，又要排除是否複習造成成績的疑慮，我們利用了 Z 分數轉換，將各組依照各組的平均數與變異數進行轉換後，再利用轉換之值運算其信度，所得為 0.536，如表 4-3，屬於信度值尚可之選擇題試卷。

2.折半信度：本測驗之折半信度與 Rulon 信度如表 4-4 所示，介於 4.15 至 4.89 之間，為偏低之折半信度，主要原因在於奇數試題與偶數試題的試題類似性太底而造成，而後測信度值偏高之原因，主要還是受複習與未複習的關係所影響，本測驗並非適用於折半信度，在此只作為參考之用。

3.庫李信度：庫李信度在未複習的前測與後側甲高中裡皆為較低的數值為 0.44 與 0.41，但在經過複習後的乙高中信度卻為 0.592，顯示在未了解題目觀念之時，學生的作答，反應了試題之間的同值性偏低(因可能利用猜測作答)，而在經過講解複習後，卻反應出試題之間確實存在了一定的同值性，以測量類似的概念。

(三)測驗效度

1.內容效度：本測驗敦請資深高級中學數學科教師、數學系教授與研究生數人針對本測驗之試題，配合雙向細目表進行試題內容效度之考驗。除了專家判斷外，我們也可以利用複本信度來作為估計內容效度之用，其內容效度為 0.536。在製作內容效度同時，我們邀請高級中學教師數名針對試卷進行表面效度分析，表面效度值為 0.825，具有頗高的表面效度。

2.效標關聯效度：同時效標效度約介於 0.3 至 0.6 之間。由表 4-5 可發現相同範圍之校度值為最高，為 0.596，最近考試成績為 0.509，顯示本試卷的效度值還算適當，而月考成績與模擬考成績之效度值偏低，主要原因在月考成績所考範圍與試題完全沒有關聯，而模考成績為大範圍考試，所以皆不適合用來作為效度之效標。而學科能力測驗也因為大範圍考試與本次測驗範圍不同，所以運算出來的預測效度也屬於過低之傾向。

(四)學生問題表

1.差異係數：本次測驗差異係數值皆介於 0.50 至 0.63 之間，如表 4-7，皆屬可接受差異之範圍，表示試卷沒有包含太多的異質，屬於穩定狀態。

2.問題係數：兩次測驗的試題注意係數介於 0.34 至 0.99 之間；學生問題注意係數介於 0 到 1.61 之間。表示有許多同學概念不清。其中約有 10%之學生屬於學習良好之學生(指圖 4-4 中 A 類與 A'類學生)，38%之學生屬於學習穩定但仍需要用功之學生(B 類學生)，20%學生準備不是很充分且偶而會粗心犯錯(B'類學生)，其餘 32%學生對於此次試卷之概念屬非常不清楚，或是針對此次試卷隨意作答，需要再加以注意，其詳細資料可見附錄三。

(五)各類統計檢定

1.性別檢定：利用 Pearson 相關運算得值為 0.004，顯示在進入高級中學後之學生，性別與測驗分數已沒有明顯的差異。

2.各校類組差異檢定：利用平均數檢定，如表 4-9。在前測時，甲高中自然組明顯較甲高中社會組與乙高中自然組來的高，檢定值分別為 3.322 與 3.363，而甲高中社會組與乙高中自然組則顯示出沒有明顯差異。但因乙高中經過複習後，在後側成績可明顯的發現乙高中皆較其他兩組來的高上許多，顯示複習教學是應該加以進行。

3.建立模型檢定差異：建立迴歸模型，如 4-1，加以檢試性別與成績差異和複習與非複習差異。利用逐步迴歸檢定的到最適模型，如 4-2 所示，發現性別之變數(W)需要被刪除，表示性別與成績之間真的沒有顯著差異，而是否進行複習教學變數(U)則需要被加以考慮進入，且其係數為正值，如表 4-11(五)所示，表示經過複習教學後，真的能提高學生的成績。

第二節　研究建議

一、測驗使用之限制

(一)本測驗是以高級中學數學甲為出題依據，所以在進行測量時，應選取已經學習過機率統計相關內容的學生作為測驗對象。不適合使用本測驗來分析學生是否具有機率統計之先備知識。

(二)本測驗主要在了解學生對於機率統計方面的概念知識認知是否清楚，而不適用於在考驗學生的計算能力，如教師想了解學生的運算能力，應利用其他的試卷再加以分析。

(三)本測驗因為非隨機抽樣，所以測驗結果所求得之常模只適用於甲、

乙兩間高中，而非適用於全國各高級中學，需請教師特別注意。

二、本次研究後之建議

(一)本次測驗因考量學校教師在編製試題卷時，常無法先行實施預測而需直接針對學生加以測驗，故研究建議，教師可聯繫其他教師來一起進行試題的測驗分析，先針對部分同學進行預測後，再為其他同學進行測驗。在本測驗中因受到樣本取樣的限制，而未進行預測，故對於試題的選取更加嚴格，煩請各類專家針對試題部分仔細的挑選，但仍不屬於良好之測驗方式。

(二)本測驗的相關分析繁多。教師在自行進行分析時，並非真需將本研究中所提到之各類分析方式皆進行分析後才能判定試題的優良與否。教師可自行選擇適當之方式加以分析即可，本研究建議至少需進行 S-P 表分析，以了解學生的作答情況。

(三)教師再對學生進行測驗時，應讓學生重視此份測驗，否則容易造成學生隨意作答或胡亂猜題現象，導致試題分析的不準確。

三、未來研究之建議

(一)針對其餘機率統計試題，能利用隨機選取測驗學生樣本，進行各試題之分析，將分析結果加以修改，製作最適當之問答方式，並將試題難度加以分類，最為教學之題庫。

(二)期望建立廣泛全國或全高雄市地區的常模參照，以供各地區的教師作為參考。

(三)配合電腦網路教學，將分析過後之試題，依隨機方式加以出題，並配合常模參照比對出學生的學習成果。

(四)製作非選擇題部分試題，並利用其他適當分析工具加以分析，而非只有利用選擇試題進行測驗。

參考文獻：

Ahmanan, J. S. ＆ Glock, M. D. (1981). Evaluating Student Progress：Principles of Tests and Measurements.(6th ed.) Allyn and Bacon：Boston.

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. Educational and Psychological Measurement, 45, 131-142.

Allen, W. J. ＆ Yen, W. M. (1979). Introduction to measurement theory. Brooks /Cole：Monterey, C.A..

American Psychological Association, American Education Research Association, ＆ National Council on Measurement in Education.(1985). Standards for educational and Psychological testing. American Psychological Association：Washington, D.C..

Anastasi, A. (1988). Psychological Testing.(6th ed.) Macmillan Publishing：New York.

Benbow, C. P. ＆ Stanley, J. C. (1983). Sex difference in mathematical reasoning ability ：more facts. Science, 222, 1029-1031.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3, 296-322.

Brueckner, L. J. ＆ Band, G. L. (1955). The diagnosis and treatment of learning difficulties. Appleton-Century-Crofts：New York.

Cronbach, L. J. (1971). Test validation. In R.L. Thorndike, Educational measurement (2rd ed.). American Council on Education：Washington, D.C..

Cronbach, L. J., Gleser, G. C.,Nanda, H., ＆ Rajaratnam, N. (1972). The dependability of behavioral measures: Theory of generalizability for scores and profiles. John Wiley ＆ Sons：New York.

Ebel, R. L. ＆ Frisbie, D. A. (1979). Essentials of educational measurment(3rd ed.). Prentice-Hall：Englewood Cliffs, N.J..

Fan, C. T. (1952) Item analysis table. Educational Testing Service：Princetion, N.J.

Fennema, E., Carpenter T. P. ＆Lamon S. J. (1991) Integrating research on teaching and learning mathematics. State University of New York Press：Albany, NY.

Flanagan, J. C. (1937). A proposed procedure for increasing the efficiency of objective tests. Journal of Educational Psychology, 28,17-21.

Gardner, H. (1983). Frams of mind: The theory of multiple intelligences. Basic Books: New York.

Gronlund, N. E. (1976). Measurement and Evaluation in Teaching.(3rd ed.). Appeton -Century-Crofts：New York.

Gronlund, N. E. (1993). How to make achievement tests and assessments(5th ed.). Allyn ＆ Bacon：Boston.

Gulliksen, H. (1950). Theory of mental test. John Wiley ＆ Sons：New York.

Guttman, L. A. (1944). A basis for scaling Qualitative data. Amreican Sociological Review, 9,179-190.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., ＆ Black, W. C. (1995). Multivariate data analysis with reading(4th ed.). Prentice-Hall：Englewood Cliffs, NJ.

Haladyna, T. M. (1994). Developing and validating multiple-choice test items. Lawrence Erlbaum Associates：Hillsdale, N.J..

Hopkins, K. D., Stanley, J. C.,＆ Hopkins, B. R. (1990). Educational and

Psycho-logical measurment and evaluation(7th ed.). Prentice Hall：Englewood Cliffs, N.J..
Jacklin, C. N. (Ed.) (1992). The Psychology of gender. New York University Press：New York.
Janes, C. J. (1992). An investigation of differences in verbal ability and mathematical ability among talented male and female high school students. UMI：Ann Arbor, Mich.
Kelly, T. L. (1939). The Selection of Upper and Lower Groups for the Validation of Test Items. Journal of Educational Psychology, 30,17-24.
Kerlinger, F. N. (1986). Foundations of behavioral research (3th ed.). CBS Publishing Japan：New York.
Kubiszyn, T. & Borich, G. (1996). Educational Testing and Measurement. Harper Collins College Publishers：New York.
LeMahieu, P. G., Gitomer, D. H., & Fresh, J. T. (1995). Protfolios in large-scale assessment: Difficult but not impossible. Educational Measurement: Issues and Practice, 14(3), 11-28.
Lindquist, E. F. (Ed.) (1951). Educational measurment. American Council on Education：Washington, D.C..
Linn, R. L. (Ed.) (1989). Educational measurment(3th ed.). American Council on Education：Washington, D.C..
Linn, R. L.(1994). Performance assessment:Policy promises and technical measurement standards. Educational Researcher, 23(9), 4-14.
Linn, R. L. & Gronlund, N.E.(2000). Measurement and Assessment in Teaching. Merrill：Upper Saddle River, N.J..
Liu, H. J. (1998). A cross-cultural study of sex differences in statistical reasoning for collage students in Taiwan and the United States. Doctoral dissertation, University of Minnesota, Minneapolis.
Lord, F. M. (1980). Applications of item response theory to practical testing problems. Lawrence Erlbaum Associates：Hillsdale, N.J..
Maccoby, E. E. & Jacklin, C.N. (1974). Psychology of sex differences. Stanford University Press：Palo Alto, CA.
Mehrens, W. A. & Lehmann, I. J. (1991). Measurement and evaluation in education and psychology. Holt,Rinehart & Winston：New York.
Montgomery D. C. (2001). Design and analysis of experiments. John Wiley：New York.
Paulsen, K. & Johnson, M. (1983). Sex role attitude and mathematical ability in 4th-, 8th-, and 11th grade students from a high socioeconomic area. Development Psychology, 19, 210-214.
Payne, D. A. (1974). The assessment of learning: Cognitive and Affective. Heath and Company：Lexigton, Massachusetts D.C..
Philippa, P. & Norma, G. (1984). Do spatial skills contribute to sex differences in different types of mathematical problems？ Journal of Educational Psychology, 76,678-689.
Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split halves. Harvard Educational Review, 9,99-103.
Sato, T. (1969). A method of analyzing data gathered by the Response Analyzer for diagnosis of student performance and the quality of instructional sequence.

Proceedings of IECE of Japan annual conference S12-1.
Sato, T. (1975). The construction and interpretation of S-P table. Meiji Tosho：Tokyo.
Sato, T. (1985). Introduction to student-problem curve theory analysis and evaluation. Meiji Tosho：Tokyo.
Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271-295.
Sharon, S. ＆Zalman, U. (1983). Geometry proof writing：A new view of sex differences in mathematics ability. American Journal of Education, 91,187-201.
Suen, H. K. (1990). Principles of test theories. Lawrence Erlbaum Associates：Hillsdale, N.J.
Thorndike, R. L. (Ed.) (1971). Educational measurment(2th ed.). American Council on Education：Washington, D.C..
Thorndike, R. L. ＆ Hagen, E. P.(1977). Measurement and evaluation in psychology and education.(4th ed.) Wiley：New York.
何英奇 (民國 78 年) 精熟學習策略配合微電腦化 S-P 表分析診斷對學生學習效果的實驗研究。師大教育心理學報，第 22 期，191-214。
余民寧、陳嘉成、潘雅芳（民國 85 年）。 概念構圖法在測驗教學上的應用。測驗年刊，43 輯，195-212 頁。
余民寧 (民國 91 年) 教育測驗與評量—成就測驗與教學評量(第二版)。台北：心理出版社。
林清山 (民國 84 年) 心理與教育統計(第五版)。台北：東華書局。
林清山 (民國 69 年) 多變量分析統計法。台北：東華書局。
教育資料文摘 (民國 87 年) 中小學生最怕數學最愛體育。教育資料文摘，第 250 期，191-192 頁。
郭生玉 (民國 85 年) 心理與教育測驗(第十版)。台北：精華書局。

附錄一　S-P 表製作方式

(選取乙學校 3 年 12 班 15 位學生與前十題試題為例)

(a)S-P 原表

學生座號	試題號碼 1	2	3	4	5	6	7	8	9	10	總分
2	0	1	1	1	0	0	0	1	0	1	5
3	1	1	0	1	1	1	1	1	0	1	8
4	0	1	0	1	0	1	0	0	0	0	3
5	1	1	0	1	1	1	1	1	1	1	9
6	0	1	0	1	1	0	0	0	0	0	3
7	0	1	0	0	0	1	0	1	0	0	3
8	0	1	1	1	0	0	0	1	0	1	5
9	1	1	1	1	1	0	0	1	0	1	7
10	0	1	0	0	1	1	0	0	1	1	5
11	1	0	0	1	0	0	0	1	1	1	5
12	1	1	0	1	0	1	1	1	1	1	8
13	1	0	0	1	0	1	0	1	1	1	6
14	1	0	0	1	1	0	1	1	1	1	7
15	0	1	0	1	0	1	1	1	1	1	7
16	1	1	0	1	1	1	0	1	0	1	7
答對人數	8	12	3	13	7	9	5	12	7	12	88

(b)按學生總分高低，由上往下排列

學生座號	試題號碼 1	2	3	4	5	6	7	8	9	10	總分
5	1	1	0	1	1	1	1	1	1	1	9
3	1	1	0	1	1	1	1	1	0	1	8
12	1	1	0	1	0	1	1	1	1	1	8
9	1	1	1	1	1	0	0	1	0	1	7
14	1	0	0	1	1	0	1	1	1	1	7
15	0	1	0	1	0	1	1	1	1	1	7
16	1	1	0	1	1	1	0	1	0	1	7
13	1	0	0	1	0	1	0	1	1	1	6
2	0	1	1	1	0	0	0	1	0	1	5
8	0	1	1	1	0	0	0	1	0	1	5
10	0	1	0	0	1	1	0	0	1	1	5
11	1	0	0	1	0	0	0	1	1	1	5
4	0	1	0	1	0	1	0	0	0	0	3
6	0	1	0	1	1	0	0	0	0	0	3
7	0	1	0	0	0	1	0	1	0	0	3
答對人數	8	12	3	13	7	9	5	12	7	12	88

(c)按試題答對人數多寡，由左往右排列

學生座號 \ 試題號碼	4	2	8	10	6	1	5	9	7	3	總分
5	1	1	1	1	1	1	1	1	1	0	9
3	1	1	1	1	1	1	1	0	1	0	8
12	1	1	1	1	1	1	0	1	1	0	8
9	1	1	1	1	0	1	1	0	0	1	7
14	1	0	1	1	0	1	1	1	1	0	7
15	1	1	1	1	1	0	0	1	1	0	7
16	1	1	1	1	1	1	1	0	0	0	7
13	1	0	1	1	1	1	0	1	0	0	6
2	1	1	1	1	0	0	0	0	0	1	5
8	1	1	1	1	0	0	0	0	0	1	5
10	0	1	0	1	1	1	1	1	0	0	5
11	1	0	1	1	0	0	0	1	0	0	5
4	1	1	0	0	1	0	0	0	0	0	3
6	1	1	0	0	0	1	1	0	0	0	3
7	0	1	1	0	1	0	0	0	0	0	3
答對人數	13	12	12	12	9	8	7	7	5	3	88

(d)畫出 S 曲線(粗線)與 P 曲線(細線)

學生座號 \ 試題號碼	4	2	8	10	6	1	5	9	7	3	總分
5	1	1	1	1	1	1	1	1	1	0	9
3	1	1	1	1	1	1	1	0	1	0	8
12	1	1	1	1	1	1	0	1	1	0	8
9	1	1	1	1	0	1	1	0	0	1	7
14	1	0	1	1	0	1	1	1	1	0	7
15	1	1	1	1	1	0	0	1	1	0	7
16	1	1	1	1	1	1	1	0	0	0	7
13	1	0	1	1	1	1	0	1	0	0	6
2	1	1	1	1	0	0	0	0	0	1	5
8	1	1	1	1	0	0	0	0	0	1	5
10	0	1	0	1	1	1	1	1	0	0	5
11	1	0	1	1	0	0	0	1	0	0	5
4	1	1	0	0	1	0	0	0	0	0	3
6	1	1	0	0	0	1	1	0	0	0	3
7	0	1	1	0	1	0	0	0	0	0	3
答對人數	13	12	12	12	9	8	7	7	5	3	88

附錄二　試題項目分析

一、前測各試題之分析

		選答人數百分比					難度		鑑別度	
		1	2	3	4	其他	P	Δ	D	γ
1	選項率	0.11	0.3	0.05	0.55	0	0.5645	12.3	0.5806	0.58
	高分組	0.05	0.08	0.02	*0.85	0				
	低分組	0.18	0.45	0.1	0.27	0				
2	選項率	0.09	0.01	0.06	0.84	0	0.8226	9.1	0.2581	0.43
	高分組	0.02	0	0.03	*0.95	0				
	低分組	0.19	0.03	0.06	0.69	0.02				
3	選項率	0.39	0.04	0.55	0.01	0	0.5484	12.5	0.0968	0.10
	高分組	0.37	0.03	*0.6	0	0				
	低分組	0.45	0.02	0.5	0.03	0				
4	選項率	0.11	0.57	0.15	0.15	0.02	0.5403	12.6	0.4355	0.44
	高分組	0.03	*0.76	0.05	0.15	0.02				
	低分組	0.11	0.32	0.32	0.19	0.05				
5	選項率	0.04	0.02	0.05	0.88	0.01	0.8548	8.7	0.0968	0.16
	高分組	0.05	0.02	0.03	*0.9	0				
	低分組	0.06	0.02	0.1	0.81	0.02				
6	選項率	0.18	0	0.01	0.81	0	0.1532	17.6	0.2742	0.54
	高分組	*0.29	0	0	0.71	0				
	低分組	0.02	0	0.02	0.97	0				
7	選項率	0.39	0.13	0.42	0.06	0	0.4355	13.7	0.4839	0.50
	高分組	*0.68	0.08	0.23	0.02	0				
	低分組	0.19	0.16	0.53	0.11	0				
8	選項率	0.08	0.85	0.02	0.06	0	0.8387	8.9	0.1613	0.28
	高分組	0.03	*0.92	0	0.05	0				
	低分組	0.13	0.76	0.02	0.1	0				
9	選項率	0.06	0.25	0.5	0.16	0.02	0.1774	16.8	0.1613	0.25
	高分組	0.02	0.15	0.56	*0.26	0.02				
	低分組	0.15	0.32	0.42	0.1	0.02				
１０	選項率	0.62	0.2	0.14	0.02	0.01	0.1129	17.8	0.0323	0.06
	高分組	0.68	0.21	*0.13	0	0.02				
	低分組	0.61	0.23	0.1	0.02	0.02				

１１	選項率	0.13	0.27	0.54	0.06	0.01	0.2742	15.6	0.3548	0.44
	高分組	0.08	*0.45	0.45	0.02	0				
	低分組	0.19	0.1	0.56	0.13	0.02				
１２	選項率	0.08	0.18	0.46	0.27	0.02	0.2984	15.3	0.3710	0.44
	高分組	0.08	0.11	0.31	*0.48	0.02				
	低分組	0.08	0.24	0.55	0.11	0.02				
１３	選項率	0.49	0.14	0.2	0.15	0.01	0.5242	12.7	0.4032	0.41
	高分組	*0.73	0.1	0.13	0.05	0				
	低分組	0.32	0.15	0.27	0.23	0.03				
１４	選項率	0.22	0.48	0.22	0.06	0.01	0.500	13.0	0.5806	0.57
	高分組	0.1	*0.79	0.1	0.02	0				
	低分組	0.29	0.21	0.37	0.11	0.02				
１５	選項率	0.05	0.28	0.38	0.27	0.02	0.4355	13.7	0.3548	0.36
	高分組	0	0.35	*0.61	0.03	0				
	低分組	0.05	0.24	0.26	0.44	0.02				
１６	選項率	0.09	0.51	0.1	0.29	0.01	0.3065	15.1	0.2581	0.30
	高分組	0.03	0.47	0.06	*0.44	0				
	低分組	0.13	0.5	0.18	0.18	0.02				
１７	選項率	0.24	0.17	0.45	0.12	0.02	0.4194	13.9	0.4839	0.49
	高分組	0.15	0.13	*0.66	0.05	0.02				
	低分組	0.27	0.27	0.18	0.24	0.03				
１８	選項率	0.09	0.27	0.23	0.4	0.01	0.4274	13.8	0.4355	0.45
	高分組	0.02	0.18	0.16	*0.65	0				
	低分組	0.16	0.29	0.31	0.21	0.03				
１９	選項率	0.42	0.23	0.17	0.16	0.02	0.4435	13.6	0.5323	0.53
	高分組	*0.71	0.06	0.13	0.1	0				
	低分組	0.18	0.31	0.24	0.24	0.03				
２０	選項率	0.13	0.63	0.2	0.02	0.02	0.6290	11.6	0.3548	0.39
	高分組	0.08	*0.81	0.1	0	0.02				
	低分組	0.18	0.45	0.31	0.03	0.03				

*表正確答案選項

附錄二　試題項目分析

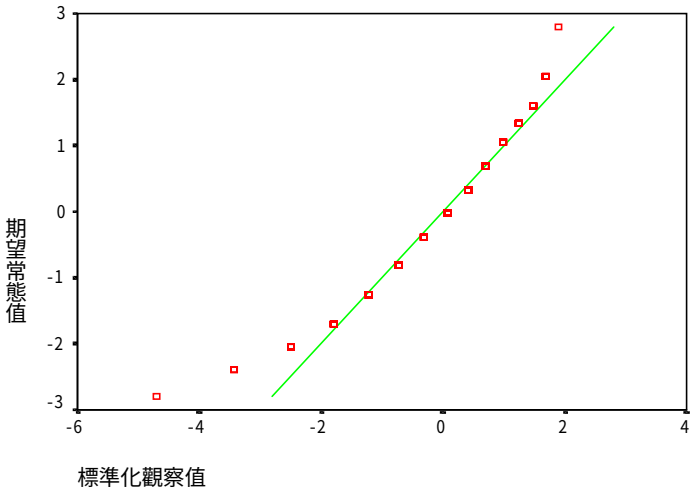
二、後測各試題之分析

		選答人數百分比					難度		鑑別度	
		1	2	3	4	其他	P	Δ	D	γ
1	選項率	0.2	0.38	0.2	0.22	0	0.44	13.7	0.52	0.52
	高分組	0.08	*0.7	0.08	0.14	0				
	低分組	0.26	0.18	0.34	0.22	0				
2	選項率	0.06	0.02	0.11	0.81	0	0.81	9.3	0.22	0.34
	高分組	0.04	0	0.04	*0.92	0				
	低分組	0.14	0.06	0.1	0.7	0				
3	選項率	0.12	0.04	0.08	0.73	0.01	0.12	17.9	0.16	0.35
	高分組	*0.2	0	0.1	0.7	0				
	低分組	0.04	0.08	0.06	0.82	0				
4	選項率	0.06	0.11	0.09	0.73	0	0.73	10.3	0.38	0.48
	高分組	0.02	0.04	0.02	*0.92	0				
	低分組	0.14	0.2	0.12	0.54	0				
5	選項率	0.21	0.61	0.15	0.03	0	0.54	12.6	0.24	0.25
	高分組	0.2	*0.66	0.12	0.02	0				
	低分組	0.4	0.42	0.16	0.02	0				
6	選項率	0.05	0	0.34	0.6	0	0.43	13.8	0.54	0.55
	高分組	0.02	0	*0.7	0.28	0				
	低分組	0.06	0.02	0.16	0.76	0				
7	選項率	0.24	0.58	0.01	0.16	0	0.30	15.5	0.48	0.58
	高分組	*0.54	0.38	0	0.08	0				
	低分組	0.06	0.68	0.06	0.2	0				
8	選項率	0.08	0.04	0.08	0.79	0	0.72	10.4	0.40	0.50
	高分組	0.02	0.04	0.02	*0.92	0				
	低分組	0.18	0.12	0.18	0.52	0				
9	選項率	0.16	0.22	0.29	0.32	0	0.25	16.0	0.38	0.51
	高分組	*0.44	0.2	0.14	0.22	0				
	低分組	0.06	0.32	0.32	0.3	0				
１０	選項率	0.52	0.17	0.3	0.01	0	0.41	14.1	0.54	0.56
	高分組	0.26	0.04	*0.68	0.02	0				
	低分組	0.66	0.18	0.14	0.02	0				

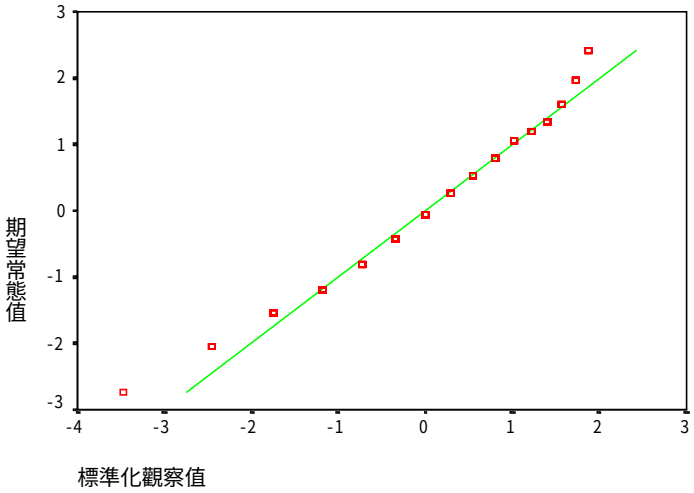
１１	選項率	0.2	0.33	0.4	0.06	0.01	0.36	14.6	0.44	0.48
	高分組	0.16	*0.58	0.22	0.04	0				
	低分組	0.24	0.14	0.54	0.08	0				
１２	選項率	0.1	0.33	0.34	0.22	0.01	0.30	15.1	0.12	0.14
	高分組	0.04	0.42	*0.36	0.18	0				
	低分組	0.24	0.28	0.24	0.22	0.02				
１３	選項率	0.16	0.12	0.36	0.34	0.02	0.33	14.8	0.26	0.29
	高分組	0.22	0.08	0.24	*0.46	0				
	低分組	0.1	0.26	0.42	0.2	0.02				
１４	選項率	0.15	0.32	0.14	0.38	0.01	0.36	14.7	0.56	0.61
	高分組	0.14	0.2	0.02	*0.64	0				
	低分組	0.24	0.44	0.22	0.08	0.02				
１５	選項率	0.24	0.42	0.14	0.17	0.03	0.28	16.0	0.52	0.69
	高分組	*0.54	0.3	0.08	0.08	0				
	低分組	0.02	0.5	0.16	0.28	0.04				
１６	選項率	0.33	0.09	0.46	0.1	0.01	0.39	14.3	0.50	0.52
	高分組	*0.64	0	0.26	0.1	0				
	低分組	0.14	0.2	0.52	0.12	0.02				
１７	選項率	0.08	0	0.16	0.74	0.01	0.71	10.4	0.46	0.57
	高分組	0.04	0	0	*0.94	0.02				
	低分組	0.18	0	0.32	0.48	0.02				
１８	選項率	0.06	0.53	0.17	0.23	0	0.52	12.7	0.64	0.63
	高分組	0	*0.84	0.04	0.12	0				
	低分組	0.12	0.2	0.2	0.46	0.02				
１９	選項率	0.16	0.51	0.15	0.15	0.03	0.56	12.3	0.60	0.60
	高分組	0.1	*0.86	0	0.04	0				
	低分組	0.22	0.26	0.26	0.24	0.02				
２０	選項率	0.42	0.16	0.26	0.15	0.01	0.45	13.5	0.26	0.27
	高分組	*0.58	0.08	0.22	0.12	0				
	低分組	0.32	0.24	0.2	0.22	0.02				

*表正確答案選項

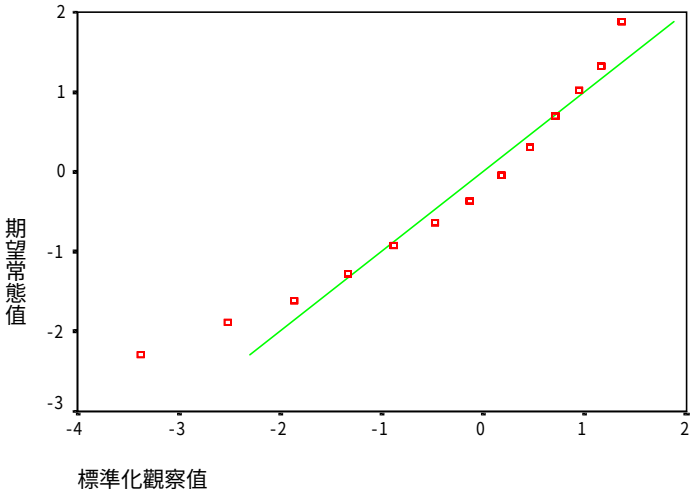
附錄三　常態分配Q-Q圖



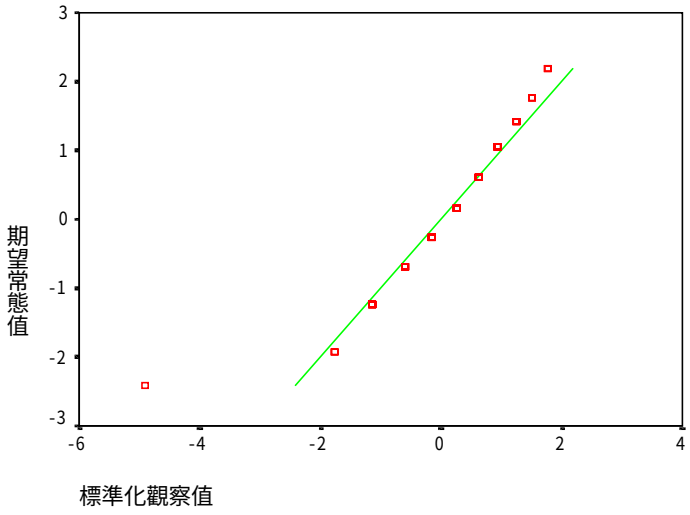
圖一　前測整體學生常態分配Q-Q圖



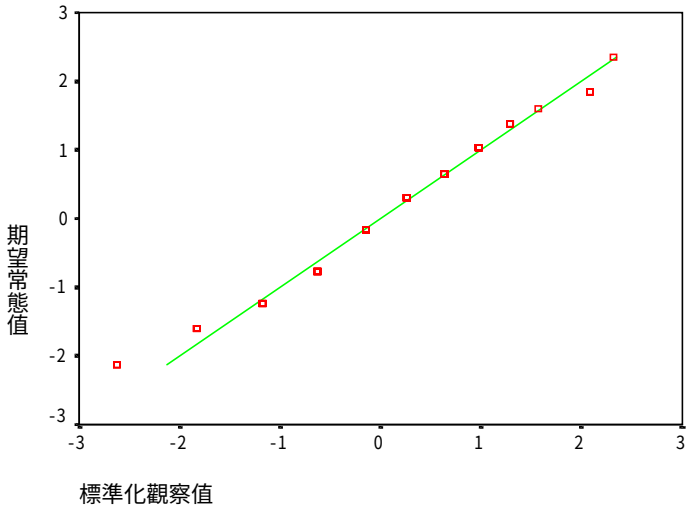
圖二　後測整體學生常態分配Q-Q圖



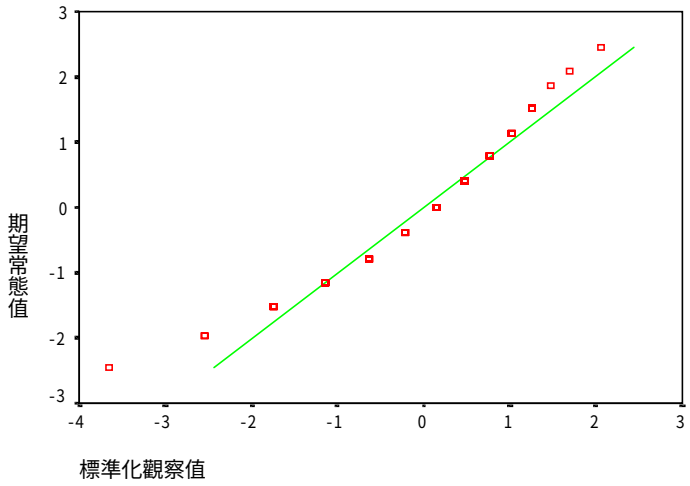
圖三　前測甲高中自然組學生常態分配Q-Q圖



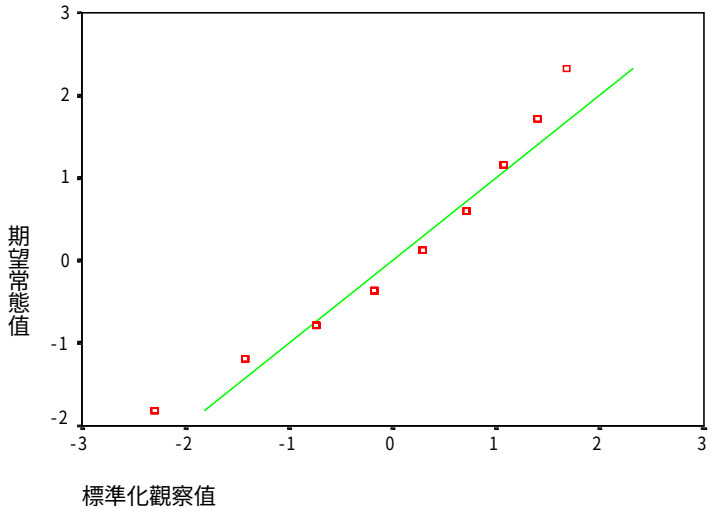
圖四　前測甲高中社會組學生常態分配Q-Q圖



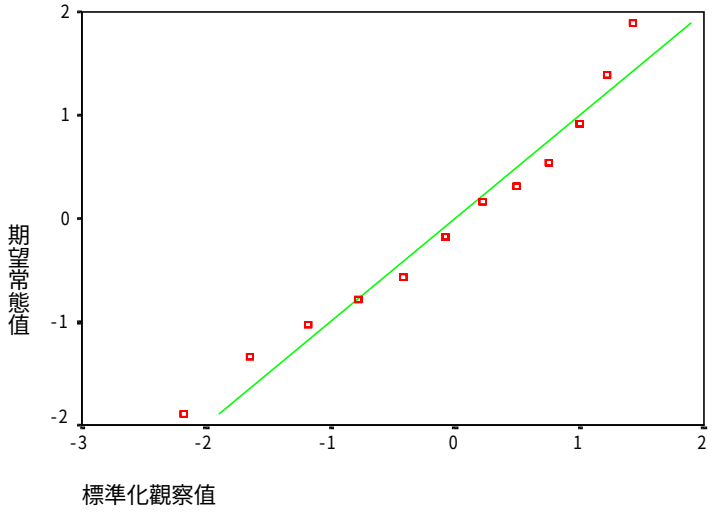
圖五　前測乙高中自然組學生常態分配Q-Q圖



圖六　後測甲高中自然組學生常態分配Q-Q圖



圖七　後測甲高中社會組學生常態分配Q-Q圖



圖八　後測乙高中自然組學生常態分配 Q-Q 圖

附錄四　注意係數分析表

一、前測學生注意係數診斷表

編號	答對率	注意指標	判定類別	編號	答對率	注意指標	判定類別
301-1	55	0.66	B'	301-38	70	0.45	A
301-2	55	0.24	B	301-40	40	0.87	B'
301-3	55	0.3	B	301-42	50	0.24	B
301-4	45	0.09	B	301-43	40	0.15	B
301-5	45	0.37	B	301-45	20	0.71	C'
301-6	50	0.13	B	304-1	60	0.4	B
301-7	50	0.58	B'	304-2	40	0.59	B'
301-8	55	0.06	B	304-3	70	0.35	A
301-9	35	0.56	C'	304-4	60	0.69	B'
301-10	55	0.35	B	304-5	55	0.82	B'
301-11	75	0.3	A	304-6	50	0.43	B
301-12	30	0.54	C'	304-8	45	0.2	B
301-13	70	0.56	A'	304-9	70	0.48	A
301-14	40	0.68	B'	304-10	65	0.56	B'
301-16	60	0.4	B	304-11	55	0.74	B'
301-17	75	0.68	A'	304-13	70	0.66	A'
301-18	45	0.58	B'	304-14	65	0.38	B
301-19	50	0.41	B	304-15	55	0.2	B
301-20	45	0.57	B'	304-16	45	0.29	B
301-21	30	0.48	C	304-17	55	0.39	B
301-22	40	0.5	B	304-18	70	1.31	A'
301-23	55	0.54	B'	304-19	75	0.32	A
301-25	70	0.21	A	304-20	45	0.48	B
301-27	35	0.52	C'	304-21	60	0.36	B
301-28	55	0.24	B	304-22	65	0.61	B'
301-30	60	0.22	B	304-23	45	0.39	B
301-31	45	0.38	B	304-25	75	0.31	A
301-32	55	0.33	B	304-27	30	0.46	C
301-33	65	0.21	B	304-28	60	0.18	B
301-34	55	0.42	B	304-29	65	0.56	B'
301-35	55	0.14	B	304-31	60	0.31	B
301-36	60	0.2	B	304-32	60	0.31	B
301-37	65	0.19	B	304-33	60	0.41	B

編號	答對率	注意指標	判定類別	編號	答對率	注意指標	判定類別
304-34	50	0.36	B	305-42	50	0.21	B
304-35	60	0.36	B	309-1	35	0.32	C
304-36	45	0.18	B	309-2	65	0.21	B
304-38	75	0.63	A'	309-3	50	0.32	B
304-40	55	0.61	B'	309-4	35	0.43	C
304-41	45	0.26	B	309-5	10	0.02	C
304-42	50	0.59	B'	309-7	35	0.21	C
305-1	40	0.1	B	309-8	40	0.98	B'
305-2	45	0.67	B'	309-9	40	0.56	B'
305-4	35	0.82	C'	309-10	45	1.1	B'
305-5	35	0.44	C	309-11	30	0.55	C'
305-6	50	0.21	B	309-12	40	0.19	B
305-8	30	0.76	C'	309-13	40	0.18	B
305-10	55	0.51	B'	309-14	55	0.41	B
305-11	35	0.44	C	309-15	30	0.16	C
305-12	35	0.71	C'	309-16	30	0.38	C
305-13	50	0.98	B'	309-17	60	0.62	B'
305-14	60	0.29	B	309-18	40	0.31	B
305-15	50	0.93	B'	309-19	40	0.4	B
305-16	25	0.17	C	309-20	45	0.11	B
305-17	50	0.54	B'	309-21	35	0.43	C
305-18	15	0.88	C'	309-22	30	0.47	C
305-19	25	0.59	C'	309-23	60	0.3	B
305-20	35	0.11	C	309-24	55	0.32	B
305-22	50	0.3	B	309-25	50	0.32	B
305-23	20	1.09	C'	309-26	60	0.33	B
305-24	40	0.53	B'	309-27	50	0.52	B'
305-25	30	0.77	C'	309-28	55	0.34	B
305-26	15	0.26	C	309-29	35	0.25	C
305-27	35	0.21	C	309-30	45	0.22	B
305-30	25	0.63	C'	309-31	70	0.9	A'
305-32	30	0.71	C'	309-32	55	0.21	B
305-33	75	0.11	A	309-33	45	0.38	B
305-35	35	0.36	C	309-34	45	0.07	B
305-36	50	0.52	B'	309-35	45	0.42	B
305-37	40	0.34	B	309-36	65	0.23	B
305-38	30	0.39	C	309-37	35	0.32	C

編號	答對率	注意指標	判定類別	編號	答對率	注意指標	判定類別
309-39	35	0.32	C	311-36	55	0.4	B
309-41	40	0.42	B	311-37	80	0.3	A
311-1	45	0.26	B	311-38	30	0.65	C'
311-2	55	0.44	B	311-39	55	0.68	B'
311-3	35	0.42	C	311-40	75	0.38	A
311-4	70	0.58	A'	311-41	40	0.8	B'
311-5	30	0.65	C'	312-1	50	0.38	B
311-6	45	0.54	B'	312-2	70	0.38	A
311-7	50	0.29	B	312-3	50	0.95	B'
311-8	60	0.73	B'	312-4	55	0.37	B
311-9	50	0.47	B	312-5	55	1.21	B'
311-10	40	0.65	B'	312-6	55	0.59	B'
311-11	25	0.81	C'	312-7	75	0.91	A'
311-12	55	0.53	B'	312-8	70	0.1	A
311-13	30	0.38	C	312-9	70	0.85	A'
311-14	30	0.2	C	312-10	75	0.48	A
311-15	50	0.37	B	312-11	65	0.45	B
311-16	30	0.15	C	312-12	75	0.29	A
311-17	40	0.76	B'	312-13	55	0.5	B
311-18	45	0.57	B'	312-14	35	0.7	C'
311-19	30	0.38	C	312-15	70	0.37	A
311-20	40	0.54	B'	312-16	65	0.61	B'
311-21	50	0.26	B	312-17	70	0.5	A'
311-22	40	0.31	B	312-18	80	0.38	A
311-23	35	0.38	C	312-19	85	0.71	A'
311-24	45	0.25	B	312-20	45	0.55	B'
311-25	35	0.27	C	312-21	20	0.11	C
311-26	55	0.45	B	312-22	20	0.82	C'
311-27	40	0.29	B	312-23	50	1.18	B'
311-28	45	0.35	B	312-24	50	0.42	B
311-29	30	0.3	C	312-25	50	0.74	B'
311-30	30	0.17	C	312-26	60	0.63	B'
311-31	45	0.53	B'	312-27	45	0.37	B
311-32	35	0.27	C	312-28	45	0.55	B'
311-33	50	0.47	B	312-29	45	0.59	B'
311-34	50	0.36	B	312-30	35	0.86	C'
311-35	40	0.14	B	312-31	30	0.38	C

編號	答對率	注意指標	判定類別	編號	答對率	注意指標	判定類別
316-1	55	0.62	B'	316-20	30	0.14	C
316-2	75	0.65	A'	316-21	60	0.34	B
316-3	25	0.11	C	316-22	55	0.65	B'
316-4	40	0.42	B	316-23	45	0.34	B
316-5	25	0.64	C'	316-24	35	0.49	C
316-6	50	0.59	B'	316-25	50	0.21	B
316-7	40	0.42	B	316-27	50	0.35	B
316-8	35	0.72	C'	316-28	40	0.76	B'
316-10	45	0.56	B'	316-29	80	0.2	A
316-11	35	0.56	C'	316-30	40	0.45	B
316-12	50	0.72	B'	316-31	45	0.78	B'
316-13	30	0.21	C	316-32	55	0.56	B'
316-14	35	0.62	C'	316-33	45	0.48	B
316-15	30	0.41	C	316-34	65	0.17	B
316-16	35	0.43	C	316-35	25	0.68	C'
316-17	35	0.3	C	316-36	50	0.29	B
316-18	40	0.15	B	316-37	40	0.52	B'
316-19	55	0.45	B				

附錄四　注意係數分析表

二、後測學生注意係數診斷表

編號	答對率	注意指標	判定類別	編號	答對率	注意指標	判定類別
301-01	45	0.22	B	304-04	50	0.16	B
301-02	40	0.14	B	304-05	30	0.39	C
301-03	35	0.46	C	304-06	30	0.74	C'
301-05	25	0.92	C'	304-08	55	0.23	B
301-06	50	0.02	B	304-09	60	0.29	B
301-07	60	0.56	B'	304-10	45	0.79	B'
301-08	30	0.29	C	304-11	40	0.73	B'
301-09	30	0	C	304-13	50	0.02	B
301-10	40	0.48	B	304-14	60	0.41	B
301-11	70	0.47	A	304-16	40	0.37	B
301-12	25	0.72	C'	304-17	40	0.34	B
301-13	55	0.36	B	304-18	55	0.43	B
301-14	20	0.57	C'	304-21	40	0.56	B'
301-16	45	0.35	B	304-22	35	0.61	C'
301-18	30	0.14	C	304-23	40	0.19	B
301-19	30	0.33	C	304-27	25	0.8	C'
301-21	15	0.28	C	304-28	45	0.12	B
301-22	50	0.17	B	304-31	35	0.32	C
301-23	35	0.61	C'	304-32	45	0.34	B
301-25	45	0.56	B'	304-33	50	0.2	B
301-27	20	0.43	C	304-34	50	0.41	B
301-30	45	0.13	B	304-35	60	0.04	B
301-31	40	0.61	B'	304-36	35	0.23	C
301-32	50	0.31	B	304-38	80	0.98	A'
301-34	40	0.46	B	304-40	30	0.29	C
301-35	30	0.06	C	304-41	40	0.29	B
301-36	45	0.05	B	304-42	35	0.55	C'
301-40	40	0.37	B	305-04	35	0.35	C
301-42	55	0.28	B	305-06	35	0.66	C'
301-43	60	0.22	B	305-08	45	0.72	B'
301-45	30	0.48	C	305-10	50	0.13	B
304-01	50	0.16	B	305-11	45	0.69	B'
304-02	55	0.37	B	305-12	25	0.42	C

編號	答對率	注意指標	判定類別	編號	答對率	注意指標	判定類別
305-13	45	0.42	B	309-33	35	0.2	C
305-14	35	0.23	C	309-34	25	0.28	C
305-15	40	0.52	B'	309-35	20	0.57	C'
305-16	20	0.76	C'	309-36	45	0.44	B
305-17	35	0.68	C'	309-37	30	0.54	C'
305-19	20	0.69	C'	309-39	30	0.47	C
305-20	35	0.44	C	309-41	45	0.63	B'
305-22	30	0.52	C'	311-01	25	0.27	C
305-24	25	0.53	C'	311-03	30	0.61	C'
305-25	15	1.61	C'	311-04	40	0.38	B
305-26	10	0.09	C	311-05	15	0.38	C
305-27	40	0.31	B	311-06	50	0.58	B'
305-30	30	0.98	C'	311-07	40	0.54	B'
305-32	40	0.45	B	311-09	40	0.38	B
305-33	45	0.58	B'	311-10	20	0.3	C
305-35	35	0.13	C	311-11	15	0.65	C'
305-36	45	0.52	B'	311-12	40	0.68	B'
305-37	55	0.6	B'	311-14	35	0.48	C
305-38	20	0.73	C'	311-15	35	0.28	C
305-42	65	0.7	B'	311-16	40	0.2	B
309-02	45	0.26	B	311-17	35	0.79	C'
309-03	40	0.73	B'	311-19	20	0.57	C'
309-04	35	0.41	C	311-20	30	0.43	C
309-05	30	0.85	C'	311-21	30	0.18	C
309-09	45	0.41	B	311-22	15	0.06	C
309-10	15	1.19	C'	311-23	30	0.09	C
309-12	30	0.29	C	311-24	35	0.32	C
309-14	35	0.45	C	311-25	30	0.25	C
309-16	50	0.5	B	311-26	55	0.09	B
309-19	40	0.33	B	311-27	20	0.24	C
309-21	20	0.23	C	311-28	25	1.16	C'
309-23	45	0.61	B'	311-29	20	0.46	C
309-24	40	0.1	B	311-30	25	0.31	C
309-26	45	0.23	B	311-31	35	0.29	C
309-27	30	0.65	C'	311-32	40	0.43	B
309-29	25	0.19	C	311-33	40	0.99	B'
309-31	50	0.5	B	311-34	35	0.21	C

編號	答對率	注意指標	判定類別	編號	答對率	注意指標	判定類別
311-35	40	0.35	B	312-28	55	0.34	B
311-36	25	0.39	C	312-29	55	0.76	B'
311-37	35	0.1	C	312-30	75	0.89	A'
311-38	30	0.37	C	312-31	75	0.75	A'
311-39	35	0.5	C'	316-01	55	0.24	B
311-40	45	0.18	B	316-03	55	0.4	B
311-41	30	0.26	C	316-04	80	0.3	A
312-02	55	0.6	B'	316-05	30	0.65	C'
312-03	65	0.22	B	316-06	55	0.68	B'
312-04	35	0.59	C'	316-08	75	0.38	A
312-05	80	0.53	A'	316-10	40	0.8	B'
312-06	30	0.24	C	316-11	50	0.38	B
312-07	30	0.6	C'	316-12	70	0.38	A
312-08	45	0.73	C'	316-13	50	0.95	B'
312-09	65	0.22	B	316-14	55	0.37	B
312-10	60	0.94	B'	316-15	55	1.21	B'
312-11	35	0.76	C'	316-16	55	0.59	B'
312-12	85	0.71	A'	316-17	75	0.91	A'
312-13	45	0.85	B'	316-19	70	0.1	A
312-14	55	0.9	B'	316-20	70	0.85	A'
312-15	60	0.51	B'	316-21	75	0.48	A
312-16	60	0.2	B	316-22	65	0.45	B
312-17	55	0.3	B	316-24	75	0.29	A
312-18	40	0.48	B	316-28	55	0.5	B
312-19	40	1.12	B'	316-31	35	0.7	C'
312-20	75	0.47	A	316-32	70	0.37	A
312-21	50	0.59	B'	316-33	65	0.61	B'
312-22	35	0.92	C'	316-34	70	0.5	A'
312-24	50	0.33	B	316-35	80	0.38	A
312-25	85	0.71	A'	316-36	85	0.71	A'
312-26	75	0.36	A	316-37	45	0.55	B'
312-27	45	0.62	B'				

附錄五　S-P 表分析程式碼(運用 S-Plus 撰寫)

input\輸入方式如下

題號	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
正解	4	4	3	2	4	1	1	2	4	3	2	4	1	2	3	4	3	4	1	2
S1	1	4	1	2	4	4	1	1	3	1	1	4	1	2	3	4	1	4	3	2
S2	4	4	1	2	4	4	1	2	3	1	3	1	3	2	3	4	3	2	4	2
S3	4	4	3	2	4	4	1	2	2	2	3	4	1	3	3	2	3	1	3	1
S4	2	4	3	2	4	4	3	2	2	1	3	3	1	2	2	2	1	4	2	2
S5	4	4	1	2	4	4	1	2	3	2	1	4	1	3	3	2	2	3	3	1
S6	4	4	3	2	4	4	3	2	2	1	2	3	2	2	4	3	3	1	2	2
S7	4	3	1	3	4	4	3	2	3	2	2	2	3	2	3	1	3	4	1	2
S8	4	4	3	2	4	4	1	2	3	1	3	2	3	2	2	2	3	3	1	2
S9	2	4	3	2	4	4	3	1	2	1	2	3	4	3	3	3	3	3	4	1
S10	4	3	3	2	4	4	1	2	2	2	3	3	3	2	3	3	3	4	2	2
S11	4	4	3	2	4	1	1	2	2	3	1	3	1	2	3	2	3	4	2	2
S12	2	4	1	3	0	4	3	2	3	1	1	3	1	4	3	4	1	1	4	2
S13	4	4	3	2	4	4	1	2	4	2	2	4	3	2	2	2	3	4	1	3
S14	3	4	1	2	4	4	1	4	3	1	2	4	4	1	3	1	2	2	4	2
S15	4	4	1	0	4	4	3	2	0	1	2	4	1	2	3	1	3	4	4	2
S16	4	1	3	2	4	1	1	2	3	1	2	3	1	3	3	4	3	4	1	2
S17	4	4	3	1	1	4	3	2	3	2	2	2	1	2	3	2	3	2	4	3
S18	4	4	3	1	4	4	2	2	2	1	2	4	1	2	2	2	2	2	1	1
S19	4	4	2	1	4	4	3	2	3	3	2	4	4	3	4	2	3	2	3	2
S20	2	4	1	2	1	4	3	2	2	2	2	3	3	1	4	2	4	2	1	2
S21	1	4	1	2	4	1	4	2	2	1	3	4	3	2	4	2	3	2	2	3
S22	4	3	2	2	4	4	1	2	3	1	2	3	1	2	2	4	2	3	1	2
S23	4	4	3	2	4	4	1	2	3	2	2	3	1	2	3	4	1	2	1	2
S24	4	4	4	1	3	4	2	2	3	1	3	4	1	1	3	2	1	3	2	2
S25	4	4	3	2	4	4	1	2	3	1	3	3	1	2	2	4	4	2	1	1
S26	4	4	3	2	4	4	1	2	3	1	2	3	3	1	3	2	3	4	3	2
S27	4	3	3	2	4	4	1	2	0	0	1	3	0	2	0	4	0	1	0	2
S28	4	4	1	2	4	4	1	2	3	1	2	2	1	3	2	4	2	3	1	2
S29	4	4	3	2	4	4	3	2	3	1	1	4	1	1	2	4	3	4	1	2
S30	4	4	1	2	4	4	1	2	3	1	2	2	1	2	2	4	1	3	1	1
S31	4	4	3	1	4	4	1	2	3	1	3	3	1	2	2	2	1	4	1	2
S32	4	4	2	2	4	4	1	2	3	1	2	3	1	2	2	2	3	4	4	2
S33	4	4	3	2	4	4	1	2	3	2	2	3	1	2	2	4	2	2	1	2
S34	4	4	3	2	4	4	1	2	3	3	2	4	4	2	4	3	4	4	1	2
S35	4	4	1	0	0	4	2	2	4	1	2	2	1	4	1	4	3	2	3	1
S36	4	4	3	1	4	4	3	2	1	1	2	3	1	4	3	1	3	1	4	2
S37	4	4	1	1	4	4	3	2	3	1	3	3	2	2	2	2	3	2	1	2
S38	2	1	1	2	4	4	1	1	3	1	3	2	2	4	2	2	2	4	3	1

\以下為程式碼部分

```
n<-輸入人數;m<-輸入題數;
output[,m+1]<-0;output[n,]<-0
for(i in 2:n){
      for(j in 1:m){
   if(input[i,j]==input[1,j])
         output[i-1,j]<- 1
      else if(input[i,j]!=input[1,j])
            output[i-1,j]<- 0
      }
}
for(i in 1:(n-1)){
            output[i,(m+1)]<-sum(output[i,1:m])
}
for(j in 1:(m+1)){
            output[n,j]<-sum(output[1:n-1,j])
}
```

output **\S-P 原表製作完成**

```
SP_output
SP[1:(n-1),]_-SP[1:(n-1),]
SP<-SP[order(SP[,m+1]),1:(m+1)]
SP[1:n-1,]_-SP[1:n-1,];
SP<-t(SP)
SP[1:m,]_-SP[1:m,]
SP<-SP[order(SP[,n]),1:n]
SP[1:m,]_-SP[1:m,]
SP<-t(SP)
```

SP **\S-P 表完成**

```
csp_SP
ssp_SP
for(i in 1:m){
      a<-SP[n,i]
      csp[1:a,i]<-1
      csp[(a+1):(n-1),i]<-0
}
for(j in 1:(n-1)){
```

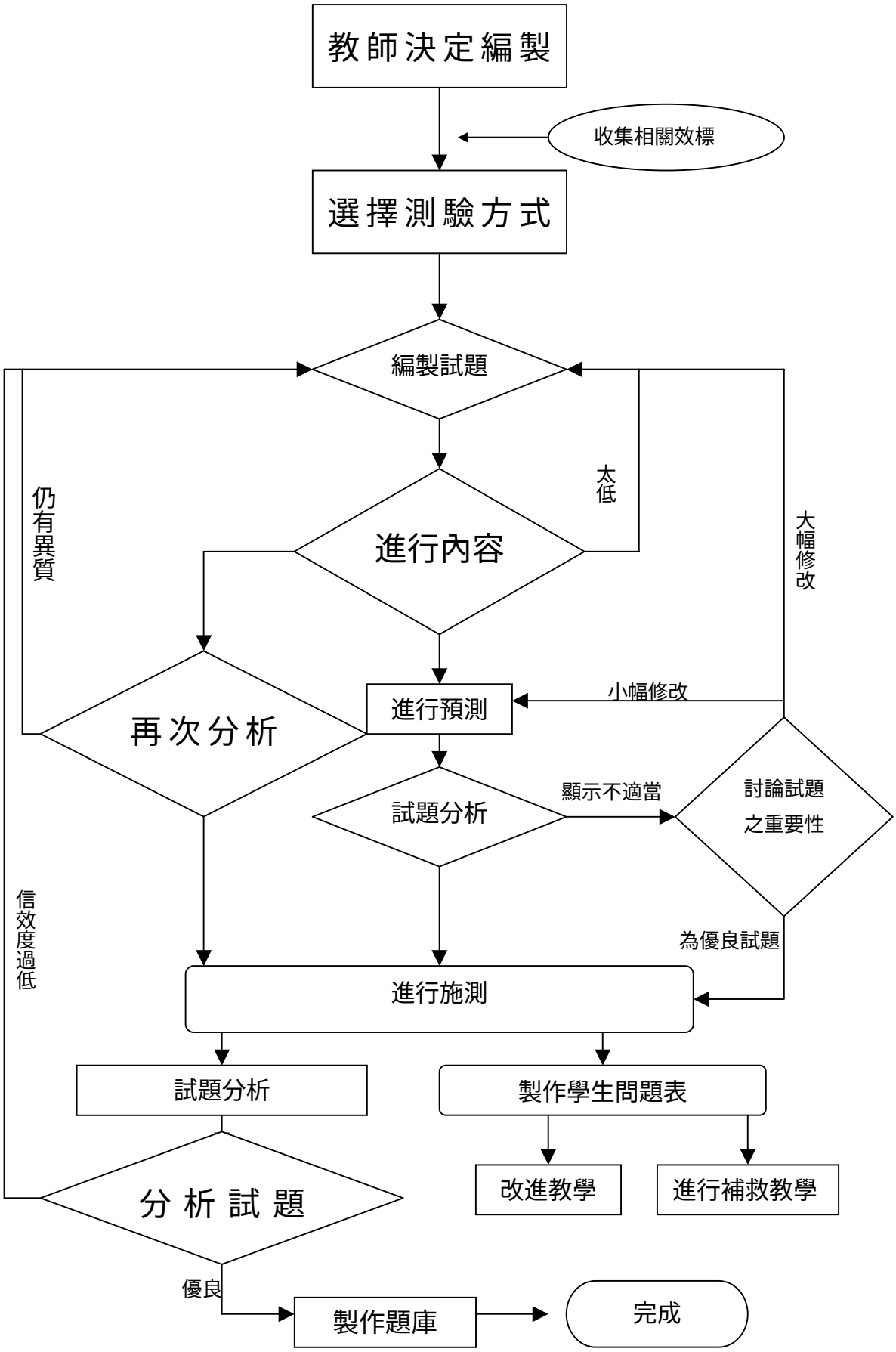
```
    b<-SP[j,m+1]
    ssp[j,1:b]<-1
    ssp[j,(b+1):m]<-0
}
Ccov_var(SP[1:(n-1),1:m],SP[1:(n-1),m+1])
Cpcov_var(csp[1:(n-1),1:m],csp[1:(n-1),m+1])
cp_1-(Ccov/Cpcov)
ssp_t(ssp)
SP_t(SP)
Scov_var(SP[1:m,1:(n-1)],SP[1:m,n])
Spcov_var(ssp[1:m,1:(n-1)],ssp[1:m,n])
cs_1-(Scov/Spcov)
cp \問題注意係數向量
cs \學生注意係數向量
SP_t(SP)
ssp_t(ssp)
```

備註：

input：資料輸入矩陣。 output：S-P 原表輸出。

SP：完成之 S-P 表。cp：問題注意係數。 cs：學生注意係數。

附錄六　試題編製流程圖


教師決定編製
收集相關效標
選擇測驗方式
編製試題
太低
仍有異質
進行內容
大幅修改
再次分析
進行預測
小幅修改
試題分析
顯示不適當
討論試題之重要性
信效度過低
為優良試題
進行施測
試題分析
製作學生問題表
分析試題
改進教學
進行補救教學
優良
製作題庫
完成