2 教育名詞的介紹

本節中教育名詞的介紹，是主要參考自余民寧教授(2002)所著：教育測驗與評量-成就測驗與教學評量與鄒慧英教授(2003)所著：測驗與評量-在教學上的應用。

2.1 古典測驗理論的概述

解釋測驗分數意義的理論學說，可以分成兩大學派：一爲「古典測驗理論」($classical\ test\ theory$,簡稱CTT)，另一爲「試題反應理論」(即IRT)($Crocker\ and\ Algina$, 1986 ; $McDonald$, 1999 ; $Suen$, 1990)。本文是建立在古典測驗理論下，故不討論試題反應理論。

古典測驗理論($Gullikeen$,1950/1987)是最早的測驗理論，至今仍是最實用的理論，且許多測驗編制以此爲依據，建立了測驗資料的實際數據。此理論又稱爲「真實分數理論」，主要目的是希望可以估計實得分數與真實分數間的相關程度。受測者在測驗中得分即爲實得分數，若對同一位受測者施予同樣測驗(理論上是無窮次)，我們可以得到該位受測者多次的實得分數，而這些實得分數的平均數即代表該位受測者能力的不偏估計值，則該估計值視爲「真實分數」。

古典測驗理論主要是以整份測驗觀點來加以討論，並解釋測驗結果後所得分數之意義。對於受測者所得分數的看法，是以整份試題得分的加總作爲依據，因此在試卷中的單題得分是不具任何意義。

理論依據的數學假設爲：

$$\chi = T + e \tag{2.1}$$

其中χ是實得分數，T是真實分數(即代表該測驗所欲測得學生真正能力或潛在特質的部分)，e爲誤差分數(即代表該測驗無法測得學生真正能力或潛在特質的部分)。式子(2.1)中的真實分數T是無法觀察，但卻是施測者想要測量的

潛在特質部分；誤差分數e亦是無法觀察，且是施測者想要降低的部分。將這兩部分合併之後，即是所謂的實得分數。理論上，真實分數不受測驗次數的影響，這是屬於「不變」的部分；實際上，單獨一次測驗所得的實得分數會與真實分數有些差異，即爲誤差分數，而誤差分數受到測驗工具之精確度影響甚大，這代表測驗結果是「可變」。

2.2 測驗

Mehrens 與Lemann(1991) 明白指出測驗的定義就是指測量的工具。而另一方面的看法則是認爲測驗是採用一套標準的刺激，對個人特質作客觀測量的有系統程序(郭生玉，民79)。然而，上面所說的兩種看法早已將測驗與測量混爲一談，這也是導致測驗與測量之間關係不易區分的主要原因。而本文則是以討論測驗爲出發點，不會對測量有多餘之敘述，並且是以Mehrens 與Lemann(1991) 的定義爲基準點。若是將測驗視爲一種收集資料的工具，那測驗即是提供受測者一套完整的問題給予回答，以便提供施測者感興趣的部分，並研究其特性。

當施測者收到測驗結果之後，從微觀分析的角度來看，試題本身已經具有特別的實證特徵，包括：難度指標、鑑別度指標等。此兩指標均有特別意義存在，並且進行分析時，通常是兩指標一起進行運算。

2.3 選擇題

由於本文的研究對象是以選擇題爲主，所以必須給予「選擇題」一個較明確的定義。選擇題是由一個問題與可能的選項所組成，而受測者的反應是選擇提供問題正確或是最佳答案的選項。其問題模式可以是直接問句或是一個不完全敘述句加以陳述，不管是使用哪一種的敘述方式，都應該是一個能讓受測者清楚瞭解其意義的一個問題。選擇題的形式是屬於較彈性的，可用於測量知識和理解層次各種不同的學習成效。

選擇題的優點，它可廣泛應用於各種不同階段的測驗上，評分方式較快速。由於選項多達四至五個，所以得分受猜測影響的程度較低。缺點則爲它是一種選擇型的紙筆測驗，只能測驗語文層次的問題。因爲選擇題只需要正確答案即可，所以呈現的學習結果並不明顯。易引起作弊行爲，影響測驗真實性。

選擇型試題具有絕對正確的標準答案，受測者作答反應只有正確、答對(即記錄爲「1」)，或錯誤、答錯(即記錄爲「0」)，不允許有半對或半錯的模糊情形發生，且易於由電腦來計算成績。

2.4 題組

一群被編制用來測量某些特殊目標的試題集合體，便可以稱做「測驗」(*test*)。測驗的子集合，可用來測量不同的子目標者，同時描述受測者在某些測量目標的剖面結構，即被稱做「分測驗」(*subtest*)或是「題組」(*testlet*)。同一題組的試題會被認爲比不同題組的試題具有較高的相似性或是同質性。

試題分析的部分，只介紹「難度指標」與「鑑別度指標」定義。

2.5 難度指標

難度指標是用來表示每道試題難易程度的指標。定義如下：

$$P_i = \frac{P_{iH} + P_{iL}}{2} \qquad i = 1, \ldots, n. \tag{2.2}$$

其中P_{iH}與P_{iL}分別代表高、低分群的學生在第i道試題上答對人數百分比。實務上，高(低)分群是成績爲前(後) $q \times 100\%$ 的受測學生，其中 $\frac{1}{4} \leq q \leq \frac{1}{3}$。

式子(2.2)即表示每道試題的難度指標是以高分群與低分群答對百分比之平均。P_i值越大高，則表示該試題較容易；反之，P_i值低時，即表示試題較困難。若是P_i是接近0.50時，則表示該試題是難易適中。倘若試題困難度達到高低分群學生均答錯時，此時難度指標值是爲0。試題簡單程度到達高低分

群學生均答對時，此時難度指標值是爲1。因此，難度指標是介於0與1之間的數值。由此可知，此數值解讀方式與一般常理正好相反，這是需要特別注意的地方。

測驗最主要目的即是區分受測者能力高低，所以施測者希望的最佳難度指標是接近0.50，也就是試題是屬於難易適中，此時試題最能區分出受測者的能力差異。但施測者不必爲了達到難度指標是接近0.50這標準，反而只給予受測者一些難易適中的試題；應該是依據受測者的實際情況，施予測驗，以免失去測驗的主要目的。

2.6 鑑別度指標

一份測驗主要是用來區分受測者能力高低，所以有鑑別度指標的產生。定義如下：

$$D_i = P_{iH} - P_{iL} \quad i = 1, \ldots, n. \tag{2.3}$$

其中P_{iH}與P_{iL}代表的意義同難度指標中所給定之意義。

由式子(2.3)可知D_i值域是介於± 1.00之間。若試題過於容易時，高分組與低分組的學生均答對，即表示高、低分組的答對百分比值均是1，差值將會等於零；相對地，試題是屬於困難時，高分組與低分組學生均答錯，即表示高、低分組的答對百分比值均是0，差值亦等於零。就此而言，過於簡單或是困難的試題，都不具有優良的鑑別度。

試題鑑別度是指試題的品質優劣，而這試題品質優劣的判斷是屬於主觀方面，並沒一定標準可供施測者做爲參考。因此，一個常用來挑選試題標準，是「先挑出鑑別度指標較高的試題，然後，從中挑選出難度指標較爲適中的試題。」(郭生玉，民79)，這項建議可提供施測者在選題時的參考。

美國的測驗學者Ebel與Frisbie(1991)提出了一套鑑別度的判斷標準，如表1所示，值得施測者們做爲參考。可依照這套判斷標準與測驗目的，從分析試題的過程中，挑選出較具有良好鑑別度指標的試題，這些試題多半是有較

高品質的試題,當然可以做爲題庫試題之用。

大多數的測驗專家建議挑選難易適中的試題，這是較爲恰當的作法，因爲試題在難易適中的情況下，它的鑑別度指標值可以達到最大。但於實務上不易達到這種微妙的平衡。因此有些學者主張以0.40至0.70之間的難度指標值範圍作爲選擇題的挑選準則(Ahmann與Glock，1981)；而有些學者主張以0.40至0.80之間的難度指標值範圍作爲標準。但就整份試卷而言，難度指標值仍是以接近0.50作爲共同的試題挑選標準。

表1　鑑別度的評鑑標準

鑑別度指標	試題評鑑結果
0.40以上	非常優良
0.30至0.39之間	優良，但可能需要修改
0.20至0.29之間	尚可，但須做局部修改
0.19以下	劣，需要刪除

余民寧教授(2002)所著「教育測驗與評量-成就測驗與教學評量」一書中，附有一套現今實務上計算難度指標與鑑別度指標值之電腦程式$TESTER\ for\ Windows$程式2.0版，本套軟體於高(低)分群之內建設定值如下：當受測人數達四十人時，高(低)分群是成績爲前(後)四分之一之受測學生；當受測人數在四十人以下時，高(低)分群是成績爲前(後)三分之一之受測學生。

2.7　難度指標與鑑別度指標間之關係

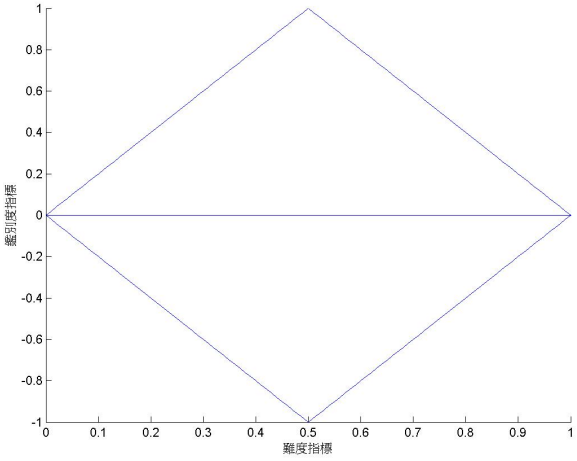
Ebel(1967)提到說：以統計學角度來看，試題的難度指標與鑑別度指標之間具有密切的關係。一份測驗試題若是偏難時，受測者得分大部分會集中在低分區域之內，整個測驗分數的曲線會呈現正偏態分配，導致無法判別能力較低的受測者作答情況。若是試題偏易時，受測者得分大部分會集中在高分區域

內，這使得測驗分數的曲線會呈現負偏態分配，此時無法輕易判別能力較高的受測者作答差異。

難度指標與鑑別度指標兩者間的關係如圖1所示。由圖1可知兩者關係是成菱形分佈，若難度指標值趨向於兩極端（即P_i值爲0.00或是1.00），則鑑別度指標趨近於0。特別的是，當難度指標值接近0.50時，鑑別度指標將可能會到達最大（即D_i值爲$\pm$1.00）。

一般而言，負的鑑別度指標值試題是捨棄之，因爲這些試題可能具有某程度上的誘答能力，但誘答能力不爲本文探討範圍，故不多加論述。以圖1而言，只採用圖形的上半部。若要使試題維持在一定的品質水準的話，在試題選擇時應該挑選難度指標值接近0.50的試題較爲適合。

圖1　難度指標與鑑別度指標的關係



基本上而言,古典測驗理論所使用的難度指標與鑑別度指標，都是一種樣本依賴(sample dependent)的指標，這意味著試題分析的結果會隨著施測樣本而有所不同。試題分析的結果，只能獲得一個「暫時性」的統計指標，因此它們的特質不會固定不變。此外，受測者的人數、教育背景、能力水準的高低與學習型態等因素，均都影響試題分析的結果。

2.8 範例

本節將舉一簡單例題說明如何以傳統模式(即目前實務上所採用之計算方式)計算難度指標與鑑別度指標。

例 1

假設有十二位學生回答一份試卷，此試卷共有六道試題，且每道試題均有四個選項。學生作答情形依照答對題數由高至低依序排列，如表2。

表 2　例 1 作答情況

學生編號		題號 1	2	3	4	5	6	答對題數
高分群	1	C	B	D	A	D	B	6
	2	C	B	D	A	D	D	5
	3	A	B	D	A	D	B	5
	4	C	B	C	A	D	B	5
中段生	5	C	B	D	A	A	D	4
	6	C	D	C	A	D	D	3
	7	C	B	C	A	A	D	3
	8	A	D	C	A	D	D	2
低分群	9	A	B	C	B	A	D	1
	10	C	D	C	B	A	D	1
	11	A	D	C	A	A	D	1
	12	A	D	C	B	A	D	0
	正確答案	C	B	D	A	D	B	
	答對率	$\frac{7}{12}$	$\frac{7}{12}$	$\frac{4}{12}$	$\frac{9}{12}$	$\frac{6}{12}$	$\frac{3}{12}$	

爲了將傳統模式的計算值與電腦程式*TESTER for Windows*程式2.0版的輸出結果作一對照，本例之高(低)分群是成績爲前(後)三分之一的受測學生。

本例僅以第一道題說明難度指標與鑑別度指標的計算過程，其餘試題均仿照

此計算方法。

高(低)分群於第一道題的答對率計算過程如下：

$$\begin{aligned} P_{1H} &= \frac{\text{高分群中答對第一題的人數}}{\text{高分群中的總人數}} \\ &= \frac{3}{4} \\ &= 0.7500 \end{aligned}$$

$$\begin{aligned} P_{1L} &= \frac{\text{低分群中答對第一題的人數}}{\text{低分群中的總人數}} \\ &= \frac{1}{4} \\ &= 0.2500 \end{aligned}$$

由式子(2.2)與(2.3)可求得第一道題之難度指標與鑑別度指標值，如下：

$$\begin{aligned} P_1 &= \frac{1}{2}(P_{1H} + P_{1L}) \\ &= \frac{1}{2}(0.7500 + 0.2500) \\ &= 0.5000 \\ D_1 &= P_{1H} - P_{1L} \\ &= 0.7500 - 0.2500 \\ &= 0.5000 \end{aligned}$$

同理，可獲得各試題之高(低)分群答對率、難度指標與鑑別度指標值，整理如表3。

表3　例1中各試題之難度指標與鑑別度指標值

題號	高分群答對率	低分群答對率	難度指標	鑑別度指標
1	0.7500	0.2500	0.5000	0.5000
2	1.0000	0.2500	0.6250	0.7500
3	0.7500	0.0000	0.3750	0.7500
4	1.0000	0.2500	0.6250	0.7500
5	1.0000	0.0000	0.5000	1.0000
6	0.7500	0.0000	0.3750	0.7500

上表與電腦程式$TESTER\ for\ Windows$程式2.0版所輸出之結果均相符。