

1.語文測驗

2.數字測驗

3.圖形測驗

(ㄅ)學術性向測驗 (Scholastic Aptitude Test (SAT))

1.語文 (反義字、語句完成、語文類推、閱讀理解等)

2.數學

(ㄆ)美國大學入學測驗 (American College Testing Program (ACT))

1.英語用法

2.數學

3.社會科閱讀測驗

4.自然科閱讀測驗

包含傳統的性向和成就測驗

(ㄇ)研究所入學測驗 (Graduate Record Examination (GRE))

1.語文 (語文推理和閱讀理解)

2.數量 (算術和代數推理、圖形推理、資料描述)

歸納上述 13 種測驗，紙筆式學術性向測驗內容宜包括下列三大類 (不包括操作測驗)

1.語文測驗 (含語文理解及語文推理)

2.數字測驗 (含數字應用及數字系列測驗)

3.圖形測驗 (含圖形類推及圖形歸類，每種再分成具體與抽象兩類形)

## 六、題目偏失 (item bias) 研究方法

(一)傳統測驗理論題目偏失研究方法

1.兩個群體的平均數有差異，那題目或測驗就是有偏失。此定義顯然不妥，因為影響兩個群體平均數差異的因素很多，並不是只有題目偏失這個因素。依照這個定義，那體重計與身高計都是有偏失的測量工具，因為男、女的身高和體重平均數均有明顯差異。

2.兩個群體所得到的題目難度指數 (或 P 值) 相同，就是沒有題目偏失。此定義與上述第一種定義有相同的問題。兩個群體在所有

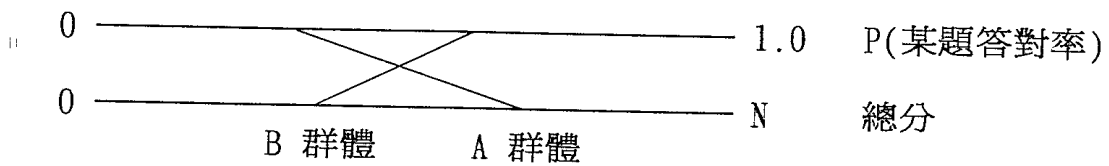
題目的 P 值均相同，那兩個群體的平均得分也必然相同，但是，兩個群體的平均數相同，其各題的 P 值卻不一定相同。

3. 題目與受試群體有交互作用，就是有偏失。

圖2.3的(1)及(2)，題目與群體有交互作用，是有偏失的題目；圖2.3的(3)是沒有偏失的題目。

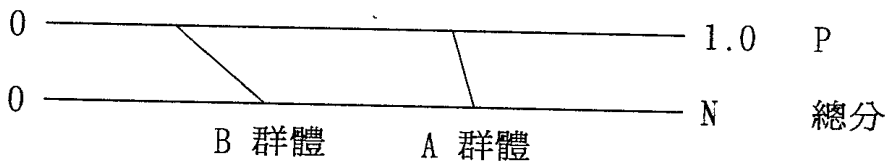
圖 2.3 題目與群體有交互作用及無交互作用的例子

(1) 無次序性交互作用



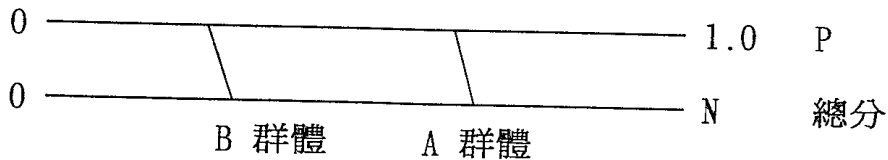
A 群體的總分比B群體高，但是某一題的P反而比B群體低。該題對B群體有利。

(2) 有次序性交互作用



A 群體的總分及某題之答對率雖然均高於B群體，但是P的差距比總分的差距大或小，也就是不平行。該題對A或B群體有利。

(3) 沒有偏失



A、B兩個群體總分之差距與P之差距相同，兩條線相互平行，此為沒有偏失的題目。

- 4.兩個群體 P 之相關為 1，那題目就沒有偏失（或是有相同程度之偏失）。題目難度散佈圖上偏離最適合線的題目就是有偏失的題目。此種方法主要的問題是 P 值之相關不是直線的，即使是沒有偏失，只要兩個群體的能力不等，仍然得不到完全相關。
- 5.△值 (P 值之反轉常態轉換其平均數為 13，標準差為 4。如 p=.84，△=9；P=.50，△=13；p=.16，△=17 偏離雙變項散佈圖主軸之垂直距離，即為該題之偏失值。Lord (1980) 和 Angoff (1982) 發現猜測、題目鑑別力不同以及兩個群體能力之差異，均會使△值偏離主軸。
- 6.具有相同潛在能力的兩組受試，其答對題目的概率相等，就是沒有偏失。  
具有相同總分的不同群體受試，如果答對某題之概率不同，即為有偏失。此類方法均採用  $\chi^2$  法，其中充足  $\chi^2$  法 (full chisq-uare method) 是將總分分成 J 組 (j=1, ..., J)。

圖 2.4 用來計算 chi-square 的交叉分類表

		A 群體	B 群體	C 群體
題 目 i	答對	$N_{11j}$	$N_{12j}$	$N_{1,j}$
	答錯	$N_{21j}$	$N_{22j}$	$N_{2,j}$
	全體	$N_{.1j}$	$N_{.2j}$	$N_j$

第 j 組，第 i 題的 chi-square 值為

$$\chi_{ij}^2 = N_j (N_{11j} N_{22j} - N_{21j} N_{12j})^2 / (N_{1,j} N_{2,j} N_{.1j} N_{.2j})$$

所有各能力組第 i 題的 chi-square 值為

$$\chi_i^2 = \sum_{j=1}^J \chi_{ij}^2,$$

只有兩個群體時自由度為 J。如果有 K 個群體，則自由度為 J(K-1)。此種方法主要的問題有二：

(1)分成多少組以及如何分組均相當主觀，會影響所得結果。

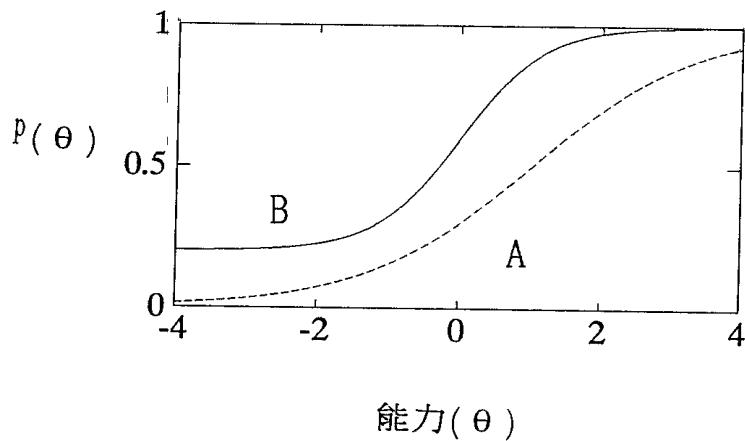
(2)chi-square對樣本數及細格數很敏感，只要人數夠多， $\chi^2$ 值均會達顯著水準。

### (二)題目反應理論 (IRT) 題目偏失研究方法

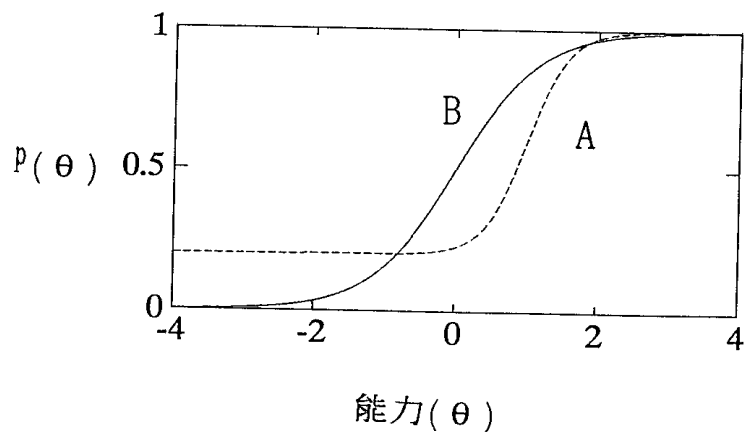
能力相同的不同群體受試，如果其答對某一題目的概率不同，那該題就是有偏失。也就是說，不同群體之題目特徵曲線 (ICC)或是題目參數如果完全相同，那就沒有偏失。

當二個群體某一題目之 ICC 相同時，其題目參數也相同。

圖 2.5 兩個群體題目特徵曲線差異情形



所有能力水準均對A群體不利的題目



低能力水準不利於B群體，高能力水準不利於A群體的題目

以 IRT 為基礎所進行的題目偏失研究，通常採用下列二種方法：

1. 題目特徵曲線比較法

此種方法是在計算兩條題目特徵曲線所涵蓋的面積之差異情形，面積差異越大，代表題目的偏失越大。完全沒有偏失的題目，其 ICC 完全相同，因此，面積之差異也等於 0。計算面積的方法有下列三種：

(1) 絕對值法

$$A_{1i} = \sum_{\theta=-5}^{+5} | P_{i1}(\theta_k) - P_{i2}(\theta_k) | \Delta \theta$$

$$\Delta \theta = .005$$

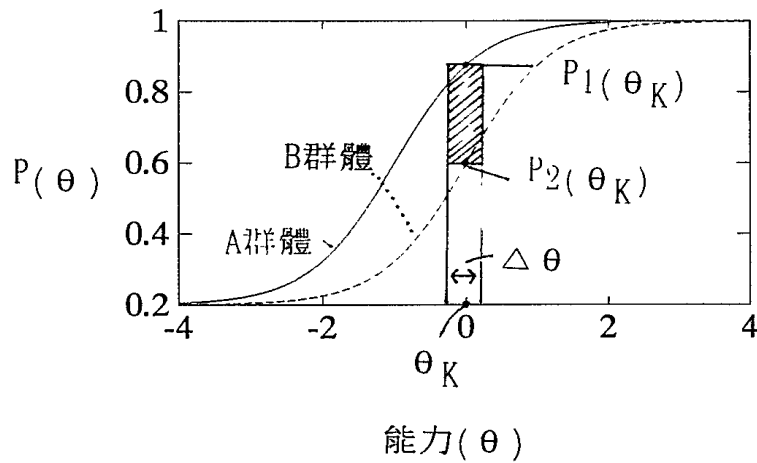
(2) 平方根法

$$A_{2i} = \sum_{\theta=-5}^5 \{ [ P_{i1}(\theta_k) - P_{i2}(\theta_k) ]^2 \Delta \theta \}^{1/2}$$

(3) 偏向法

$$A_{3i} = \sum_{\theta=-5}^5 \{ [ P_{i1}(\theta_k) - P_{i2}(\theta_k) ] \Delta \theta \}$$

圖 2.6 兩個群體的 ICC 差異面積計算法



第(1)及第(2)兩種面積可找出有偏失的題目，第(2)種先平方再開二次方，加重極端差異值的重要性，此兩種均能找出有偏差的題目，二者所得的面積也均永為正值，無法知道對那個群體較不利（或有利）。第(3)種面積值可提供題目偏失方向之訊息（從正負號判斷），但是較無法有效偵測有偏失之試題，像上圖有一段對 B 群體有利，也有一段對 A 群體有利，正負相抵，所得到的面積值可能很小，但實際上是有明顯偏失。

## 2. 題目參數比較法

此種方法類似多變項變異數分析法（MANOVA），考驗每一個題目的兩組題目參數（a, b, c）有無顯著差異，其公式為

$$Q = (X_1 - X_2)' (V_1 + V_2)^{-1} (X_1 - X_2)$$

$X_1$  及  $X_2$  分別為兩個群體的題目參數之向量， $V_1$  及  $V_2$  是兩組題目參數的估計誤差變異數 - 共變數矩陣。此項考驗的假設為

$$H_0: \tau_{1i} = \tau_{2i} \quad (i=1, \dots, n)$$

$$H_1: \tau_{1i} \neq \tau_{2i}$$

$$\tau_{1i} = \begin{bmatrix} a_{i1} \\ b_{i1} \\ c_{i1} \end{bmatrix}, \quad \tau_{2i} = \begin{bmatrix} a_{i2} \\ b_{i2} \\ c_{i2} \end{bmatrix}$$

題目參數估計誤差的變異數 - 共變數矩陣等於訊息矩陣 (I) 的反矩陣，

即

$$V_{1i} = [ I_1 ( X_i ) ]^{-1} \equiv I_{1i}^{-1}$$

$$V_{2i} = [ I_2 ( X_i ) ]^{-1} \equiv I_{2i}^{-1}$$

因此，

$$Q_1 = ( X_{1i} - X_{2i} )' ( I_{1i}^{-1} + I_{2i}^{-1} )^{-1} ( X_{1i} - X_{2i} )$$

此項考驗所得 Q 值近似  $\chi^2$  分配，其自由度等於所要比較的參數數目。如採單參數模式，那就只有一個 b 參數，因此考驗的公式可簡化為

$$z_i = ( b_{1i} - b_{2i} ) / ( V_{1i} + V_{2i} )^{1/2}$$

### (三)IRT 題目偏失研究之步驟

任何 IRT 模式題目參數比較法之步驟如下：

1. 選擇一個適當模式。
2. 分別為每一個群體估計題目和能力參數。
3. 由於兩個群體的參數是分開估計的，因此，必須安置在共同量尺上。參數估計時，如選擇將  $b_i$  標準化，那參數就在已同一量尺上。否則就要採用前面介紹的題目參數連結法來將題目參數安置於同一量尺上。