

試題分析軟體 MicroCAT 簡介

*黃啟彥、**蕭顯勝

*台灣師大工業科技教育系碩士

**台灣師大工業科技教育系副教授兼進修推廣部資訊組組長

本文旨在介紹適用於試題反應理論的試題分析軟體 - MicroCAT(3.5 版)，並簡介試題反應理論，說明如何使用 MicroCAT 來估計試題參數，並分析題庫性質。

壹、前言

試題反應理論(Item Response Theory, 以下簡稱 IRT)自 70 年代起開始成為測驗學者的研究焦點，在此之前的測驗領域仍然以傳統測驗的強真分數理論為主。雖然傳統測驗理論(Classical Test Theory, 以下簡稱 CTT)的假設弱，但是較容易滿足，因此相較於數學模式複雜、嚴謹的 IRT，CTT 反而獲得大多數測驗使用者的青睞。但是 IRT 特性中的試題參數不受樣本影響、試題能提供每個受試者個別差異的測量誤差指標、以試題訊息量取代信度等，都是 CTT 所無法達到的。時代進步日新月異，在資訊科技發達的今日，過去 IRT 困擾測驗使用者的繁雜數學模式與假設，已經可以交給電腦處理，也因而發展出所謂的電腦適性化測驗(Computerized Adaptive Test, 以下簡稱 CAT)，至目前為止，一些常見的測驗工具如托福測驗等都以 CAT 的方式實施。而在國內，IRT 一直多停留在研究階段，應用在各領域測驗上的研究反而較為少見，實為可惜。

比較 IRT 與傳統測驗理論，其數學化、數量化可說是其理論上的一大特色，尤其當我們要應用 IRT 的方法來分析某份測驗資料時，首先必須估計所選用的試題反應模式的參數，而這些參數的估計常涉及艱深難懂的數學公式及繁瑣的計算過程(劉湘川、施淑娟，1995)。不過近年來資訊科技發展迅速，各種適用於 IRT 的程式軟體都相繼問世，使得試題參數估計變得輕鬆許多，也鼓舞許多測驗使用者研究 IRT 並進而促進 CAT 的推廣。因此本文將簡介 IRT 理論，並介紹一套美國 Assessment Systems Corporation 發展的 MicroCAT 測驗軟體，說明如何操作以估計試題參數。期能對 IRT 及 CAT 有興趣者提供實務上的參考。

貳、試題反應理論

製作一個 CAT 的程序，可以分為以下步驟(何榮桂，1999)：

1. 選擇一個合適的 IRT 模式。
2. 建置一個已校準的題庫。
3. 決定測驗的起始點。
4. 選擇有效率的適性策略及能力估計方法。
5. 決定測驗終止的標準。

CAT 施測過程中重要的選題策略以及能力估計是依據 IRT 發展而來。在實施前必須先建立已校準的題庫，而題庫的校準必須通過試題參數分析，是以欲瞭解 CAT 原理必需從 IRT 開始。

一、IRT 概要

受試者在某測驗上的表現情形,可由一組因素來加以預測或解釋，這組因素稱為潛在特質(latent traits)或能力(abilities)。受試者的表現和潛在特質之間的關係，可透過一條連續嚴格遞增(monotonically increasing)的函數來解釋，此函數稱為試題特徵函數(item characteristic function)，然後再把不同能力的學生在某試題的得分期望連結成線，所形成的曲線則稱為試題特徵曲線(item characteristic curve，簡稱 ICC)。IRT 的目地就在於提供計的試題不變量(invariant)和能力估計值。由於 IRT 試題反應理論是針對古典測驗理論的缺失而發展出來的，所以它有幾項特色(Hambleton&Swaminathan,1985)：

1. 在受試母群中，試題參數估計值是不受取樣波動(sampling fluctuation)的影響，不同的取樣結果，只呈現 ICC 的局部區間，連結了不同的取樣結果，則可呈現較完整的 ICC 圖。
2. 在試題的選擇中，受試者能力估計值不受樣取波動的影響。
3. 受試者的能力可以確切估計求出。

二、基本假定

任何測驗都可以透過現代測驗的試題反應模式的基本假定 - 單向性與局部獨立的檢驗，以了解模式與資料的符合程度。另外有些模式則有其他的特殊假定，如 Rasch 模式，則不在本文討論之列。

(一)單向性 (unidimensionality)

單向性指的是測驗只測一種能力或特質。在此「測驗」只表示測量同一特

質或能力的題目組合。最常用於考驗測驗單向性的方法就是因素分析法。IRT 屬於強假定不同於古典測驗的弱假定，所以測驗資料很難完全符合單向性的假定，但項目反應模式有其強韌性 (robustness)，也就是允許某種程度的違反假定。因此，假定的考驗不能做為判定現代測驗生死的依據，應將之視為題目的性質 (王寶庸，1995)。

使用因素分析考驗單向性時，Lord(1980)認為第一個特徵值比第二個特徵值大很多，且其他特徵值都差不多時，測驗大致呈單向性。如表 1 中，第一個特徵值佔 14.55%，且明顯大於其他特徵值，表示有一明顯的主控因素。

表 1 特徵值及比例

因素	特徵值	百分比
1	34.629	14.550
2	4.327	1.818
3	3.734	1.569
4	3.644	1.531
5	3.028	1.272
6	2.838	1.192
7	2.749	1.155
8	2.538	1.066
9	2.480	1.042
10	2.412	1.013

(二)局部獨立 (local independence)

局部獨立指的是對某受試能力而言，試題之間無相關存在。局部表示只針對某一受試能力 θ 而言，而非對整體受試。獨立表示試題間無相關，也就是統計獨立 (statistical independence)。也就是某試題不能為另一個試題提供線索。局部獨立可以由因素分析考驗或使用卡方考驗每一配對題目間的顯著性(王寶庸，1995)。由於 IRT 屬於強勢假定的理論，也就是允許某種程度的違反假定，因此假定的考驗不能做為判定現代測驗生死的依據，應將之視為題目的性質。若卡方值達到顯著水準，表示該題不適合反應理論。事實上，只要人數夠大，所有試題都可以達到 0.05 或更高的顯著水準，所以很少有試題是非常適合模

式的。所以卡方值往往用來剔除極端不佳參數用，此類题目的卡方值通常會比臨界值高上數倍，表示參數估計不精確，不應當使用。如表 2 所示，以 $df=17$ 而言， $Chi\ square \geq 27.59$ ，即達到 0.05 的顯著水準。因為樣本數較大($N=479$)，所以 $Chi\ square$ 本來就會比較大，範例中有四題達到 0.05 的顯著水準。但整體而言， $Chi\ square$ 平均值低於顯著水準，且最大值只有 34.49，並未出現高於臨界值數倍的試題，故可說明此份試卷大致符合 IRT 的局部獨立假定。

Item	N	Chi square	Df	達顯著
1	479	28.074	17	✓
2	479	34.49	17	✓
3	479	15.331	17	
4	479	11.941	17	
5	479	43.6	17	✓
6	479	21.886	17	
7	479	28.74	17	✓
8	479	16.484	17	
9	479	14.418	17	
10	479	21.564	17	

三、參數模式

IRT 以試題特徵函數來表示受試者答題反應與能力值之間的關係，因為函數中採用的參數不同，而有單參數、雙參數、三參數等三種模式之分。常見模式之試題特徵函數如下所示(Hambleton, Swaminathan & Roger,1991)：

$$P_{ij}(\theta_j) = \frac{1}{1 + e^{-D(\theta_j - b_i)}} \quad (\text{公式一}), \text{單參數模式}$$

$$P_{ij}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (\text{公式二}), \text{雙參數模式}$$

$$P_{ij}(\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (\text{公式三}), \text{三參數模式}$$

- 常數 $D=1.702$ ；
- e 為自然對數；
- i 為試題編號；
- j 為受試者編號；
- θ_j 為受試者 j 的能力值；

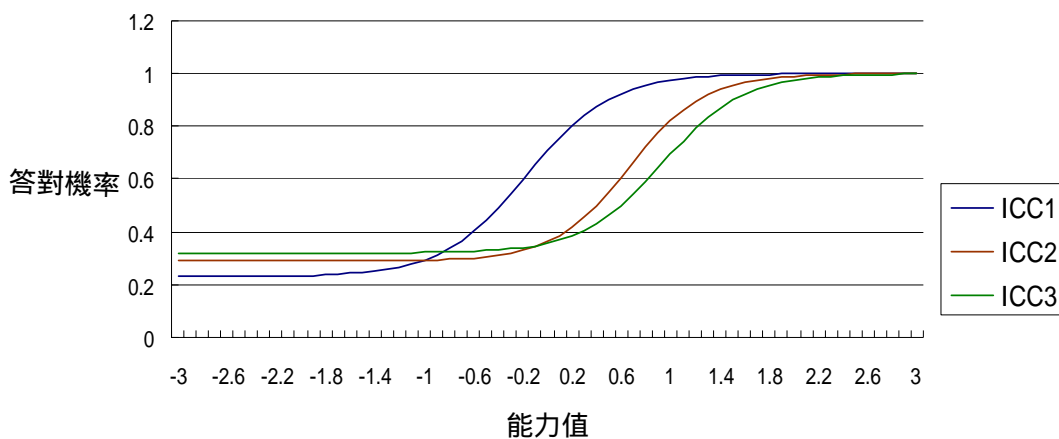
- a_i 為試題 i 的鑑別參數；
- b_i 為試題 i 的難度參數；
- c_i 為試題 i 的猜測參數；
- $P_{ij}(\theta_j)$ 為能力值為 θ_j 者答對試題 i 的機率函數；

試題特徵曲線就是將 $P_{ij}(\theta_j)$ 與 θ_j 的關係圖，圖一所示為三參數模式 ICC，橫軸表示受試群體之能力，縱軸表示答對機率。

$$ICC1(a,b,c)=(1.694,-0.17,0.23),$$

$$ICC2(a,b,c)=(1.919,0.667,0.29),$$

$$ICC3(a,b,c)=(1.803,0.932,0.32)。$$



圖一、三參數試題特徵曲線

透過特徵曲線圖，我們可以發現 ICC1 的難度較低，表示試題較容易，所以曲線圖偏左，即難度參數落在能力值較低的範圍；而從曲線圖的左端終點機率，我們可以看出 ICC3 的猜測度較高，即試題較容易被猜中。表示即使能力值低到 -3 也有一定的機率達對該題。而試題三參數在 IRT 的應用上，不僅止於分析試題，在能力估計上也必須使用三參數以計算能力。

四、能力估計

IRT 的能力估計是從受試者現有的反應組型裡，去推估適當的考生能力值。假設某一受試者在一份 n 題的測驗上，其反應組型為 $(u_1, u_2, \dots, u_j, \dots, u_n)$ ，其中 u_j 只有 0 (答錯) 與 1 (答對) 兩種情形。依據局部獨立的假設，反應組型的聯合機率(joint probability)就是每個試題反應機率的連乘積。

$$\text{聯合機率 } P(U_1, U_2, \dots, U_n | \theta) = \prod_{j=1}^n P(U_j | \theta) \text{ (公式四)}$$

又 u 值(答題反應)非 0 即 1, 所以我們可以將之改寫為近似值函數(likelihood function), 並表示為

$$P(U_1, U_2, \dots, U_n | \theta) = \prod_{j=1}^n P_j^{U_j} Q_j^{1-U_j} \text{ (公式五)}$$

其中 $P_j = P(U_j | \theta), Q_j = 1 - P(U_j | \theta)$ (公式六)

這種聯合機率的表示公式便稱為概似函數, 並記作 $L(u_1, u_2, \dots, u_j, \dots, u_n | \theta)$, 因此可得到以下公式,

$$L(u_1, u_2, \dots, u_j, \dots, u_n | \theta) = \prod_{j=1}^n P_j^{u_j} Q_j^{1-u_j} \text{ (公式七)}$$

再將概似函數取對數後, 當對數概似函數達到最高點(最大值)時的能力值 θ , 便定義成該受試者的 θ 的最大近似估計值(Maximum Likelihood Estimate, 簡寫成 MLE), 也就是我們要求的估計能力值。

而取得對數概似函數最高點的能力值, 必須經過牛頓法的迭代(iteration)過程, 而最大近似估計值無法估計全對或全錯的能力值, 因此必須要收集受試者的答題反應直到不出現全對或全錯的情況下, 才能取其最初能力估計值 θ_0 , 其中 n 為答題數, r_a 為答對題數。

$$\theta_0 = \ln\left(\frac{r_a}{n-r_a}\right) \text{ (公式八)}$$

接著進入迭代過程, 求最佳估計值。若 θ_m 為一能力估計值, 則 θ_{m+1} 為更精確的解, 利用侏代法求解, 直到能力值的改變量 h 趨於穩定或小於預設值, 如 .001。

$$\theta_{m+1} = \theta_m - h \text{ (公式九)}$$

$$h = \frac{f(\theta_0)}{f'(\theta_0)} \text{ (公式十)}$$

$f(\theta)$ 是對數概似函數取一次導數後的函數, $f'(\theta)$ 則是二次導數, 轉換成三參數模式, 可表示成下列兩式:

$$f(\theta) = D \sum_{i=1}^n \frac{a_i [r - P_i(\theta)] [P_i(\theta) - c_i]}{P_i(\theta)(1 - c_i)}, \text{ 其中 } r \text{ 為答題反應}(0 \text{ 或 } 1) \text{。 (公式十一)}$$

$$f'(\theta) = D^2 \sum_{i=1}^n \frac{a_i^2 [P_i(\theta) - c_i] [rc_i - P_i^2(\theta)] Q_i(\theta)}{P_i^2(\theta)(1 - c_i^2)} \quad (\text{公式十二})$$

利用牛頓法的迭代過程，每一階段答完一題後，都可以求得一新的能力精確估計值，除了第一次的最初能力估計必須以公式代入，之後各階段估計都會使用前次的估計值作為起始點來作一次新的估計。

五、試題訊息函數

IRT 利用試題訊息函數(item information function)來作為描述試題效能、挑選試題的方法，試題訊息函數定義如下：

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad i = 1, 2, \dots, n \quad (\text{公式十三})$$

其中，

- $I_i(\theta)$ 代表試題 i 在能力為 θ 上所提供的訊息；
- $P_i'(\theta)$ 為在 θ 點上的 $P_i(\theta)$ 值的導數；
- $P_i(\theta)$ 為能力 θ 在試題 i 上的試題反應函數；
- $Q_i(\theta) = 1 - P_i(\theta)$ 。

將公式三代入，則可化簡為

$$I_i(\theta) = \frac{a_i^2(1 - c_j)}{[c_j + e^{D_a(\theta - b)}] \cdot [1 + e^{-D_{a_i}(\theta - b_i)}]^2} \quad (\text{公式十四})$$

六、終止條件

所謂一份測驗的訊息函數(test information function)是指它在某一個 θ 值上所提供的訊息量，該訊息量剛好是在 θ 值上的試題訊息函數之總和，記作 $I(\theta)$ ：

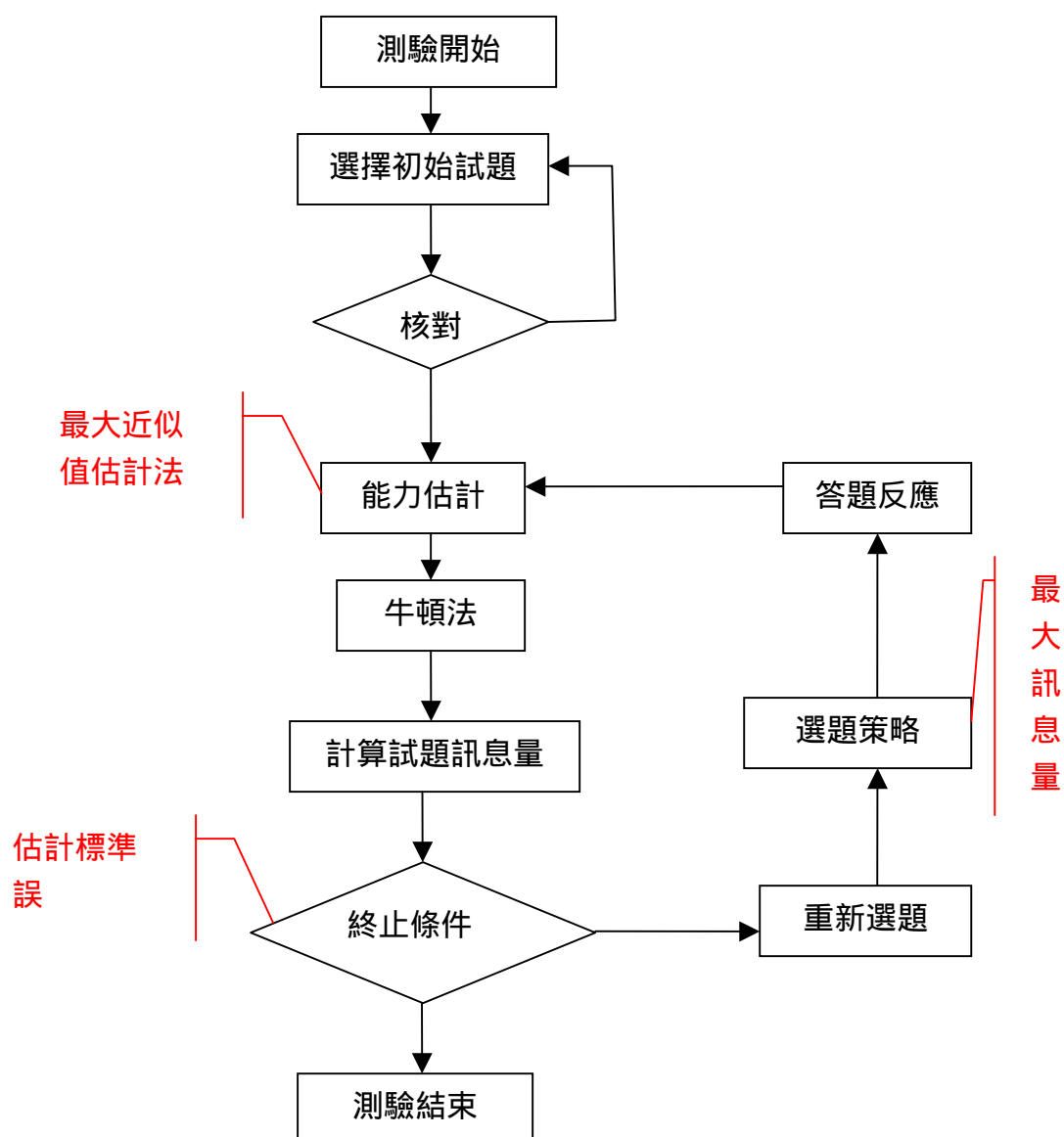
$$I(\theta) = \sum_{i=1}^n I_i(\theta) \quad (\text{公式十五})$$

在 θ 值上的測驗訊息量與能力的估計值的精確性成平方根反比，記作：

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \quad (\text{公式十六})$$

其中， $SE(\theta)$ 稱作估計標準誤(standard error of estimation)。該項指標只要在能力參數的最大近似估計值求出後，便可計算得出。有了能力參數的最大近似估計值，並且也求出在 θ 值上的測驗訊息之後，我們便可以估計信賴區間的方式來解釋能力估計值的涵義(余民寧，1992)。亦即當估計標準誤之變化量低於預設的標準

後(如.01),便可確認找到受試者真正的能力值並終止測驗。上述 IRT 之執行流程如圖二所示。



圖二 IRT 執行流程

以上能力估計的計算步驟中，我們可以發現再實際進行 CAT 時，受試者的最初估計能力可以從公式八求得，但是計算試題訊息量以及能力估計步驟中的牛頓法，都需要先取得試題參數，是以在實施 CAT 之前，必先建立具備試題參數的題庫。在過去試題分析軟體尚未出現之前，估計試題參數必須將三參數與能力同時進行迭代才能取得，但是過於繁雜的運算過程使得 IRT 有其實際應用上的困

難，至今資訊科技發達，已可程式運算來代替人工計算過程，使得 IRT 逐漸撥雲見日，不再難以得其門而入。

參、試題分析軟體 MicroCAT

試題分析工作是測驗編制或建立題庫的必經過程。如本文前面所述，當我們要應用 IRT 來分析某份試卷時，首先必須估計使用模式下的試題參數，而這些參數的取得必須經過複雜的運算，目前已有多套軟體可供進行試題分析工作，如美國 Assessment Systems Corporation 於 1985 年發展的 MicroCAT、Mislevy & Bock 於 1990 年發展的 Bilog 等。其中 MicroCAT 不僅提供傳統試題分析，也提供 IRT 參數估計，使得 CAT 更容易實施。以下擬簡介 MicroCAT3.5，並舉例說明如何應用在三參數模式之試題分析，協助對 CAT 有興趣者進行實際測試。

一、系統簡介

MicroCAT 主要是由 ITEMAN、RASCAL、ASCAL 等軟體組成，ITEMAN 是一套傳統測驗試題分析軟體，提供古典測驗項目內容分析，試題分析上限是 500 題試題，受試人數 15000 人。RASCAL 是 Rasch 模式試題校準程式，也就是適用於單參數模式試題分析，限制題數及上限人數方面，則與 ITEMAN 相同。ASCAL 則是試題參數估計程式，可以分析雙參數及三參數模式，試題分析上限是 250 題試題，受試人數 30000 人。系統所需的最低軟硬體環境為 DOS3.3 及 640KB 的記憶體。由於三參數模式為最常應用的一種，故以下主要針對 ASCAL 做說明。

二、準備受試資料

在開始使用 MicroCAT 之前，必須先將受試資料按照規定建立資料檔（如圖三所示），資料檔的格式必須為純文字，其中包括五個主要部分：

（一）資料控制列(the control line)

寫在資料檔的第一列，必須依據表二所示欄位順序鍵入指定資料。

表二 MicroCAT 受試資料檔格式表

欄位	資料	說明
1-3	欲分析的試題數目	ASCAL 最多到 250，ITEMAN 與 RASCAL 則可到 500，第一欄表示百位，第二欄表示十位第三欄表示個位。
4	空白	
5	省略未答反應(omitted response)代碼	數字或英文字母皆可，如 0 或 N。在 ASCAL 運算中，省略未達反應與未答完反應均視為答錯。
6	空白	
7	未答完反應(not reached)代碼	數字或英文字母皆可，如 1 或 Q。
8	空白	
9-10	受試者身份代碼數目	最多到 80 碼，第 9 欄是十位，第 10 欄是個位。如以身份證號碼作為代碼，則在此輸入 10。若 9-10 欄是空白或 0，受試者資料就得從第一欄開始輸入。

(二) 答案列(the keyed responses)

資料檔的第二列就是答案列，第一欄就是第一題的正確答案，依此類推。所有答案只能輸入在第二列上。就二元計分而言，可以輸入數字 1-9 字母 A-I，字母大小寫視為相同，且 $A=a=1$ ， $B=b=2$ ，依此類推到 $I=i=9$ 。

(三) 選項數(the numbers of alternatives)

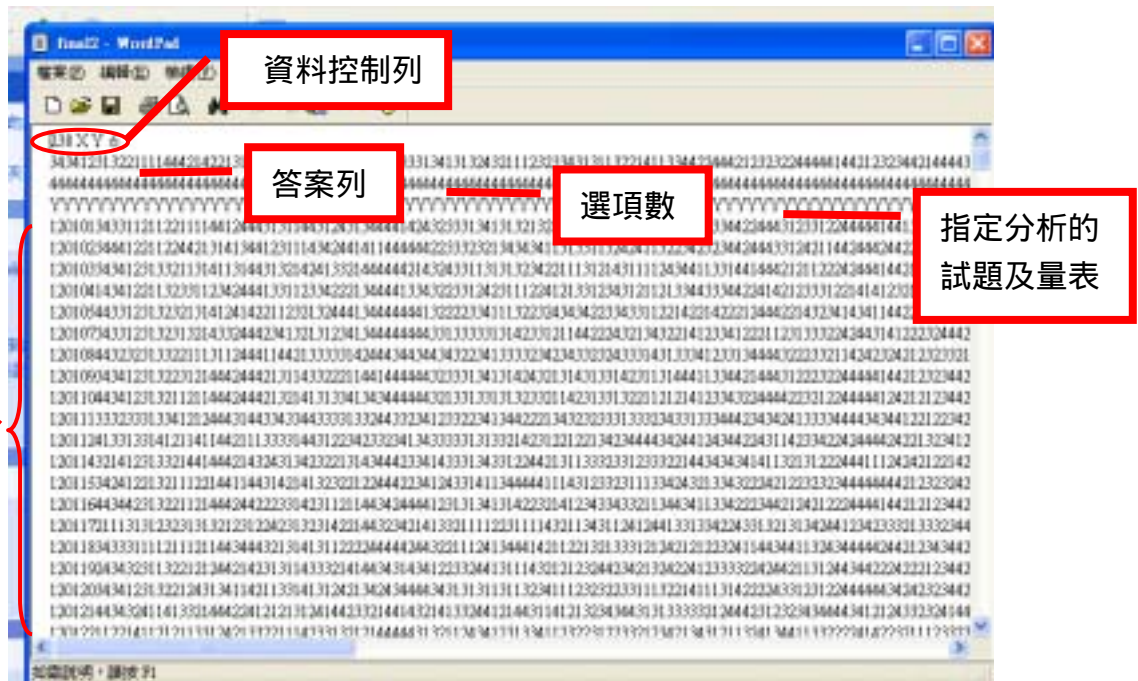
資料檔的第三行指定的是每一題的選項數，就二元計分而言，四選一的選擇題選項數就是 4，選項數的範圍介於 2-9 之間。在 ASCAL 中，選項數用在最初的猜測參數 c 的估計，如選項數是 4，則初始的 c 值則為 0.25。若在 ASCAL 把選項數設為 0，則表示不估計 c 值，也就是代表進行雙參數模式的試題分析。

(四) 指定分析的試題及量表(the scale inclusion codes)

資料檔的第四列的字母，在 ASCAL 中如果是 Y 表示該試題要分析，若是 N 則表示不分析。

(五) 受試者資料(the examinee data lines)

資料檔的第五列之後的每一列都代表一筆受試者資料。呈現的順序是先受試者身分代碼，後面接著受試者答題資料。試題數與身分代碼所占欄位上限為 255，亦即試題數若達上限 250 題，身分代碼只能有 5 欄。



圖二 資料檔格式

另外如需要在資料檔中加入備註說明，則說明文字必須放在每一列的最右邊，且與資料碼中間必須有一空白，即第一列說明從第 12 欄開始。第 2.3.4 列的說明從 $n+1$ 欄開始， n 為試題數。第 5 列之後的說明從 $m+1$ 開始， m 為試題數加上身分代碼數。

三、三參數試題分析軟體 ASCAL 操作

使用 ASCAL 前，得先了解其估計範圍。就 a 參數來說，限制在 0.4 到 2.5 之間。而 b 參數限制在 -3.0 到 3.0 之間。 C 參數限制在 0 至 $2/N$ 間， N 選項數。

(一) 準備資料檔

資料檔格式如前段所整理格式。需要注意者在於省略未答與未答完試題反應的差別。ASCAL 將省略未答者視為能力不足只能猜測答案，亦即其答對機率等於 $1/N$ ， N 為選項數。而未答完反應則不在該受試的分析之中。剩餘的試題反應只要不是正確答案，亦非以上兩者時，一律視為答錯。

(二) 執行程式

執行步驟如圖三所示。

```
MicroCAT (tm) Testing System
Copyright (c) 1982, 1984, 1986, 1988 by Assessment Systems Corporation
Unalis Corp; Milpitas, CA

Item Parameter Estimation Program -- ASCAL (tm) Version 3.20
Enter the name of the input file: c:\data.txt
Enter the name of the output file: c:\outdata.txt

The file already exists -- OK to erase? (Y/N) y
Do you want scores written to a file? (Y/N) y
Enter the name of the score file: c:\out.txt

The file already exists -- OK to erase? (Y/N) y
Maximum number of loops for estimation: 20_
```

圖三 MicroCAT 三參數試題估計軟體 ASCAL 操作過程

步驟一：進入 MicroCAT 主程式，選擇 ASSESS 選項，選擇 ESTIMATE 選項。

步驟二：出現 Enter the name of the input file:

輸入資料檔檔名以及路徑，如，c:\data.txt 或 c:\data.dat。資料檔必須以純文字格式輸入。

步驟三：出現 Enter the name of the output file:

輸入試題分析完的輸出檔檔名及路徑，如 c:\outdata.txt，若不輸入副檔名時，系統會自動加上.out。

步驟四：出現:Do you want scores written to a file?<Y/N>

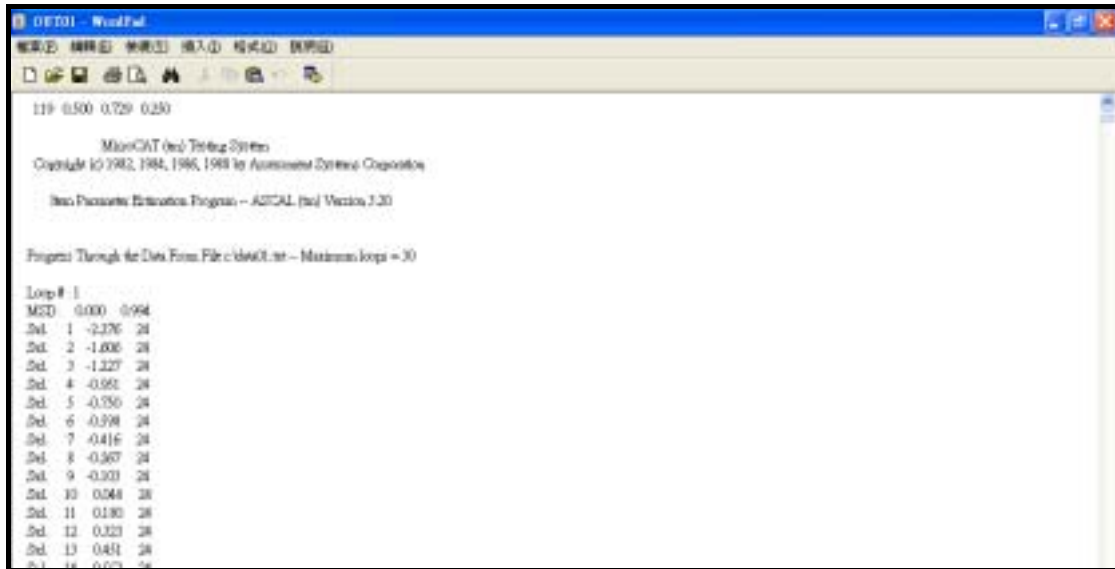
詢問是否要將受試者能力值存檔，如要則輸入 Y，並輸入檔名及路徑，如 c:\out.txt。若不需要，則輸入 N。

步驟五：出現:Maximum number of loops for estimation:

詢問參數估計最大迭代次數。因為試題參數的估計和能力估計一樣都必須經過牛頓法的迭代過程，每次都會使用前次的估計值作為起始點來作一次新的估計，一直到前後兩次估計值的差異收斂到預定的範圍為止。良好的估計通常是 5 至 10 次，大約 15 次便可滿足所有試題(本範例設定為 20 次)，如果在指定的迭代次數之前便已經收斂，或是重複迭代也無法降低差異時，程式便會停止估計，輸出估計結果。

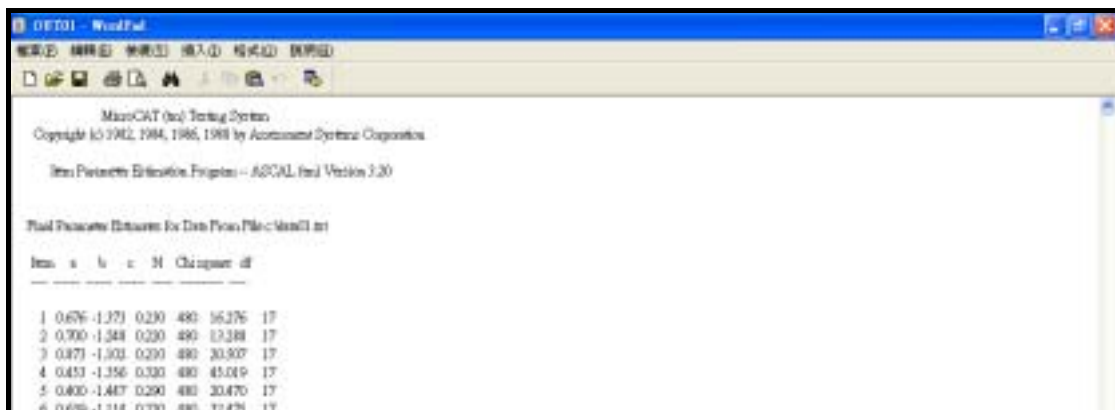
步驟六：估計完成出現:**ITEM CALIBRATION IS COMPLETE**

再下一段資料則是每次試題參數估計迭代過程的參數差異值，如圖六所示。每次迭代過後的參數差異值會逐漸降低，直到低到設定標準或設定的迭代次數為止，一般設定約 15 次即可(本範例設定為 20 次)。



圖六 輸出檔之三

在檔案的最後才是我們要的精確參數（如圖七所示），並顯示各試題卡方值(chi-square)，而當人數夠多之後，自由度通常都是 17。



圖七 輸出檔之四

最後的輸出結果可以參考前面所述，利用卡方值檢驗 IRT 假定的符合程度，以說明題庫的性質。估計出的三參數則可應用於 CAT 中，用以建立題庫。

肆、結論

由於資訊時代的來臨，教學環境的電腦化是未來的趨勢之一，因此評量系統的電腦化也是必然的，IRT 具備有強勢假設，也能因應受試者能力做出適性的試題選擇，不僅提高測驗效率也能符合適性需求，更能確實掌握受試者能力，這也正是所有測驗的共同目標，所以我們可以預期將來測驗發展趨勢，將朝向電腦

化、適性化的 CAT 為主要趨勢。

採用 IRT 的 CAT 在實際推廣上最大的困擾之一在於複雜的理論與數學公式，而 IRT 的核心觀念便是利用能力估計與選題策略來確認受試者能力，這兩部分都需先求得試題參數，否則無法進行。一般教師（施測者）若要使用 CAT，必須要先建立符合 IRT 假設及具備參數的題庫。但是依據公式計算參數對於大多數的教師來說是一項相當沈重的負擔，目前此一繁雜步驟可以透過簡單的電腦軟體處理。完成試題分析的工作後，才能真正跨出 CAT 的第一步。

MicroCAT 是一套兼具傳統試題分析與 IRT 試題分析的軟體，操作簡易的介面讓使用者可以輕鬆處理 IRT 的試題參數估計，省去繁雜的計算，希望能藉由本文的介紹，提供給對 CAT 及 IRT 有興趣的教師一個實用的試題分析工具及一些實例說明，使其能有具體的瞭解與體驗。

參考文獻

- 王寶壙(1995)。現代測驗理論。台北：心理出版社。
- 余民寧(1992)。試題反應理論的介紹(四)能力與試題參數的估計。研習資訊, 9(3), 6-10。
- 何榮桂(1999)。量身訂製的測驗-適性測驗。測驗與輔導, 157, 3288-3293。
- 劉湘川、施淑娟(1995)。試題反應理論參數估計電腦程式之簡介。測驗統計簡訊, 6, 7-12。
- Hambleton R.K., Swaminathan H.(1985).Item Response Theory:Principles and application. Boston, Kluwer Nijhoff.
- Hambleton R.K., Swaminathan H.,& Rogers H.J.(1991).Fundamentals of item response theory. Newbury Park, CA: SAGE.
- Load F.M.(1980).Application of item response theory to practical testing problems. Hillsdale, N.J.:Lawrence Erlbaum.