

# 遠距測驗—Dear CAT 的設計與實施\*

何榮桂

國立台灣師範大學資訊教育研究所

(最後修改日期 86 年 4 月 8 日

接受刊登日期：86 年 4 月 16 日)

## 摘要

本研究旨在結合項目反應理論(item response theory, IRT)及適性測驗(adaptive testing)之理論與技術,以三層式主從遠距測驗系統架構(three-tier client server distance testing system architecture),在 Internet 上發展遠距適性測驗系統(remote computerized adaptive testing),以克服時空的限制,提供遠端使用者一個開放的施測環境。在適性測驗的過程中,受試者可接受適合自己能力範圍的題目,以較少的試題獲得更準確的結果。本研究分為三個步驟:首先規畫一個具平台獨立之遠距測驗系統架構,其次,依據項目反應理論設計適性測驗系統與流程,最後將系統實作於 Internet 的環境中。希望能夠透過此設計架構,為遠距適性測驗提出一個可供參考的設計模式,也為未來做進一步研究預先架設一個實驗平台,以利進行各種電腦模擬研究。

**關鍵詞:** 項目反應理論、適性測驗、網際網路、遠距測驗

## 一、前言

遠距適性測驗 (Distance Evaluation: A Remote Computerized Adaptive Testing ; Dear CAT) 是一個建構於 Internet 的全球廣域網路(world wide web ; WWW)環境下的測驗,旨在提供學習者一個評量學習成效的環境。學習者只要透過網路就可以由遠端的個人電腦,連線到主機端接受測驗,學習者可以自行選擇適合的能力測驗、態度量表、或是練習題目,這些測驗單元全部都必須融合適性測驗的觀念,也就是要依據學習者的能力給予適當程度的題目。

---

\* 本系統係筆者所執行之國科會科教處個別型計畫(網路環境題庫與測驗之整合系統)(NSC86-2511-S-003-017-cl)之一部分。本文係在台灣師大科教研究所(85.11.28)及中山大學資工所做過專題報告(86.3.26)後改寫而成。感謝李田英所長及楊竹星所長的邀請與其師生同仁的指教。本系統之所以順利完成,得力於我實驗室(CAT Lab)內所有的大貓小貓(郭再興、蘇建誠、賴信仁、陳麗如、秦靜儀、簡月梅、王鼎中、籃玉如、王緒溢及顏龍源),她/他們長期給我的支持與協助,由衷感激。物理教育學會會刊主編台灣師大物理系楊遵榮教授邀我寫這篇稿子,也令我受寵若驚,筆者也希望此系統會對物理科的評量有所助益,謝謝他對我們做的研究有興趣。Dear CAT 的誕生是筆者這一、二年研究過程中感到最興奮的事,這個暱稱希望能給冷冰冰的測驗帶來些許較親和的氣氛。

遠距測驗系統最大的優點就是提供了一個不受時空限制的施測環境，在此種開放性的環境中，學生不僅可適時、適地、適性的接受測驗，老師也可根據其豐富的教學評量經驗，提供品質優良的測驗題目，如此即可集眾人之力不斷的更新測驗系統的內涵，教、學雙方都可以獲得資源共享所帶來的便利。此外，當「終身學習」的觀念逐漸被一般人接受時，遠距教學已是必然的學習模式，相對的也彰顯出遠距測驗的必要性。因此考慮到目前的需求與未來發展的趨勢，因而有規畫遠距適性測驗系統之構想(何榮桂、蘇建誠、郭再興，民 85)。

## 二、測驗的發展趨勢

在開始進行系統規畫之前，宜先針對測驗的發展趨勢略作說明，以掌握正確的研究方向。目前測驗的發展趨勢可概略敘述如下：

### (一) 由古典或傳統測量理論轉變為現代測量理論

難易適中的題目最能夠評鑑出受試者的能力，但是應用傳統測量理論的紙筆式測驗卻無法同時顧及到每位受試者的能力水準，以至於受試者不論能力高低都必須接受同樣的題目，這樣的設計往往不能精確的評估考生的能力，爲了要使題目的難度能夠隨著受試者的個別差異而調整，適性測驗的概念便受到了重視。項目反應理論(item response theory; IRT)的提出(Hambleton & Swaminathan, 1985)，爲適性測驗提供了嚴謹的理論基礎，受試者只需接受適合其能力的題目，就可以精確的估算出能力值。此外，IRT 還有下列優點：題目參數不因受試樣本的不同而改變；受試者能力估計不受題目樣本不同的影響；以及測量誤差的估計因受試者之能力水準而異。因此，可以預期 IRT 將會廣泛的應用在各個測驗領域上。

### (二) 由紙筆式測驗轉變為電腦化測驗

一般常用的測驗方式多爲傳統的紙筆式(paper-and-pencil)施測，但隨著電腦的快速普及，借重電腦的快速、精確特性，學生可在測驗結束時，立即獲知自己的受測結果，節省了閱卷、評分的程序，電腦也可以搜集紙筆式測驗難以記錄的反應資料(例如學生的作答時間、反覆修改次數、答題順序等)，以供更進一步的分析。另外在適性測驗的過程中，每一階段(即每作答一次)需要不斷進行大量的數值計算以推估學生的能力、選取下一個最適當的題目，這都必須以電腦作爲施測工具才得以實施。

### (三) 由單機的施測環境轉變為網路化的施測環境

隨著網路技術的快速發展，電腦逐漸深入社會的各個階層，透過網路的連線，地域的界限已經逐漸的模糊，所有的網路用戶都能以最便捷的方式獲得資訊。就測驗環境而言，也是由單機施測慢慢演變至區域網路的施測環境，而有更進一步的發展出以網際網路(Internet)爲骨幹的遠距測驗的必要，其目的都是希望能結合網路化的優點，提供更開放的空間、更彈性的施測環境。

### 三、系統發展目的

基於上述，發展遠距適性測驗系統實為正確且必要的工作，因此本系統主要發展方向包括下列三點：

(一) 規畫遠距測驗系統架構

此架構可適用於各種作業平台，同時也考慮不同種類的測驗(如傳統成就測驗、適性測驗、性向測驗或是一般的問卷量表)。

(二) 依據項目反應理論設計適性測驗系統。

(三) 實作遠距適性測驗系統

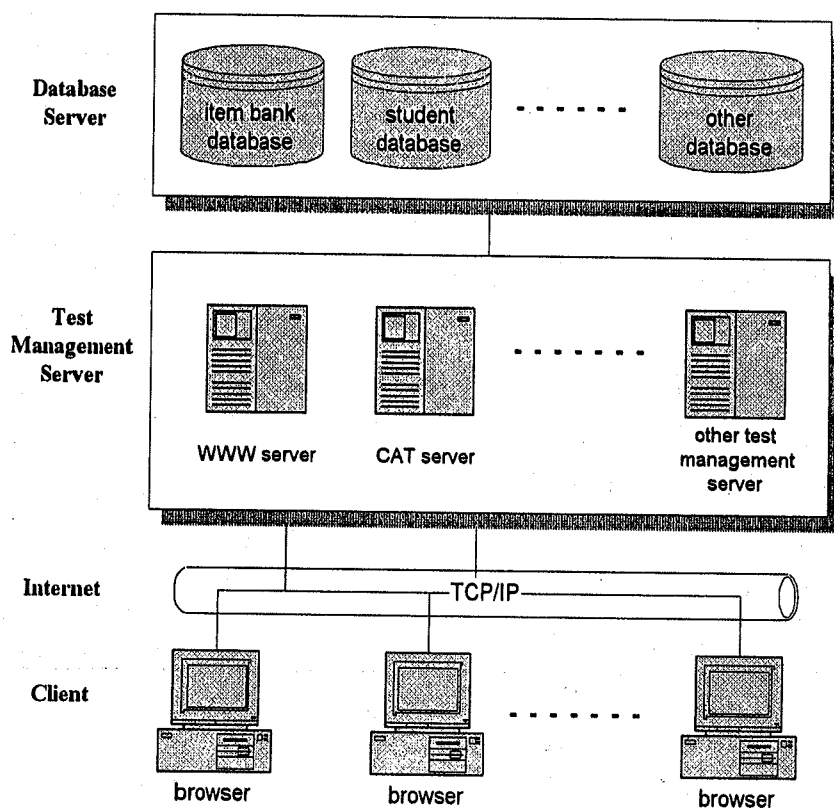
根據前述的遠距測驗系統架構與適性測驗系統，建構出遠距適性測驗系統。

### 四、系統發展方法與步驟

系統發展主要可分為三個步驟：首先規畫一個具有平台獨立性的遠距測驗系統架構，然後再設計適性測驗系統，最後將系統實作於 Internet 的環境中(何榮桂等，民 85)。茲將方法與步驟依序說明如下：

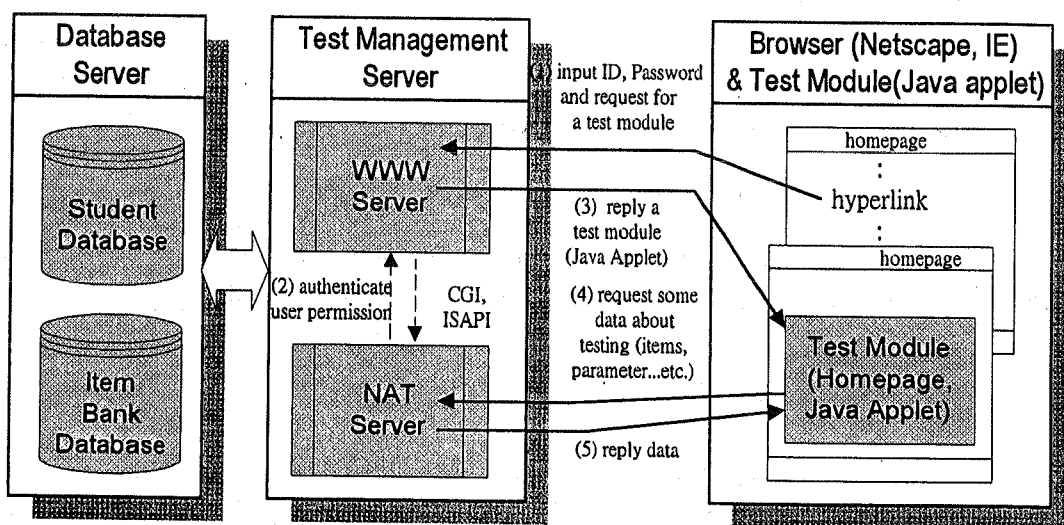
(一) 三層式主從遠距測驗系統架構與實施方式

此系統為一三層式主從遠距測驗系統架構(three-tier client server distance testing system architecture)，以 Internet 為網路骨幹之系統架構如圖一所示：



圖一 三層式主從遠距測驗系統架構

在圖一中，資料庫伺服器(database server)職司管理不同種類的題庫以及受試者資料庫；測驗管理軟體(test management server)則針對某特定目的測驗，負責處理題目擷取、管理上線受試者所設置的 server，包括 WWW server、CAT server (computerized adaptive testing server)以及其他 test management server，這些 servers 都需具備互相通訊的能力，以便進行資料交換。Client 端因考慮到使用者的方便性、測驗的安全性以及適性測驗需用到大量數值計算，因此採用一般的 browser (如 Netscape, Internet Explorer 等)，並結合 Java 技術來規畫。此種設計使系統更具彈性，client 端的測驗軟體、主機端的 test management server 或是 item bank server 皆可獨立開發，整個系統可隨時加掛不同的 item bank server 或 test management server。如果 server 的數量太多使得系統效能降低，也可以改由多台電腦分別執行 server 軟體，達到分散處理的功效。詳細的系統實施方式如圖二所示：



圖二 遠距測驗的實施方式

首先在主機設置 database server，負責管理題庫與使用者資料，並架設 WWW server 和 CAT server；然後 client 端的使用者利用一般的 browser 連上 WWW server，在 homepage 所列示的各種測驗中，點選要作的測驗(或問卷、量表等)，透過 hyper linker 由主機傳回測驗模組。這些測驗模組可能只是一個簡單的 homepage，使用者填答完 homepage 上的問題後，再將結果送回 WWW server，交由 CGI(common gateway)或 ISAPI(Internet service application program interface)處理；如果是適性測驗，則測驗模組就是一個 Java applet，使用者藉由此測驗模組直接與 CAT server 通訊，進行適性測驗，而 CAT server 根據測驗模組的要求，到題庫中選擇適當的題目回傳給使用者。在測驗的過程中，CAT server 必須負責管制受測人數、檢查使用者身份、記錄使用者上線次數、記錄受測時數、錯誤檢查(error detection)、錯誤回復(error recovery)等工作，並隨時監控使用者的施測情形，以保護題庫安全。此種步驟將一直持續到適性測驗結

束為止，使用者可以再選擇其他測驗或是離開系統。

## (二) 遠距適性測驗系統之設計

### 1. 適性測驗的基本原理

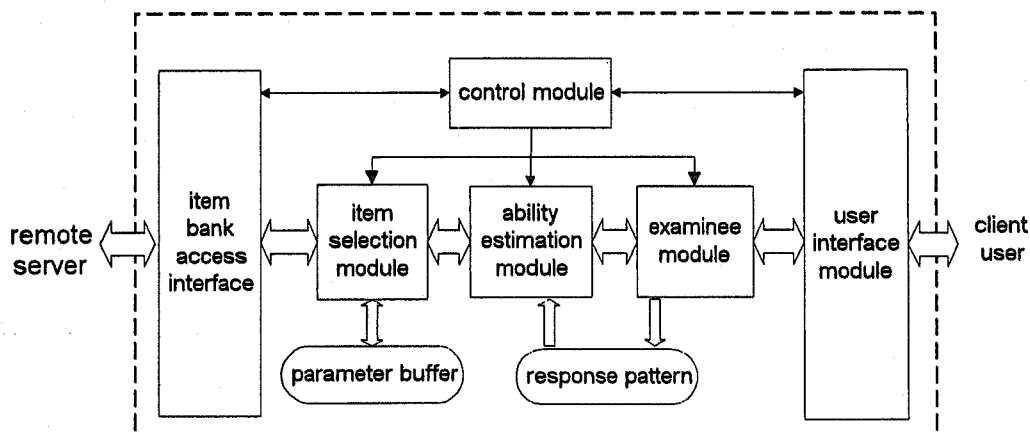
適性測驗係應用項目反應理論的原理，藉電腦來施測的一種個別化測驗的實施策略。基本上其施測程序僅根據一個簡單的原理：倘受試者答對了一題則接下來的題目會比前一題難一點；若答錯了一題則後一題會比前一題簡單一些。循此程序，直到被估計的能力特質達到足夠精確為止。發展電腦化適性測驗與傳統的紙筆測驗迥異，主要須考慮下列要項(Ho, 1989)：

- (1) 採用一個適當的項目反應模式(例如 three-parameter logistic model)，在 CAT 的研究裡，使用三參數模式較多。
- (2) 建立題庫(calibrated item bank)，原則上所含題數至少 100 題以上，試題參數(item parameters)之分配型態視測驗的目的而定。
- (3) 決定測驗的起始點(starting point)；也即推估受試者起始能力估計(initial ability estimate)。
- (4) 選擇一個有效率且精確度高的計分與選題策略。目前被採用較廣的有 Bayesian, modal Bayesian, 與 maximum likelihood 等策略。
- (5) 決定測驗的終止點(termination criterion)。施測的結束條件可由估計標準誤(standard error of estimate; SE)或測驗訊息(test information;  $I(\theta)$ )量推算而得；也可由所欲施測的題目數(如最少或最多幾題)來決定。

有關適性測驗的理論與技術皆已相當成熟(Mills & Stocking, 1996; Dorans, Flaughter, Green, Mislevy, Steinberg & Thissen, 1990)，其細節不在此贅述，本文僅就建置於網路環境之系統架構與實施加以敘述。

### 2. 遠距適性測驗系統架構

如圖三所示，遠距適性測驗系統可分為六個模組(module)，各模組均可獨立設計，並可更換模組內容以改變適性策略，茲將圖三中各模組之功能及其相互間之關係，說明如下：

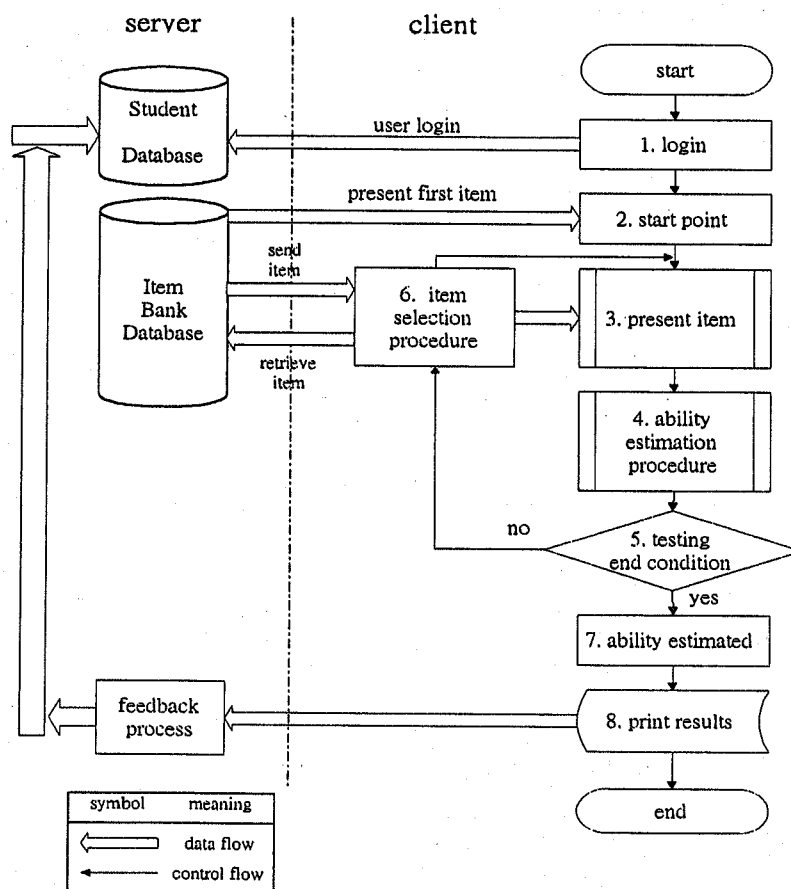


圖三 適性測驗系統架構

- (1) 題庫存取介面(item bank access interface)係以 Internet 標準之 TCP/IP (transmission control protocol/internet protocol)為通訊協定，訂定與 CAT server 間的通訊基本用語。
- (2) 選題模組(item selection module)採用最大訊息法(maximum information strategy)，根據能力估計值(estimated ability)在參數緩衝區中取出每一題的 a,b,c 參數值，並逐一計算該題的訊息量，以找出最適合的題目，再透過題庫存取界面從題庫選取題目。
- (3) 能力估計模組(ability estimation module)採用貝氏估計法(Bayesian procedure)，依據受試者的反應組型(response pattern)估算能力值，再經選題模組取得次一題目。
- (4) 受試者模組(examinee module)負責記錄受試者的個人資料、測驗的反應組型、及施測時間等資訊。
- (5) 控制模組(control module)負責協調各模組之間的動作、決定測驗起始點(start point)、判斷是否已達測驗結束條件、並與 CAT server 保持聯繫。
- (6) 使用者介面模組(user interface module)則負責處理系統與使用者間的溝通，以提供一個親和的操作介面。

### 3.適性測驗流程

根據前述適性測驗的原理，本系統之適性測驗流程可以圖四表示。圖中左半部有關學生資料庫、題庫均由 server 負責管理，右半部則屬於 client 的工作。中間虛線部份即為兩者的溝通介面。



圖四 遠距適性測驗流程

首先從圖四右邊 client 的順序，說明本系統之適性測驗流程：

- (1) 使用者輸入自己的姓名及密碼等身分資料，經 server 端確認後則開始測驗。
- (2) 電腦可參考學生過去的紀錄，或是其他策略(如受試者自陳方式)決定施測起始點，然後由題庫中呈現第一個題目給受試者。
- (3) 將題目呈現在螢幕上等候受試者作答。
- (4) 受試作答後依反應組型估計其能力。
- (5) 判斷是否符合測驗結束的條件，若尚未滿足終止條件，則繼續下一步驟。
- (6) 根據受試者能力估計值，在題庫中擷取符合其能力的題目。接著重複步驟(3)、(4)、(5)及(6)。
- (7) 若受試者的能力估計已經達到測驗終止標準，便以最大概似估計法(maximum likelihood estimation; MLE)重新估計其能力不偏估計值。
- (8) 顯示測驗成績，將測驗結果輸出給受試者並送回學生資料庫中儲存，以便分析。

#### 4. 遠距適性測驗系統的基本雛形設計

為測試本系統架構之可行性，本研究首先依據架構，設計系統的基本雛形，包括 IRT-based 之三參數題庫及其他資料庫，CAT server 及 WWW server，以及 Java applet 之適性測驗模組，茲分述如下：

- (1) 建置三參數(three-parameter logistic model)模式的 IRT 題庫(在評估系統雛形時，係以模擬之參數資料進行測試)、使用者資料庫以及反應組型資料庫。
  - a. 題庫的內容包括題號、題幹、選項、答案、提示語、題目屬性類別、試題參數(a: 鑑別度參數 b: 難度參數 c: 猜測度參數)及題目使用次數等。
  - b. 使用者資料庫應記錄使用者 ID、密碼、個人基本資料、每次上線受測的記錄、使用時間、及次數等。
  - c. 每位受試者每次施測所產生的反應組型都儲存在反應組型資料庫中，以作為爾後進行線上參數校正(on-line calibration)(何榮桂、賴信仁，1997)。
- (2) CAT server 的設計：以 Delphi 程式語言設計，其功能包括檢查使用者身分、可供多人上線使用、依照需求傳回題目、資源分配與管制、監控使用者的施測情形、錯誤檢查、錯誤回復、施測結果解釋以及維護題庫安全。
- (3) 建置 WWW server：架設 WWW server 以測試 CGI, ISAPI 和 Java applet 程式。
- (4) 依照適性測驗系統之規畫，設計出 Java applet 之適性測驗模組。

上述系統之基本雛形設計完成後，繼續進行整體測試，包括 a. 製造各種錯誤狀況(例如不正常離線、使用者重複登入等)以驗證系統是否正確執行， b. 針對各種可能影響題庫的行為，進行安全性測試， c. 以電腦模擬大量上線者，檢

驗系統的最大承受力。

## 五、系統效能試行評估

遠距適性測驗系統基本雛形設計完成後，則先以模擬資料評估系統效能(何榮桂，1991)，評估方式如下所述。

### (一) 模擬評估

在評估系統效能方面，係採用電腦模擬題庫中參數資料與受試者能力值，以產生適性測驗的反應組型，以下即依序說明模擬資料的產生方式，以及分析的方法。

(1) 以電腦產生一個題數為 1000 題，參數特性為均勻分配(uniform distribution)之模擬題庫，題庫中的試題參數產生方式如下：

$$a_i = 0.5 + (2 - 0.5) \times p_1,$$

$$b_i = -3.0 + (3 - (-3)) \times p_2,$$

$$c_i = 0 + (0.35 - 0) \times p_3,$$

$$p_1, p_2, p_3 \xrightarrow{iid} U(0,1),$$

其中  $i = 1, 2, \dots, 1000$  代表題目代碼，

$p_1, p_2, p_3$  為彼此獨立的均勻分配的隨機變數，

$a_i, b_i, c_i$  代表第  $i$  題的參數值，且

$$0.5 \leq a_i \leq 2, \quad -3 \leq b_i \leq 3, \quad 0 \leq c_i \leq 0.35$$

(2) 以電腦模擬產生界於 -3.0 到 3.0 呈常態分配之受試者能力值  $q_j$ ，即：

$$\theta_j \sim N(0,1), \quad -0.3 < \theta_j < 0.3, \quad j = 1, 2, \dots, 1000$$

其中  $j$  為受試者代碼

(3) 產生反應組型資料

- a. 將前述所產生的題庫與受試者資料，進行電腦模擬測驗。當第  $j$  位受試者接受第  $i$  題目的模擬測驗時，其能力值  $q_j$  與所受測的題目參數( $a_i, b_i, c_i$ )代入下式中，可以得到一預期答對機率  $p_{ij}$ ，即：

$$P_{ij}(\theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{1}{1 + e^{-D a_i (\theta_j - b_i)}}$$

其中  $i = 1, 2, \dots, 1000$ ， $D = 1.702$

- b. 再由電腦自 0 到 1 的均勻分配中，隨機產生一個反應係數  $r$ ，當此  $r$  值小於預期答對機率  $p_{ij}$ ，則表示第  $j$  名受試者答對該題；反之，若  $r$  值大於預期答對機率  $p_{ij}$ ，則代表第  $j$  名受試者答錯該題，其反應情形  $u_{ij}$  如下：

$$U_{ij}(p_{ij}) = \begin{cases} 1, & r \leq p_{ij} \\ 0, & r > p_{ij} \end{cases} \quad r \sim U(0,1)$$

- c. 以向量  $U_j = [u_{1j}, u_{2j}, \dots, u_{ij}]$  ( $i=1000$ )，代表第  $j$  位受試者的反應組型，將  $U_j$  代入能力估計器，以 Bayesian MAP(maximum a posteriori)方式估計出該名受試者的能力估計值  $\hat{\theta}_j$ 。



(4)將適性測驗系統的終止條件，設定為估計最小標準誤(minimum standard error of estimator; SE)等於 0.5, 0.4, 0.3, 0.2 的情況下時，計算其平均施測題數，並採用相依樣本 t 檢定(pair t-test)來檢定 $\theta_j$ 與 $\hat{\theta}_j$ 的差異情形。

(5)分析結果(何榮桂、郭再興，1997)

經模擬測試結果顯示，當終止條件 SE 設為 0.5、0.4、0.3、0.2 時，測驗平均長度分別是 4.49、6.67、11.21、25.94 題，此結果顯示確實可以有效的縮短測驗長度，結果如表一所示。

表一：不同終止條件之施測題數分析

終止條件	施測題數			
	平均數	標準差	最小值	最大值
SE ≤ 0.5	4.4860	0.7826	4	10
SE ≤ 0.4	6.6650	1.2249	5	14
SE ≤ 0.3	11.2070	1.4326	9	18
SE ≤ 0.2	25.9420	2.3082	22	41

(模擬受試者人數為 1000 人)

由表一中的數據可知，當終止條件較為寬鬆時(SE ≤ 0.5)，在較短的施測題數下(平均 4.4860 題)能力估計值就能夠收斂到終止條件所要求的標準，而且在所有模擬受試者的反應資料中，最多也只要作 10 題便可結束測驗。終止條件如果比較嚴格時(SE ≤ 0.2)，受試者所要接受的施測題數也隨之增加(平均 25.9420 題)，但相對的能力估計值也會比較準確。

本研究中，能力估計係採用 Bayesian MAP 估計法，在不同的終止條件下所得到的能力估計值，經檢定與模擬資料(即受試者能力值)均無顯著差異，結果如表二所示。

表二：不同終止條件之能力估計值與模擬能力值之差異性檢定

終止條件	能力差值平均	標準誤	t 值	p 值
SE ≤ 0.5	0.0050	0.0143	0.3484	0.7276
SE ≤ 0.4	-0.0015	0.0121	-0.1249	0.9006
SE ≤ 0.3	-0.0067	0.0092	-0.7361	0.4619
SE ≤ 0.2	-0.0026	0.0064	-0.4001	0.6891

(模擬受試者人數為 1000 人)

由表二資料可知本系統在估計受試者能力時，確能有令人滿意的效果。至於在實際使用時終止條件應該設定為多少，則可由系統管理者視情況需要指定之。

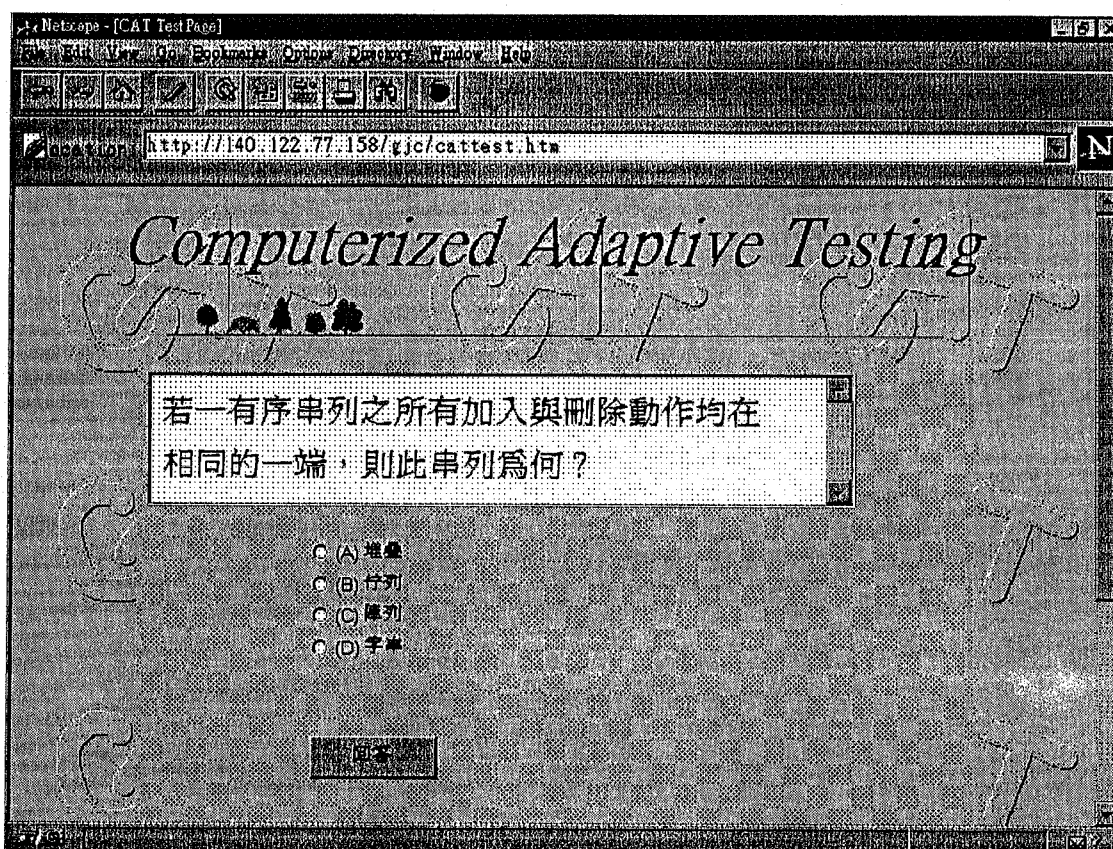
## (二) 系統實作

目前本研究已經完成建置一個內容為「高級中學電腦概論」的題庫系統，題目數量為 50 題，題目參數暫時以隨機方式產生，以便進行模擬測試驗證系統效能。俟實際搜集到受試者作答資料後，再以參數校正軟體 MicroCAT (Assessment Systems Corporation, 1987)估算實際的參數值。

在 server 端則是架設 Microsoft NT server 3.51、WWW server (Internet information server)、撰寫 homepage 以及 ISAPI 程式，以供使用者上線選擇欲作的測驗(參考網址: <http://140.122.77.158/cat/index.htm>)；另外 CAT server 的基本

功能也已經完成，可以隨時監控上線使用者的施測情形，如果超過 5 分鐘都未收到 client 端任何訊息，則視為通訊中斷，系統將立即把該名受試者目前的測驗資料備份，等下次重新恢復連線時可以再繼續進行測驗。

在 client 端已經設計完成 Java applet 適性測驗模組(詳見圖五)，可以與 server 端的 CAT server 通訊取得測驗題目，並控制測驗進行的過程。在系統的整體表現上，若同時有 20 名模擬受試者由遠端上線測試，在能力估計與選擇下一題目的速度上，系統的平均反應時間在 2 秒之內，已可以滿足基本的要求。不過系統反應時間與 client 所使用的電腦速度，以及當時的網路負荷量有密切的關係，因此將來實際開放使用時，可能會有些需的差異，但不會影響施測的進行。



圖五 Java applet 適性測驗模組  
(參考網址: <http://140.122.77.158/cat/index.htm>)

## 六、可能遭遇的問題與瓶頸

### 1. 施測模式

目前本系統採用的題庫為單純的文字模式，其優點在於題目的編輯、轉換都非常簡便，且在網路傳輸時因資料量小，所以速度快。但目前在 WWW 上均已提供多種媒體呈現方式的使用環境(如音效、動畫、影像、虛擬實境等)，因此有必要將題庫的內容，提昇為支援多媒體的施測環境，以發展出多媒體的網路適性測驗。本研究小組將繼續設計多媒體的施測方式，以使本系統之施測環境更具多樣性。

### 2. 題庫安全性問題

建置一個質優量足的題庫，必須花費相當大的人力、物力。爲了維持題庫的品質與題庫的安全，需考慮下述問題。

(1) 題目曝光率

品質佳的題目常會被過度使用，增加曝光率，如果一個題目的曝光率太高，將會成爲猜題的對象，以致失去了題目的隱密性。因此必需控制題目的使用率，以保障題庫的品質(Stocking & Lewis, 1995)。

(2) 題目失竊

一個具有代表性的測驗系統，往往會成爲有心人士窺探的目標，彼等可能會藉由不同的使用者身分，不斷的上線使用測驗系統以獲取題庫的內容。針對這個問題，一方面需加強檢核使用者身份，確認上線者爲合法使用者才提供服務，如果在施測的過程中出現不尋常的反應組型(例如連續的答對或答錯)，系統也必須及時發現，以防止系統遭人侵入。另一方面也應時時擴充題庫的內容，並控制題目的出現率，使得題目數量夠多，如此就算試題庫完全的被公開，也因爲題目呈現的隨機性，而仍可維持題庫的安全性。

(3) 通訊加密

網路是一個開放的環境，在傳輸題目資料時也有必要將題目予以編碼加密，防止題目在傳輸過程中被攔截。常用的編碼技術如 DES (data encryption standard)，以及在網路上所使用的通訊安全規格：如 Netscape 公司提出的 SSL (secure sockets layer) 以及 Microsoft 公司所發展的 PCT (private communication technology) 均可作爲系統實作上的參考 (Chapman, D., Igal, S. W., Beem, W. R., Sdler, K., Dumbrill, D., Thompson, D. & Medinets, D., 1996, pp.428-443)。

### 3. 線上參數連結

建置符合現代測量理論模式的題庫，首先必須找到大量的受試者進行預試，搜集足夠的反應組型，以估計每一題的參數值，此過程十分耗費成本，如果能利用網路的特性，直接在線上陸續搜集上限受試者的反應組型，進行參數校正及連結的工作，就可以動態的調整題庫的內容加入新的題目，也可以將過時、不適用或曝光太高的題目予以剔除，節省下可觀的施測成本(何榮桂、賴信仁，1997)。

## 七、結論

結合項目反應理論及適性測驗，以三層式主從遠距測驗系統架構，在 Internet 上發展遠距適性測驗系統，是本研究之主要目的。除了希望能夠透過此種設計架構，爲遠距適性測驗提供一個可供參考的設計模式外，也爲未來做進一步研究預先架設一個實驗平台，以利進行各種電腦模擬研究。至目前爲止，筆者對本系統之運作頗有信心，但是網路技術變化快速，我們須時時注意各種理論與技術的更新，使系統維持在一個最佳的狀態。

以上所述僅是我們研究小組在執行「網路環境題庫與測驗之整合系統」之前期研究成果的一部分，適性測驗的理論與技術雖已相當成熟，但將此理論與技術建構在網際網路環境裡，我們小組的研究應是草創的工作；換言之，在網際網路裡，我們尚未看到類似的研究或設計，這是讓我們感到相當興奮的工作。

由於本文所述之系統是針對最常用的能力測驗而設計，因此採用最常見之兩點計分模式(binary scored item) 為主，而其他如多點計分之IRT模式(如 graded response, partial credit, and rating scale 等模式)也漸被使用，事實上我們並沒有忽略這些理論與技術的發展，在發展此系統時，我們也同時進行「遠距適性態度測驗系統設計」(何榮桂、蘇建誠，1997)的研發，此系統即以多點計分模式設計，以適用於多種典型表現測驗(typical performance tests)在線上施測。俟這些系統皆測試完成後，將會整合在 Dear CAT 系統中，以使 Dear CAT 能適用各種類型測驗之線上適性測驗。

最後要說明的是本系統具有一般性或普遍性的性質，也即不限於某一學科的測驗，我們也期望本系統對各學科之教學評量能有實質的助益。

## 參考文獻

1. 何榮桂，〈電腦化測驗概述〉，《現代教育》18期121-129頁，1991年。
2. 何榮桂，〈題庫中項目參數分配型態對電腦化適性測驗選題的影響〉，《中國測驗學會測驗年刊》，第38輯71-96頁，1991。
3. 何榮桂、郭再興，〈網路化適性測驗系統〉，《第六屆國際電腦輔助教學研討會大會論文集》186-196頁，1997年。
4. 何榮桂、賴信仁，〈遠距適性態度測驗系統設計〉，《第六屆國際電腦輔助教學研討會大會論文集》454-463頁，1997年。
5. 何榮桂、蘇建誠，〈項目參數之線上校準研究〉，《第六屆國際電腦輔助教學研討會大會論文集》175-185頁，1997年。
6. 何榮桂、蘇建誠、郭再興，〈遠距適性測驗系統架構〉，《資訊與教育雜誌》42期29-35頁，1997年。
7. Assessment Systems Corporation., User's manual for the MicroCAT Testing System (2<sup>nd</sup> ed.), St. Paul, MN: Author, 1987.
8. Chapman, D., Igal, S. W., Beem, W. R., Sdler, K., Dumbrell, D., Thompson, D. & Medinets, D., Building Internet applications with Delphi. IN: Que, 1996.
9. Hambleton, R. K. & Swaminathan, H., Item responst theory: Principles and applications, Boston: Kluwer-Nijhoff, 1985.
10. Ho, R.G., Computerized adaptive testing, Psychological Testing, Vol. XXXVI: 117-130, 1989.
11. Gosling, J. & McGilton, H., The Jave(tm) language environment:A white paper, [http://java.sun.com/doc/language\\_environment/\(URL\)](http://java.sun.com/doc/language_environment/(URL)), 1996.
12. Mills, C. N. & Stocking, M. L., Practical issues in large-scale computerized adaptive testing, Applied Psychological Measurement, Vol.9(4):287-304, 1996.
13. Roberts, D., Developing for the Internet with Winsock, NH: Coriolis Group, 1995.
14. Stocking, M. L. & Lewis, C., A new method of controlling item exposure in computerized adaptive testing, ED389739, 1995.
15. Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., Mislevy, R., J., Steinberg, L., and Thissen, D., Computerized adaptive testing:A primer, Hillsdale, NJ: Lawrence Erlbaum Associates, 1990.

\* 致謝：楊遵榮教授及三位匿名審查者對本文的評述及建設性之建議，使本文之缺點或錯誤減至最少，特此致謝。