

中文電腦化適性測驗系統軟體之研發

李 茂 能

國立嘉義師範學院初等教育學系副教授

摘 要

本文旨在簡介中文電腦化適性測驗(CCAT)系統之研發目的、架構與使用方法。首先，說明本測驗系統之研發目的與所涉及之相關理論及技術。其次，分析學校中常用之測量方式，以決定本測驗系統之主要架構：電腦化適性成就測驗(CAAT)、電腦化適性精熟測驗(CAMT)、電腦化適性等第評量測驗(CAGT)與電腦化紙筆式測驗(CPPT)；並陳述各子系統相關之Visual Basic程式物件的結構與功能。最後，概述本測驗系統之各項操作步驟、建立題庫之方法、與安裝本適性測驗系統之注意事項。

本測驗系統之基本特色為(一)測驗中可隨時變換文字與圖形大小，(二)圖文可以整合，(三)可提供診斷性資訊，(四)搭配Otek International公司所發行之「語音精靈」軟體，可以朗讀測驗題幹與選項，(五)具偵測「異常作答行為」之機制，(六)同時具有題庫「參數估計」與「適性施測」之功能，(七)可以播放Midi音樂與聲波檔，(八)題庫中之難度可以自動重估，(九)題庫中題目之曝光率可以自動檢查。

版權與商標聲明

- 1、MS-Windows 95中文版係台灣微軟公司的註冊商標
- 2、MS-Visual Basic 6.0係美國微軟公司的註冊商標
- 3、Speech Wizard 1.0係Otek International公司的註冊商標

壹、緒論

「因材施教」的教育理念自古即普遍受到重視，而「因材施教測」的評量理念則到晚近適性測驗(adaptive testing)興起之後，才受到注意。傳統式之紙筆測驗通常是一個尖峰分配的測驗(a peaked test)，中等難度的測驗題目居多，較難或較易的測驗題目則居少數。考試時，每一位考生不管能力之高低皆需做完試卷上之每一道題目。事實上，此種尖峰分配式的測驗只適合於中等能力之考生。對於那些能力較高或較低者來說，他們常需做一些太容易或太困難的題目，這對於他們能力之估計與區辨常是既不精確又不經濟的作法。因此假如考試時，能夠依照考生的性向或能力高低，挑選不同難度之試題給與施測，必能使考試更精確而有效率。此種適性施測的測驗的先河，可溯至西元1905年時，Binet & Simon 所編製的第一份智力量表。惜因施測時費力費時，此種人工式之適性測驗並未普及。直到1950年代之後，由於IRT (Item Response Theory) 測驗理論的出現，使得測驗客觀性之兩個基本要求：(1)題目參數之估計不因考生樣本的變動而變動(person-free item calibration)，(2)考生能力之估計不受題目之不同而變動(item-free person measurement)，得以容易實現，也使得適性測驗的效度得到理論上之支持。但IRT理論涉及較深奧之測驗模式，其實際之應用則一直到既便宜且高效率之個人電腦普及的九〇年代才變得具體可行。電腦化適性測驗(Computerized Adaptive Testing，簡稱CAT)亦成為當代測驗界之寵兒，這可從眾所皆知的GRE與TOFEL的測驗趨勢加以驗證，ETS已決定從1999年秋季起全球大部份國家與地區將同步實施電腦化適性測驗。

二十一世紀是「資訊科學」顯揚之年代，「試題反應理論」(簡稱IRT理論)與之相結合不只促使了適性測驗變為具體可行(Sands, W. A., Waters, B. K., & McBride, J. R. (Eds.), 1997; Wainer, et al., 1990)，而且讓測驗方式得以動態化與遠距化(何榮桂、蘇建誠、郭再興，民85；孫天光、謝凱隆、鄭海東、陳新豐，民86；Bunderson, C. V., Inouye, D. K., & Olsen, J. B., 1989)，也使得測驗結果更

具診斷性與時效性(王力行，民86；林世華、劉子鍵，民86；Bunderson, C. V., Inouye, D. K., & Olsen, J. B., 1989)。學校教師如能運用CAT快速、精確之作答評分、記錄、分析與多重媒體整合之功能，不僅可以使測驗結果立即迴饋、降低測量誤差、提高多元化教學評量之成效與時效，而且教師之工作負擔亦可減輕不少。對學生來說，考試可因材施教、評量個別化、與縮短考試時間。爲了改善學校情境中慣行之尖峰型紙筆式測驗之缺失與迎合電腦科技的多媒體功能，以利教師改善多元化教學評量之品質與效率，分析與發展一套適合於學校情境之中文電腦化適性測驗系統，以便利進行各科成就測驗、等第分派、精熟評量，實有其必要性與迫切性。

貳、研究目的與適用範圍

本電腦化適性測驗系統之開發係根據以下四個研究目的：第一、設計與發展一套適合於中文環境且能圖文整合之電腦化適性測驗系統，以利各級學校教師或研究者進行適性化評量或電腦化測驗，改善各類教學評量(含成就測驗、診斷測驗、等第分派、精熟評量)之品質與效率。第二、尋求會說話之「語音軟體」與適性測驗系統搭配，以利幼兒或視障兒童之施測。第三、設計與發展測驗題庫之參數估計與測驗訊息計算程式，以便建立與評鑑測驗題庫。第四、檢驗「異常作答型態」指標在電腦化適性測驗上之偵測力，以強化電腦化適性測驗之效度與診斷功能。

本電腦化適性測驗系統適用範圍爲(一)選擇式與是非型成就測驗、與(二)對、錯計分式態度量表，其它題型暫不適用。由於目前資訊網路之塞車問題尚無法有效解決，而且遠距施測時並無法有效確認作答者，本系統之開發暫將以單機使用與校園網路上之班級或測驗中心的運用爲主。

參、理論依據

本電腦化適性測驗系統之開發與應用涉及試題反應理論、適性測驗理論、異常作答型態(unusual response patterns)之偵測、與題庫(item banking)建立之方法。為便利一般讀者對於CAT之了解與應用，茲簡介紹這些相關理論如下：

一、試題反應理論模式

到底要選擇何種試題反應理論模式，單參數、雙參數、或三參數？這是一個複雜且見仁見智的問題。Hambleton(1992)在AERA年會上所主持的IRT模式研討會，主題為：IRT in the 1990s: which models work best?，就有極精彩之論戰。論戰中偏愛單參數者認為此模式只涉及題目難度與學生能力之估計，參數估計值具有統計上所謂之充分性(sufficiency)；且因其所加諸於模式之參數估計上的限制相對的少於三參數的IRT模式，實用性因而更佳。過去Andersen(1977)亦證明了唯有題目難度參數是一致性與充份性的統計估計值(consistent and sufficient estimates)。Lord(1980)指出即使三參數IRT模式比單參數Rasch模式更適配，當樣本較小時，Rasch參數估計值亦可能比三參數的估計值更正確(P. 190)。Forsyth, Saisangian, & Gilmer(1981)在一項實徵性的研究中證實：即使違反Rasch模式之基本假設，單參數Rasch模式仍具有合理的參數值不變之特性。此外，此模式已被證實不僅在小規模的測驗情境中，所估計的參數較具穩定性(Lord, 1983；Wright & Stone, 1979)；而且Olsens, Maynes, Slawsen, & Ho(1986)針對不同IRT模式適性測驗與不同型態之小學數學測驗成績之比較，指出使用單參數IRT模式與使用雙參數或三參數之IRT模式估計的電腦化適性測驗能力值具有高度相關，三年級與六年級均為 $r = .99$ 。Wright(1995)亦撰文說明單參數與三參數IRT模式所獲得之能力與難度估計值，在統計的意義上是相等的(statistically equivalent)，而鑑別度之估計會更加複雜化參數估計與參數不變性(parameter invariance)之問題，而猜測度之估計對於未猜測之考生是一種不當之懲罰，而且也無法區辨隨機猜測(random

guessing)與明智猜測(informed guessing)。因此，基於上述之理由與為使本適性測驗系統更有效適用於較小規模的教室測驗情境中，本系統將採用單參數試題反應理論模式(one-parameter IRT model)，亦稱 Rasch測驗模式。

二、偵測「異常作答型態」

為使本適性測驗系統具有偵測「異常作答行為」的功能，將植入一偵測異常作答型態之個人適合度模組，使測驗品質、猜測行為、異常作答行為、與學習困難得以有效加以控制或偵測出來(Li, 1992；Li & Olejnik, 1997)。該模組提供了經證實偵測力較強的二個指標：LZ&WSR，並能記錄考生偏好某一選項之比率，比率超過0.5時會在成績單中註記。

三、建立題庫方法

建立電腦化適性測驗題庫與傳統之測驗題庫大同小異，基本流程為(Wainer, et al. 1990)：

- 1、確定測驗目的與範疇。
- 2、界定測驗內容之明細表。
- 3、草擬試題。
- 4、預試。
- 5、利用試題之藍圖、難度、鑑別度、與試題訊息量進行試題分析，以便篩選試題。如有必要，可運用測驗題庫之等化與連結技術，以便建立足夠之題目。
- 6、測驗編輯。
- 7、測驗之信、效度檢驗。

而電腦化適性測驗題庫與傳統測驗題庫不同的是，它應具有下列特性：題庫夠大，測驗題目難度與內容均勻分佈，測驗題目之估計參數很穩定，且符合IRT之主要基本假設：測驗題目主測單一重要(dominant)能力與各題答對機率之獨立性。為達到上述目標，測驗編製者可先使用國內外IRT測驗分析軟體(如BILOG中之三參數試題作答理論模式)進行題庫中各題目之項目分析。為使題庫中之試題適用於Rasch測驗模式，應刪除猜測估計值較大之題目(檢查低能力組考生在較困難試題上之答對機率是否趨近於零)，選取題目鑑別度較大且類似(譬如介於0.3~0.8之間)之題目，並注

中文電腦化適性測驗系統軟體之研發

意測驗題目難度與內容是否均勻分佈。為使此測驗題庫具有內容效度，各題目之命題均將依據命題雙向細目表之計畫而命題，題庫也儘可能增大至100題以上。如欲使紙筆式與電腦化適性測驗式測驗在內容上與測量特質上平行，Stocking(1994)建議電腦化適性測驗題庫之長度為紙筆測驗之6至8倍。最後之題庫可利用因素分析方法，檢驗其是否具有單向性(unidimensional)，即只有一個明顯之主要因素(a dominant factor)存在。筆者建議使用圖1之流程建立RASCH測驗題庫。

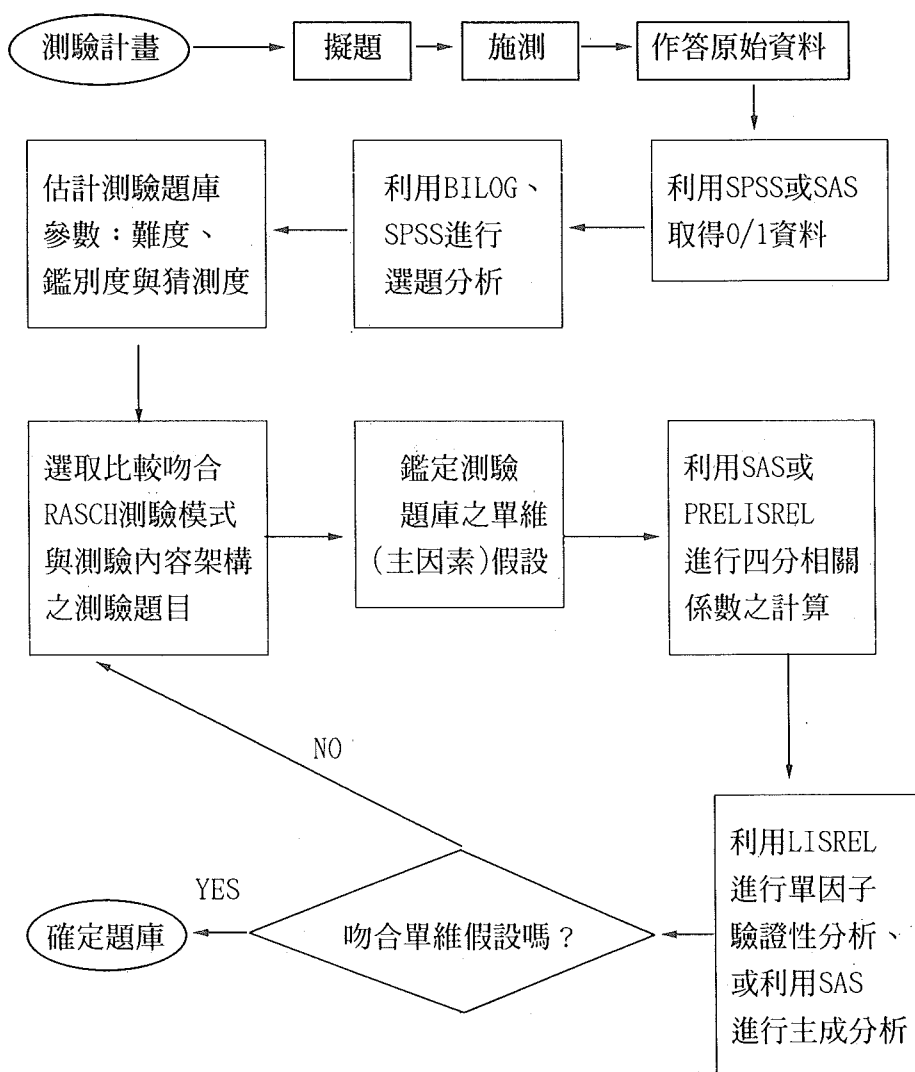


圖 1. 建立RASCH測驗題庫之流程

四、估計受試者之起始能力

在各類適性成就測驗中，本測驗系統均以中等難度(45-55)之題目隨機施測起，再依據其作答結果適性選題施測。

五、適性選題施測策略

本測驗系統為兼顧每位考生測驗內容之平衡性，每一考題可設定「內容類別代號」與「必考題」，選題施測時，系統會依代號順序在該類別內選取具有最大訊息(maximum item information)的題目，即當與受試者能力相當時，該題目之訊息會趨於最大。當受試者Rasch之能力估計值經電腦自動計算後，本測驗系統會在兼顧測驗內容均勻性之下，隨機選擇與其能力相匹配之題目給予作答。題目如被標定為「必考題」，則每位考生都要考這一題。題庫中如果不標註「內容類別代號」與「必考題」，系統則隨機選擇與其能力相匹配之題目進行施測。

六、計分方法

本測驗系統採用最大概率估計(maximum likelihood estimation)法，以重估題目與能力參數。在成就測驗上呈現分數時，為讓一般學生與家長容易了解，題目難度(B)與受試者能力(Q)之估計值均經直線轉換為T分數，並限制其值介於0~100之間，大於100者以「100+」表示之，小於0者以「0-」表示之。在精熟與否之評量時，學生之成績以「通過」與「未通過」呈現。在等第分派之評量時，學生之成績以「優等」、「甲等」、「乙等」、「丙等」、「丁等」呈現。

七、終止適性測驗之標準

本適性測驗系統在一般適性成就測驗中，採用「固定估計標準誤」方法，每一能力水準之估計誤差(SE)均限在0.6個logit(亦即6個T分數點)之內，以均等各能力值間之測量精確度。因為在電腦化適性測驗中，受試者可能接受不同長度與難度之測驗，本系統並未設定任何時間之限制。但受試者在回答問題後，如其估計標準誤達到預期之精確度或題庫題目已耗盡時即終止測驗。而在精熟測驗系統中，本系統採用Weiss & Kingsbury(1984)之方法，考生之估計能力的 $\cdot 95$ 信賴區間($1.96 \times SE$)必須完全在事先所設定之「精熟標準」之上或之下，測驗才會終止(參見圖4)。

在等第分派之評量時，本系統亦採用Kingsbury & Weiss(1984)之方法，主試

者必須事先設定四個能力分割點，以作為五等第評量之依據。「優等」之.95下信賴區間必須完全在事先所設定之最高分割點之上。「甲等」之.95上信賴區間必須完全在事先所設定之最高分割點之下，而.95下信賴區間必須完全在事先所設定之次高分割點之上。「乙等」之.95上信賴區間必須完全在事先所設定之次高分割點之下，而.95下信賴區間必須完全在事先所設定之第三高分割點之上。「丙等」之.95上信賴區間必須完全在事先所設定之第三高分割點之下，而.95下信賴區間必須完全在事先所設定之第四高分割點之上。「丁等」之分派則需其.95上信賴區間完全在事先所設定之最低分割點之下(參見圖5)。

根據筆者初步之實驗，發現在精熟測驗與等第分派的測驗結果中，受試者之精熟或等第之評定可能會因「測量誤差過大」而有誤判之情形。為避免此一不正常後果與齊一各能力水準之測量信度，本系統設定每一考生之估計標準誤亦需小於事先設定之0.6個logit，才進行精熟與否之判定與等第之分派。而精熟或等第通過百分比的期望值與潛在能力之關係乃透過公式一加以界定(Kingsbury, & Weiss, 1979)。

$$E(p|\theta) = \sum_{i=1}^n [e^{1.7(\beta_i - \theta)} / (1 + e^{1.7(\beta_i - \theta)})] / n \dots\dots\dots \text{公式一}$$

公式一中， θ 表能力估計值， β 表難度估計值， n 表題庫中題目之大小。透過此一關係，可以算出在每一能力水平上之精熟水準，以決定精熟水準或等第水準之 θ 能力分割點。傳統測驗所採取之精熟通過百分比與電腦化適性測驗能力量尺之轉換，可利用本系統所提供之正確百分比與精熟水準互轉功能，以確立精熟水準與等第分割點(參見圖2)。圖2之左下角提供常用之精熟通過百分比與電腦化適性測驗能力量尺之轉換表，教師與研究者透過此一簡表當已能滿足一般測驗之所需；例如，教師如認為正確百分比(P)達0.688以上，始認定為「精熟」時，轉換為電腦化適性測驗能力(θ)量尺即為0.5，此 θ 值為「精熟與不精熟」之臨界點。各等第分割點之 θ 亦依前述方法加以確定。如仍未能符合實際評量的需要，可透過圖2中特徵曲線找到X軸與Y軸之對應點，即可確立精熟水準或等第分割點。

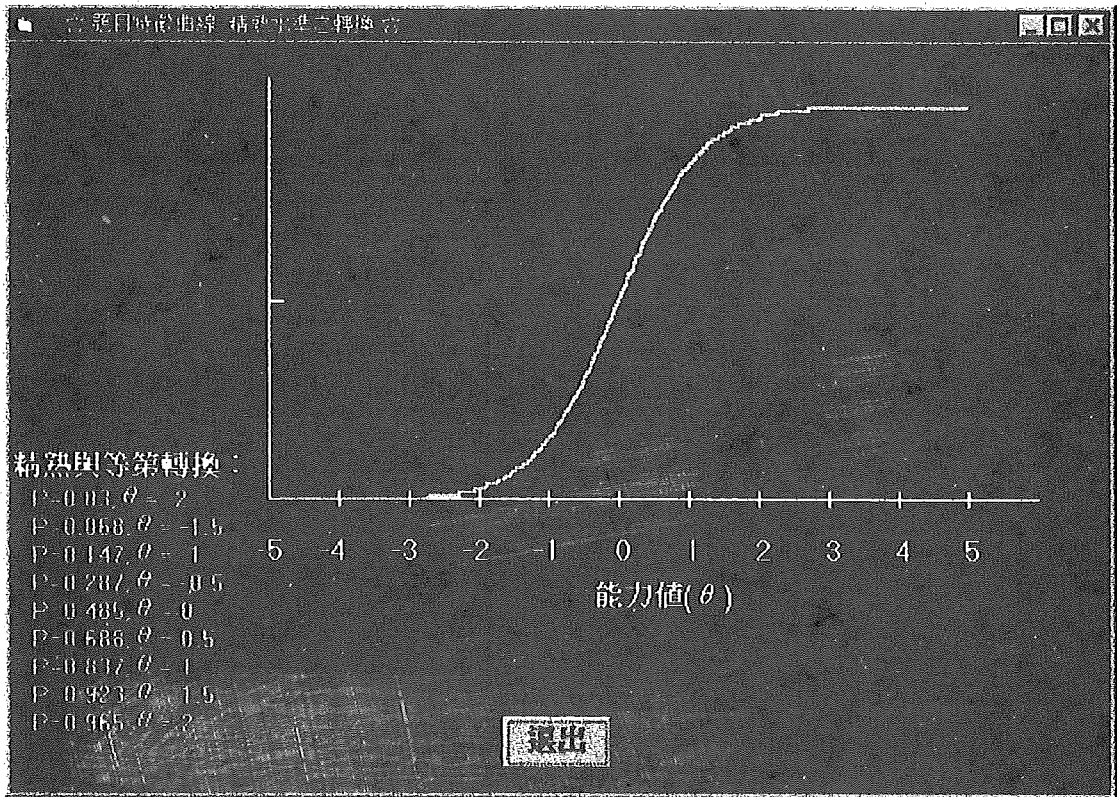


圖 2. 測驗特徵曲線與能力值、答對百分比關係圖

肆、發展本電腦化適性測驗系統之工具軟體

撰寫主體程式語言為VISUAL BASIC 6.0，所搭配之合成語音系統則為Otek International公司所發展之語音精靈：中英文朗讀系統，此系統可以不必事先錄音而可以直接朗讀測驗題幹與選項。傳統之Basic、Quick Basic等均屬於「程序導向」的程式語言，且常必須花許多時間去設計與發展程式界面。VISUAL BASIC 6.0工具軟體，係屬「物件導向」、「事件驅動」的程式語言。其最大好處是程式界面可輕易設計完成，設計者可以用大部分時間去思考系統規劃、撰擬程式碼、與分析如何強化副程式之可再用性、與可維護性。因此，本著副程式本身內之凝聚力(cohesion)應越強與副程式間之關聯性(coupling)應越弱之原則；每一副程式盡可能朝著單一功能去設計，使其可為其它副程式所用，副程式與副程式間之關聯則盡可能使資料關聯、控制關聯與全域變數關聯明確化，以利程式日後之維護。

伍、參個電腦化適性測驗子系統之主要流程

圖 3 說明了一般電腦化適性成就測驗之施測流程，系統設計者必須根據此步驟，決定(1)如何預估受試者起始能力，以便開始施測、(2)如何適性選題施測、(3)如何計分/估計能力方法、與(4)如何終止測驗等方法，方能著手撰寫程式與設計使用者界面。圖 4 與圖 5 之流程旨在說明精熟測驗與等第分派之過程。在本系統中，精熟之決定與等第分派是在「滿足測驗終止標準」之後才進行，以確保評定結果之信度。

一、適性成就測驗之主要流程

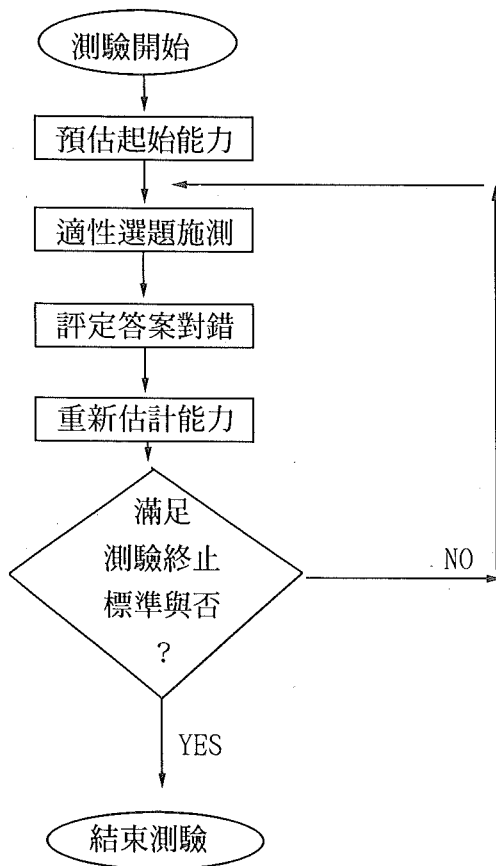


圖 3. 電腦化適性成就測驗之主要流程

電腦化適性成就測驗(Computerized Adaptive Achievement Testing)說明了「適性選題施測」之能力估計過程。這是一般適性測驗之主要核心概念。

二、適性精熟測驗之主要流程

適性精熟測驗(Adaptive Mastery Testing)主要包含五大步驟：(1)將精熟水準(如通過之百分比)轉換為IRT量尺，(2)設定施測起點，(3)適性選題施測，(4)估計新能力，(5)決定適性選題施測是否終止，(6)檢驗該能力之信賴區間(如.95水準)是否包含精熟分割點，以確定測驗是否終止。圖4為A、B兩位學生在一精熟測驗上適性選題施測之過程(修訂自圖1, p. 369, Weiss & Kingsbury, 1984)。圖中精熟與否之決斷 θ 水準設定為0.5，考生能力估計值之信賴區間如果全在決斷點以上，即屬「通過」，如果全在決斷點以下，即屬「未通過」，此時測驗即可終止。例如甲生在做完第一題之後，其能力之估計值為-0.5，而且其.95之信賴區間亦仍包含了0.5之精熟水準。因此，該生必須再接受第二題之施測，新能力估計值之信賴區間亦同時計算出來，以便判定是否繼續施測。在本例子中，甲生一直做完第十二題之後，其能力估計值之信賴區間才明確完全落在精熟水準之下。因此，甲生被判為「未通過者」。乙生依此規則，在做完第七題之後已能完全判定其能力之信賴區間完全落在精熟水準之上。因而，判定乙生為「精通者」。

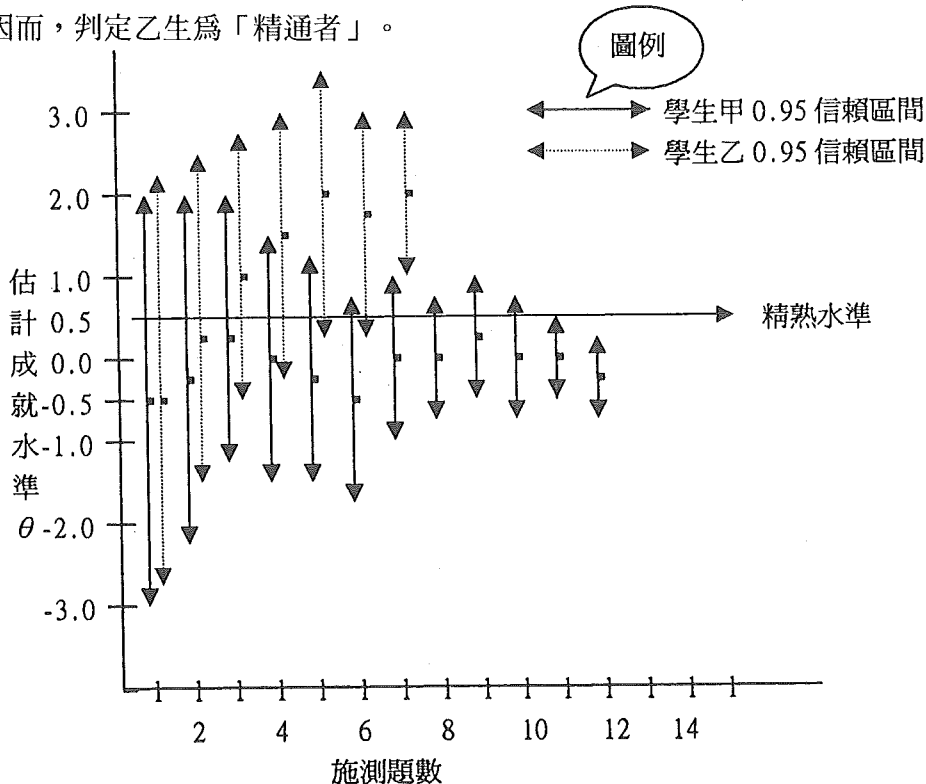


圖 4. 適性精熟測驗施測之主要流程

三、適性等第分派測驗之主要流程

適性等第分派測驗之主要流程與適性精熟測驗類似，只是分割點由一點增加為數點而已。目前學校評量系統中，常需將學生各科之成績轉換成等第。目前國小的各科成績報告，分為「優等」、「甲等」、「乙等」、「丙等」、與「丁等」五等。因此，共需四個等第之分割點(Cut-off Points)，以作為等第分派用。『優等』之分派，係指考生之成績的.95信賴區間完全在最高分割點之上，「丁等」之分派，係指考生之成績的.95信賴區間完全在最低分割點之下，「甲等」、「乙等」、「丙等」之分派，係指考生之成績的.95信賴區間完全落在上、下分割點之中。圖5係A、B兩位學生在一適性等第分派測驗上適性選題施測之過程(修訂自圖3，p. 369, Weiss & Kingsbury, 1984)。從圖中可清楚看出，甲生在做完第八題之後其能力之.95信賴區間已完全高於優等之最低水準，因此，甲生被判為「優等」；而乙生在做完第九題之後其能力之.95信賴區間已完全落於乙等之上下區間，因此，乙生被判為「乙等」。

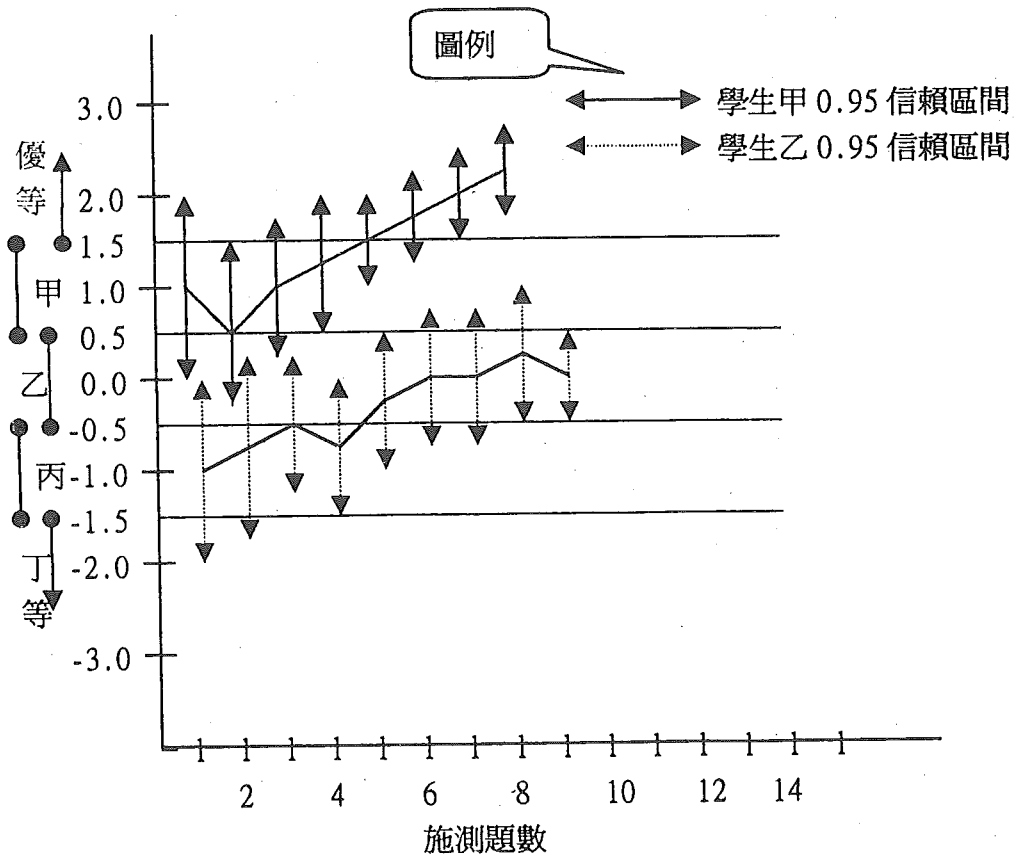


圖 5. 適性等第分派測驗施測之主要流程

陸、中文電腦化適性測驗系統之結構與程式物件

一、規劃中文電腦化適性測驗系統之基本界面與物件結構

根據目前學校常見之評量形式與需要，本中文電腦化適性測驗系統(CCAT)含有三個適性測驗子系統：電腦化適性成就測驗(簡稱CAAT)系統、電腦化適性精熟測驗(簡稱CAMT)系統、電腦化適性等第評量(簡稱CAGT)系統與一個附屬之電腦化測驗(係將傳統紙筆式測驗電腦化，簡稱CPPT)系統。為幫助初次使用本系統之受試者熟悉操作方法，系統中提供一練習題庫供其考前練習。另外，為便於進行本測驗系統之實驗研究，筆者另備有實驗版可直接進行CAAT與CPPT測驗結果之比較，而不必花更多時間去做相同試題。實驗版上亦備有電腦化問卷調查，可直接進行受試者對於CAAT與CPPT兩種測驗模式之態度與感受調查。

根據圖3～圖5的流程圖及前述測驗系統所需之功能，整合建立如圖6之基本測驗架構，繼而決定所需之使用者界面、物件與事件，最後利用美國微軟公司之Visual Basic 6.0結構化及物件導向(object-oriented)之特性，撰寫所需之程式碼即可進入測試階段。而相關之副程式(sub-program)、函數(function)或模組(module)皆可由界面上之物件加以驅動。為強化本系統之效率性、維修性與結構化，部分測驗流程已加上必要之控制，但仍不失物件化視窗程式設計風格。

由圖6知，本中文電腦化適性測驗系統具有三個「電腦化適性測驗」子系統：CAAT、CAMT、與CAGT及具有一個附屬「電腦化測驗」子系統(CPPT)，可供使用者選擇使用。此測驗系統總共使用到四個表格(form)、二個模組(module)及圖片框、文字框、按鈕、計時器、目錄表列框等控制元件(control)。這幾個表格及模組分別命名為：Module32、Fadeform、Loadpic、Frmrecord、Frmdoc、Frmmain。各表格及模組所用到的副程式或函數及其功能分別簡述如下：

1、Module32中使用了五個Win32應用程式介面(API)及一個計算題庫題數之testlen副程式。這五個Win32應用程式介面為：GetFocus、與

中文電腦化適性測驗系統軟體之研發

SendMessage 函數、及 MciExecute、snd PlaySound，與 SetWindowPos 等副程式。GetFocus 函數用來指定焦點視窗，SendMessage 函數用來傳送設定表列框中之捲軸參數與捲動範圍，以便利閱讀視窗中之表列文件，SetWindowPos 副程式用來決定圖形在題幹右側之顯示位置。MciExecute 與 snd PlaySound 等副程式分別用來執行*.Mid 與*.Wav 語音檔。

- 2、FadeForm 中使用一個函數及三個副程式，分別命名為 CreateSolidBrush、DeleteObject、FadeColor、與 FillRect。本模組旨在讓 Frmrecord 表單具有漸層顏色之背景。
- 3、Loadpic 表格顯示 Visual Basic 所能接受之圖形，無副程式或函數。
- 4、Frmdoc 表格，含一 Form_Load 副程式，內含一個 OLE 物件，用以呼叫 WINDOWS 95 系統中之 word/WordPad，用以顯示 word/WordPad 所能顯示之文字檔與圖形檔。
- 5、Frmrecord 表格含有一表列框控制元件(Listbox control)作為顯示測驗成績用。本表格上具有一個命令按鈕副程式：Command1_Click 退出鍵，以便結束測驗結果顯示，退回至施測主畫面。
- 6、Frmmain 表格含有 CPPT、CAAT、CAMT 與 CAGT 四個子測驗流程中所需之主要副程式與函數；各副程式與函數之物件功能，請詳見附錄B。

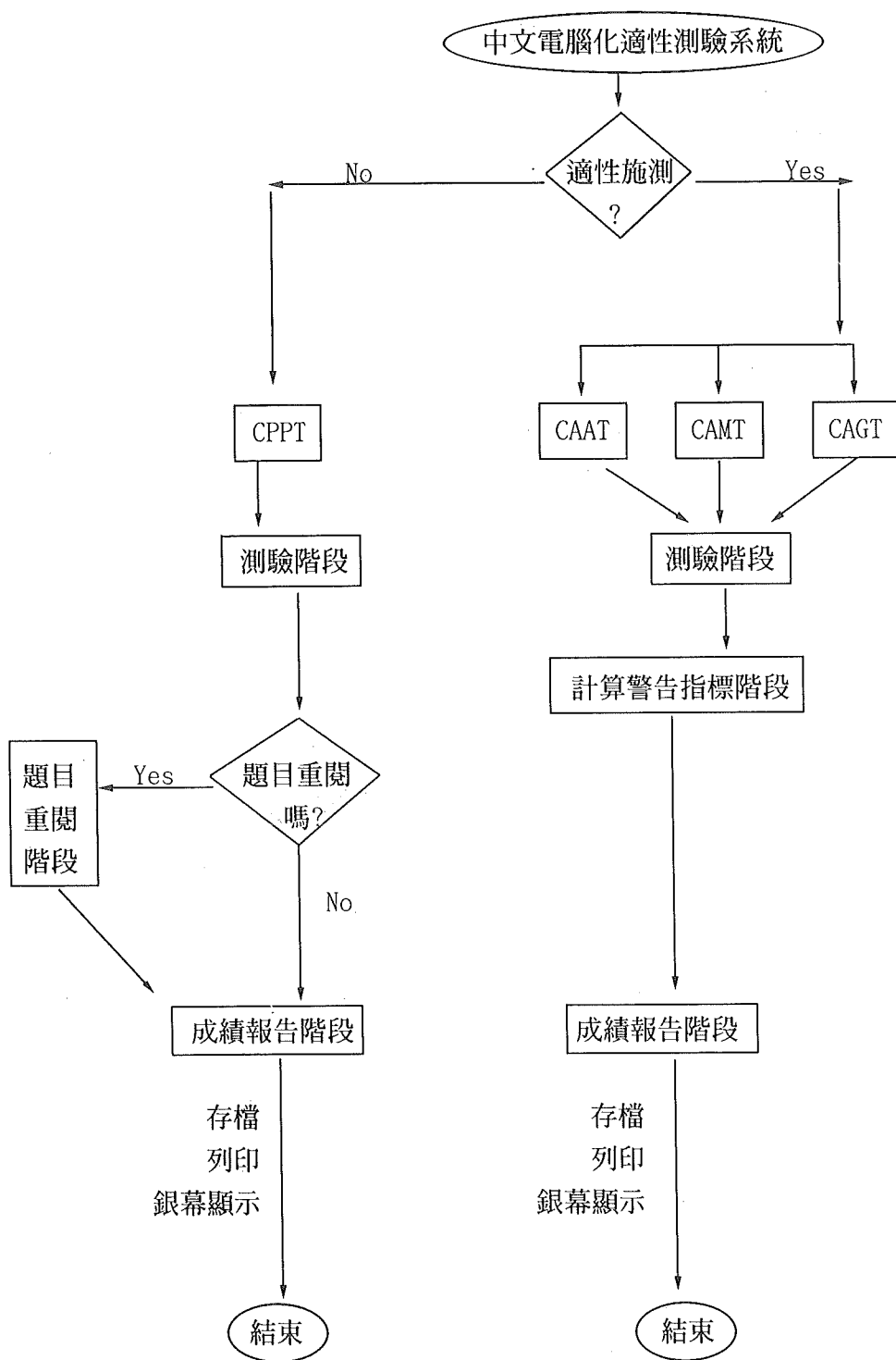


圖 6. 中文電腦化適性測驗系統之基本架構

二、研擬與測試各物件與功能所需之程式碼

本系統中，各物件之相關驅動程式均經實地數次測試無誤後，再予以整合測試，每一程式段亦稍加說明其用途以強化其維護性。為使本軟體能順利在校園網路上使用，程式編譯時採共享模式而非單機使用模式。

柒、考生作答與中文適性測驗系統之運作流程

本適性測驗系統之運作流程與前節之基本架構息息相關，涉及主試者、受試者、與測驗系統。主試者在考生施測之前，需檢查測驗題庫中各題之撰寫格式是否正確，與相關之圖形、OLE檔或語音檔是否存在正確之路徑。主試者在考生施測之後，需決定是否重新估計 β 值、並更新受試者之 θ 值、儲存更新記錄、與檢查曝光率。而一個考生在作答時，其基本施測過程簡述如下：

- 1、由功能表單「測驗系統」中，選擇施測子系統。
- 2、輸入測驗題庫之基本訊息：題庫名稱，測驗長度，作答記錄檔名稱。
- 3、決定欲進行何種測驗。
- 4、輸入受試者之名字。
- 5、輸入答案。
- 6、關閉測驗系統。

而系統之主要運作工作依序有：計時、選擇適性題目施測、顯示題目文字、顯示題目圖形或語音、估計受試者之能力、決定是否滿足測驗終止條件、決定精熟與否、或等第分派、計算作答時間、計算作答異態指標：LZ/WSR、作答記錄之存檔、顯示測驗結果、作答時間、作答異態指標。

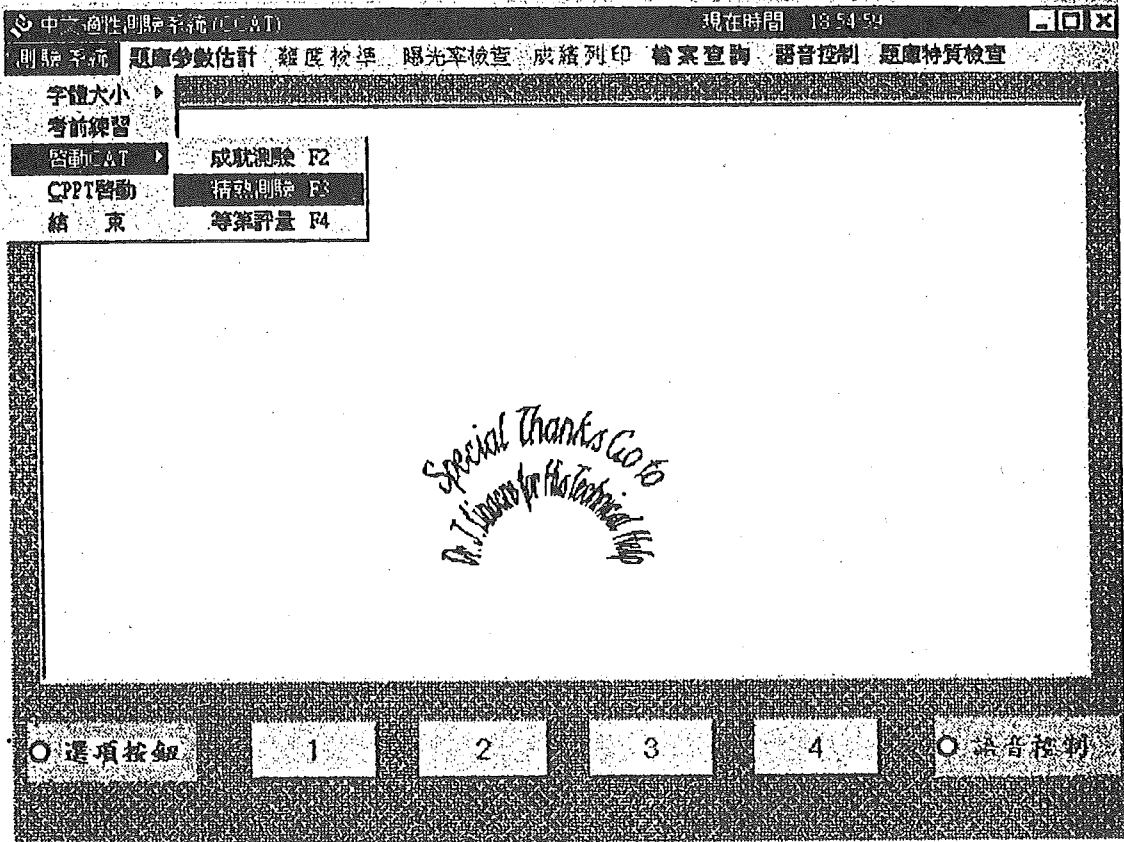


圖 7. 中文電腦化適性測驗系統之主要畫面

捌、中文電腦化適性測驗系統簡介

中文電腦化適性測驗系統之視窗中共有八個主要功能表單：測驗系統、題庫參數估計、難度校準、曝光率檢查、成績列印、檔案查詢、語音控制、題庫特質檢查。測驗系統表單內有字幕大小之變換、考前練習、CAT、CPPT、與結束等功能(參見圖7)。CAT內有三個子系統：適性成就測驗系統、適性精熟測驗系統、與適性等第評量系統。題庫參數估計提供PROX、UML兩種估計方法，以方便研究者進行題庫參數之估計。難度校準提供重估CAT難度與能力，或重估CPPT難度之功能。曝光率檢查旨在了解各個題目被使用之頻率。檔案查詢則在便利使用者查詢題庫與記錄檔案。語音控制則提供語音之播放與停止之功能。題庫特質檢查內，含有題庫與圖檔格

中文電腦化適性測驗系統軟體之研發

式之檢查、題庫測驗訊息之檢查，與精熟或等第分派時 θ 水準之設定。

以下將簡單介紹本測驗系統之基本特色、操作方法、與題庫建立方法，這些說明亦可從系統之線上協助功能中「按F1」取得。

一、中文電腦化適性測驗系統(CCAT)基本特色

- 1、本CCAT系統含有電腦化適性成就測驗(CAAT)系統、電腦化適性精熟測驗(CAMT)系統、電腦化適性等第評量(CAGT)系統、與一個附屬之電腦化傳統紙筆式測驗(CPPT)系統。極適合運用於一般學校之成就測驗、精熟測驗(mastery testing)、與等第分派(grades assignment)。
- 2、具有OLE與圖文整合能力，可讀進*.DOC之圖文檔與*.BMP,*.ICO,*.WMF等256色以上之圖形檔。
- 3、中英文試題均可適用。
- 4、測驗結果可立即顯示與列印，達到立即回饋與增強之效果。
- 5、難度與能力可隨考生作答記錄之增加予以重估，而使題目之難度與考生之能力估計值逐漸趨於穩定。
- 6、本系統中之電腦化適性測驗，可提供兩種高偵測力之作答異態指標:WSR & LZ, 可作為測驗效度之指標或學習異常之警告指標。
- 7、每一考生適性成就測驗所估計之能力、等第分派與精熟測驗之測量標準誤均在預先設定之範圍內，各考生之能力估計值之精確度得以提高；並報告上下一個測量標準誤之信賴區間(約正負6個百分點)。
- 8、自動記錄考生之作答資料與時間，以供教師資料分析與參數重估用。
- 9、能播放聲波檔(*.Wav)與Midi(*.Mid)音樂檔，便於語文與音樂等學科之應用。
- 10、各題曝光率之檢查，可以確保測驗資料庫之安全性。
- 11、為避免題庫中某些題目過分使用或內容不當，而需更換或修正，題庫可隨時增添、修正、或編輯。新增添之題目難度可暫定為50,當受試者夠多時，再利用難度更新功能予以校準，節省事前估計難度之麻煩。
- 12、配合Otek International公司之語音精靈：中英文朗讀系統，可朗讀試題中之文字與測驗結果；對於低年級學生與弱視者亦甚方便。

- 13、題庫建立者如在診斷訊息欄中，輸入與受試者選錯答案有關之診斷訊息，本測驗系統將兼具有錯誤診斷之功能。
- 14、在一般學校之測驗情境中，時間是必須考慮之因素。為避免增加作答時間，而降低電腦化適性測驗之效率與產生考生舞弊之情況 (Gershon & Bergstrom , 1995 ; Wainer, 1993) ，在本電腦化適性測驗系統中，並未設計考生重新審閱所作過之題目與答案(item review)之功能。
- 15、CPPT測驗系統中，雖無法計算作答異態指標:LZ/WSR，但具有重新審閱作過之題目與答案之功能。這使得作答方式與傳統紙筆式測驗相同，受試者可以擁有一次檢視與修改答案之機會。
- 16、提供線上協助說明與考前作答練習。
- 17、電腦化適性測驗系統之題庫中，可設定每位考生共同「必考」題目與「內容分類」之設定，以確保測驗內容之代表性與周延性。
- 18、電腦化適性測驗(CAAT、CAMT、CAGT)系統採適性選題施測，能提昇測驗效率與精確度，而電腦化紙筆式測驗(CPPT)系統隨機選題施測，以避免或降低考生舞弊之機率。
- 19、測驗題目的文字與圖形大小，可於測驗之前或測驗當中讓考生自行調整，便利各考生閱讀。
- 20、自動檢查考生選擇每一選項之百分比，作為作答異常之參考指標。

二、CCAT操作方法

中文電腦化適性測驗使用前，請先利用系統內「題庫特質檢查」功能下之「題庫格式檢查」，檢查欲使用之題庫是否有格式錯誤，及相關之圖檔、語音檔、OLE檔是否存在、或存在正確路徑(這些相關之檔案應存在與本系統執行檔所在之路徑內)。如有錯誤，必須找出錯誤所在並加以更正，否則系統可能會出現錯誤資訊或無法正常運作。中文電腦化適性測驗系統操作之基本原則，簡述如下：

- 1、本測驗系統係WINDOWS 95視窗版軟體，啟動時連續按CCAT圖像兩次，即可進入本測驗系統。
- 2、本測驗系統內各視窗之放大、縮小、還原、與關閉完全與WINDOWS 95之操作

中文電腦化適性測驗系統軟體之研發

相同。

- 3、本測驗系統之功能表列開啓，悉依WINDOWS 95之操作方法，作答時主要以滑鼠操作爲主，直接點選選項按鈕(參見圖7)即可；如欲使用鍵盤，請按數字鍵亦可。
- 4、考生與電腦之溝通管道乃透過訊息框進行溝通，如需要輸入動作，需鍵入所需之資訊後，再按ENTER鍵或使用滑鼠左鍵點選OK小框，才能關閉訊息框。
- 5、作答時依銀幕上之指示與說明，即可順利完成測驗。
- 6、考生每一題目均需作答，無法跳過不答。
- 7、按F1可出現線上協助視窗(參見圖8)，快速了解本系統之主要功能、基本特色、題庫建立方法、與操作方法。

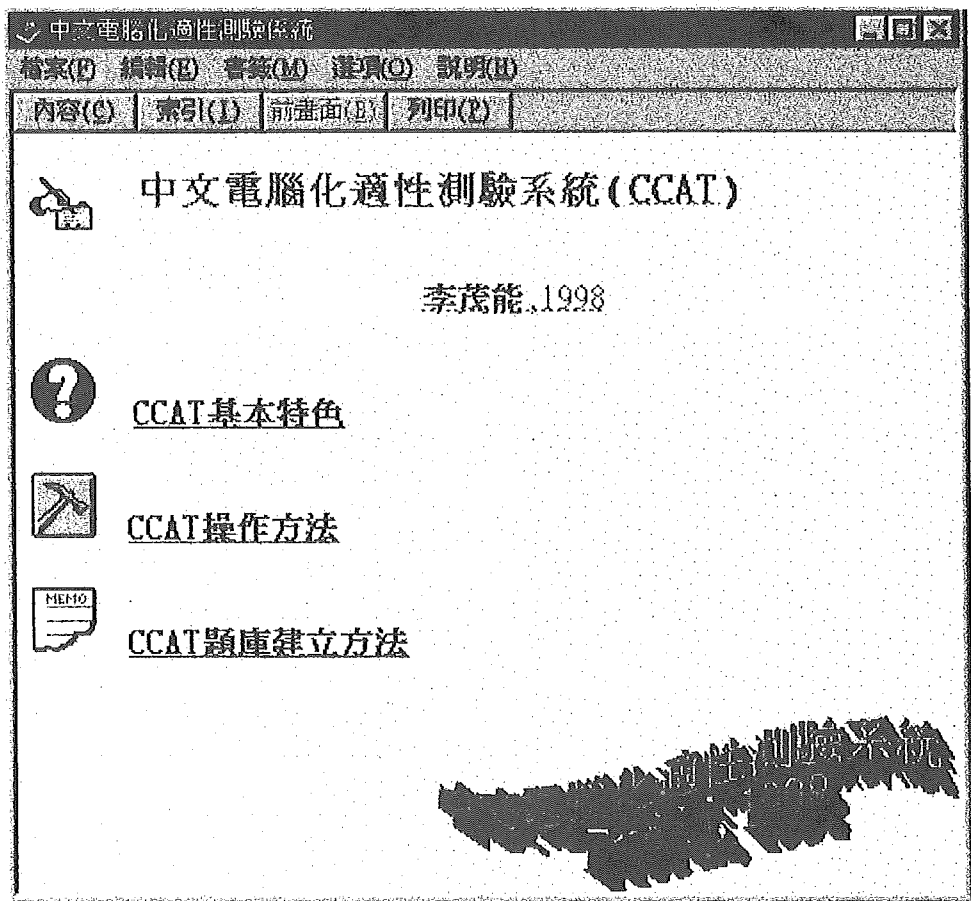


圖8. 線上協助視窗

8、遇題目內容過長，超過一個銀幕之容量時，可利用滑鼠，點取視窗邊緣之垂直捲動軸向下捲動，以便向下觀看題目內容。

以下將針對各類測驗系統與相關之附屬功能的操作步驟，簡要說明如下：

(1) 電腦化紙筆式測驗(CPPT)

使用滑鼠直接點選「測驗系統」之功能表單，再按「CPPT啟動」啟動之。

(2) 電腦化適性成就測驗(CAAT)

使用滑鼠直接點選「測驗系統」之功能表單，點選「啟動CAT」、再按「成就測驗」啟動之；或直接按「F2」啟動之。

(3) 電腦化適性精熟測驗(CAMT)

使用滑鼠直接點選「測驗系統」之功能表單，點選「啟動CAT」、再按「精熟測驗」啟動之；或直接按「F3」啟動之。

(4) 電腦化適性等第評量測驗(CAGT)

使用滑鼠直接點選「測驗系統」之功能表單，點選「啟動CAT」、再按「等第評量」啟動之；或直接按「F4」啟動之。

當啟動上述這些測驗系統之後，系統會要求考生輸入「測驗題庫名稱」、「名字」、「測驗名稱」、「作答記錄檔案名稱」。考生必須一一填入後，方能繼續下一步驟：進行正式測驗。當測驗結束之後，會立即顯示成績摘要報告，並立即自動將成績存入先前所設定之記錄檔案內；之後，亦可將此成績摘要報告列印出來。

(5) 字幕大小控制功能

在進行測驗之前或測驗之中，均可直接點選「測驗系統」下之「字幕大小」，改變顯示字型之大小，共有大、中、小三種供選擇。

(6) 作答練習功能

為使不善於運用電腦之考生熟習電腦之基本操作，考生可在考前利用此一功能，練習運用電腦作答。

(7) 題庫參數估計功能

本系統提供PROX與UML(Wright & Stone, 1979)兩種題庫參數估計方法。使用滑鼠點選欲估計之方法後，輸入題庫名稱，即立刻將題目之難度與能力估計值顯示

中文電腦化適性測驗系統軟體之研發

出來。待估計之原始資料的建檔格式，請參見附錄A之範例。

(8) 難度校準功能

使用滑鼠點選「難度校準」表單之後，再用滑鼠點選「重估CAT」或按下F5，可以重新估計CAT題庫中之難度參數，考生之能力也可因之而調整，使用滑鼠點選「重估CPPT」或按下F6，可以重新估計CPPT題庫中之難度參數。

(9) 曝光率檢查功能

按「曝光率檢查功能」表單後，系統會將剛施測之題庫內各試題的過去使用率顯示出來，作為研究者或老師決定是否更新題庫之參考。

(10) 成績列印功能

使用滑鼠點選「成績列印」表單之後，再用滑鼠點選「列印CAT」或按下F7，可以列印CAT測驗結果，使用滑鼠點選「列印CPPT」或按下F8，可以列印CPPT測驗結果。

(11) 檔案查詢功能

使用滑鼠點選「檔案查詢」表單，可以查詢磁碟上之檔案名稱。

(12) 語音控制功能

使用滑鼠點選「語音控制」表單後，點選「語音播放」、「語音停止」，以進行題目所連結之語音檔案的語音控制。

(13) 題庫特質功能

本功能表單提供「題庫格式檢查」、「題庫訊息檢查」、「精熟與等第分界點查詢」等三個功能。點選這些功能之後，系統會要求您輸入題庫檔案之名稱。當您使用滑鼠點選「題庫格式檢查」，可以檢查您所建立之題庫中各題之格式是否符合規定，各題所連結之圖形檔、語音檔與OLE檔是否存在。使用滑鼠點選「題庫訊息檢查」，可以了解待使用題庫最適合於施測之能力範圍。使用滑鼠點選「精熟與等第分界點查詢」，可以幫助您快速決定精熟與否之 θ 值分割點、各等第分派之 θ 值分界點的設定。

三、CCAT題庫格式

1、本測驗系統題庫可用一般文書處理軟體(如PE3,HE4,Word)等建立與編輯，惟存

檔時需以「文字檔」轉存。如欲以WORD文書處理較複雜之特殊圖文題目，則可直接以*.DOC或*.RTF檔案格式存檔，亦或透過剪貼簿之媒介轉存成*.bmp圖檔，以便讓程式呼叫使用，此方式較費時與佔用磁碟空間，若非必要，仍以文字模式之文書軟體加以處理為佳。

- 2、測驗題庫之大小、難度分配與能力估計之速度與正確性具有密切關係。測驗題目之難度分佈以矩形(等機率)分配為最佳(為降低中等難度題目之曝光率，可酌于增加中等難度題目)，否則有導致考試時間拖長與測驗精確度下降之虞，而各科之測驗題數最好大於100。
- 3、為便於檢查題目建檔格式與內容，使建立題庫簡便易行，本測驗題庫採Linacre (1987)的固定格式，各行之內容亦不得對調。題庫內容與所規定之格式不符時，考前系統會立即要求更正，才能繼續作答。每一測驗題目固定使用10行，欲看詳細格式請參閱圖9說明，實際範例請參看圖10。文字內容與相關之圖形採分離處理，呈現圖文時再予以整合及調整呈現位置。

由圖9之第一行(1st row)可知，第1~3欄位(1st~3rd column)之位置為各題題號之輸入位置。第4欄位之位置為各題內容分類與必考設定之輸入位置。第5~25欄位之位置為各題圖檔名稱之輸入位置(可設定存檔之路徑，以確保圖檔可以被讀取)。第30~50欄位之位置為各題語音檔名稱之輸入位置(可設定存檔之路徑，以避免找不到語音檔之困境)。第二行到第四行係選擇題之題幹內容的顯現位置，每行請勿超過127個中文字。各行之文字超過銀幕所能容納之長度時，會自動換行顯示。如果題幹文字較短而僅使用一行時，第三與第四行仍需以空白行空著。第五行到第八行係選擇題或是非題選項內容(1.上述敘述正確 2.上述敘述錯誤)之顯現位置(最多只能提供四個選項)。每一選項之長度最長為127個中文字，而各選項之代號不必輸入，系統會自動提供。第九行之第1~2欄位為正確答案之輸入位置，第九行之第3~80欄位為提供考生訊息回饋之輸入位置。第十行之第1~3欄位為難度指標(1~100)之輸入位置。

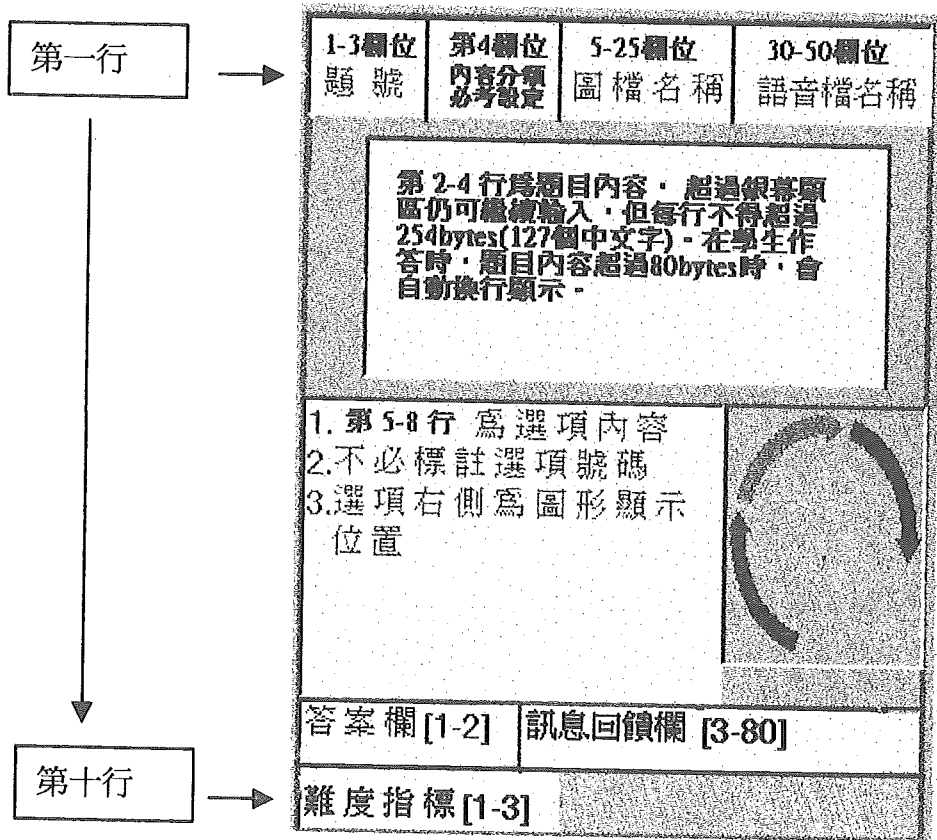


圖 9. 題庫格式基本設定圖

1	2
下列敘述何者是錯誤?(甲)百分等級可告訴我們熟練程度(乙)百分等級是順序量尺(丙)百分等級不可相加減(丁)原始分數分配會影響百分等級之正確解釋	
甲乙	
乙丙	
丙丁	
甲丁	
4	【百分等級之意義與特質需再復習】
70	

圖 10. 測驗題庫範例

- 4、本測驗系統可以整合*.BMP,*.ICO, & *.WMF等圖型檔。如為圖文整合之題目，各圖型檔之顯示位置，雖然會依選項之長短而在適當之顯示位置，但原則上其大小請勿超過6 x 10平方公分，以免發生圖文錯置或被截斷等現象。
- 5、圖型檔固定顯示於題幹之下方，選項之右側。
- 6、題目文字如欲以Word、WORDPAD等軟體處理(例如題目中需使用到特殊符號或色彩)，則需以「RTF」、「DOC」文字檔案格式或以圖型模式儲存成一獨立之檔案。此種呼叫「*.RTF」、「*.DOC」檔案與圖形檔之模式，於題庫中仍應將題幹及各選項之處，皆以空行留置，但試題難度與正確答案仍須於正確位置設定之。
- 7、考生之能力與題目之難度採T分數制，最簡單題目的難度為0，中等難度之題目為50，最困難題目的難度為100；而受試者能力最低的為0，中等能力的為50，能力最強的為100。題庫中各題目之難度必須事先估計好，使用者可利用本軟體中之Rasch 參數估計功能，或使用像BILOG、LOGIST、RASCAL、BIGSCALE、ASCAL、MSCALE等IRT參數估計電腦軟體進行估計。如果已有傳統之難度指標，可暫以 $(100 \times (1-p))$ 代替，俟作答人數足夠時，再以本軟體之難度重估功能逕行調整。
- 8、圖型檔、語音檔與題庫需同時存在本測驗系統執行檔之目錄中，最好在指定圖型檔名時亦加上路徑名稱，否則會找不到圖型檔而無法繼續作答，如能在autoexec.bat檔中加入下列指令：「APPEND %PATH%」亦可解決找不到圖型檔或語音檔之困擾。
- 9、本測驗系統之題庫內定最大可處理999個題目，但仍需視主記憶體多寡而定。
- 10、錯誤分析所需之診斷訊息請儘量簡潔、扼要，最好勿超過40個中文字。
- 11、如欲設定所屬測驗內容之分類或必考題目，請在該題第一行行號之後，即該行第四欄位輸入「0」、「1」、「2」等類別代號，電腦化適性測驗系統就會在考試過程當中，按此順序循環選取測驗題目施測，以平衡考試內容；或輸入「R」或「r」標示為「必考題」，測驗系統就會在考試過程當中，優先將之納入考題中(測驗內容類別代號之設定，須介於0~9之間)。注意：測驗系統選題施測之順序

為「必考題」、「內容分類題」、「未設定題」。

玖、安裝中文電腦化適性測驗系統之注意事項

一、軟、硬體基本配備

CCAT的使用者必須檢查以下提及的電腦基本配備，是否符合最低要求：

- 1、主機需求：中央處理器486以上。
- 2、記憶體容量：8MB RAM以上的主記憶體(視題庫大小而定)。
- 3、顯示器與印表機：印表機與VGA以上之輸出設置。
- 4、音效卡與音箱：8 Bit以上(如題庫中具有語音檔才需要)之音效輸出設備。
- 5、中文WINDOWS 95與WORD97以上。
- 6、滑鼠與鍵盤等之輸入設備。
- 7、中英文朗讀系統(Speech Wizard)：Make Your Computer Talk(Optional)，由Otek International公司所發行之「語音精靈」軟體。

二、中文電腦化適性測驗系統之安裝步驟

- 1、執行WINDOWS 95。
- 2、在硬碟中開一次目錄，以便CCAT程式之解壓縮與存放。
- 3、將中文電腦化適性測驗系統的安裝磁片上所有相關程式，拷貝至硬碟中所設定之次目錄。
- 4、從「開始」的功能表單中選擇「執行」後，點選或輸入CCAT安裝 程式所在次目錄中的SETUP.EXE並加以執行。
- 5、依照銀幕上之指示即可順利安裝本程式。

三、題庫、圖型檔與語音檔之安裝

需與主程式之執行檔安裝在同一目錄中，否則會有找不到題庫、圖型檔與語音檔之困擾。如果題庫中之題目係以*.DOC或*.RTF檔案格式存檔，可先開啓WORD文書處理軟體，以加快秀出題目時間。

拾、參考書目

- 王力行(民86)。電腦化、動態化、智慧化測驗方式之探討。台南：南師主辦教育測驗新進發展趨勢學術研討會論文集，217-227。
- 林世華、劉子鍵(民86)。整合認知心理學、心理計量學與教學的理想模式：結合認知設計系統、反應產生模式、認知診斷評量系統、以及動態評量系統。台南：南師主辦教育測驗新進發展趨勢學術研討會論文集，229-235。
- 何榮桂、蘇建誠、郭再興(民85)。遠距適性測驗系統架構。資訊與教育，53, 29-35。
- 孫天光、謝凱隆、鄭海東、陳新豐(民86)。智慧型線上適性測驗系統。國科會專案 NSC-87-2511-S-024-008-ICL。
- American Council in Education. (1995). Guidelines for computerized-adaptive test development and use in education. Washington, DC: Author.
- Andersen, E. B. (1977). Sufficient statistics and latent trait models. Psychometrika, 42, 69-81.
- Bunderson, C. V., Inouye, D. K., & Olsen, J. B.(1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), Educational Measurement(3rd ed., pp. 367-408), New York: Macmillan.
- Forsyth, R., Saisangian, U., & Gilmer, J. (1981). Some empirical results related to the robustness of the Rasch model. Applied Psychological Measurement, 5, 175-186.
- Green, B. F., Bock, R. D., Humphrey, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. Journal of Educational Measurement, 21, 347-360.

- Hambleton, R. K. (1992). Hambleton's 9 theses. Rasch measurement transactions, 6(2), 215-217.
- Kingsbury, G. G., & Weiss, D. J. (1979). An adaptive testing strategy for mastery decisions(Research report 79-5). Arlington, VA: Office of Naval Research.
- Li, M. N. (1992). Index robustness and effectiveness of Rasch appropriateness measurement in unusual response pattern analysis (Doctoral dissertation, University of Georgia).
- Li, M. N., & Olejnik, S. (1997). The power of Rasch person fit statistics in detecting unusual response patterns. Applied Psychological Measurement, 215-231.
- Linacre, J. M. (1987). A computer program for adaptive testing. (ERIC Document Reproduction Service No. ED 280 895)
- Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, New Jersey: Lawrence Erlbaum.
- Lord, F. M. (1983). Small N justifies Rasch model. In D. J. Weiss(Ed.), New Horizons in Testing. New York: Academic Press.
- Olsens, J. B., Maynes, D. D., Slawson, d., & Ho, K. (1986, April). Comparison and equating of paper-administered, computer-administered and computerized adaptive tests of achievement. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- Sands, W. A., Waters, B. K., & McBride, J. R. (Eds.) (1997). Computerized adaptive testing: From inquiry to operation. Washington, DC: APA.
- Stocking, M. L. (1994). Three practical issues for modern adaptive testing item pools(Report No. ETS-RR-94-5). Educational Testing Service. Princeton, N. J. (ERIC Document Reproduction Service No. ED 385

551)

- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., Mislevy, R. J., Steinberg, L., & Thissen, D. (1990). Computerrized adaptive testing: A primer. Hillsdale, N. J.: Lawrence Erlbaum associates.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. Journal of Educational Measurement, 21, 361-375.
- Wright, B. D., & Stone, M. H. (1979). Best test design. Chicago: MESA press.
- Wright, B. D.(1995). 3PL or Rasch? Rasch measurement transactions, 9 (1), 408-419.

附錄 A

題庫參數估計之資料矩陣範例

11111111111111111111
11111110000000000000
11111111111000000000
11111111100100000000
11110010100000000000
11111111110000000000
11111111110000000000

← 本資料為批改過之矩陣，故答案列均以1表示。
↪ 第二列起為學生之作答資料，因係批改過之矩陣，故以1、0表示之。

附註：1. 假如左列資料為作答之原始資料，在答案列中請輸入各題之正確答案，本系統將會自動批閱與估計試題之難度與受試者之能力。
2. 當每一筆資料超過80欄位時，請繼續往下輸入，不要按Enter鍵換行。最大輸入題數為254題。
3. 存檔時請以純文字檔儲存。

附錄 B

CPPT、CAAT、CAMT與CAGT中副程式與函數之各物件功能

- (1) Adjust_Click: 進行題目難度與受試者能力之重估。
- (2) CAT_Click: 啟動電腦化適性測驗系統、輸入受試者名字、測驗名稱、及記錄檔案名稱。
- (3) CautionIndex: 計算警告指標WSR、LZ。
- (4) CheckDisp: CPPT測驗系統檢核階段時，呈現題目與決定是否更正答案。
- (5) Command1_Click 第一選項按鈕。
- (6) Command2_Click 第二選項按鈕。
- (7) Command3_Click 第三選項按鈕。
- (8) Command4_Click 第四選項按鈕。
- (9) Command1_KeyUp 第一選項按鈕後顯示時間。
- (10) Command2_KeyUp 第二選項按鈕後顯示時間。
- (11) Command3_KeyUp 第三選項按鈕後顯示時間。
- (12) Command4_KeyUp 第四選項按鈕後顯示時間。

- (13) CTT_Click: 啓動CPPT測驗系統、輸入受試者名字、測驗名稱、及記錄檔案名稱。
- (14) Dril_Change: 尋找檔案時，決定檔案所在之路徑。
- (15) Drive1_Change: 尋找檔案時，決定檔案所在之磁碟機。
- (16) Ending_Click: 退出本測驗系統。
- (17) ErrorCheck: 檢核題庫檔案之格式是否正確。
- (18) Exit_Click: 退出檔案查詢。
- (19) Expocheck_Click 電腦化適性測驗各題目曝光率之檢查。
- (20) Filing_Click: 電腦化適性測驗成績之存檔。
- (21) Filing2: CPPT測驗成績之存檔。
- (22) FNL: 進行CAT測驗時，轉換Rasch能力與題目難度為T分數量尺之函數。
- (23) FNU: 進行CAT測驗時，轉換T分數為Rasch能力與難度量尺之函數。
- (24) FNT12_Click 測驗題目的字型設定為12號。
- (25) FNT14_Click 測驗題目的字型設定為14號。
- (26) FNT16_Click 測驗題目的字型設定為16號。
- (27) FORMATCHECK_Click: 檢查本測驗題庫格式與圖檔、語音檔是否正確、及是否存在正確路徑。
- (28) Form_KeyPress 利用鍵盤上1~4按鈕進行選項之選擇。
- (29) Form_eyUp 按鍵後顯示時間。
- (30) Grading_Click 進行適性等第評量。
- (31) Information 檢查測驗題庫之訊息曲線。
- (32) ItemDisp: 顯示題目文字。
- (33) Keyin: 進行電腦化適性測驗時輸入答案。
- (34) Keyin2: 進行CPPT測驗時輸入答案。
- (35) Mastery_Click 進行適性精熟測驗。
- (36) Modify_Click 進行CPPT測驗時，重新審題時顯示文字。
- (37) MusicStart_Click 聲波檔(*.Wav)與Midi(*.Mid)音樂檔啓動。

中文電腦化適性測驗系統軟體之研發

- (38) MusicStop_Click 結束聲波檔(*.Wav)與Midi(*.Mid)音樂檔。
- (39) NewAns:CPPT中重新檢核題目時，更正舊有答案與總分。
- (40) NOCONV：當使用unconditional maximun likelihood 估計Rasch題目參數法時，遇到超過10個迴圈仍無法聚斂時所使用之估計法。
- (41) Options_Click:在螢幕上呈現測驗結果。
- (42) PicItem2：進行CPPT測驗時選擇題目。
- (43) PickItem：進行CAT測驗時選擇題目。
- (44) Practice_Click 本測驗系統作答練習。
- (45) PROX_Click：使用PROX法估計Rasch題目參數法。
- (46) Prt_Click: 啟動功能表上之印表機。
- (47) PrtCat_Click：列印CAT測驗結果。
- (48) PrtCtt_Click：列印CPPT測驗結果。
- (49) Rasch：進行CAT測驗時，估計受試者之能力用。
- (50) Reestimate_Click：啟動功能表上之重估功能。
- (51) Reestimation：進行CAT測驗時，題目難度與受試者能力之重估與存檔。
- (52) RevDiff_Click：進行CPPT測驗時，題目難度之重估與存檔。
- (53) Search_Click：啟動功能表上之檔案尋找功能。
- (54) SortString：決定題目與題幹之長度以便決定圖形顯示位置。
- (55) Tccheck_Click：答對百分比與測驗訊息之對照，用以決定精熟水準與等第分派之 θ 決斷點。
- (56) Testing：CAT測驗系統之流程控制。
- (57) Testing2：CPPT測驗系統之流程控制。
- (58) Textbox_Click：此為Rich TextBox之物件，呈現題目文字與圖形用。
- (59) TimeEnd：決定測驗結束時間。
- (60) TimeStart：決定測驗開始時間。
- (61) UML_Click：使用Unconditional Maximun Likelihood法估計Rasch題目參數法。

(62) WavPlay_Click/DbClick/KeyUp 聲波檔(*.WAV)與Midi(*.Mid)音樂檔之
啓動、結束與顯示時間。

A Chinese Computerized Adaptive Testing System Developed for Classroom Uses

Fred Mao-neng Li

Abstract

The main purposes of the present article are (1)to analyze which components of a Computerized Adaptive Testing (CAT) are most needed in school testing settings ,(2)to describe briefly related theoretic foundations required for developing a CAT, (3)to explain how CAT is developed or programmed using Microsoft Visual Basic 6.0, and (4)to introduce the necessary steps to run the 4 sub-systems of CAT: Computerized Adaptive Achievement Testing, Computerized Adaptive Mastery Testing, Computerized Adaptive Grading Testing, as well as Computerized Paper-Pencil Testing, (5)to guide the user to build a test bank, and (6)to explain how to set-up the Chinese CAT system(CCAT).

The CCAT was designed with the following special features:

- (1) Font size or picture size may be adjusted during testing,
- (2) Graphics can be embedded in the text,
- (3) Diagnostic information can be obtained with an immediate test result,
- (4) Test items can be read aloud with the help of Otek Speech Wizard,
- (5) Student's unusual response patterns can be detected,
- (6) Parameter estimation and adaptive testing can be easily done within the CCAT,
- (7) Sound or music in midi- or wave-form can be played with the text,
- (8) Item difficul estimates in the item bank can be automatically recalibrated after enough data collection of examinee's responses,
- (9) Item exposure rates can be automatically checked for test security.