

多媒體英語文能力檢定暨適性化網路評量系統之建置

學生：黃吉楠

指導教授：陳登吉博士

國立交通大學理學院網路學習碩士在職專班

摘 要

本研究的目的是建立一套多媒體英語文能力檢定暨適性化網路評量系統，此系統功能如下所示：

- 符合英語聽、說、讀、寫四種能力的檢定機制，並以多媒體試題呈現之。
→改進傳統英語評量系統中缺乏支援多樣化多媒體試題的能力。
- 題庫適性值，將隨著不同受試樣本做動態的調整。
→改進傳統英語評量系統中缺乏動態適性值能力的調整機制。
- 將提供網路介面適性測驗的試題特徵曲線，方便研判試題適性值。
→改進傳統英語評量系統中缺乏試題特徵即時統計與曲線呈現機制。
- 整合不同的 IRT 分析器，進行適性測驗。
→改進傳統英語評量系統中缺乏整合不同特性的 IRT 分析器的能力。

本系統在題庫的建立上可做同等級試題適性值的驗證與兩段式的動態調整；並可多樣化呈現多媒體試題，為一多媒體適性化測驗平台。在聽力測驗與閱讀測驗中，由電腦自動評分，學生可立即知道成績，得到回饋。在口說測驗與寫作測驗中，更提供友善的介面，方便教師線上閱卷評分。測驗後的結果，亦可提供學生、教師及家長作成績查詢。

本系統可針對學生的作答反應資料，作 IRT 參數估算的分析，分析試題適性值，並分別產生試題的 ICC 圖、IIC 圖及試題難易度、鑑別度及猜測度等相關資料以進行研判。此系統不僅可動態地調整題庫之適性值，同時亦可編製成適性化測驗。當 IRT 試卷編製完成後，可以透過系統，進行試測，而系統將可立即性畫出作答情形、能力估算及標準誤圖形曲線，方便使用者研判。

關鍵字：試題反應理論、電腦適性化測驗、英語文能力檢定系統

The Design and Implementation of a Computerized English Test Delivery System and Corpus Analysis System based on IRT Model

Student: Jih-Nan Huang

Advisor: Dr. Deng-Jyi Chen

Degree Program of E-Learning
College of Science
National Chiao Tung University

Abstract

Recently, test items based on multimedia presentation for computerized test approach have received considerable attention due to the use of animation and video presentation features for the stimuli in the test question. The dynamic features incorporated in these multimedia test items can simulate the daily life that we encountered in the real world. This is particularly important for the language learning area such as English language ability evaluation. Traditional computerized English test questions lacking of the use of animation presentation features makes it a counterpart of the pen-paper based test approach except that it is conducted in the computerized environment. Also, the current IRT English test delivery systems do not support test analyzers the real-time display of Item Characteristic Curve (ICC), Item Information Curve (IIC), and Test Information Curve (TIC). Furthermore, only single IRT analyzer (such MLE or EAP) is used to estimate the test taker's ability, which may become inaccurate when abnormal response patterns are detected during the test.

This thesis study is to design and implement a multimedia computerized English test delivery system that overcomes those problems mentioned above. Specifically, the proposed system supports the following features: 1) it is equipped with listening, speaking, reading, and writing tests in multimedia presentation; 2) it is equipped with a two-staged IRT item attribute adjustment feature to normalize test question attribute indexes; 3) the analysis system allows analyzers to upgrade, downgrade, or modify the item attributes such as difficulty level of the analyzed test question by using the real-time display curves: Item Characteristic Curve (ICC), Item Information Curve (IIC), and Test Information Curve (TIC); and 4) it integrates several IRT analyzers to give a more accurate estimation of the test taker's ability when abnormal response patterns are detected. Application examples are studied and illustrated using the proposed approach. Besides, the proposed test delivery system and corpus analysis system has been used for the National English Test of Proficiency For All on the Web (NETPAW) to demonstrate the applicability and feasibility.

Key words: IRT, CAT, Computerized English Test Delivery System

誌 謝

本論文承蒙恩師陳登吉教授耐心的指導及教誨，恩師不僅在學術上給予指導，在待人處世方面及做研究的態度上也深深受恩師影響，本論文得以順利完成，在此對恩師致上無限的感謝。

此外，感謝所有曾教導我、幫助我的師長、朋友和同學，尤其是基隆市教育網路中心系統管理人員的大力協助，讓自己在程式的開發技術上，有了豐沛的支援。

感謝養育我、栽培我的父母親在背後的大力支持，使我得以專心完成論文，也才能有今天的我，從小一直支持我讀書的兄弟吉慧、吉助，如果當初沒有他們的支援，亦無現在的我。最後，感謝內人詩音照顧家庭，讓我無後顧之憂順利完成學業，這分辛勞與親情我銘記在心，終生的感激。



目錄

中文摘要	I
英文摘要	II
誌 謝	III
目 錄	IV
表 目 錄	VI
圖 目 錄	VII
一、 緒論	1
1.1 研究背景與動機	1
1.2 研究目的	2
1.3 研究方法、步驟	2
1.4 名詞定義	3
1.5 章節概要	4
二、 文獻探討	5
2.1 英語檢測相關探討	5
2.2 試題反應理論	7
2.2.1 試題特徵函數	8
2.2.2 試題參數估計	10
2.2.3 受試能力估計	10
2.3 電腦化適性測驗	11
2.3.1 電腦化適性測驗的特性	11
2.3.2 電腦化適性測驗的流程與方法	11
三、 系統需求、分析與設計	14
3.1 系統需求	14
3.1.1 同等級試題參數的驗證與兩段式的動態調整	14
3.1.2 多媒體適性化測驗平台	14
3.1.3 多樣化呈現之多媒體試題	15
3.2 系統架構	16
3.3 系統模組	17
3.4 資料庫結構	18
3.4.1 資料庫規格(PK 代表主鍵)	18
3.4.2 資料關聯圖(Entity- relation diagram)	23
四、 系統實作	25
4.1 使用者功能	25
4.2 試題上傳與維護	25
4.2.1 試題的上傳	25
4.2.2 試題的維護	26
4.3 線上試卷管理	27
4.3.1 新增試卷	27
4.3.2 同一等級隨機命題	29

4.3.2	試卷維護	30
4.4	線上測驗	31
4.4.1	登入測驗	31
4.4.2	多媒體線上測驗	32
4.4.3	場次管理與線上閱卷	35
4.5	IRT 測驗管理	37
4.5.1	試題的參數估算	37
4.5.2	試題特徵曲線(ICC)與試題訊息曲線(IIC)呈現	39
4.5.3	IRT 試卷電腦試測	41
五、	系統測試與結論	43
5.1	系統測試	43
5.1.1	線上多媒體測驗	43
5.1.2	IRT 適性分析	44
5.2	研究結論	48
5.3	對未來研究的建議	49
	參考文獻	50



表目錄

表 1 國內英語檢定比較表.....	7
表 2 本系統所提供的試題類型與傳統紙筆測驗之比較.....	16
表 3 檢測場次資料表 tcool_test_turn.....	18
表 4 題庫資料表 tcool.....	19
表 5 試卷資料表 tcool_exam.....	20
表 6 試卷明細資料表 tcool_qus	20
表 7 學生參加檢測明細 tcool_signup	20
表 8 答案卷 tcool_test.....	21
表 9 答案卷明細 tcool_test_ans.....	22
表 10 IRT 測驗卷資料表 tcool_irt_level	22
表 11 IRT 試題資料表理 tcool_irt.....	23



圖目錄

圖 1 研究步驟流程圖	3
圖 2 試題特徵曲線 ICC 圖	3
圖 3 試題訊息曲線 IIC 圖	4
圖 4 測驗訊息曲線 TIC 圖	4
圖 5 單參數模式之 ICC 圖	8
圖 6 雙參數模式之 ICC 圖	9
圖 7 三參數模式之 ICC 圖	9
圖 8 適性測驗施測流程	13
圖 9 同等級試題適性值的驗證與兩段式的動態調整流程圖	14
圖 10 多媒體測驗的流程與適性化測驗的的流程圖	15
圖 11 系統架構圖 (System Architecture)	17
圖 12 系統模組圖	17
圖 13 資料關聯圖	24
圖 14 使用者功能	25
圖 15 試題瀏覽器上傳	26
圖 16 試題維護畫面	27
圖 17 規劃新測驗試卷	28
圖 18 選擇測驗試題	28
圖 19 試卷與試題預覽	29
圖 20 同一等級試題隨機命題	29
圖 21 試卷維護	30
圖 22 試卷試題查詢	30
圖 23 測驗選項	31
圖 24 一般測驗登入畫面	31
圖 25 英檢錯誤訊息	31
圖 26 英檢錯誤訊息 2	32
圖 27 聽力測驗畫面	32
圖 28 英檢電腦評分	33
圖 29 閱讀測驗	33
圖 30 口說能力測驗	34
圖 31 錄音元件	34
圖 32 寫作能力測驗	35
圖 33 新增檢測場次	36
圖 34 口說測驗能力的閱卷	36
圖 35 寫作能力的閱卷	37
圖 36 新增 IRT 試卷	38
圖 37 試題參數的估算	38
圖 38 試題特徵曲線(ICC)與試題訊息曲線(IIC)圖	39
圖 39 即時 ICC 圖的比較	39
圖 40 即時的 IIC 圖比較	40

圖 41 即時試卷的全部試題的 ICC 與 IIC 圖比較.....	40
圖 42 試卷訊息曲線圖(TIC).....	41
圖 43 不同 IRT 分析器的能力估算.....	41
圖 44 整合不同特性的 IRT 分析器作答曲線圖	42
圖 45 作答的標準誤曲線呈現	42
圖 46 初級的通過率自動分析	44
圖 47 以 IRT 分析之多媒體試題 1.....	45
圖 48 以 IRT 分析之多媒體試題 1 之 ICC 圖與 IIC 圖.....	45
圖 49 以 IRT 分析全部 ICC 圖	46
圖 50 以 IRT 分析全部 IIC 圖	47
圖 51 以 IRT 分析 TIC 圖.....	48



一、緒論

本章共分為三節，分別為(一)研究背景與動機(二)研究目的(三)研究步驟與方法(四)名詞定義(五)章節概要。

1.1 研究背景與動機

隨著我國加入世界貿易組織(WTO)，我們所需面對的是全球化市場的競爭，因此具備良好的外語溝通能力，已成為職場競爭力之要件。鑑於此行政院於 2002 年 5 月宣布「挑戰 2008 國家發展重點計畫」，而其中「E 世代人才培育計畫」排名首位也最為大家所重視。此計畫之四大議題的第一項即為營造國際化生活環境，提昇全民英語能力[1]。

現今許多企業為为了提高其國際競爭力，對於員工之英語能力亦多加重視與要求，招聘時多加註求職者需提出全民英檢、托福等具體之能力證明，或是鼓勵員工在職進修，以取得相關文件。嚴格者甚至要求每隔幾年參加檢定考試，並設定最低分數門檻，以提升其英語應用能力，因此發展一套具公信力的英語文能力檢定系統是必要的。

在現今資訊科技與網際網路快速發展下，測驗亦已由傳統的紙筆測驗逐漸 e 化。且經由多媒體電腦呈現多樣化的試題，並依照考生能力出題的適性化測驗亦成趨勢。但研究者比較國內較著名的英語文檢定測驗如全民英檢(GEPT)、多益測驗(TOEIC)、托福檢定(TOEFL)及 ILTES 發現僅有 TOEFL 以電腦適性化方式進行[2,3,4,5]，而對於目前國內的許多英語文能力檢定部份，仍存在著以下問題：

- 缺乏支援多樣化多媒體試題的能力。
- 缺乏試題特徵即時統計與曲線呈現機制。
- 缺乏動態適性值能力的調整機制。
- 缺乏整合不同特性的 IRT 分析器的能力。

因此研究者針對上述問題建立一套多媒體英語文能力檢定暨適性化網路評量系統，以符合下列功能：

- 符合英語聽、說、讀、寫四種能力的檢定機制，並以多媒體試題呈現之。
- 作為適性化題庫適性值，隨著不同受試樣本做動態的調整。
- 將提供網路介面適性測驗的試題特徵曲線，方便研判試題適性值。
- 整合不同的 IRT 分析器，進行適性測驗。

1.2 研究目的

本研究的目的是在於結合適性化理論及網路技術，建置多媒體英語文能力檢定暨適性化網路評量系統，並完成下列目的：

1. 符合聽、說、讀、寫能力檢測多媒體試題
2. 試題參數可隨不同樣本數作動態調整
3. 以即時視覺化圖形曲線表示試題特徵曲線(ICC&IIC)
4. 以即時視覺化圖形曲線表現適性化施測流程與測驗訊息(TIC)
5. 統整不同的 IRT 分析器進行適性測驗

1.3 研究方法、步驟

爲了達成製作三階層網路適性測驗，研究者將以動態網頁(PHP)技術結合資料庫(MySQL)及網路伺服器(Apache)的方式來設計與規劃整個系統，研究步驟可分爲下列六點：

1.收集分析現有適性測驗平台系統架構:

首先研究分析現有的適性測驗平台架構，整理歸納其優缺點，並由相關文獻的閱讀中，整理出目前英語能力檢定之方式及不足之處；並了解試題參數可隨不同樣本數作動態調整及其在適性測驗中的重要性。

2.需求分析:

根據系統建構的基礎理論探討，及對現有測驗平台系統架構作研究需求分析的過程中，逐步歸納出適性測驗的系統規格需求。

3.規劃系統架構:

以互動式網頁(PHP) 的技術，設計規劃系統整體架構，以確保符合適性測驗之需求。

4.系統發展與實作:

根據所規劃出之系統架構，逐步將需求完成，並設計人性化介面以管理、應用系統。

5.由系統實際實作後所收集之測試結果，評估系統可行性。

6.提出論文結論與未來發展方向。

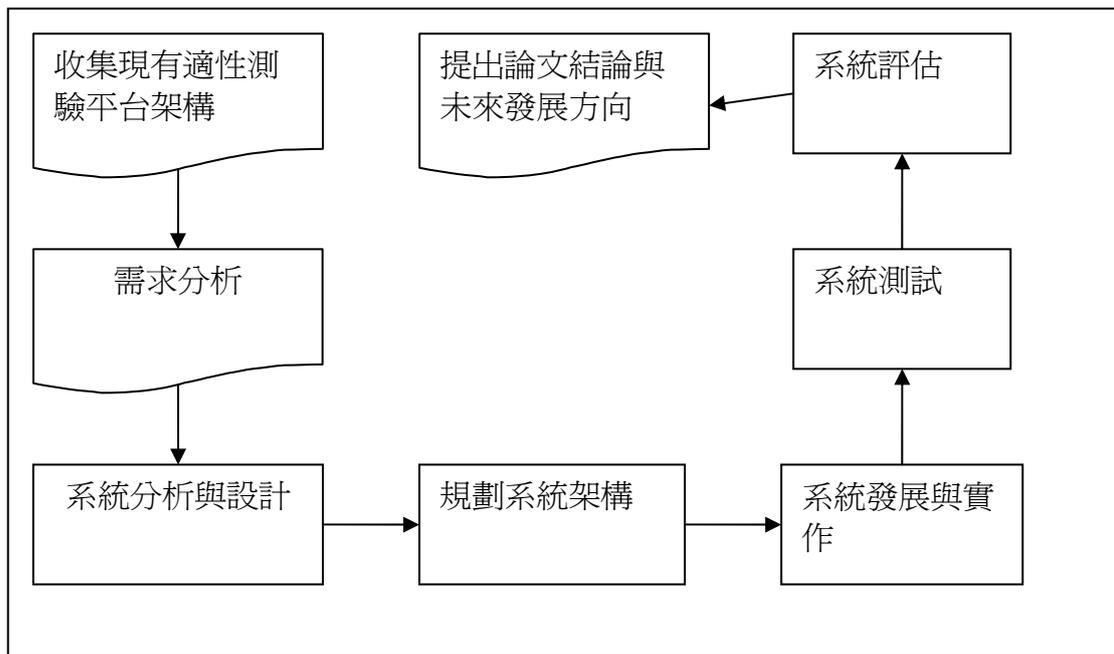


圖 1 研究步驟流程圖

1.4 名詞定義

1. 試題反應理論 IRT (Item Response Theory)

以機率來解釋考生能力與試題反應間的關係，能夠評估受測者的能力值範圍，並且以鑑別度(a 值)、難易度(b 值)及猜測度(c 值)提出以試題特徵、試題訊息及測驗訊息的概念，來作為評定某試題或某測驗的施測準確性。

2. 試題特徵函數(Item Characteristic Function)

用來解釋考生的表現情形與潛在特質（能力）間的關係，是一條連續性遞增的曲線，又稱試題特徵曲線(Item Characteristic Curve, ICC)，如下圖所示：

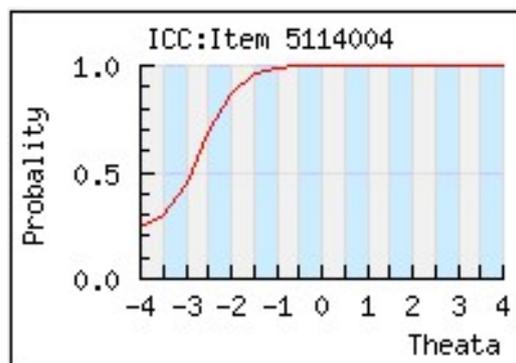


圖 2 試題特徵曲線 ICC 圖

3. 試題訊息函數(Item Information Function)

試題訊息函數(Item Information Function)可作為試題優劣的一個判定函數，在某一能力值的試題訊息，代表該試題所能提供給某能力值的貢獻量，試題訊息量越高，其誤差越小，可用來解釋試題與考生潛在特質（能力）

間的關係，將其繪出即試題訊息曲線(Item Information Curve, IIC)，如下圖所示：

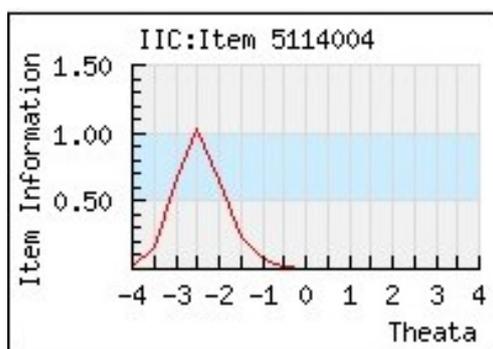


圖 3 試題訊息曲線 IIC 圖

4. 測驗訊息(Test Information Function)

具有加總的特性，為所各施測試題在相同的能力值上的試題訊息總和，而將其繪出的圖形則為測驗訊息曲線 (Test Information Curve, TIC)。

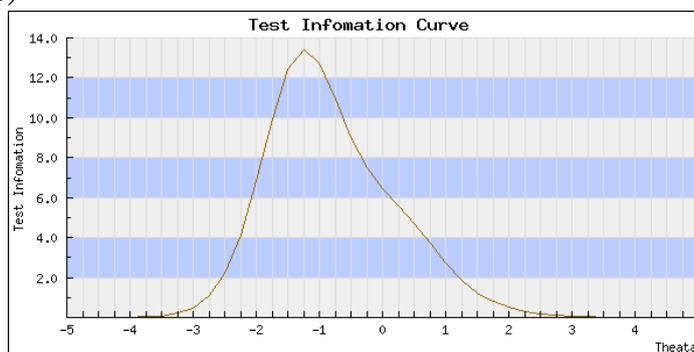


圖 4 測驗訊息曲線 TIC 圖

5. 電腦適性化測驗(Computerized Adaptive Testing, CAT)

電腦化適性測驗就是將原本各項考試的題目儲存在電腦，形成題庫(item bank)，CAT 是以題庫為核心發展而成的，而且其選題是根據試題參數來進行，並依受試者之答題反應，估計受試者能力值，選擇難易適中的題目，以獲得描述個體成就水準的訊息。

1.5 章節概要

在第二章中，為使建構的系統能切合適性測驗的需求，將探討系統建構的基礎理論，包括相關英語能力檢定方式之探討、試題反應理論及電腦化適性測驗。

在第三章中，依照所提出系統架構功能需求分析，規劃製作出一個三階層網路適性測驗系統架構並符合所需。

在第四章中，說明系統的建置過程與實作情形。

在第五章內，將會對系統的實際運作情形做評估及總結，並對未來的發展方向提出建議。

二、文獻探討

本章共分三節，分別為（一）英語檢測相關探討（二）試題反應理論（三）電腦適性化測驗。

2.1 英語檢測相關探討

本節將以近年來較受重視之四項英語檢測—TOEIC、TOEFL、IELTS、GEPT 做相關文獻探討。

1. 國際溝通英語測驗(TOEIC, Test Of English for International Communication)

TOEIC 主要是評量母語非英語人員在國際商務環境中實際運用英語的能力，該測驗由美國教育測驗服務社（Educational Testing Service, ETS[®]）於 1979 年所設計，且由 ETS 的子公司(Chauncey Group International Ltd)負責發行製造。因其測驗能做出公正客觀的評量，具有高度可信度，故已成為當今頗具公信力的職業英語能力測驗，至今已有數千家公司根據 TOEIC 成績評估員工的英語能力，受測人數亦高達二百多萬人。

目前 TOEIC 施測方式仍為傳統的紙筆測驗，內容以國際商業活動與日常使用的英語為主，題型則分為聽力與閱讀二大部份，共計有二百題之單選題，測驗進行時間約兩小時，考試分數由答對題數決定，加總聽力與閱讀得分即為總分。其考試場地本僅有台北與高雄兩考區，今年更增加桃園及台中考區，場次部份台北每月一場，台中及高雄每二個月一場，桃園則三個月一場，2005 年總計有 22 場[3]。

2. 托福測驗(TOEFL, Test Of English as a Foreign Language)

TOEFL「托福」主要為測驗應試者在學業環境中的英語能力，此測驗由美國教育測驗服務社(Educational Testing Service, ETS[®])於 1965 年起承辦，並在全世界舉行。至今美加已超過 2,400 所大學和學院承認該考試成績，並規定申請者的最低 TOFEL 錄取分數。

台灣地區「托福」測驗自 2000 年 10 月起已由傳統的紙筆測驗全面改為電腦化測驗（Computer- Based Testing, CBT），分為聽力(Listening)、文法結構(Structure)、閱讀(Reading)及寫作(Writing)等四個項目。其中聽力及文法結構是採用電腦化適性測驗即考題難易度與應試者程度相配合，作答時應試者一次只能看到一道考題，待輸入答案後，電腦會立刻根據答題的對錯情況，決定下一題的難易度，通常第一道考題為中等難度，若答對，下一題難度會提高；答錯了，下一題難度會降低。此方式在應試者輸入答案後，無法檢查或更改答案，且不能略過任何一題不作答。閱讀項目則採用傳統的線型測驗，應試者可以略過某題不作答，也可回頭更改已答過的答案。寫作測驗為必考項目，應試者必需以打字或手寫方式完成一篇文章。

目前 TOFEL 測驗台灣考區僅有台北與高雄兩地，但測驗中心全年開放，

每週一至週五有 2 場次，每月第三個週六則開放 1 場次[4]。

3. 國際英語測驗(IELTS, The International English Language Testing System)

IELTS 的前身原是英國當局針對留學生而設的英語能力檢定測驗，自 1990 年起改稱為 IELTS (The International English Language Testing System)。因 IELTS 注重學生在專業範疇內需具備之英語能力，故成為國際認可的高等教育入學所需之英語能力證明。IELTS 測驗是由英國文化協會 (the British Council)、劍橋大學考試委員會(Cambridge ESOL) 及澳洲大學聯盟(IDP Education Australia) 共同舉辦，而由劍橋大學考試委員會統籌管理。在臺灣則由英國文化協會和澳洲大學台灣分公司負責定期舉辦測試。

IELTS 目前仍以紙筆測驗為考試方式，內容分為聽、說、讀、寫四大部份，其中讀與寫測驗採分組方式進行，分別為適用於申請正式課程者所需的 Academic 組與適用於申請移民者之 General Training 組；說的部份是屬於一對一的口試方式，分為自我介紹、個人表達及雙方討論三階段進行。在臺灣 IELTS 有台北、台中及高雄三考區，每月進行至少一場次的測驗[5]。

4. 全民英語檢定(GEPT, General English Proficiency Test)

全民英語能力分級檢定測驗(以下稱「全民英檢」)係教育部補助研發，自民國 89 年起由財團法人語言訓練測驗中心主辦，本測驗參考當前我國英語教育體制，訂定分級標準，共分為初級、中級、中高級、高級及優級等五級，每級各有其明確能力標準，適合各階段英語學習者依能力選擇適當的級數報考。

全民英檢的檢測對象為一般社會人士及在校學生，測驗之目的在評量一般英語能力，因此命題不侷限於特定領域或教材。目前測驗仍以傳統紙筆方式進行，內容均包含聽、說、讀、寫四大項，其中聽力及閱讀測驗採選擇題方式，口說及寫作測驗則為非選擇題型式，各級並針對能力指標分別設計題型及命題內容。除優級外，各級初、複試的考試場次，每年各舉辦二次，只要通過所報考級數之初試與複試之標準即可取得該級的合格證書。此成績除可用來評量個人的英語能力，亦可作為公私立學校教學成果評鑑及改進的依據，並提供公民營機構招募人才、升等評鑑及甄選派駐國外人員等之參考[2]。

茲將以上四種英語檢定考試依測驗目的、項目、方式、題型及考區場次等作一整理如下表所示。由表 1 可看出，傳統紙筆測驗考試考區及場次都較電腦化測驗受限制，而適性化測驗更是國際英語檢測未來的趨勢。

表 1 國內英語檢定比較表

測驗名稱	測驗目的	測驗項目	測驗方式	測驗題型	測驗考區/場次
TOEIC 多益	評量個人在國際商務環境中實際運用英語的能力，是企業評估員工英語能力標準之一。	聽力、閱讀	傳統紙筆測驗	單選題	台北：每月一次 台中、高雄：二月一次 桃園：三月一次
TOEFL 托福	評量母語非英語者之英語能力，主要為申請進入美加地區之大學或研究所之學生的語文能力證明。	聽力、文法結構 閱讀、寫作	1.全球之測驗方式有紙筆與電腦化。 2.台灣區則以電腦化為測驗方式，其中聽力及文法結構為電腦化適性測驗。閱讀項目則採傳統的線型測驗。	選擇題 插入文句 作文等	考區僅台北與高雄兩地，場次週一至週五各二場，每月第三個週六一場
IELTS 國際英語測驗	評量學生在專業範疇內的英語能力。	聽、說(不分組) 讀、寫(分Academic 及 General Training 兩組)	傳統紙筆測驗	填空、選擇、問答、配合、作文及口試	考區分台北、台中與高雄三地，場次為每月至少一次
GEPT 全民英檢	為國內評量個人英語能力之工具，並可作為學校教學成果評鑑及改進的依據。	初試(聽力、閱讀) 複試(寫作、口試)	傳統紙筆測驗	選擇題 單句寫作 朗讀短文等	除優級外，各級每年舉辦二次

2.2 試題反應理論

試題反應理論(Item Response Theory) 是由 Lord[6] 於 1980 年提出，是以機率來解釋考生能力與試題反應間的關係；亦即以一數學模式，推估受試者的能力或潛在特質，此數學模式稱為試題特徵函數(Item Characteristic Function, ICF) 余民寧[7]認為 IRT 建立於兩個基本概念上：(1) 考生在某一試題上的表現情形，可由一組因素來加以預測或解釋，這組因素叫作潛在特質或能力(ability)；(2) 考生的表現情形與潛在特質間的關係，可透過一條連續性遞增的曲線來加以解釋，這個曲線便叫作試題特徵曲線(Item Characteristic Curve, ICC)。

2.2.1 試題特徵函數

IRT 以 ICF 表達受試者能力和測驗反應間之關係，因函數中所採用的參數個數不同，通常被分成單參數、雙參數以及三參數等三種模式。茲將各模式簡述如下：

1. 難度參數又稱單參數模式

$$P_i(\theta) = \frac{1}{1 + e^{-1(\theta - b_i)}} \quad i = 1, 2, \dots, n \quad (1)$$

- $P_i(\theta)$ 表示任何一位能力為 θ 的考生答對試題 i 或在試題 i 上正確反應的機率
- b 表示試題難度(*difficulty*)參數
- n 是該測驗的試題總數
- e 代表以底為 2.718 的指數
- $P_i(\theta)$ 是一種 S 形曲線，其值介於 0 與 1 之間。

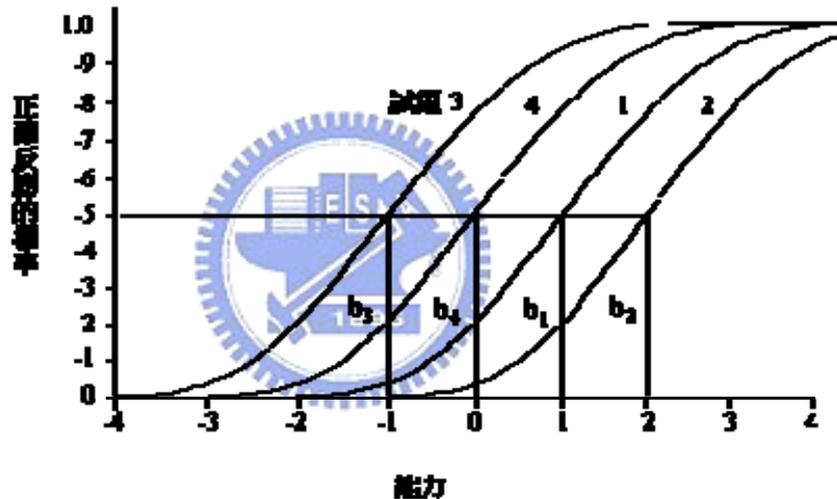


圖 5 單參數模式之 ICC 圖

說明：根據公式 1 的定義，試題難度參數(b)正好落在正確反應機率為 0.5 時的能力量尺上的點（即能力值）；即當能力值和試題難度相等時($\theta - b = 0$)，考生答對該試題的機率只有 50%。由此可知圖 5 中試題 3 的難度為 -1，試題 4 的難度為 0，試題 1 的難度為 1，試題 2 的難度為 2，曲線愈往右試題愈難。

2. 鑑別力參數又稱雙參數模式

$$P_i(\theta) = \frac{1}{1 + e^{-a_i(\theta - b_i)}} \quad i = 1, 2, \dots, n \quad (2)$$

比公式 1 多了一個參數：試題鑑別度(item discrimination) a_i ，用來描述試題 i 所具有的鑑別力。

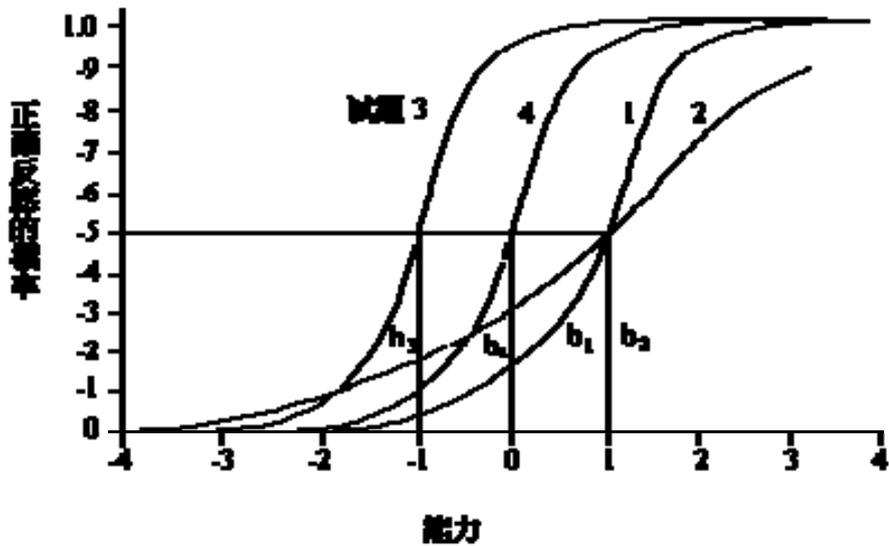


圖 6 雙參數模式之 ICC 圖

說明：試題鑑別度參數 a 值，正好與在 b 點的試題特徵曲線的斜率成某種比例。 a 值愈大，代表試題特徵曲線愈陡，試題愈有良好的鑑別度；故由圖 6 中可看出試題 1,3,4 的鑑別度相似，且比試題 2 高。

3. 猜測參數又稱三參數模式

$$P_i(\theta) = C_i + (1 - C_i) \frac{1}{1 + e^{-a_i(\theta - b_i)}} \quad i = 1, 2, \dots, n \quad (3)$$

猜測參數(pseudo-chance parameter) C_i 。這個參數提供試題特徵曲線一個大於零的下限，它代表著能力很低的考生答對某試題的機率。

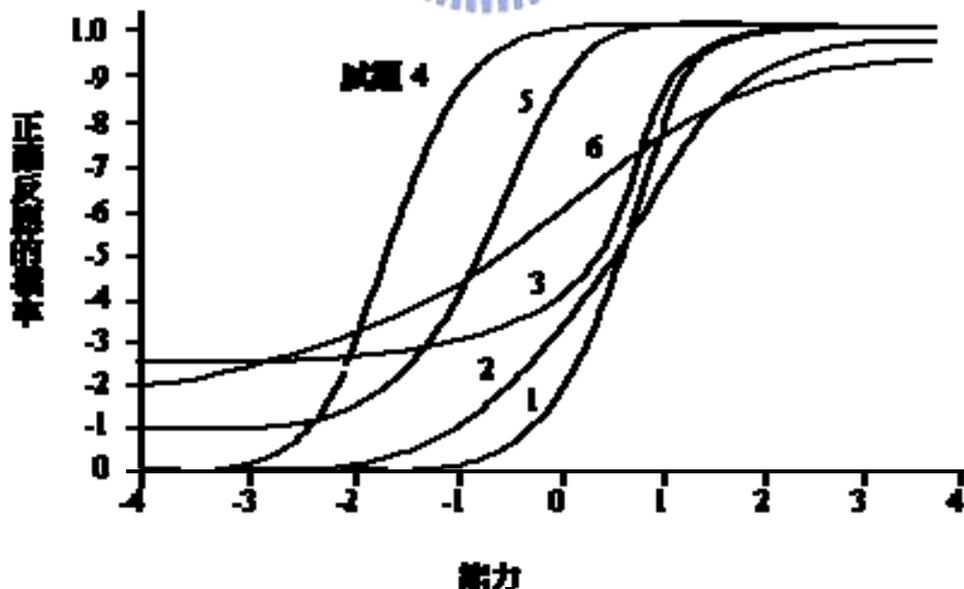


圖 7 三參數模式之 ICC 圖

說明：參數 C 表示能力值很低的考生答對試題的機率，即猜測率。由圖 7 可知，試題 1,2,4 當考生能力值低至 -1,-2,-3 時，答對機率為 0，故可謂無猜測度。

2.2.2 試題參數估計

IRT 以試題為測量受試能力的基本單位，並以試題參數來描述試題的特質。認為試題參數資料配合受試反應組型之函數運算後，即可估計出受試者之能力，亦即試題參數估計之準確性，會直接影響受試能力估計。

引用劉長萱[8]所提試題參數之建立通常是先經傳統施測過程，收集受試者的反應組型，再進行試題參數的推估。倘若受試者的能力值為已知時，可使用最大概似估計法(Maximum Likelihood Estimation；MLE)、貝氏估計法(Bayesian Parameter Estimation)直接估計試題的 a 、 b 、 c 參數；當受試能力和試題參數皆未知的情況時，則可使用聯合最大概似估計(Joint Maximum Likelihood Estimation；JMLE)、邊際最大概似估計(Marginal Maximum Likelihood Estimation；MMLE)、最大後驗估計(Bayesian Modal or Maximum a Posterioriestimation；MAP)、期望後驗估計(Bayesian Mean or Expected aposteriori Estimation；EAP)等方法，對試題和能力參數進行同時推估(Hambleton & Swaminathan[9])。

目前已有許多軟體提供試題參數估計的程式，大多數的測驗編製者也使用它們來進行試題分析和參數估計。這些軟體是由不同的學者所發展，所採用的參數估計方法也略有差異，例如常用的電腦程式，MicroCAT 所發展的估計軟體 X-Calibrate 則使用貝氏估計法，另一估計軟體BILOG 則採用MLE、MAP、EAP。

2.2.3 受試能力估計

訊息函數(Information Function)在電腦適性測驗中，扮演一個重要的角色，在測驗進行中，藉由試題訊息函數，挑選適合於受測者的試題，可求得最大測驗訊息，使其能力估計的標準誤為最小，達到提高受測者能力估計的精確度，並且能提高測驗的效率。

試題訊息函數可以應用到前面所談到的單參數、雙參數、三參數對數形試題反應模式，這些模式都適合用於二元計分(Dichotomously Scored)的測驗資料。例如，以三個參數對數形模式為例，公式為：[6]

$$L_i(\theta) = \frac{a_i^2(1-C_j)}{[C_j + e^{a_i(\theta-b_i)}][1 + e^{-a_i(\theta-b_i)}]^2} \quad (4)$$

從此公式裡，很容易便可推知 a 、 b 、和 c 參數在試題訊息函數中所扮演的角色：(1)當 b 值愈接近 θ 時，訊息量較大；(2)當 a 參數較高時，訊息量也會較大；(3)當 c 參數接近 0 時，訊息量則會增加。

試題訊息函數能表示出試題對能力估計的正確性，而正確性的大小端受兩個主要因素的決定：一為試題的鑑別度參數大小，亦即 a 值愈大，試題特徵曲線即愈陡， $P(\theta)$ 的斜率愈大，所以訊息量愈高；另一為試題的難度參數，它的位置會決定訊息量的高低。Birnbaum[10]指出，某個試題所提供的最大訊息量，剛好出現在能力參數為 θ_{\max} 的點上， θ_{\max} 的值為：

$$\theta_{\max} = b_i + \frac{1}{a_i} \ln[0.5(1 + \sqrt{1 + 8C_i})] \quad (5)$$

如果猜測機率為最小時（亦即，當 $C=0$ 時），則 $\theta_{\max}=b_i$ 。一般而言，當 $C>0$

時，某個試題在能力水準比其難度值稍高的位置上，所提供的訊息量會達到最大。訊息量愈大，能力估計標準誤就愈小；當信賴區間愈小，能力的估計即愈精確。

2.3 電腦化適性測驗

2.3.1 電腦化適性測驗的特性

適性測驗的觀念係源於現代心理計量學者所提出的一種較新的測量概念：測量個體能力時若能施以此個體難易適中的題目，最能獲得描述個體成就水準的訊息[11]。因此適性測驗實施時，須依據受試者之答題反應，立即估計出受試者目前能力值，並依此能力值，選擇難度相當之題目；即答對後選擇較難試題，反之，則選擇較易試題[12]；如此一來，測驗就可以適應個別的學習情境與能力，可說是一種量身定做的測驗[13,14]。但在施測過程須繁覆計算受試者能力值，以及各試題之訊息量等，以人工方式進行可謂費時費力，因此適性測驗之實施須在電腦上執行，故常稱為電腦化適性測驗(omputerized Adaptive Testing, CAT)。

根據學者們的研究[8,15,16]，電腦化適性測驗和傳統紙筆式測驗相較，具有以下優點：

1. 電腦化適性測驗施測之題目符合受試者能力水準，施測完畢後能即時評分，克服紙筆測驗評分速度慢及選題上的困難；
2. 不同受試者作答不同的試題，其能力估計值可以互相比較；
3. 適性測驗是依據受試者在答題時的表現，估計其能力，再從題庫中選擇能提供最大訊息量的試題來施測，因此施測題數少於紙筆測驗的一半以上，故可以節省測驗的時間，但又不影響對受試者能力估計的準確度。

學者們亦指出其缺點如下：

1. 受限於電腦等資訊設備的效能；
2. 考生要具備電腦的基本操作常識；
3. 一次只能作答一題，且不能回頭修改。

從以上相關的研究可知電腦化適性測驗主要的優點是能為考生量身定做合適的測驗，而主要的限制為設備與操作電腦能力的問題，但這些限制在資訊科技日益發達，資訊教育往下紮根的現代，已不復存在。因此在講求適性化教育與評量多樣化的今日，電腦化適性測驗實值得採用。

2.3.2 電腦化適性測驗的流程與方法

電腦化適性測驗是基於受試者的個別差異，選擇最適合其能力水準的項目來施測。在施測時，不同的學者有不同的見解，但綜合而言，仍不脫以下六個步驟[8,15,17]：

1. 試題反應模式的選擇：在現代測驗理論中，包括多種試題反應模式，如單參數、雙參數、或三參數對數型模式，每種模式各有其適用的範圍及基本假設，故針對不同類型資料及應用目的，可選擇最適合的模式作為適性測驗進行時的基本根據。其中，三參數對數型模式是被最常選用的模式。

2. 題庫的建立：建立題庫是電腦化適性測驗的首要條件，題庫的性質具備以下特點[15]：
 - (1) 題庫題目的參數應包括難度、鑑別度及猜測機率三種參數；
 - (2) 難度參數的範圍應能涵蓋所有受試能力；
 - (3) 鑑別度參數應以0.8~1.25 為宜，猜測機率則應在0.3 以下；
 - (4) 題庫的題數最好在100 題以上；
 - (5) 題庫的題目均在測相同特質的能力。
3. 測驗的起始點：常用的方法有：(1)自難度適中的試題中隨機抽取；(2)由系統完全隨機抽取；(3)先調查受試者的測驗記錄，估算其能力，然後再決定出那一類的試題。由於電腦化適性測驗試題的難度會隨受試者的作答反應做調整，所以偏差的起始點不至於影響測量結果，故Lord[18]認為，只要測驗的題數不少於25 題的話，以那一個試題做為起點的影響不大。
4. 選擇題目的方式：一般而言，常用的試題選擇的方法有三種：(1)最大訊息選題法：選取一題能提供受試者目前能力水準最豐富訊息的試題施測；(2)貝氏選題法：選取一題能使受試者能力估計的事後變異數為最小的試題施測。(3)挑選難度最接近考生現階段能力估計之試題。
5. 終止的標準：終止適性測驗的方法，和前述的選題與計分方法間有很密切的關聯。若以最大訊息選題法作為選題標準的話，可設定整個測驗的訊息達到某一特定標準，測驗即可終止。若以貝氏估計法來選題的話，則可設定估計能力之變異數小到某一特定值時，終止測驗。此外，如果前述兩種標準均很慢才達到的話，也可以預設施測試題的上限，只要題數一測完，便終止施測，以避免測驗漫無止境地繼續下去。
6. 計分方法：其實也就是學生的能力估計方法，常用的兩種能力估計法為(1) 最大概似值估計法(Maximum- Likelihood Proficiency Estimate)；(2) 貝氏估計法(Bayesian Estimate)，來進行估算受試者的能力，據此選擇下一個試題，直到估算的標準誤控制在設定範圍內為止最大概似值估計法的估計效能很好，但遇到題數少或估計值無法收斂時，都會產生很大的問題；貝氏估計法雖能克服這些困難，但對事前分配的假設如果不當的話，卻會產生有所偏差的能力估計值。

除上述六個步驟外，何榮桂[13]亦指出適性測驗施測方式係根據一個簡單的原則，即：受試者每做完一個題目，其能力會依答題反應重新被估計，然後再依據此能力值，從題庫中選取難度合適的題目施測，直到此估計能力值之精確度達到一個預先設定的標準時，測驗即結束，整個流程如圖8所示。

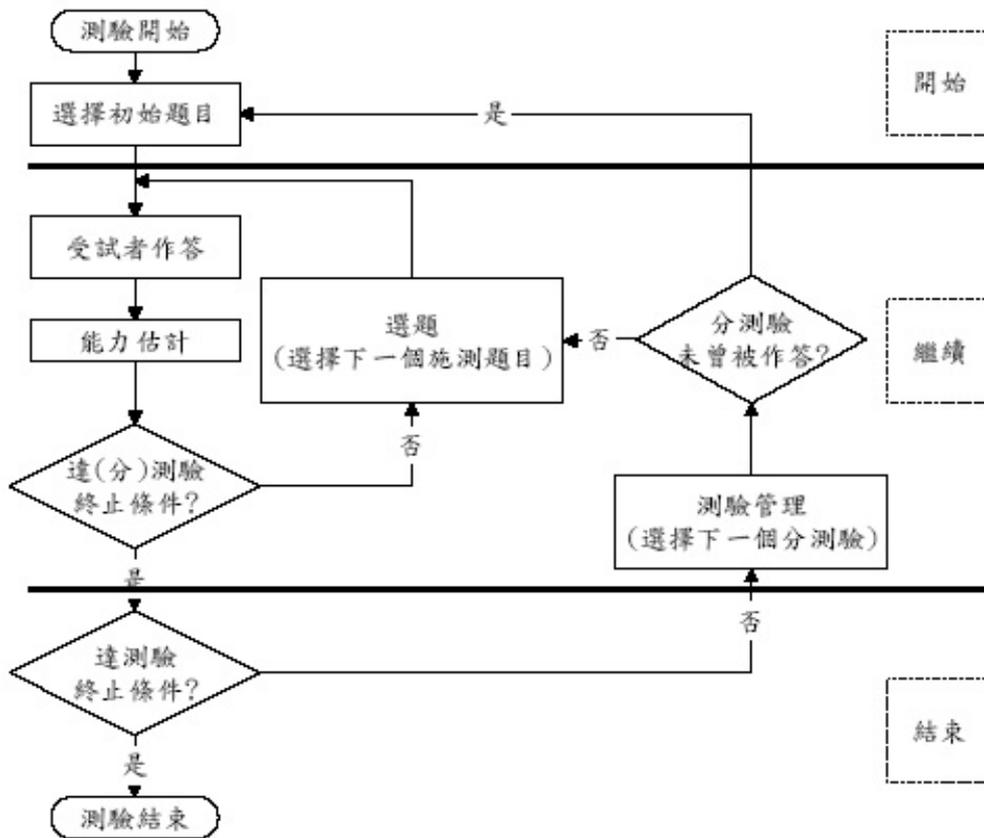


圖 8 適性測驗施測流程



三、 系統需求、分析與設計

本章將針對系統的需求，系統的分析，以及系統設計，利用軟體工程的方法，逐步的加以建立分析文件，分別提出系統需求、系統架構、系統模組與資料庫設計。

3.1 系統需求

3.1.1 同等級試題參數的驗證與兩段式的動態調整

題庫品質的是否具有鑑別度及符合現況，常常是一個測驗系統發展的重要關鍵。故本系統之題庫建立機制，詳如下圖：

由命題委員編製完成試題後，須先經驗證系統進行驗證，以確保其相關的字頻、字表、文法構句等相關適性值，然後將其上傳至試題庫，而進行試題預試後，收集相關之答題反應結果，即啟動 IRT 分析器的機制，立即進行分析，分別求出其相關 a,b,c 等參數值，透過立即性視覺化圖形呈現，便可以進行研判，進行第一次動態調整將相關適性值回存儲至試題庫。之後再進行 CAT 施測後，便可以完成第二次的適性值動態調整。

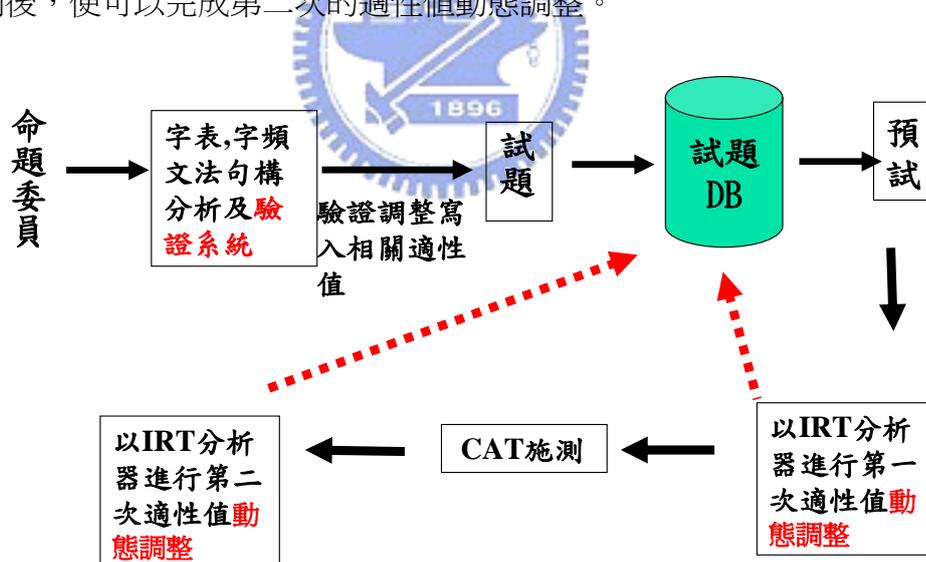


圖 9 同等級試題適性值的驗證與兩段式的動態調整流程圖

3.1.2 多媒體適性化測驗平台

不管是多媒體試題測驗或是適性化測驗，其測驗流程如下圖所示：

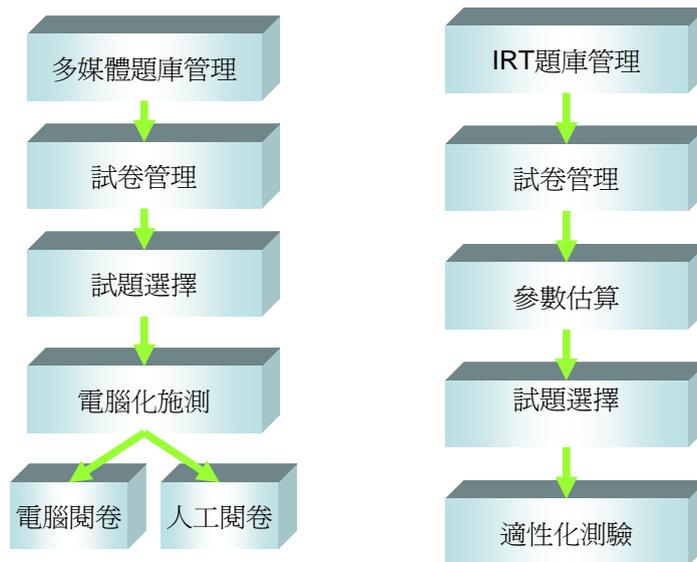


圖 10 多媒體測驗的流程與適性化測驗的的流程圖

故系統須能提供支援多媒體題庫的的題庫管理，而老師可以透過它來分別製作符合不同等級的聽力、口說、閱讀、寫作四種能力英語檢測多媒體的試卷，以提供學生做英語能力多媒體試卷的測驗，受測者完成測驗中之聽力與閱讀能力部份，可以立即知道測驗結果(電腦閱卷)，加強學生學習的效果。

在口說與寫作能力測驗部份並提供教師一個良好的介面，可以直接進行線上閱卷(人工閱卷)並加以評分。測驗後的結果，亦立即提供給學生、老師與家長作成績的查詢。

聽力與閱讀之測驗試題，將可以啟動 IRT 的分析機制，進行分析，分析完後便可以重新調整適性值，然後進行適性化測驗。

3.1.3 多樣化呈現之多媒體試題

目前現有的測驗軟體系統，在試題內容上，都偏重於文字測驗，至於多媒體試題非常少，本系統試題可以支援內容包括：文字、圖片、聲音、影片等多樣化的組合，以達到以多媒體呈現之測驗的目的。

表 2 本系統所提供的試題類型與傳統紙筆測驗之比較

	本系統支援題型	傳統紙筆測驗題型
聽力測驗	聲音播放試題（多種題型樣板） 視訊播放試題	聲音播放試題(要用錄音機)
閱讀測驗	圖片展示試題 文字展示試題 動畫試題 視訊播放試題	圖片展示試題 文字展示試題
寫作測驗	圖片展示試題 文字展示試題 動畫試題 視訊播放試題	文字展示試題
口說測驗	圖片展示試題 文字展示試題 視訊播放試題 動畫播放試題	口說能力測驗 圖片展示試題 文字展示試題
其它	影片視訊（視訊播放試題） 歌謠（聲音播放試題） 歌劇（聲音播放試題） 演講或詩詞朗誦聲音播放試題） 口說翻譯（視訊播放試題）	紙筆測驗無此呈現方式

3.2 系統架構

經過系統的分析後，我們繪出多媒體英語能力檢定暨適性化網路評量系統架構圖如下圖，並說明如下：

1. CAT Server：為適性化測驗系統，主要負責線上命題，線上測驗、作答記錄及測驗流程的控管。
2. Item Bank Server：為題庫系統，由資料庫記錄其相關試題的索引值及屬性，而每一道題庫由一個資料夾放置其實體檔案，並能即時繪出相關的試題特徵曲線圖(ICC)、試題訊息曲線圖(IIC)等。
3. FTP Server：為檔案伺服器，負責作答時如需要上傳檔案(如口說測驗時作答的錄音檔)的上傳或是由其它的題庫編輯軟體(如編輯手)上傳，再由 CAT Server 將它記錄。
4. Parameters Estimating Server：為參數做估算系統，主要的功能為試題參數的估算與作答時能力值的估算。

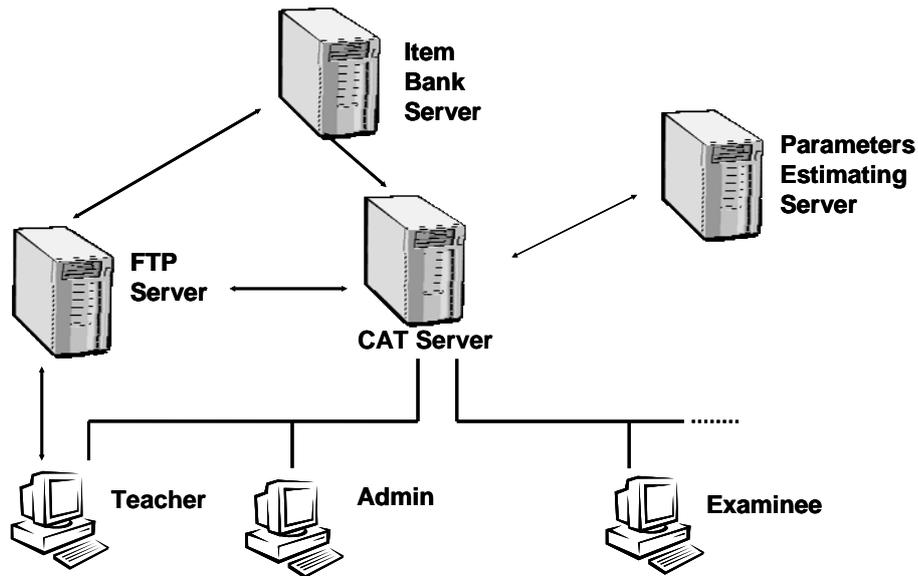


圖 11 系統架構圖 (System Architecture)

3.3 系統模組

本系統可以分成使用者的管理模組，多媒體題庫管理模組、一般測驗管理模組、一般測驗施測模組、及 IRT 的施測模組。詳細各子模組如下圖所示：

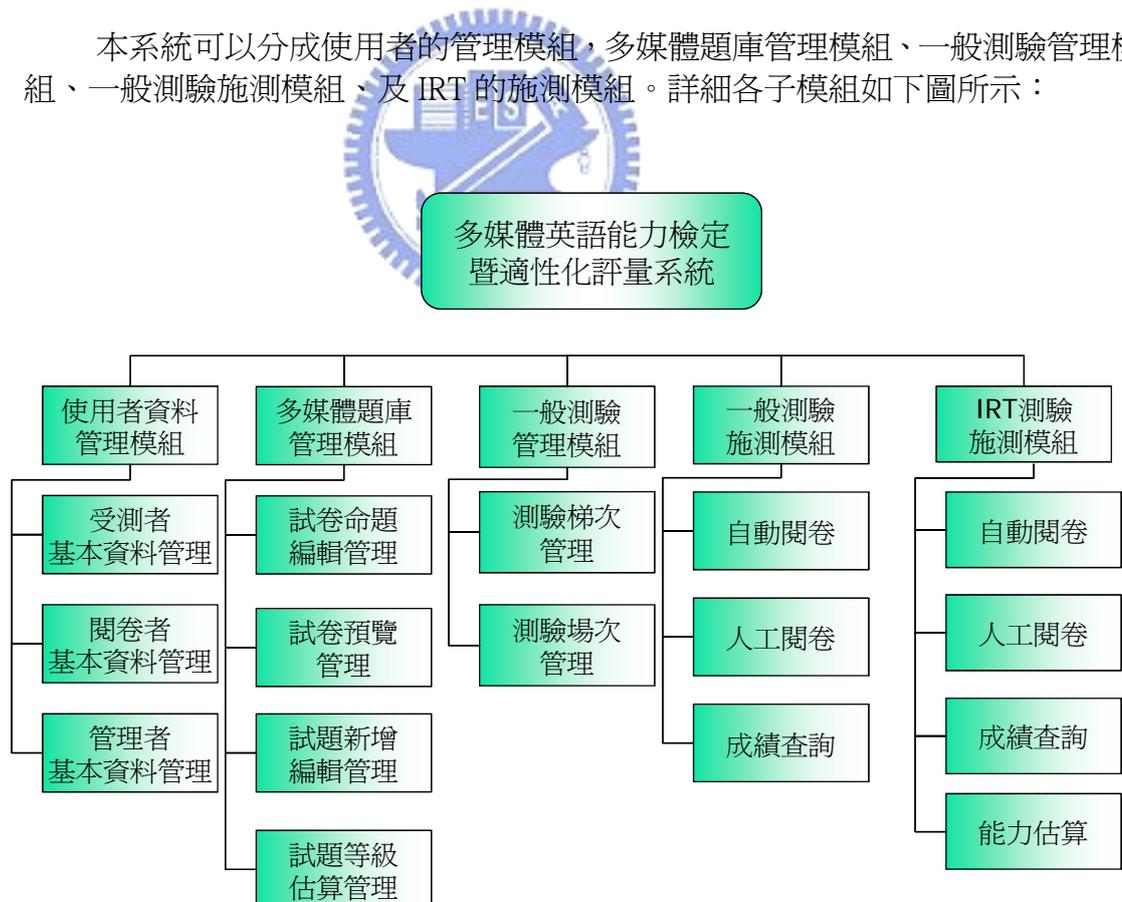


圖 12 系統模組圖

1. 使用者的資料管理：
此模組主要為管理受測學生、閱卷或命題教師、及系統管理者的權限，即使用

- 者資料的設定與管理。
2. 多媒體題庫管理模組：
此模組主要為管理多媒體試題，並且提供教師作線上的題庫管理及命題。並且由試題的作答作適性值調整。
 3. 一般測驗管理模組：
主要安排相關英語之不同等級(入門級、基礎級、初級、中級、中高級等)，與不同的測驗項目(聽、說、讀、寫四種能力)的測驗場次。
 4. 一般測驗的施測模組：
提供不同等級，不同測驗項目的受測者線上測驗機制，測驗完後，可進行線上電腦閱卷及人工閱卷，同時亦提供成績查詢。
 5. IRT 測驗模組：
包括受測者的線上能力估算，及試題的參數估算，並立即繪出相關試題特徵曲線，試題訊息曲線等作試題研判。

3.4 資料庫結構

由於 PHP 已經內建相關資料庫存取函數庫，可以用來存取各種資料庫內容，不同的資料庫須使用不同的函式庫，所以為了以後可以支援不同的資料庫如 MSSQL 等，將重新寫一個 DB CLASS 當作資料庫存取的中介層，當本系統使用 DB CLASS 連結資料庫，對於所使用到的資料庫表格有：題庫檔，組合好的試卷登錄檔，學生測驗後的答案登錄檔，IRT 測驗試卷等。詳述如下：

3.4.1 資料庫規格(PK 代表主鍵)

1. 檢測場次資料表，設定好相關的舉辦檢測場次的資料，以提供欲參加報名學生參加。

表 3 檢測場次資料表 tcool_test_turn

欄位	型態	Null	預設值	連結到	註解	MIME
turn_id	int(5)	否			場次編號	PK
subject	varchar(200)	否			科目	
exam_kind	varchar(50)	否			測驗類型	聽力、閱讀、口語、寫作
level_start	tinyint(1)	否	0		級別	1=> 入門級 2=> 基礎級 3=> 初級 4=> 中級 5=> 中高級
exam_title	varchar(200)	否			場次名稱	
exam_commnet	varchar(200)	否			場次說明	

start_time	datetime	否	0000-00-0000:00:00	開始時間
end_time	datetime	否	0000-00-0000:00:00	結束時間
limit_time	tinyint(3)	否	0	時間限制
is_use	char(1)	否		是否使用
person_id	varchar(10)	否		登錄人員
unit_id	varchar(10)	否		登錄單位

2. 題庫資料表，當試題上傳後，詳細的資料皆會記錄在此資料表，以方便教師命題使用，或是進行試題維護時使用。

表 4 題庫資料表 tcool

欄位	型態	Null	預設值	連結到	註解	MIME
tcool_id	int(5)	否			題庫編號	PK
subject	varchar(200)	否			科目	
tcool_kind	varchar(200)	否			題目類型	tk1=>聽力 tk2=>閱讀 tk3=>口語 tk5=>寫作
tcool_anskind	varchar(50)	否			作答方式	1=>線上錄音 2=>點選作答 3=>輸入文字 4=>上傳檔案
sel_nums	smallint(3)	否	0		選項數	
answer_time	char(3)	否			作答時間	
tcool_level	int(1)	否	0		難易等級	
answer	varchar(200)	否			題目 - 解答	
main_txt	varchar(200)	否			題目 - 文字	
main_img	varchar(200)	否			題目 - 多媒體	
main_sound	varchar(200)	否			題目 - 聲音檔	
main_index	varchar(100)	否			題目 - 執行 html	
is_tcool_group	char(1)	否	0		是否題組	
tcool_group_id	varchar(5)	否			題組編號	
word_level	varchar(5)	否			字頻等級	
grammer_level	varchar(5)	否			文法等級	
sentence_level	varchar(5)	否			構句等級	
table_level	varchar(5)	否			字表等級	
tcool_group_id	varchar(5)	否			題組編號	
mk_time	datetime	否	0000-00-0000:00:00		登錄時間	
person_id	varchar(10)	否		person	登錄人員	
unit_id	varchar(10)	否		unit	登錄單位	

3. 當教師要進行不同等級的測驗，先要依序建立好一份試卷，並設好該測驗的級別(入門、基礎、初級、中級、中高級)及類型(聽力、口說、閱讀、寫作)並將它記錄在試卷的資料表裏面。

表 5 試卷資料表 tcool_exam

欄位	型態	Null	預設值	連結到	註解	MIME
exam_id	int(5)	否			試卷編號	PK
subject	varchar(200)	否			科目	
exam_kind	varchar(50)	否			考試類型	tk1=>聽力 tk2=>閱讀 tk3=>口語 tk5=>寫作
level_start	tinyint(1)	否	0		適用級別	1=>入門級 2=>基礎級 3=>初級 4=>中級 5=>中高級
exam_title	varchar(200)	否			試卷名稱	
exam_commmnet	varchar(200)	否			相關說明	
start_time	datetime	否	0000-00-0000:00:00		開始時間	
end_time	datetime	否	0000-00-0000:00:00		結束時間	
limit_time	tinyint(3)	否	0		施測時間(分鐘)	
mk_time	datetime	否	0000-00-0000:00:00		登錄時間	
person_id	varchar(10)	否		person	登錄人員	
unit_id	varchar(10)	否		unit	登錄單位	

4. 建立好試卷之基本資料，教師需再選擇試題並將它記錄在試卷明細的資料表裏面。

表 6 試卷明細資料表 tcool_qus

欄位	型態	Null	預設值	連結到	註解	MIME
qus_id	int(5)	否			試卷明細序號	PK
exam_id	varchar(10)	否		tcool_exam	試卷編號	
tcool_id	varchar(10)	否			題庫編號	
qus_order	char(3)	否			排序	
mk_time	datetime	否	0000-00-0000:00:00		登錄時間	
person_id	varchar(10)	否		person	登錄人員	
unit_id	varchar(10)	否		unit	登錄單位	

5. 當學生報名後，將依報名參加的等級名稱，參加的場次，記錄在學生參加檢測明細資料表，用以控制學生檢測，並且提供學生線上成績查詢的來源。

表 7 學生參加檢測明細 tcool_signup

欄位	型態	Null	預設值	連結到	註解	MIME
signupID	int(11)	否			序號	PK
subject	varchar(20)	否	english		科目	

person_id	int(11)	否	0		考生編號	
seatNumber	varchar(10)	是	NULL		座號編號	
passNumber	varchar(20)	是	NULL		登錄密碼	
testID	int(11)	否	0		考試編號	
testLevelID	tinyint(1)	否	1		等級	
testTypeID	tinyint(1)	否	1		初(複)試	
area	varchar(8)	否			考場區域	
payNumber	varchar(12)	是	NULL		付款號碼	
payStatus	enum('Pay','No','Free')	否	No		報名狀況	
isTested	char(1)	否	0		是否考完	
loginTime	datetime	否	0000-00-0000:00:00		登入時間	
fillDate	datetime	否	0000-00-0000:00:00		考完時間	

6. 一般學生參加檢測後，將其作答記錄在答案卷資料表，用以控管學生檢測即單一場次僅能考一次。

表 8 答案卷 tcool_test

欄位	型態	Null	預設值	連結到	註解	MIME
tt_id	varchar(30)	否	0		答案卷編號	PK
exam_id	varchar(10)	否			試卷編號	
turn_id	varchar(5)	否			場次編號	
s_time	datetime	否	0000-00-0000:00:00		開始作答時間	
e_time	datetime	否	0000-00-0000:00:00		結束作答時間	
total_time	int(6)	是	0		作答次數	
op_mode	char(1)	否			操作模式	
op_times	char(1)	否			操作次數	
final_grade	varchar(50)	否			得分	
final_judgement	text	否			評語	
person_id	varchar(10)	否		person	登錄人員	
unit_id	varchar(10)	否		unit	登錄單位	
check_person_id	varchar(10)	否		person	閱卷人員	
check_time	datetime	否	0000-00-0000:00:00		閱卷時間	
irt_theata	float(7,6)	否	0.000000		IRT 能力值	
irt_eap	float(7,6)	否	0.000000		EAP 能力值	
irt_mle	float(7,6)	否	0.000000		MLE 能力值	
irt_comment	varchar(50)	否			備註	
irt_level_id	varchar(10)	否			IRT 試卷 ID	

7. 一般學生參加檢測後，將其作答明細記錄在答案卷明細資料表，用以控管學生檢測，除記錄相關作答記錄外，並記錄線上閱卷後的每一題成績。

表 9 答案卷明細 tcool_test_ans

欄位	型態	Null	預設值	連結到	註解	MIME
ans_id	varchar(35)	否			序號	PK
tt_id	varchar(30)	否		tcool_test	答案卷編號	
exam_id	varchar(10)	否			試題卷編號	
tcool_id	varchar(10)	否			試題編號	
ans_txt	text	是	NULL		作答文字	
ans_file	varchar(200)	否			作答檔案	
get_grade	float(2,1)	否	0.0		得分	
ans_mode	tinyint(1)	否	0		模式	
s_time	datetime	否	0000-00-0000:00:00		開始作答時間	
e_time	datetime	否	0000-00-0000:00:00		結束作答時間	
ans_nums	char(3)	否			作答次數	
total_time	int(5)	否	0		總花費時間	
mk_time	datetime	否	0000-00-0000:00:00		登錄時間	
person_id	varchar(10)	否		person	登錄人員	
unit_id	varchar(10)	否			登錄單位	
check_person_id	varchar(10)	否			閱卷人員	
check_time	datetime	否	0000-00-0000:00:00		閱卷時間	

8. IRT 測驗試卷資料表，主要為將閱讀能力與聽力測驗之選擇題的題庫，進行參數估算，重新編輯成一份適性化試卷，並可設定通過之條件。

表 10 IRT 測驗卷資料表 tcool_irt_level

欄位	型態	Null	預設值	連結到	註解	MIME
irt_level_id	int(5)	否			IRT 試卷 ID	PK
subject	varchar(200)	否			科目	
exam_kind	varchar(50)	否			測驗類型	tk1=>聽力 tk2=>閱讀 tk3=>口語 tk5=>寫作
level_start	tinyint(1)	否	0		等級	1=>入門級 2=>基礎級 3=>初級 4=>中級 5=>中高級
irt_level_title	varchar(100)	否			試卷名稱	
tcool_nums	varchar(5)	否			題庫數量	
tcool_ids	text	否			題庫 ID 群	
allow_num	char(3)	否			取樣考過次數	
pass_se	varchar(5)	否			通過標準誤	
pass_num	varchar(5)	否			通過題數	
is_stop	char(1)	否	1		是否停用	

mk_time	datetime	否	0000-00-0000:00:00		登錄時間	
person_id	varchar(10)	否			登錄人員	
unit_id	varchar(10)	否			登錄單位	

9. IRT 測驗試題資料表，進行參數估算試題，將其 a,b,c 值作記錄。

表 11 IRT 試題資料表理 tcool_irt

欄位	型態	Null	預設值	連結到	註解	MIME
tcool_irt_id	int(5)	否			序號	PK
irt_level_id	varchar(5)	否		tcool_irt_level	IRT 試卷 ID	
tcool_id	varchar(10)	否			題庫 ID	
para_b	decimal(9,4)	否	0.0000		參數 c	
para_a	decimal(9,4)	否	0.0000		參數 a	
para_c	decimal(9,4)	否	0.0000		參數 b	
mk_time	datetime	否	0000-00-0000:00:00		登錄時間	
person_id	varchar(10)	否			登錄人員	
unit_id	varchar(10)	否			登錄單位	

3.4.2 資料關聯圖(Entity- Relation Diagram)

分析系統中各成員的屬性及關聯性，我們建立其 E-R Model。這樣我們在資料庫的設計先期可以很容易的找出一些實體成員(Entity)，如老師、學生、題庫...等，而對這些實體成員我們也很容易的可以定義出它們各自的欄位，另一方面，實體成員之間的關聯(relationship)，我們也可以很容易的定義出來，如下圖就是它們之間的資料關聯圖。

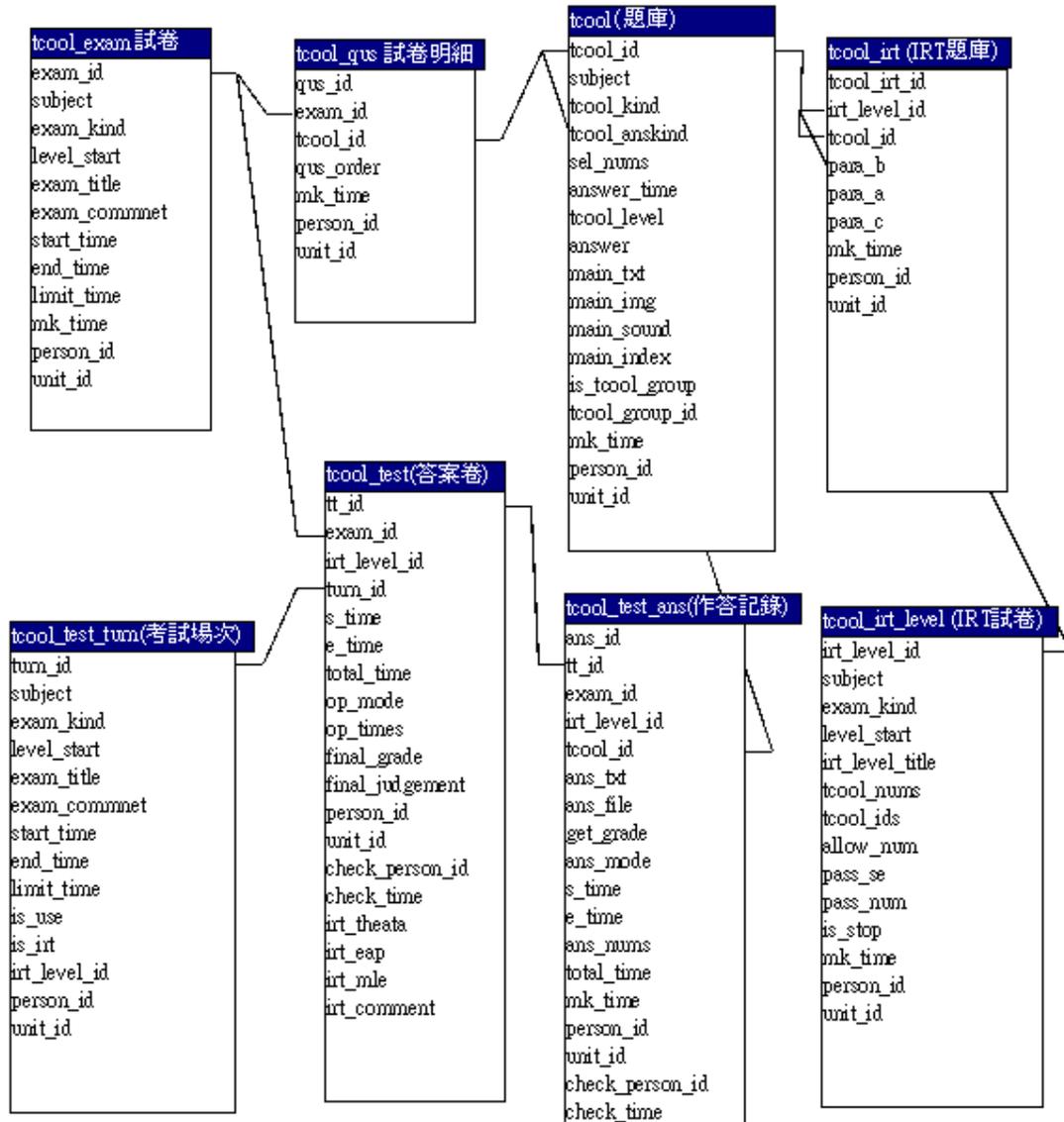


圖 13 資料關聯圖

四、系統實作

我們針對系統的需求、分析與設計，實作以多媒體試題呈現適性化網路評量系統，以驗證平台的可行性與特色。

4.1 使用者功能

系統將使用者的身份分成系統管理者，教師及學生(受測者)三種不同的身份，使用者登入系統成功後，系統會依據使用者的權限，給予不同的功能選單，這是以使用者的觀點來設計使用者介面，使用者不致於混淆不清。其完整的功能如下：

■系統管理	■教師功能	■學生功能
<ul style="list-style-type: none">■ 模組管理■ 網站樣版■ 類別設定■ 權限管理■ 系統記錄■ 流量統計■ 帳號管理	<ul style="list-style-type: none">■ 個人帳號管理■ 題庫管理■ 試卷管理■ 閱卷系統■ 答案卷管理■ 梯次管理■ 場次管理 ■ IRT-試題管理■ IRT-測驗試卷管理■ IRT-測驗施測	<ul style="list-style-type: none">■ 檢測參加明細■ 個人帳號管理■ 一般檢測■ IRT-檢測

圖 14 使用者功能

4.2 試題上傳與維護

4.2.1 試題的上傳

一份試卷的完成，首先需要完整的題庫，我們在系統的線上試題管理的功能表中，提供教師或系統管理者試題上傳的機制，以上傳試題來建立完整的題庫。並提供兩個介面，讓使用者上傳題庫。

1. 瀏覽器上傳：

- (1) 當老師完成試題之編輯後，先壓縮成一個壓縮檔，其中執行此題的題目設成 index.htm。
- (2) 上傳題庫的相關資料記錄於資料庫中，教材的實體檔案則依相關屬性，難易度、題目類型(聽力、閱讀、口語、寫作能力測驗等)、作答方式(線

上錄音 點選作答 輸入文字 上傳檔案等)、教材序號建立資料夾，存放於檔案目錄中。

- (3) 上傳後將可以啟動試題的驗證機制，將相關的字頻、文法句構等適性值寫入題庫的資料表中。

編號	<input type="text"/>
科目	<input checked="" type="radio"/> 英語
類目類型	<input checked="" type="radio"/> 聽力能力測驗 <input type="radio"/> 閱讀能力測驗 <input type="radio"/> 口語能力測驗 <input type="radio"/> 寫作能力測驗
作答方式	<input type="radio"/> 線上錄音 <input checked="" type="radio"/> 點選作答 <input type="radio"/> 輸入文字 <input type="radio"/> 上傳檔案
選項數	<input type="text"/>
作答時間	<input type="text"/>
難易等級	<input type="text"/>
題目-解答	<input type="text"/>
題目-文字	<input type="text"/>
題目-多媒體	<input type="text"/> 瀏覽...
題目-執行html	index.htm

圖 15 試題瀏覽器上傳

2. 編輯手直接上傳

- (1) 使用者使用編輯手編輯題庫，並設好相關難易度，類型等屬性後，可以直接上傳至 FTP Server。
- (2) 再通知 CAT Server 由系統自動化將題庫建入資料庫及解壓縮。
- (3) 啟動試題的驗證機制，將相關的文法、句構等適性值寫入題庫的資料表中。

4.2.2 試題的維護

對於已建立好的題庫，系統亦提供線上試題維護的功能，方便教師修改或刪除試題。試題維護畫面如下圖，教師可以按編輯來重新修改試題相關設定值或者是立即刪除試題。

搜尋表單 | 全部列表 | 新增表單 | 管理列表 | 主管理區 |

[第一頁] [上一頁] [下一頁] [最後頁] [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] 共 4642 筆 頁次: 1/233 頁 20 筆/頁

序	科目	類目類型	作答方式	選項數	難易等級	題目-解答	題目-文字	題目-多媒體	題目-執行html	管理
5402006	英語	寫作能力測驗	輸入文字	0	5402	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5402005	英語	寫作能力測驗	輸入文字	0	5402	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5402004	英語	寫作能力測驗	輸入文字	0	5402	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5402003	英語	寫作能力測驗	輸入文字	0	5402	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5402002	英語	寫作能力測驗	輸入文字	0	5402	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5402001	英語	寫作能力測驗	輸入文字	0	5402	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5401003	英語	寫作能力測驗	輸入文字	0	5401	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5401002	英語	寫作能力測驗	輸入文字	0	5401	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5401001	英語	寫作能力測驗	輸入文字	0	5401	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5307006	英語	口語能力測驗	線上錄音	0	5307	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5307005	英語	口語能力測驗	線上錄音	0	5307	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5307004	英語	口語能力測驗	線上錄音	0	5307	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5307003	英語	口語能力測驗	線上錄音	0	5307	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5307002	英語	口語能力測驗	線上錄音	0	5307	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5307001	英語	口語能力測驗	線上錄音	0	5307	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5306004	英語	口語能力測驗	線上錄音	0	5306	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5306003	英語	口語能力測驗	線上錄音	0	5306	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5306002	英語	口語能力測驗	線上錄音	0	5306	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5306001	英語	口語能力測驗	線上錄音	0	5306	請見題目	請見題目		index.htm	[編輯][刪除][預覽]
5305008	英語	口語能力測驗	線上錄音	0	5305	請見題目	請見題目		index.htm	[編輯][刪除][預覽]

圖 16 試題維護畫面

4.3 線上試卷管理

提供教師組合一份新試卷，檢視這些試卷是否恰當，也可查詢試卷或刪除不恰當的試卷。

4.3.1 新增試卷

教師組合一份新試卷，首先在「新增表單」中，選擇考試科目、及測驗的等級、及聽、說、讀、寫四種能力類型後，再進入試題勾選後即可完成試卷的編輯。(操作畫面如下，分為 3 個步驟)

步驟 1：試卷測驗項目規劃

分別填入測驗的類別，適用等級，試卷名稱等相關欄位後，即可得到一份空白的試卷(無試題)。

科目	<input checked="" type="radio"/> 英語
考試類型	<input checked="" type="radio"/> 聽力能力測驗 <input type="radio"/> 閱讀能力測驗 <input type="radio"/> 口語能力測驗 <input type="radio"/> 寫作能力測驗
適用級別	入門級 <input type="button" value="v"/>
試卷名稱	<input type="text"/> [*必填]
相關說明	<input type="text"/>
開始時間	<input type="text"/>
結束時間	<input type="text"/>
施測時間(分鐘)	60
<input type="button" value="重新填寫"/> <input type="button" value="確定送出"/>	

圖 17 規劃新測驗試卷

步驟 2：試卷試題選擇

點選新增試題後，分別選出所須要之試題，即可將試題加入，各試題皆可從線上進行預覽。

admin 登出 管理區

[搜尋表單](#) | [全部列表](#) | [新增表單](#) | [全部刪除](#) | [刪除答案卷](#) | [已命題試題排序](#) | [管理列表](#) | [主管理區](#)

序號	科目	考試類型	適用級別	試卷名稱	施測時間(分鐘)	管理
4	英語	聽力能力測驗	中級	聽力測驗(文字題 9421)	60	[秀出] - [編輯] - [刪除] -- [隨機命題] - [新增試題] - [編輯試題] -

EnglishTest - Microsoft Internet Explorer

共 26 筆 頁次 3/3 頁 10

序	科目	類
<input type="checkbox"/>	942106	英語 聽力測驗
<input type="checkbox"/>	942105	英語 聽力測驗
<input type="checkbox"/>	942104	英語 聽力測驗
<input type="checkbox"/>	942103	英語 聽力測驗
<input type="checkbox"/>	942102	英語 聽力測驗
<input type="checkbox"/>	942101	英語 聽力測驗

請注意內容聽並作答
按下喇叭可再聽一次

(1) Yes, I do.
(2) No, I am not.
(3) Yes, I am.

完成 網際網路

圖 18 選擇測驗試題

步驟 3：試卷試題預覽

試卷與試題建好之後，老師可按編輯試題，進行每一題試題的預覽及修正。



圖 19 試卷與試題預覽

4.3.2 同一等級隨機命題

爲了施測系統安全性的考量，所以每個考生將提供不同的試卷，但是每一份題目的難度等級是一致的，所以透過系統可以立即地，隨機產生多份同難度等級的測驗試卷，以方便教師測驗命題。



圖 20 同一等級試題隨機命題

4.3.2 試卷維護

對於教師命題好的試卷，系統提供教師查詢與修改試卷的功能，方便教師管理試卷，試卷維護功能選單如下圖

圖 21 試卷維護

教師可以按編輯試題後，進行編輯試題的內容。

圖 22 試卷試題查詢

4.4 線上測驗

當老師命題完成，其試卷已經放入系統的試卷庫中，分成一般測驗與 IRT 測驗，學生可以選擇參加一般測驗或是 IRT 測驗，點選後進入系統將會看到可以參加測驗畫面，列出老師指定給他的測驗試卷名稱，學生僅需輸入帳號後即可開始線上測驗。

4.4.1 登入測驗

學生先點選一般測驗，再輸入帳號登入後，立即可以進入看到老師所指定的試卷進行施測



圖 23 測驗選項

線上測驗系統-請輸入您的身份證字號	
您的帳號：	<input type="text"/>
<input type="button" value="確定登入"/>	<input type="button" value="離開系統"/>

圖 24 一般測驗登入畫面

當輸入身份證字號後，系統會進行確認，將確認該名學生是否為合法的使用者：檢查項目包括：

1. 是否為參加該場次的受測學生。
2. 學生是否已完成過該場次。

如果有錯誤發生，將秀出錯誤訊息：

英檢系統訊息
無此考生帳號！
<input type="button" value="回上頁"/>

圖 25 英檢錯誤訊息

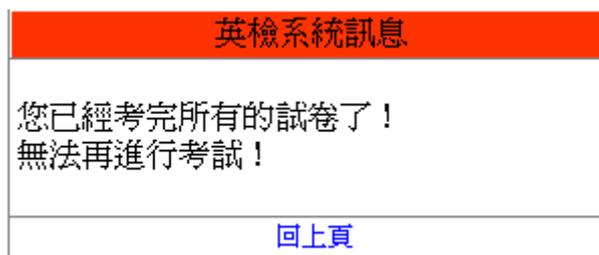


圖 26 英檢錯誤訊息 2

如果沒有錯誤訊息，系統將進行多媒體線上測驗。

4.4.2 多媒體線上測驗

登入完成後進行測驗，其測驗方式，可依聽力、口說、閱讀、寫作能力測驗進行。

1. 聽力測驗

聽力測驗的作答方式為選擇題，題目將由題目區自動播放，學生在聽完題目後，可以從作答區點選答案後，再按繳交答案，系統將自動跳至下一題。



圖 27 聽力測驗畫面

資料來源：動畫引自新竹市英語文網站

在全部作答完成後，電腦將自動評分，學生可以立即知道成績，如下圖所示：

題數	得分
26	45.6
英檢測驗系統訊息	
請回到登入畫面，按登入鈕繼續下一項測驗。	
<input type="button" value="回到登入畫面"/>	

圖 28 英檢電腦評分

2. 閱讀測驗

閱讀測驗的作答方式 皆為選擇題，題目將由題目區自動播放，學生在閱讀完題目後，可以從作答區點選答案後，再按繳交答案，系統將自動跳至下一題。在全部作答完成後，電腦將自動評分，學生可以立即知道成績。

Part III. Short Talks

In this part, there will be ten questions for ten short talks. After each talk, there will be one question about the talk. After you hear a question, read the four alternatives on your computer screen and choose the one that you think is the best based on what you hear.



(A) There were two young men killed in the fight.

(B) The two groups of people came from different countries.

(C) The drug sale took place very often.

(D) The police could not really understand the situation.

圖 29 閱讀測驗

資料來源：影片引自美國 new14 由中華民國多媒體會提供

3. 口說測驗

口說測驗的作答方式為線上錄音，題目將由題目區自動播放，學生在閱讀完題目後，可以從作答區點選，將出現錄音元件如下圖，再錄音，試聽，確認無誤後再按繳交答案，系統將自動跳至下一題。在全部作答完成後，作答結果將以人工閱卷計分。



圖 30 口說能力測驗

資料來源：動畫引自新竹市英語文網站



圖 31 錄音元件

4. 寫作測驗

寫作測驗的作答方式為線上輸入，題目將由題目區自動播放，學生在閱讀完題目後，可以從作答區進行打字，學生可以依需要將寫作區加大或是減小，再按繳交答案，系統將自動跳至下一題。在全部作答完成後，作答結果將以人工閱卷計分。



圖 32 寫作能力測驗

4.4.3 場次管理與線上閱卷

系統將提供五種等級(入門、基礎、初級、中級、中高級)及四種能力(聽、說、讀、寫)的測驗場次，教師依序填入場次名稱、開始時間及結果時間等即可完成一場檢測場次。

科目	<input checked="" type="radio"/> 英語
測驗類型	<input checked="" type="radio"/> 聽力能力測驗 <input type="radio"/> 閱讀能力測驗 <input type="radio"/> 口語能力測驗 <input type="radio"/> 寫作能力測驗
級別	<input checked="" type="radio"/> 入門級 <input type="radio"/> 基礎級 <input type="radio"/> 初級 <input type="radio"/> 中級 <input type="radio"/> 中高級 <input type="radio"/> 文字題 <input type="radio"/> 多媒體題
場次名稱	<input type="text"/> [*必填]
場次說明	<input type="text"/>
開始時間	<input type="text"/>
結束時間	<input type="text"/>
時間限制	<input type="text"/>
是否使用	<input type="checkbox"/>

圖 33 新增檢測場次

學生考完後，便可以進行人工閱卷。系統提供一個友善的介面，讓老師可以直接進行口說測驗與寫作能力的閱卷。

多媒體英語文能力檢定 暨適性化網路評量系統

admin 登出 管理區

搜尋表單 | 全部列表 | 新增表單 | 管理列表 | 主管理區

帳號列表 (88)

重新整理

考生編號:2181 / 試卷編號:7_server005_2181_20041218133941_001 號:77_server005_2181_20041218133941_001 時 90 分

序號	試題編號	試題內容	作答檔案	得分	
7_server005_2181_20041218133941_001	2301002 (...)	2301	請選擇	0.0	
7_server005_2181_20041218133941_002	2302008 (...)	2302	請選擇	0.0	
7_server005_2181_20041218133941_003	2303003 (...)	2303	請選擇	0.0	
7_server005_2181_20041218133941_004	2304001 (...)	2304	請見題目	請選擇	0.0
7_server005_2181_20041218133941_005	2305003 (...)	2305	請見題目	請選擇	0.0
7_server005_2181_20041218133941_006	2306002 (...)	2306	請見題目	請選擇	0.0
7_server005_2181_20041218133941_007	2307006 (...)	2307	請見題目	請選擇	0.0
7_server005_2181_20041218133941_008	2308007 (...)	2308	請見題目	請選擇	0.0
7_server005_2181_20041218133941_009	2309004 (...)	2309	請見題目	請選擇	0.0
7_server005_2181_20041218133941_010	2310003 (...)	2310	請見題目	請選擇	0.0
7_server005_2181_20041218133941_011	2311001 (...)	2311	請見題目	請選擇	0.0
7_server005_2181_20041218133941_012	2312002 (...)	2312	請見題目	請選擇	0.0
7_server005_2181_20041218133941_013	2313004 (...)	2313	請見題目	請選擇	0.0
7_server005_2181_20041218133941_014	2314004 (...)	2314	請見題目	請選擇	0.0
7_server005_2181_20041218133941_015	2315003 (...)	2315	請見題目	請選擇	0.0
7_server005_2181_20041218133941_016	2316003 (...)	2316	請見題目	請選擇	0.0

圖 34 口說測驗能力的閱卷

搜尋表單 | 全部列表 | 新增表單 | 管理列表 | 主管管理區

帳號列表 (47)

重新整理

考生編號: 672 / 試卷編號: 16_server006_672_20041218134143 / 共 0 / 2 題 / 得 80 分

序號	試題編號	試題等級	試題說明	作答文字	作答檔案	得分
	16_server006_672_20041218134143_001	4401004 (...)	4401	請見題目 My father conducted research at the University of Illinois in the United States four years ago. My brother and I also went to Illinois and studied there. I went to senior high school and my brother went to elementary school. We not only read different varieties of English books, but also went to the National Park in the holiday. I also made a few good friends. Although three years have already passed since we left the United States, I still miss my classmates over there.		0.0
	16_server006_672_20041218134143_002	4402001 (...)	4402	請見題目 Dear Alice: First of all, I would like to make an apology, because I can not attend your dad's company opening ceremony. I am so happy that I am invited to join it. However, something exciting happened to me last night. I got a letter from IBM company. They are going to give me an opportunity of interview. I think you probably know that it is my dream to work in that global company. And the interview time is this Sunday. I hope that you can		0.0

圖 35 寫作能力的閱卷

4.5 IRT 測驗管理

學生作答完後，透過系統即可針對收集到學生的作答反應資料，作 IRT 參數估算的分析，分析試題適性質，分別產生試題的 ICC 圖、IIC 圖及試題難易度、鑑別度及猜測度進行研判，不僅可以動態地調整題庫之適性值，同時亦可以編製成適性化測驗。

4.5.1 試題的參數估算

1. 首先新增一份 IRT 空白試卷，並決定欲分析的測驗類型(目前僅支援閱讀與聽力測驗)、並且決定好取樣人數。

科目	<input checked="" type="radio"/> 英語
測驗類型	<input checked="" type="radio"/> 聽力能力測驗 <input type="radio"/> 閱讀能力測驗 <input type="radio"/> 口語能力測驗 <input type="radio"/> 寫作能力測驗
級別	<input checked="" type="radio"/> 入門級 <input type="radio"/> 基礎級 <input type="radio"/> 初級 <input type="radio"/> 中級 <input type="radio"/> 中高級
設定考過人數	<input type="text"/>
IRT測驗名稱	<input type="text"/> [*必填]
IRT試題數量	<input type="text"/>
IRT試題編號群	<input type="text"/>
合格標準誤	<input type="text" value="0.5"/>
測試題上限	<input type="text"/>
是否停用	<input type="checkbox"/>

圖 36 新增 IRT 試卷

- 進行 IRT 的參數估算，將所獲得試題參數值及試題編號寫入資料表中。

聽力能力中高級 題目: 42 題

•

施測反應資料

題數42

樣本數60參數一覽表

題號	難度b	鑑別度a	猜測機率c
1	0.470798	0.684829	0.190110
2	3.793534	0.210618	0.253011
3	-4.796919	0.265275	0.229317
4	-1.563144	0.972732	0.218582
5	-0.206312	0.675693	0.219588
6	-1.244082	0.685011	0.245826
7	0.348966	0.911729	0.216839
8	0.221745	0.267659	0.232652
9	-0.810457	0.541675	0.219127
10	-5.831867	0.317190	0.229791

圖 37 試題參數的估算

4.5.2 試題特徵曲線(ICC)與試題訊息曲線(IIC)呈現

1. 詳如下圖，可立即獲知全部試題的試題特徵曲線(ICC)與試題訊息曲線(IIC)，可以在設定 a,b,c 值後進行篩選。

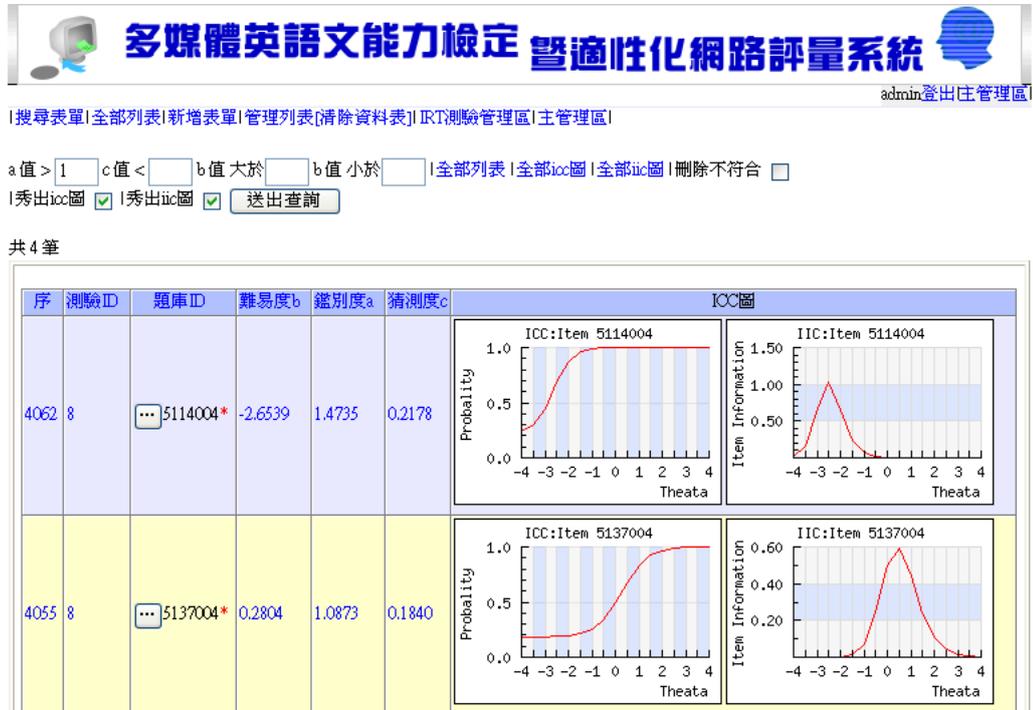


圖 38 試題特徵曲線(ICC)與試題訊息曲線(IIC)圖

2. 如下圖，可以即時獲知所有 ICC 圖比較

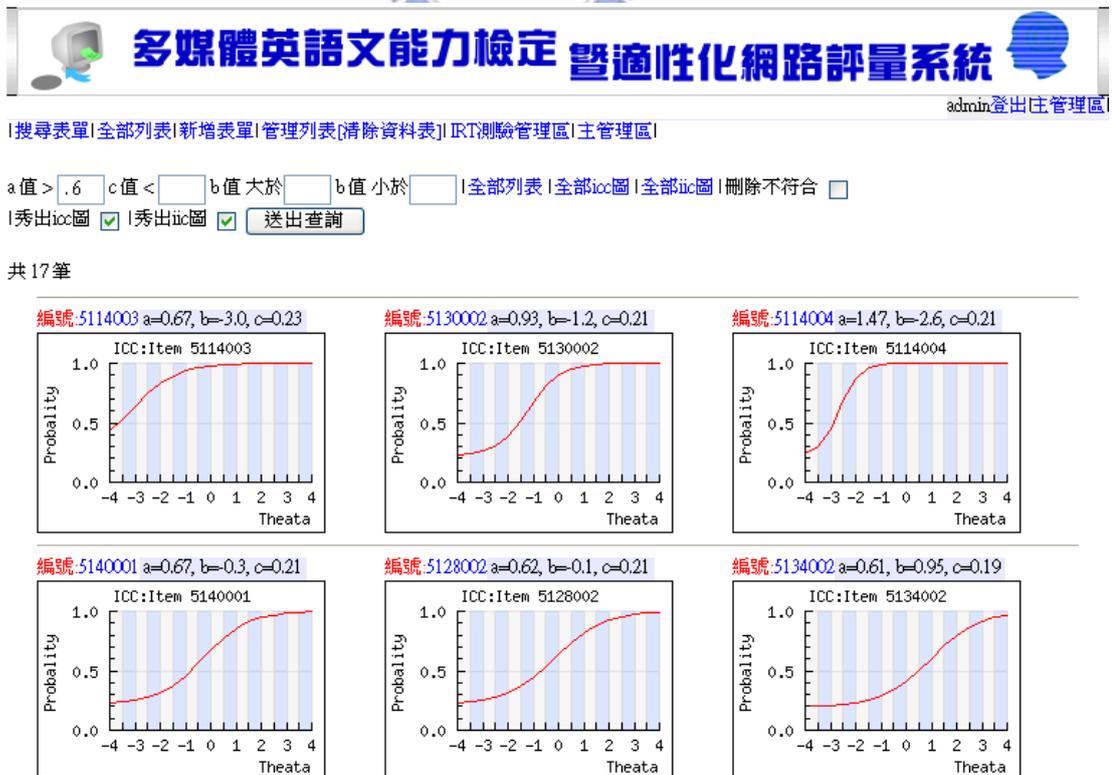


圖 39 即時 ICC 圖的比較

3. 如下圖，可以即時獲知所有 IIC 圖比較


多媒體英語文能力檢定 暨適性化網路評量系統


admin 登出 管理區

[搜尋表單](#) | [全部列表](#) | [新增表單](#) | [管理列表](#) | [清除資料表](#) | [IRT測驗管理區](#) | [主管理區](#)

a 值 > .6 c 值 < b 值 大於 b 值 小於 | [全部列表](#) | [全部icc圖](#) | [全部iic圖](#) | [刪除不符合](#)

[秀出icc圖](#) | [秀出iic圖](#) | [送出查詢](#)

共 17 筆

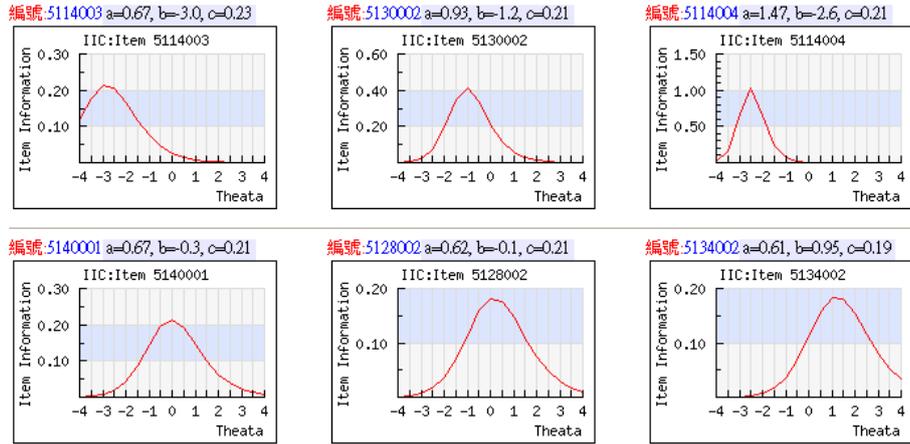


圖 40 即時的 IIC 圖比較

4. 如下圖，可以即時獲知全部試題之綜合比較(IIC,ICC,TIC 圖)

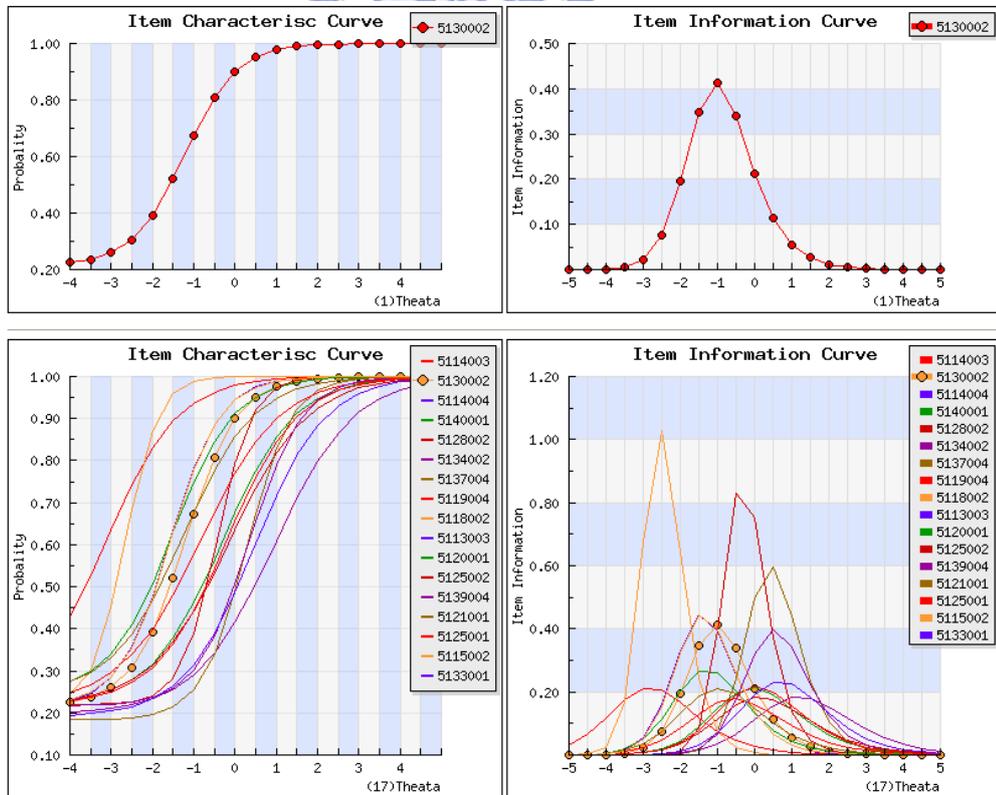


圖 41 即時試卷的全部試題的 ICC 與 IIC 圖比較

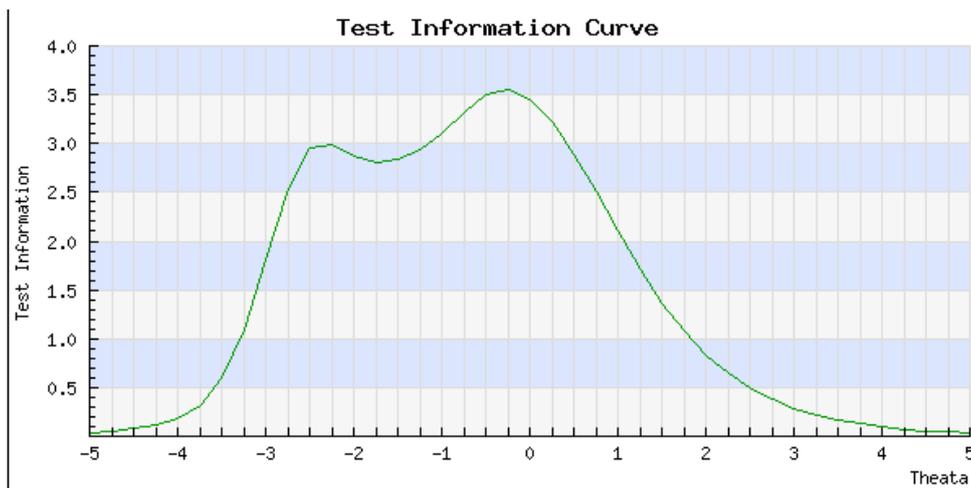


圖 42 試卷訊息曲線圖(TIC)

4.5.3 IRT 試卷電腦試測

IRT 試卷編製完成後，可以透過系統，進行試測，而系統將可以立即性劃出作答情形、能力估算及標準誤圖形曲線，方便研判。

多媒體英語文能力檢定 暨適性化網路評量系統

admin 登出 管理區

搜尋表單
全部列表
新增表單
管理列表
IRT試題管理
清除IRT作答記錄
主管理區

序號	試題編號	作答文字	得分	Theata	MLE	EAP	SE
s8_A_m_20050509144239_001	5113003	C	1.0	0.376133	3.999615	0.376133	1.732993
s8_A_m_20050509144239_002	5125002	B	1.0	0.593570	3.999552	0.593570	1.196999
s8_A_m_20050509144239_003	5137004	B	0.0	-0.288913	-0.344647	-0.288913	0.976768
s8_A_m_20050509144239_004	5139004	A	0.0	-0.158060	-0.158060	-0.226687	0.830020
s8_A_m_20050509144239_005	5125001	A	1.0	0.100546	0.100546	-0.021671	0.774337
s8_A_m_20050509144239_006	5140001	A	1.0	0.309017	0.309017	0.175156	0.601141
s8_A_m_20050509144239_007	5130002	C	0.0	-0.024893	-0.024893	-0.114839	0.530940
s8_A_m_20050509144239_008	5133001	A	1.0	0.199343	0.199343	0.106336	0.515730
s8_A_m_20050509144239_009	5128002	B	1.0	0.320636	0.320636	0.233158	0.495420

[回上一頁](#)

圖 43 不同 IRT 分析器的能力估算

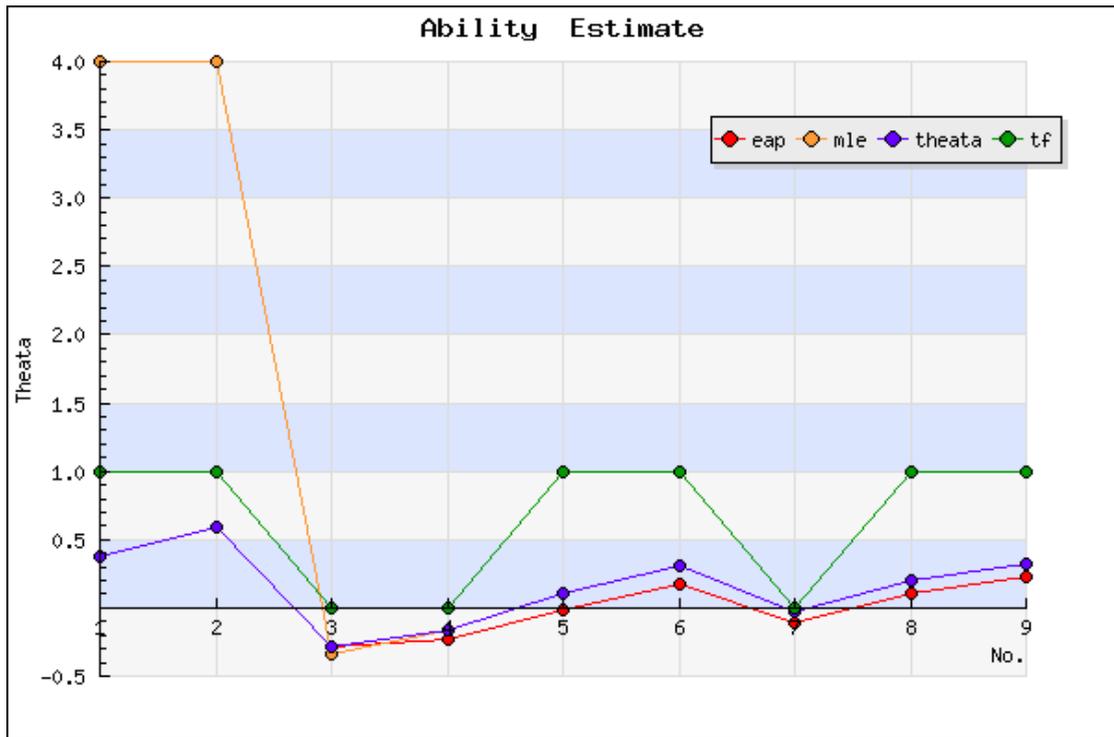


圖 44 整合不同特性的 IRT 分析器作答曲線圖

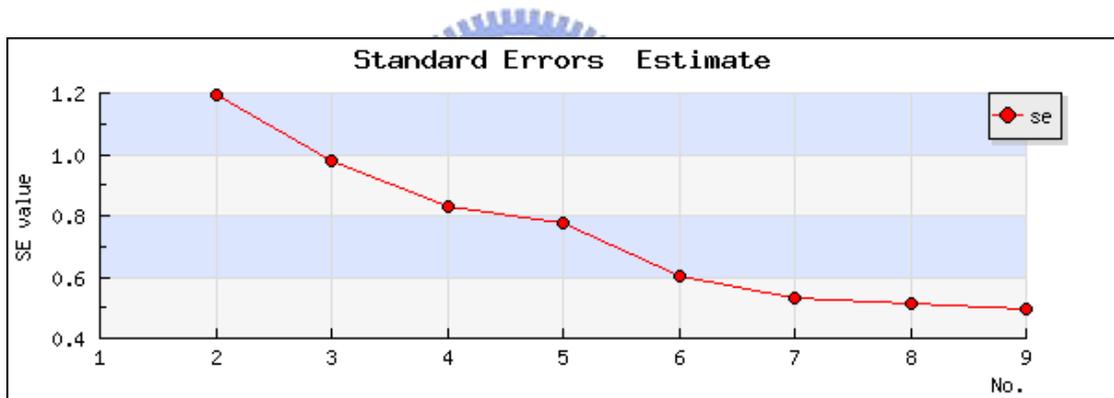


圖 45 作答的標準誤曲線呈現

五、系統測試與結論

5.1 系統測試

針對系統的功能，我們分別先以線上多媒體方式進行測驗，測驗完後再啓動 IRT 分析機制進行，以驗證平台的可行性與特色。

5.1.1 線上多媒體測驗

1. 我們分別利用台北與新竹地區國小 5-6 年級約 1000 個學生進行網路線上測驗，採用的設備與軟體如下：

(1) Server

系統為 Mandrake Linux 10.1+Apache2.0.50+PHP 4.3.8+ MySQL 4.0.20

主機為 Pentium 4 3.0G CPU*2、1GB RAM、120GB HDD

(2) Client

Windows 98 SE2(IE5.5)

AMD K6-300(IE5.5)

(3) 多媒體試題：是由智勝國際科技所出版的編輯手[19]所編輯完成的，多媒體檔案大小由 200K 到 400K 左右
測試結果，學生皆能正常完成多媒體測驗，其數據分析的結果請參閱由李玉珍老師所撰寫的碩士論文[20]。

2. 我們分別在全國分設 17 檢測中心同時施測級數分爲：

(1) 入門級

(2) 基礎級

(3) 初級

(4) 中級

(5) 中高級

進行檢測，合計共有 6000 人次，進行檢定，學生皆能正常完成測驗。其初級測驗結果分析如下：

- 初級聽力測驗：通過比率爲 68%
- 初級閱讀測驗：通過比率爲 27%
- 初級初試：通過比率爲 26%
- 初級口說測驗：通過比率爲 75%
- 初級寫作測驗：通過比率爲 77%
- 初級複試：通過比率爲 55%

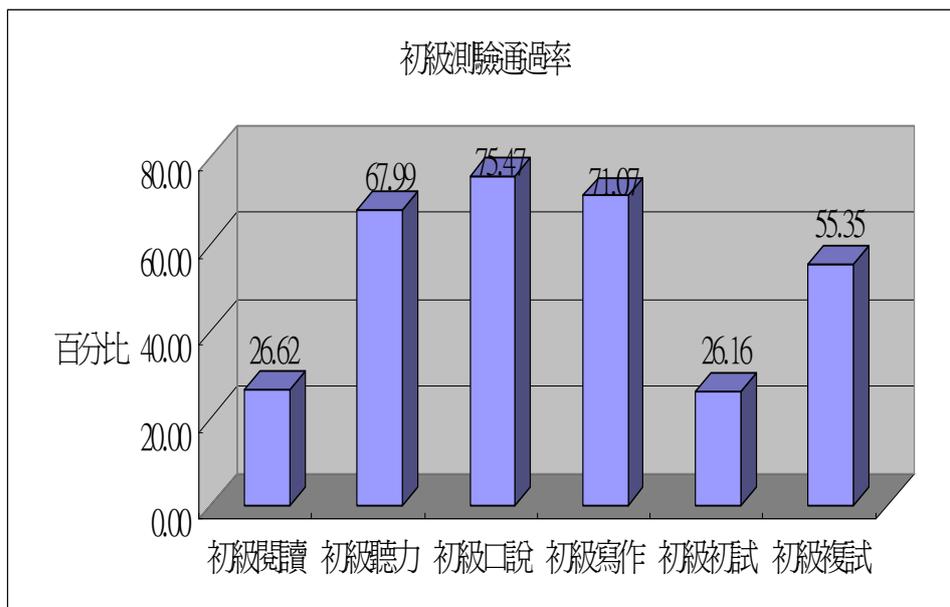


圖 46 初級的通過率自動分析

5.1.2 IRT 適性分析

將全國 17 檢測中心所得到的受測結果進行分析，以初級的聽力測驗分析為例：

1. 試題品質分析部份：

第一部份—看圖辨義

測驗目標：評量考生對淺易日常用語之聽解能力

Question 1：

A. 題目：請聽電腦播出題目和三個英語句子之後，選出與所看到的圖畫最相符的答案。

播放的聲音內容：

Look at the picture on the left. What do people usually do here?

A. They usually swim here.

B. They usually buy tickets here.

C. They usually watch movies here.

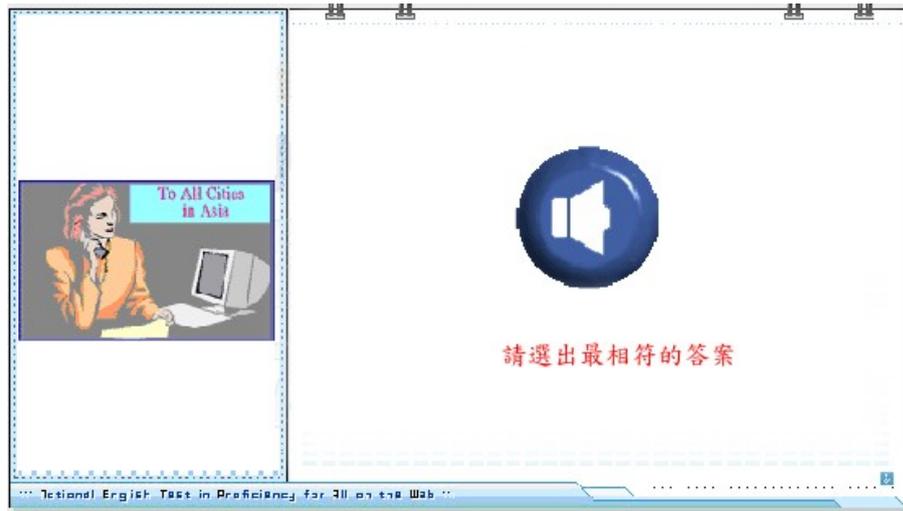


圖 47 以 IRT 分析之多媒體試題 1

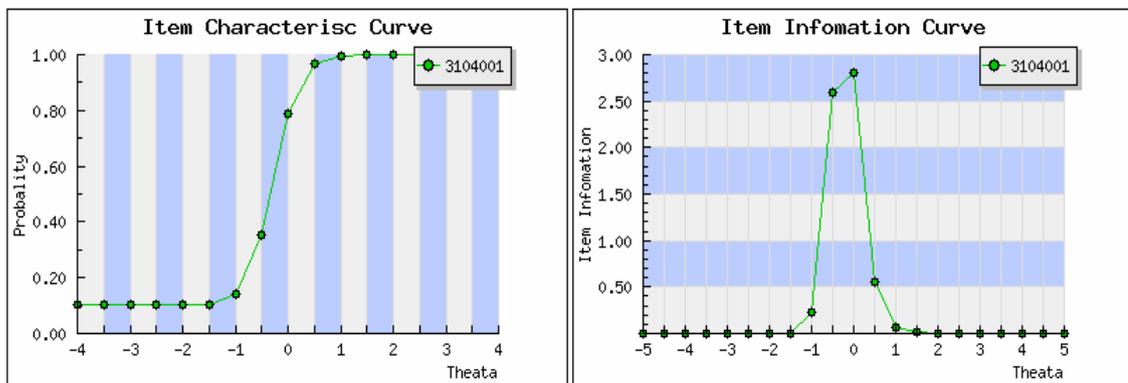


圖 48 以 IRT 分析之多媒體試題 1 之 ICC 圖與 IIC 圖

C. 測驗參數(a、b、c 值)

難易度 b	-0.28033900
鑑別度 a	2.48758000
猜測度 c	0.10083300

D. 說明：本題由 ICC 圖中得知鑑別度高、猜測度低；由 IIC 圖中得知難易度適中。故本題適合繼續保留在初級聽力測驗試題庫中使用。

2. 試卷品質分析部份：

A. 試卷各題特徵曲線(ICC)

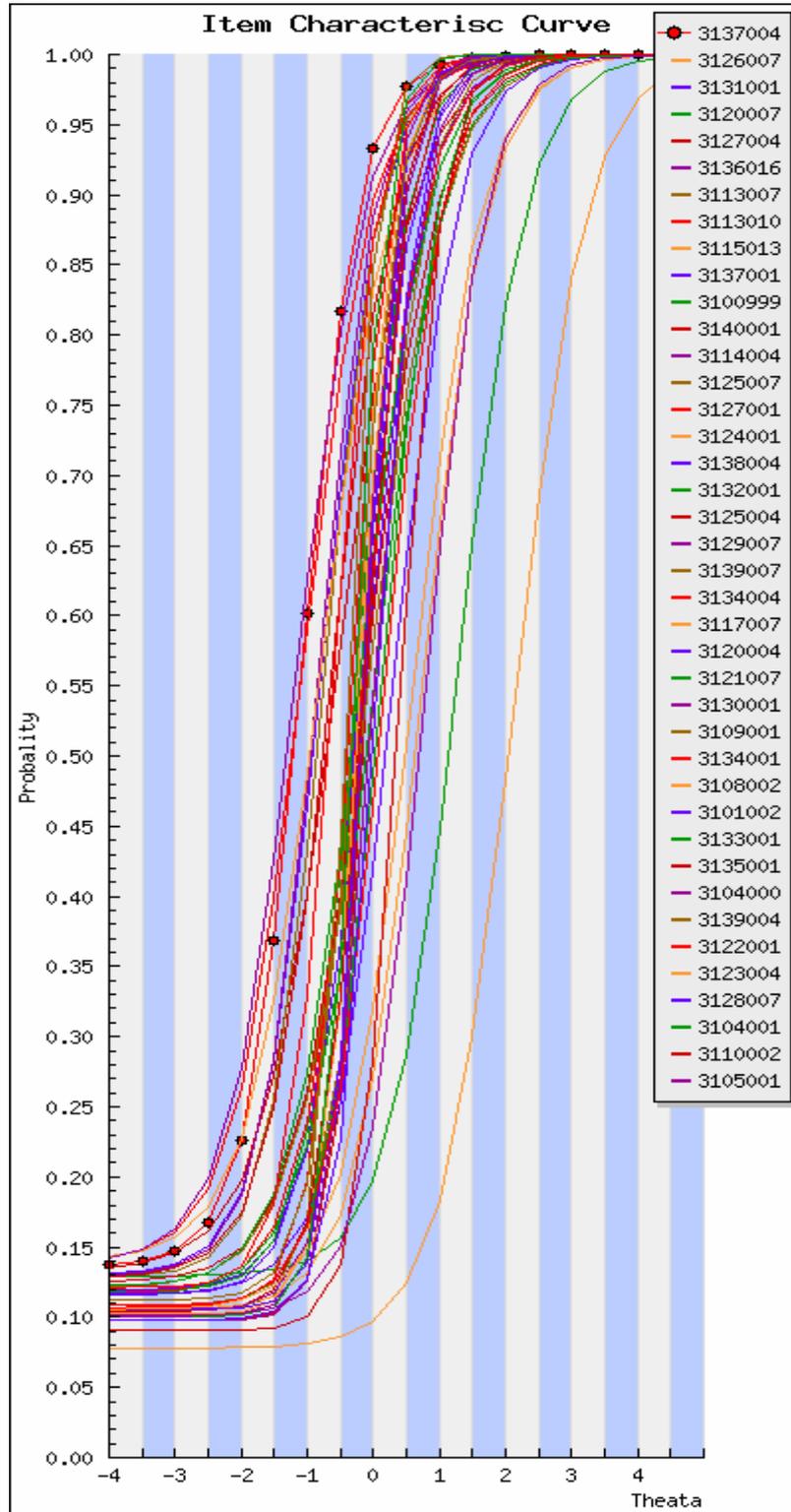


圖 49 以 IRT 分析全部 ICC 圖

由上圖得知，就鑑別度而言，整份試卷題目鑑別度高；就難易度而言，因為是針對初級能力檢測，所以大部份的題目難易度適中，偏易的題目稍多，符合該級學生能力。

B. 試卷各題訊息曲線(IIC)

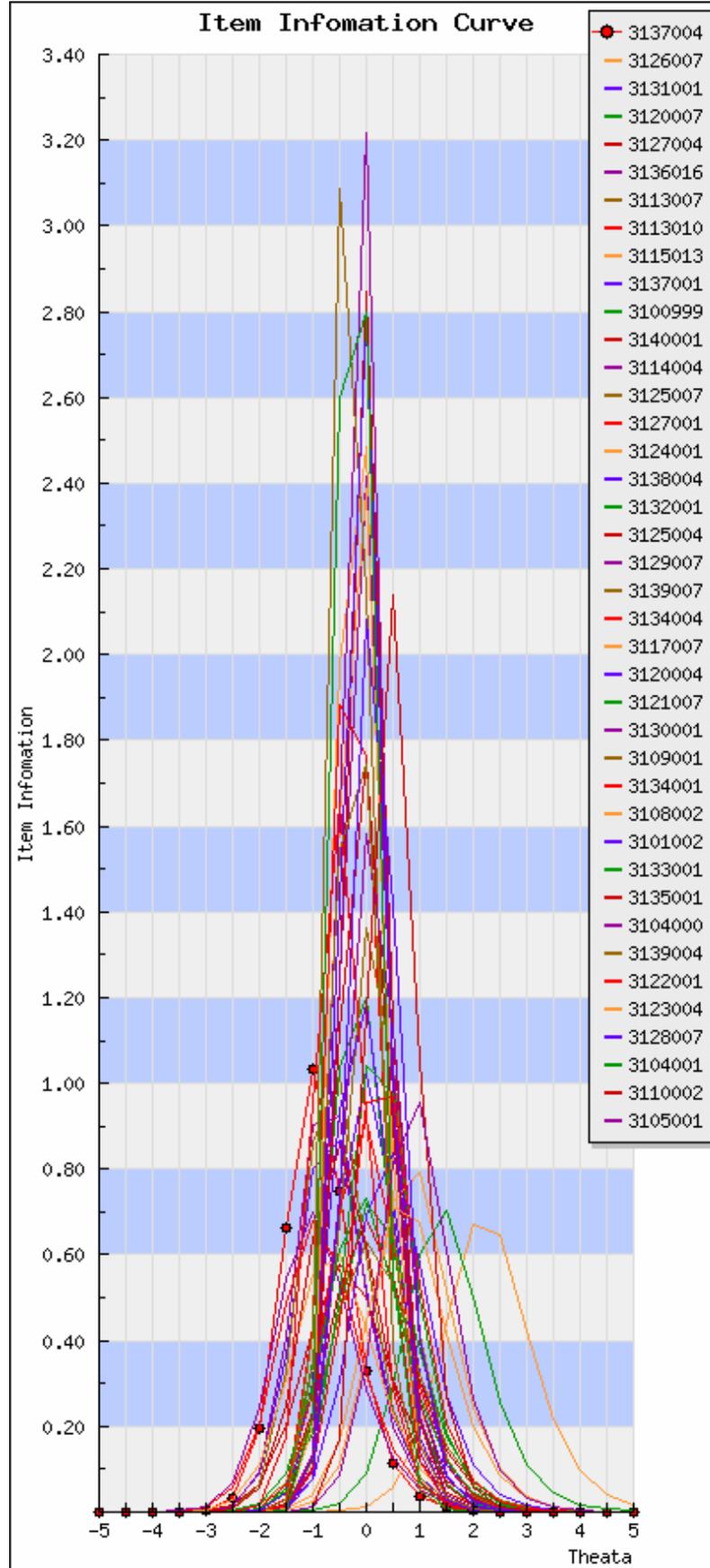


圖 50 以 IRT 分析全部 IIC 圖

C. 試卷訊息曲線(TIC)

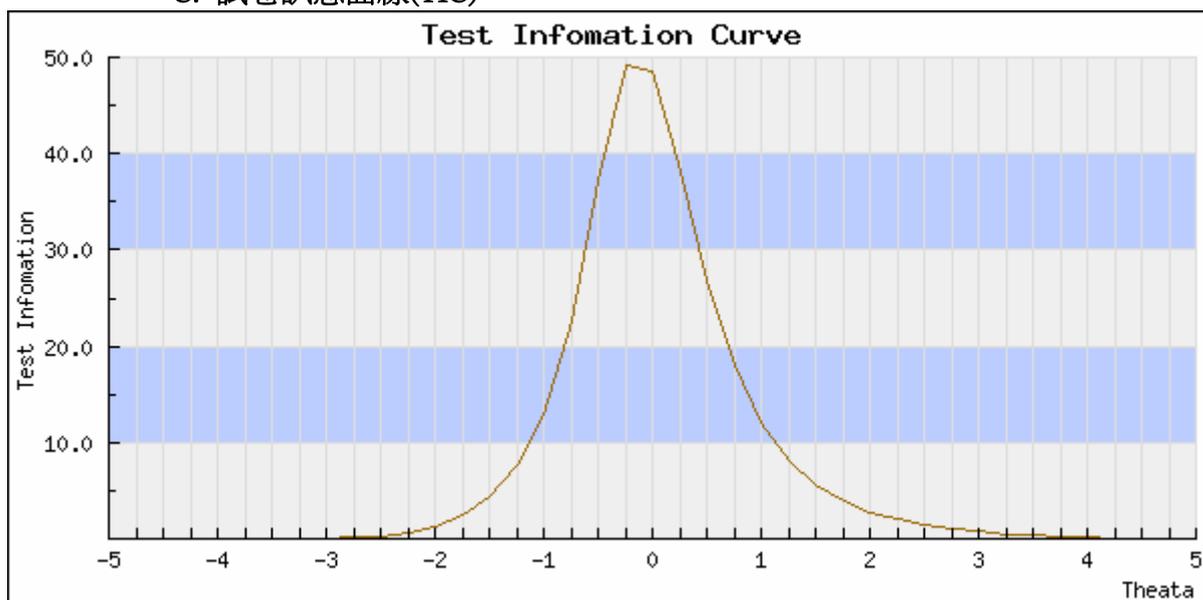


圖 51 以 IRT 分析 TIC 圖

由上圖得知，這份聽力試卷對於學生能力值(Theata)介於 -1 到 1 之間的測驗訊息量最多，所以此份試卷對能力值(Theata)介於 -1 到 1 之間的學生的能力估計越準確。測驗訊息依受試者能力水準的不同而異，可取代傳統測驗之信度指標及測量標準誤的概念。故整份試卷的品質良好。

5.2 研究結論

在訊息萬變的網路化時代裡，測驗的方式不再侷限於文字性的表達。以紙筆測驗為主傳統測驗，全班學生用同一份考卷考試，這一份試卷要適合所有學生有其困難。本研究所設計出來的多媒體測驗系統，在測驗試題呈現中表達更多的訊息，提供受試者不因對於紙筆測驗的語意解讀能力較差，以文字模式為基準的試題較不利於他們，可以更公平地進行檢定。

我們發現國內許多測驗系統，皆是以電腦適性測驗來進行核心設計，但是適性化測驗的重點是「題庫建置」，但是題庫的難易度及效度要如何鑑定？如果這個題目已經考過了數百次之後，它的難易度是否能保存像當初建立的那樣呢？

所以本系統提出「試題適性值的驗證與兩段式的動態調整」模式，讓命題委員在命題時，就可以

- 啟動試題的驗證機制

客觀地依驗證系統進行文法、句構等難易度等適性值的分析，然後在預試時，先以IRT的分析機制，並且透過系統

- 即時以視覺化圖形曲線表示題目特徵曲線(ICC&IIC)

- 即時以視覺化圖形曲線表現適性化施測流程與測驗訊息(TIC)

等的協助下，進行「動態第一次的題目適性值調整」。然後在完成不同的測驗之後，我們又可以再次地啟動IRT的分析機制，而進行「動態第二次的題目適性值調整」。

調整完後的參數，便可以放入IRT測驗的試題庫，透過電腦的模擬測驗，我們可以立即知道模擬的結果，及模擬過程中，統整不同的IRT分析器進行適性測驗，以達能力精確估算結果。

本系統從試題上傳以建立多媒體資料題庫，再利用題庫試題組合新試卷，以提供學生英語能力檢定測驗，最後將測驗的結果，診斷分析，動態調整題庫，針對適性化的英語能力測驗提出了一個可行的解決方向。

5.3 對未來研究的建議

1. 本系統針對選擇題都能自動評分，其他類型試題的評分機制尚待建立。
2. 可以結合於 S-P 表的分析理論，提供教師作為診斷試題與診斷學生的重要依據，充分達到診斷的效益。系統可以依據分析的數據，當這數據達到某一個理論值時，系統可以自動的刪除，達到系統自動修正的效果。
4. 未來亦可開發行動裝備介面，讓學生的手持式裝置也可以使用本系統，如：PDA 或手機。特別是日前學生手機持有率有逐年升高的趨勢，若在手機上也可以進行護照闖關，那將可改善設備不足以及場地限制的問題。
5. 目前本系統僅對英語文測驗有關於選擇題測驗的部份，可以進行的適性化測驗，至於寫、以及口說部份，尚未完成，未來可以加以擴充以多方面實施。



參考文獻

- [1] 行政院經濟建設委員會，[On-line]. Available:
http://www.cepd.gov.tw/business/business_sec2.jsp?parentLinkID=7&linkID=146
- [2] 全民英檢學習網站，[On-line]. Available: <http://www.gept.org.tw>
- [3] 多益測驗，[On-line]. Available: <http://www.toEIC.com.tw>
- [4] 財團法人語言訓練測驗中心，[On-line]. Available: <http://www.lttc.ntu.edu.tw/TOEFL.htm>
- [5] 英國文化協會，[On-line]. Available:
<http://www2.britishcouncil.org/zh/taiwan-exams-ielts.htm>
- [6] Lord, F. M. Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates. 1980.
- [7] 余民寧，現代測驗IRT，[On-line]. Available: <http://www.edutest.com.tw/e-irt/index.htm>
- [8] 許擇基、劉長萱，「試題作答理論簡介」，中國行為科學社，台北市，民國八十一年。
- [9] Hambleton, R. K., & Swaminathan, H. “Item response theory: Principles and applications.” *Boston: Kluwer-Nijhoff*, 1983
- [10] Birnbaum, A. Some latent trait models and their use in inferring an examinee’s ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (chapters 17-20). Reading, MA: Addison-Wesley. 1968.
- [11] 何榮桂，「題庫中項目參數分配型態對電腦適性測驗能力估計的影響」，國立臺灣師範大學教育研究所，博士論文，民國八十年。
- [12] Lord, F.M., “Robbins-Monro Procedures for tailored testing” , *Educational and Psychological Measurement*, 31, pp3-31, 1971.
- [13] 何榮桂，「量身訂製的測驗－適性測驗」，*測驗與輔導*，157期，3289~3293，民國八十八年。
- [14] Wright, B. D. and Stone, M. H. *Best Test Design*, Chicago: Mesa Press, 1979.
- [15] 王寶墉，*現代測驗理論*，心理出版社，台北，民國八十二年。
- [16] Wainer, H. *Computerized Adaptive Testing: A Primer* (2nd), New Jersey London: Lawrence Erlbaum Associates. 2000.
- [17] 余民寧，現代測驗IRT－電腦化適性測驗，[On-line]. Available:
<http://www.edutest.com.tw/e-irt/irt12.htm>
- [18] Lord, F. M. “Practical applications of item characteristic curve theory” . *Journal of Educational Measurement*, 14, 117-138. 1977.
- [19] 智勝國際創意網站[On-line]. Available: <http://www.caidiy.com/menu01/menu01.htm>
- [20] 李玉珍，「多媒體英語檢測試題與傳統試題之分析比較--以國小高年級生課程範圍為例」，國立交通大學理學院網路學習碩士在職專班，碩士論文，民國九十四年