

陸、數位化測驗與評量

程式設計教學之學習風格診斷與教學風格施測 朱君怡、廖文淵、楊子奇、黃福銘、楊鎮華(台灣)	1031
2D 動畫數位學習系統導入同儕互評機制之研究 葉嵩生、顏榮泉、林進隆(台灣)	1039
以貝氏網路為基礎進行資訊科技融入國中數學課程與評量之實施研究---以一元一次多項式單元為例 陳威聖、林士勳、施淑娟、劉育隆(台灣)	1048
遊戲互動操作式題型與劇情導向式評量題型之比較-以國中理化「溫度與熱」為例 梁家銘、曾憲雄、翁瑞鋒(台灣)	1056
「分數的乘法」數學單元之建構反應題電腦化測驗 吳任婕、鄭涵、施淑娟、郭伯臣(台灣)	1064
高校教学质量监控综合平台的研究与设计 刘英群、韩锡斌、杨娟、张明(大陸)	1072
電腦語音報讀系統對閱讀障礙考生的影響 黃怡馨、曾芬蘭、林明志、許福元、宋曜廷(台灣)	1079
特殊教育啟智班 Moodle 學習檔案評量系統之開發與可行性評估 徐國梁、劉漢欽(台灣)	1089
應用 S-P 表分析理論輔助國中學生資訊搜尋能力診斷與適性化教學之探究 王曉璿、劉宜欣、柯立芳(台灣)	1095
英文寫作自我效能量表之發展 張苑真、廖長彥、辜玉旻(台灣)	1101
高校数字校园建设水平评价指标体系的研究 韩锡斌、刘英群、张明、周潜(大陸)	1107
基於連續機率比檢定模式之電腦化自然發音精熟測驗遊戲的設計與發展 林逸農、王詩婷、林政廷、楊接期(台灣)	1113
A Folksonomy Aggregation Concept Tagging Scheme for Collaborative Test Sheet Design Ching-Yi Liao (廖經益)、Shian-Shyong Tseng(曾憲雄) Jui-Feng Weng(翁瑞鋒)、Jun-Ming Su(蘇俊銘)(台灣)	1119
鷹架式的模型追蹤於 VB 程式設計理解層級之評量 胡莉玲、曾憲雄、翁瑞鋒(台灣)	1125
應用模糊方程式近似值解法之技能評量分析 曾煥雯、吳孟軒(台灣)	1131
微观建构理论指导下协作知识建构的学习绩效评价研究 张伟、谢幼如(大陸)	1137

大專學生數位健康科技知識分享及學習生涯量表之發展	鄭金謀(台灣)	1141
基於微軟 OpenXML 的 Office 操作題自動閱卷的實現	陈杰、朱卫东(大陸)	1145
電腦化適性測驗的條件化信度區間標記法	吳佳儒、陳柏熹、周業太(台灣)	1149
專題研習中硬支架和軟支架對學生問題解決的影響	周玉霞、李芳乐、胡继昌(香港)	1151
導入知識能量地圖於中小學教育評量之研究	王宗一、蘇建元、趙偉智、張駿淵(台灣)	1156
國中社會領域模擬考之線上適性測驗研發	陳詩靖、郭伯臣、吳任婕、林文質(台灣)	1160

程式設計教學之學習風格診斷與教學風格施測

Learning Styles Diagnosis and Teaching Style-based Assessment in Programming Design

Teaching and Learning

朱君怡、廖文淵、楊子奇、黃福銘、楊鎮華

國立中央大學 資訊工程學系

【摘要】 資訊相關科系在程式設計能力的培養上極為重視，如何增進學生在程式設計課程之學習成效成為重要的議題。本研究使用學生之學習測驗結果進行學習風格的診斷，透過深入了解學生的學習情況及教學風格建議的提供，讓教師可依此設計並調整其教學內容和方法，以提昇程式設計課程之學習成效。

【關鍵詞】 學習風格、教學風格、學習診斷、學習分析系統、程式設計

Abstract: It is devoted much attention to the training of the capability of programming design in Computer Science and in other related departments. How to improve the learning of programming design has become a significant theme for discussion. This paper presents an approach showing a different way to diagnose leaning styles of students by using performances of tests. From deeply knowing learning circumstances of students and recommendations of teaching styles, teachers can revise their teaching to improve the learning effects of programming design.

Keywords: Learning Styles, Teaching Styles, Learning Diagnosis, Learning Analysis System, Programming Design

1. 前言

1.1. 研究動機與目的

程式設計的教學和能力培養在許多大學資訊相關科系中是非常重要且幾乎是必要的，經由課堂上的概念教學和評量並無法確保學生是否真正具備了程式設計的能力，因此，使學生都能擁有正確且完整的程式設計概念，以及具備良好且紮實的程式設計能力，是程式設計學習的最終目標。然而，為了要讓學生的學習可以達到最大的成效，完善的學習成效評估方法及詳細了解學生的學習狀況，更是課程教學中不可被忽略的。在以往的學習階段中，經過了課堂上老師的講授及相關練習題的撰寫和訓練，最後學習成效的診斷往往由幾次測驗的評量分數做為依據。然而，測驗的成績高不一定代表學生具備良好的程式設計能力；相對的，測驗成績差也不一定代表學生的學習成效差，在程式設計能力上仍可能具備相當的水準。對於學生程式設計能力及學習成效的評估，在此我們設計了一個學習分析系統，希望經由此系統對學生的學習狀況做深入的了解，讓學生了解自己的學習風格和學習狀況，進而可自我檢討與加強學習成效不佳之處，同時也讓老師知曉學生們的學習成效與缺失，將評定學習效果的依據由單純數字所組成的直觀的測驗分數，變成客觀、多方向因素的分析結果，為老師進行學生後續的學習輔導及補救教學提供了重要依據，教師可依此分析結果調整教學型態或擬定新的教學方案，以求對學生提供最適合的教學策略，並克服其學習障礙，協助學生作有效的學習以達到發揮學生最大潛能及自我實現之目標。

1.2. 論文大綱

本文的架構如下，第 1 節先簡述研究的動機、背景與目的；第 2 節為介紹相關的研究和概念；第 3 節會詳細敘述本研究之系統架構與所使用的方法；第 4 節將呈現系統之執行與分析結果；第 5 節則對研究結果做總結並提出未來可進行改善的方向；最後，列出本文所參考的文獻資料。

2. 文獻探討

2.1. 學習風格與教學風格

每個學生都有不同的學習方法(Felder & Silverman, 1988)，有些人是經由圖片的閱覽或流程的講解，有些人則是需要詳細的理論推導，儘管每個人的學習方法都不盡相同，但學生大都可依自己的方法和步調達到應有的學習成效。老師們在教學方面，也有屬於自己獨特的教學風格，有些老師喜歡單純的演講授課方式，有些老師喜歡透過小組討論和活動來延伸教學的進行。學生在課堂上能吸收到多少知識及其學習成效，除了和學生本身具備的學習能力和是否對課程做好預先準備等有關外，也與學生的學習風格和教師的教學方式是否相對應有關。

觀察到每個學生都具有屬於自己學習知識的型態和使用所學知識的方法，目前已有許多研究者研究並建立了不同的學生學習風格模組(Thomas, Ratcliffe, Woodbury, & Jarman, 2002)，如 The Myers-Briggs Type Indicator (MBTI)、Kolb's Learning style Model、The Herrmann Brain Dominance Instrument (HEIDI)、The Felder-Silverman Learning Style Model。由於 The Felder-Silverman Learning Style Model 在學習風格之區分與建立有較詳細的分類(Graf, Viola, Leo & Kinshuk, 2007)，所以此處我們採用 Felder 和 Silverman 提出的學習風格模組。Felder 和 Silverman(1988)根據學生不同的學習行為和表現提出了學習風格，或稱為學習類型(Learning Style)的概念，他們認為根據每一個學生所表現出的學習特性，都可以被歸類並找到一相對應的學習風格。在他們所建立的 Felder-Silverman Learning Style Model(FSLSM)中，把學習風格定義成四個象限，每一個象限皆各有兩種相對的學習風格(Felde & Spurlin, 2005; Litzinger, Lee, Wise, & Felder, 2007)，每一學生在此四象限中可分別找到較顯著的學習風格。在第一象限中，Active 表示會積極且主動的使用學習工具，；Reflective 表示根據已知的學習資料和成果來學習。第二象限中的 Sensing 表示喜歡學習能從真實世界找到事例的具體事實；Intuition 則表示較能接受抽象的學習知識，如一些理論和定義。在第三象限的 Visual 是較擅長影像圖表的記憶，如圖片、流程圖、數據圖等；Verbal 則擅長對文字性敘述之呈現方式的學習。而第四象限中，Sequential 是指要按照順序地吸收和理解；Global 則表示可隨意的吸收知識，當學到的知識足夠時會自己歸納成一完整的概念。

由於教師們皆有其慣於使用之教學方法，且每個學生在不同的教學方法下具有不同程度的學習表現，其學習成效也皆有所差異，Felder 和 Silverman(1988)因此針對此現象探討並歸納了教師們的教學行為，提出了與學生之學習風格相對應的概念，稱為教學風格或教學類型(Teaching Style)，他們認為不同學習風格的學生在相對應的教學風格下，有較理想的學習表現並具有顯著的學習成效。Felder 和 Silverman 對於教學風格的定義也分成四大部分，每一部分同樣是由兩種不同的教學風格所組成，其對應如表 1。在第一部分，Active 是讓學生們擁有主動發表意見、進行討論的方式；Passive 則讓學生們從聽講自己消化老師講授的內容。第二部分的 Concrete 為有具體實例的教學內容，讓學生可依事實情況來吸收所學的知識；Abstract 就以一些敘述性的概念和理論來做為教學內容。第三部分中的 Visual 便以圖片、示意圖表、影片等方式來進行教學；Verbal 則是採取演說講解、文獻閱讀等的教學方式。而在第四部分，Sequential 為依邏輯順序，按照步驟章節設計教學內容及進行教學；Global 為使用大綱內容及利用各概念之間的相對關係講解來進行教學。

另外，至今也已有許多將學習風格與教學風格應用在教學上的例子。Zuolkernan(2007)利用學生在各學習風格之學習特性與喜好，對教科書、模擬實驗、互動教學平台等各式教學資源做評估，判斷各式的教學資源究竟較適合哪種學習風格的學生來使用，為達到較理想的教學，將不同的教學資源做結合便可讓分屬不同學習風格的學生皆受益。Kapadia(2008)則探討當學生的學習風格和教師所擅長和採用的教學風格不同時，教師可針對不同學習風格的特性做教學風格的調整和融合，在經過教學實驗後發現儘管只是對教學做些微的調整，仍有可能讓不同學習風格的學生都受用。

表 1 學習風格與教學風格之對應

Preferred Learning Style	Corresponding Teaching Style
<ul style="list-style-type: none"> • Active • Reflective 	<ul style="list-style-type: none"> • Active • Passive
<ul style="list-style-type: none"> • Sensing • Intuitive 	<ul style="list-style-type: none"> • Concrete • Abstract
<ul style="list-style-type: none"> • Visual • Verbal 	<ul style="list-style-type: none"> • Visual • Verbal
<ul style="list-style-type: none"> • Sequential • Global 	<ul style="list-style-type: none"> • Sequential • Global

2.2.MOOSHAK

MOOSHAK 為一 Web-based 之自動評估程式碼的系統工具 (Leal & Silva, 2003)，主要被使用在如 ACM-International Collegiate Programming Contest 等的程式設計比賽中，對參賽者所繳交的程式進行批改和管理，而近年來也漸漸被應用在程式設計學習的相關課程和活動上 (Punitha, Kumar, Seshadri, & Srinivasan, 2008)。MOOSHAK 使用了符合 International Collegiate Programming Contest(ICPC)的規則進行程式的評估以及對程式結果作等級分類，其功能特色為擁有可擴充的系統架構、強大的資料管理系統、自動批改程式碼、及內建的安全機制，這些特色亦是 MOOSHAK 有別於其他程式批改工具的創新之處。

自動批改程式(automated judging)的功能可說是 MOOSHAK 系統的基石，系統根據設定好的規則去驗證參賽者其程式碼之內容和答案，並產生一個結果報告供評估人員做最後的修改和驗證。參賽者送出的資料需包含原始程式碼、隊伍名稱、所使用的程式語言名稱等項目，而在符合比賽限定時間內所送出的答案，MOOSHAK 系統會立即進行批改並顯示程式等級分類的結果，評估人員可選擇重新批改程式，再次地去進行程式的偵錯和測試，以修正最後輸出的程式等級分類結果。自動批改程式的進行，主要可分成兩部分：靜態分析(static analysis)，以及動態分析(dynamic analysis)。

在靜態分析的階段，主要為確認參賽隊伍所送出答案的完整性並進行程式碼的偵錯，若皆成功便產生程式的執行檔，及進入至動態分析。靜態分析首先會檢驗參賽隊伍的答案是否已作答完畢並提交，並再次的檢驗參賽隊伍的名稱、使用的程式語言名稱，及題目序號是否和系統介面的資料符合，之後再檢驗程式原始碼的檔案大小是否在限定的範圍內，最後，當程式碼的檔案大小符合規定，則進行程式碼的偵錯，而若程式偵錯產生錯誤(error)或警告(compiler warning)時，靜態分析結束且也不會再繼續進行到動態分析。而動態分析則為主要對參賽隊伍的程式進行執行與測試，測試的方法則是將系統擁有的標準輸入作為程式的輸入資料並執行程式，之後將執行後的輸出結果和系統的標準輸出答案作比較，最後依答案的比對情形顯示程式批改的結果。

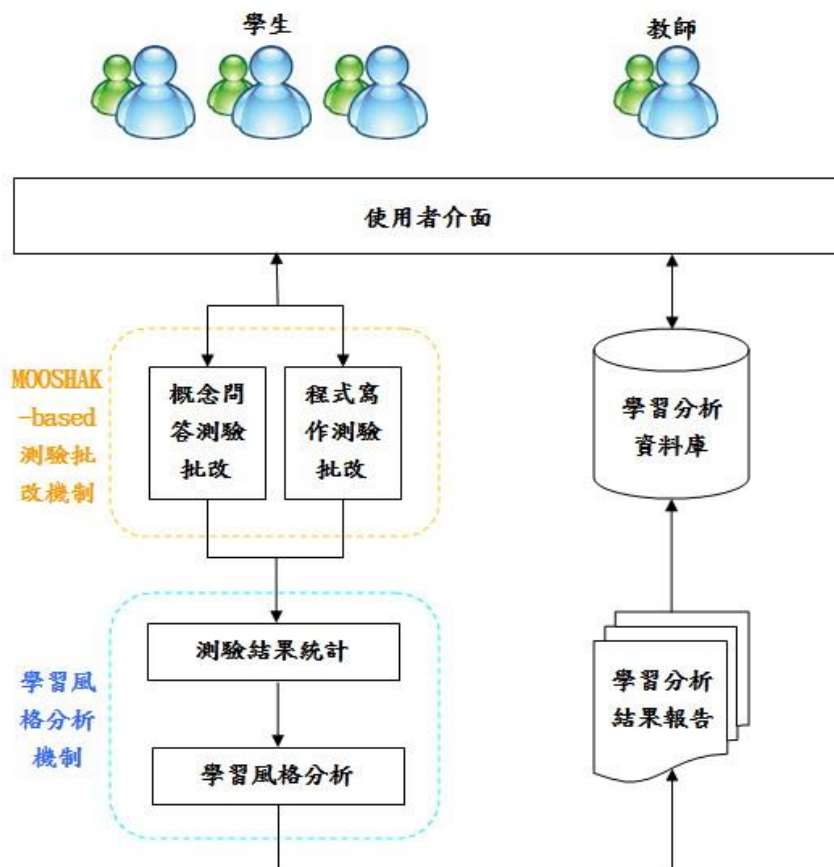


圖 1 學習分析系統架構圖

3. 系統架構與方法

3.1. 系統架構

此學習分析系統主要由三大部份所構成，分別是 MOOSHAK-based 測驗批改機制、學習風格分析機制、及學習分析報告書，如圖 1。在首先的測驗批改機制中，我們將學生的程式寫作測驗和概念問答測驗分開批改。在傳統的觀念問答測試部分，每個題目的答案皆是唯一的，因此在批改上較無困難性，我們便只針對學生回答的答案正確與否作記錄與計分；而程式設計寫作測驗的部分，為了減少批改者的負擔以及節省時間，我們採用了可以自動批改程式碼的 MOOSHAK 系統，減輕辛苦的人力批改負擔。MOOSHAK 系統屬於主從式架構的應用，它是在 Linux 的作業系統下使用 Apache HTTP 伺服器來執行，而 MOOSHAK 也是一 Web-based 的系統，因此其功能皆是經由網頁瀏覽器之界面來使用與執行。而在接下來的學習風格分析機制中，我們將前一個階段 MOOSHAK-based 機制批改出來的結果進行分析，利用學習風格分析方法計算學生在各學習測驗題型中的答題情況，並依據答題表現來歸納學生在各學習風格上之表現情況，找出最符合學生學習表現的學習風格。最後，整理此分析方法的計算結果並綜合 MOOSHAK 系統之程式批改結果，所輸出的學習分析報告書即為此分析模組之重要產物。

3.2. 學習風格分析

在此系統模組中我們設計了一學習風格分析方法，有別於原始 FSLSM 使用其專門設計之學習風格量表的判定方式(Felder & Soloman, 1997)，在此仍採用 FSLSM 對於學習風格的分類，但我們另外歸納了許多不同的學習測驗題型對應於 FSLSM 的學習風格，如表 2，本研究為利用學生們在各種題型的答題情況，進而推斷學生其表現最顯著的學習風格。

表 2 學習風格與學習測驗題型之對應

Style	學習測驗題型	Style	學習測驗題型
Active	操作應用題組	Sequential	流程順序題組
	團體程式測驗		單一概念/章節題組
Reflective	方法設計題組	Global	概念關聯架構題組
	個人程式測驗		綜合概念/章節題組
Sensing	實例應用題組	Visual	圖示圖表題組
Intuitive	理論概念應用題組	Verbal	純文字敘述題組

我們所使用的分析方法，為先對學生在各種題型下之作答結果做統計，統計其每個題型的答對和答錯之題數，並由此計算出學生在各學習風格下的表現情況，歸納出學生屬於何種學習風格，所使用的分析步驟如下：

- (1) 計算各學習題型所屬測驗題的答題狀況，統計各題型中答對的總題數 c1
- (2) 將屬於同一學習類型下之學習題型的答對題數 c1 加上對等學習類型下之答錯題數 q2-c2，再去除以屬於兩對等學習風格之學習題型的總題數 q1+q2，得到各學習風格表現之歸屬值
- (3) 從各學習風格之歸屬值中找出最大值，則為最符合學生的學習風格

接下來，我們以 Visual-Verbal 為例來說明此分析方法。依歸納結果，圖表圖示題型為 Visual 的表現，而純文字敘述題型為 Verbal 的表現，學生受測後的答題狀況如圖 2，圖表圖示題的 10 題中答對了 8 題，在純文字敘述題的 15 題中答對了 4 題，而各學習風格歸屬值的計算是以前者為主，在此處則是 Visual，則 Visual 的歸屬值為

$$R = [8 + (15 - 4)] / (10 + 15) = 19 / 25 = 76\%$$

而另一相對的學習風格，此處即 Verbal，其歸屬值為 1-R，也就是 24%，由此兩種相對的學習風格的歸屬值即可看出，在 Visual-Verbal 象限中學生偏向於 Visual 的學習風格。其他象限的學習風格分析即按照此方法與步驟進行，最後可分別算出此八種學習風格的歸屬值，在此研究我們再將此八個歸屬值做比較，從中找出擁有最大歸屬值的學習風格，並推論學生屬於此種學習風格。

Intuitive	理論概念應用題	correct	7
		wrong	3
		total	10
Visual	圖示圖表題	correct	8
		wrong	2
		total	10
Verbal	文字敘述題	correct	4
		wrong	11
		total	15
Sequential	流程順序題	correct	7
		wrong	1
		total	8

圖 2 學生答題結果統計範例

3.3. 學習分析離型系統

此處我們簡單介紹此學習分析系統的界面與使用，首先學生透過使用者介面進行測驗題的撰寫，程式寫作測驗的作答方式需另外上傳自己的程式碼，如圖 3。在學生答題完畢後，系統便進入分析階段，教師在此階段可依系統提供之答題結果統計進一步查閱學生之程式寫作答題狀況，修改系統對程式寫作的批改結果，如圖 4，修改完後系統便繼續學習風格分析機制的執行。最後系統會將分析報告顯示於使用者介面，讓學生得知自己的學習診斷結果，也

讓教師由此掌握學生之學習狀況，依學習風格與教學風格的診斷建議調整補救教學的內容或方法。

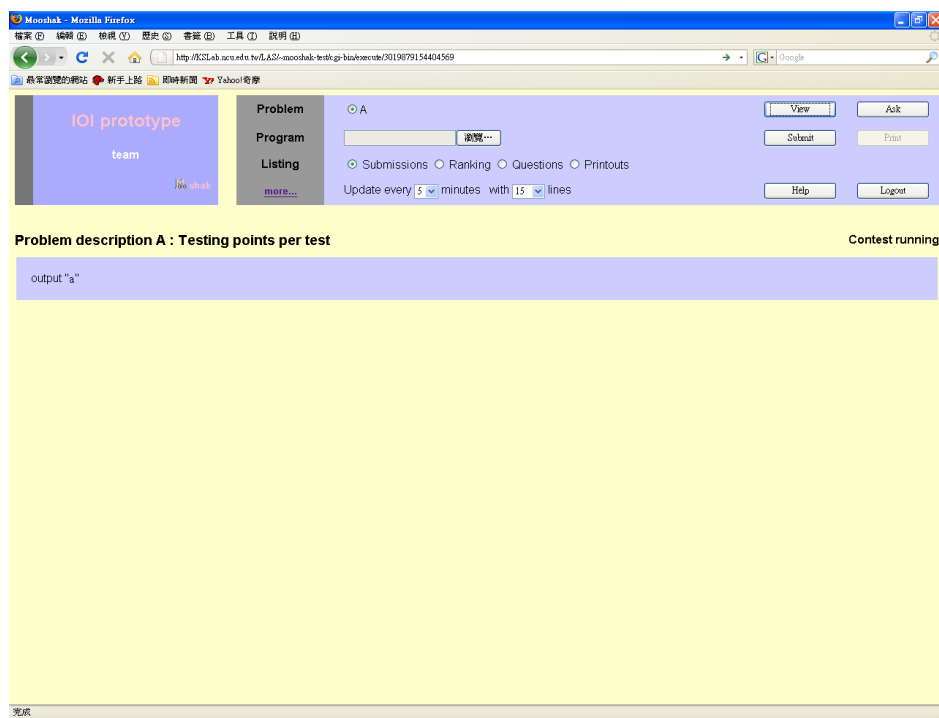


圖 3 學生測驗界面

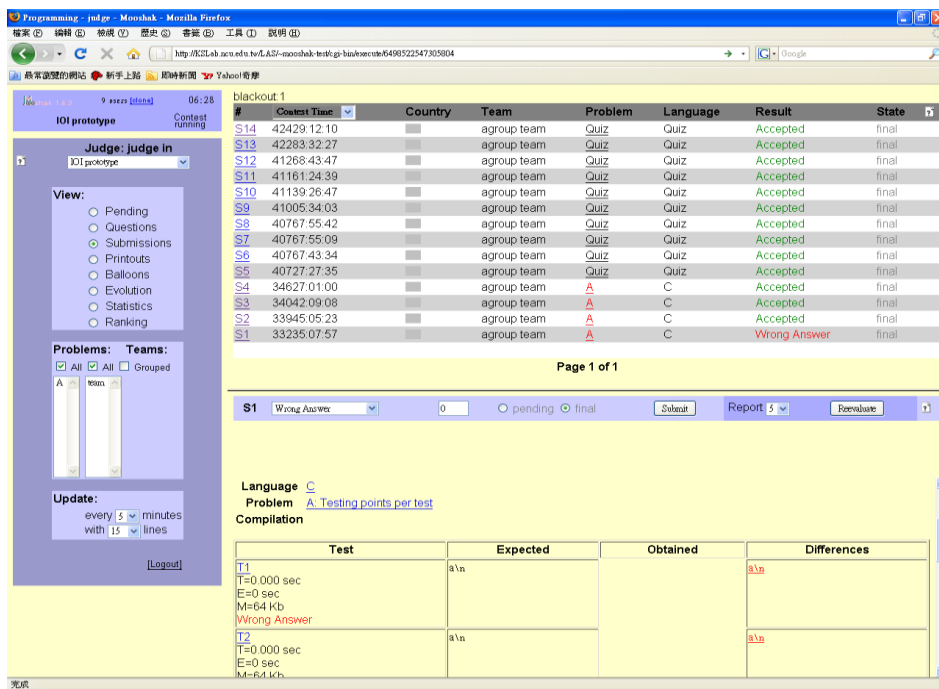


圖 4 教師批改界面

4. 研究結果與討論

此研究之實驗對象為大學部物件導向程式設計課程之所有學生，經由學習分析系統的使用，我們結合學生程式寫作及概念問答的答題狀況與結果，統計出學生的答題結果後會先產生兩份統計報告結果，一是程式能力報告書，另一是學生答題結果統計表，之後系統會再將前兩份報告結果整理，送進學習風格分析機制中進行推論學生所屬之學習風格，以及給予適當的教學風格建議，最後會把整合後的結果呈現出來，如圖 5。根據此報告書我們可得知學

生所屬的學習風格，既讓學生了解自己的學習狀況，也讓老師可以清楚掌握學生的學習成效及學習風格並選擇最適當的教學方式對學生進行補救教學，如 Visual 類型的學生可以使用流程圖、數據圖等圖表的方式來教學， Sequential 類型的學生則需按順序一步一步地慢慢教導，仔細的講解每一步驟的定義，而 Sensing 類型的學生應避免講授過多抽象的概念，適時的帶入一些實例來講解。



圖 5 學習分析報告範例

5. 結論與未來展望

有別於一般對學習風格的判定方法，本研究使用融合了程式設計概念的測驗題替代學習風格量表對學生進行學習風格的診斷，經由學習分析系統的使用後，可同時得知學生程式設計之學習結果及其所屬學習風格，教師再依照學習分析報告所顯示之教學風格建議進行教學之調整，提昇了整體程式設計課程之學習成效。經過數次的教學調整與此系統測驗之搭配使用後，學生懂得依照自己的學習風格選取適合之學習方法，不用完全仰賴教師之指引，而教師也發現學生學習成效逐漸提高的同時，藉著不同學教學風格之混合搭配，學生不再僅局限於適合固定的教學風格，對其他教學風格的適應度也提高了。

此系統的使用不但讓老師可較清楚的掌握學生的學習情況，也讓學生瞭解自己的學習成果，有助於提昇學生們在程式設計學習之成效，但我們也發現當學生沒有擁有正確的程式設計概念時，其作答結果會影響學習風格之判斷。後續我們將繼續改良此學習分析系統，除了讓使用界面可以更簡單明瞭且美觀外，預計再加入原始 FLSM 之學習風格量表，經由原始量表所得之學習風格結果搭配系統之分析結果，以更準確的診斷出學生所屬之學習風格，以及給予更恰當的教學風格建議。另外，在系統功能上希望增加課後作業之批改，將作業之回答狀況一同納入學習成效之診斷，加強系統之實用度與準確度。

誌謝

本研究感謝國科會專題研究計畫(計畫編號：NSC 95-2520-S-008-006-MY3 與 NSC96-2628-S-008-008-MY3.)的經費支持。

參考文獻

- Felder, R. M., & Silverman, L. K. (1988). Learning and Teaching Styles in Engineering Education. *Engineering Education*, 78(7), 674–681.
- Felder, R. M., & Soloman, B. A. (1997). *Index of Learning Styles questionnaire*.
<http://www.engr.ncsu.edu/learningstyles/ilsweb.html>. Retrieved January 30, 2009.
- Felder, R.M. & Spurlin, J.E. (2005). Applications, Reliability, and Validity of the Index of Learning Styles. *Intl. Journal of Engineering Education*, 21(1), 103-112.
- Graf, S., Viola, S.R., Leo, T., & Kinshuk (2007). In-depth analysis of the Felder-Silverman learning style dimensions. *Journal of Research on Technology in Education*, 40(1).
- Kapadia, R. J. (2008). Teaching and learning styles in engineering education. 38th ASEE/IEEE *Frontiers in Education Conference*, T4B-1 - T4B-4.
- Leal, J.P., & Silva, F. (2003). Mooshak: a Web-based multi-site programming contest system. *Software Practice and Experience*, 33(6), 567–581.
- Litzinger, T.A., Lee, S.H., Wise, J.C., & Felder, R.M. (2007). A Psychometric Study of the Index of Learning Styles. *Journal of Engineering Education*, 96(4), 309-319.
- Punitha, T., Kumar, R.P., Seshadri, R., Srinivasan, R. (2008). Mooshak – A Valuable Repository of Codes. 8th *IEEE International Conference on Advanced Learning Technologies*, 644-646.
- Thomas, L., Ratcliffe, M., Woodbury, J., & Jarman, E. (2002). Learning Styles and Performance in the Introductory Programming Sequence. *ACM SIGCSE Bulletin*, 34(1), 33-37.
- Zuolkernan, I. A. (2007). Using Solomon-Felder Learning Style Index to Evaluate Pedagogical Resources for Introductory Programming Classes. 29th *International Conference on Software Engineering*, 20(26), 723-726

2D 動畫數位學習系統導入同儕互評機制之研究

A Study on 2D Animation e-Learning System with Mechanisms of Peer Assessment

葉嵩生、顏榮泉、林進隆*

德明科技大學多媒體設計系

德明科技大學體育室*

【摘要】 同儕互評機制能提升學習者自我察覺、反思及學習成效之表現，已獲許多實證研究的支持。本研究之目的旨在探究同儕互評策略應用於 2D 動畫網路學習之實施成效，研究設計以二因子變異數分析探討學習者之自我調節策略與社會比較心理對採用同儕互評機制之網路學習成效的影響。本研究實施對象為北區某科技大學兩班選修「2D 創意動畫概論」之大二學生共 88 名，進行為期 18 週之實驗教學。研究結果顯示：(1)自我調節策略與社會比較心理對同儕互評學習成效之交互作用並不顯著；(2)低社會比較心理之學習者其學期總成績顯著較高；(3)學習者對同儕互評策略之有用性均抱持相當正面的看法。

【關鍵詞】 同儕互評、2D 動畫、數位學習、自我調節、社會比較

Abstract: Many studies are cited as offering opportunities for enhancing learners' self-awareness, self-reflection and learning effect through the use of peer-assessment. The purpose of this study was to examine the effects on a 2D animation e-Learning system with mechanisms of peer assessment. Through a quasi-experiment instruction, a factorial research design with totally 88 participants was employed with self-regulated strategies (high vs. low) and social comparison (high vs. low) as the independent variables and the scores of peer-assessment and term project as the dependent variables. A 2-way ANOVA was conducted to analyze the main effects of these measures, respectively. The results showed that (a) there are no significant interactions between self-regulated strategies and social comparison, (b) students of the low-group of social comparison outperformed the high-group on the assessment of term project, (c) most of the learners showed positive attitude in helpfulness toward peer-assessment.

Keywords: 2D animation, digital learning, peer-assessment, self-regulation, social comparison

1. 前言

網路科技與多媒體技術的蓬勃發展，不但改變了人類的生活方式，也大幅改變了傳統的學習方式。在數位化與教育科技軟體的協助下，結合大量的圖文、動畫與影音的數位教材逐漸演變成目前學校教育的主流。從數位學習的研究角度觀之，運用動畫科技整合教學內容中的文字、聲音、音樂、圖片、影像等多媒體後，能有效提升各個學科領域之學習成效，並成為學習者克服障礙的利器，已普獲許多實證研究上的支持(陳伶伶, 2004; 梁雪娟, 2005; 劉嘉慶, 2005; 李宜芬, 2006; 王明潭, 2007; 黃幸雯, 2007; 馬紀楨, 2008)。然而，運用網路與數位學習系統，建置適合動畫學科教學的整合式情境，則仍有待相關研究資源投入。

同儕互評(peer-assessment)係指學習程度與知識背景相近的學習者，提供其同時扮演評量者與被評量者身分之機會，促進其認知成長與觀摩學習的教學策略(Topping, 1998; Liu, Lin, & Yuan, 2002)。學習者在同儕互評機制中，通常需要瀏覽並觀察其他同儕的作品，並在教師所給予的評量準則與配分架構中，給予其他同儕學習者回饋，評量的內容可以包含作品的價值、技術層級、創造性或實用性等構面，是相當適合需要邏輯思考與創造力的教學策略(劉旨峰,

2002；吳昆庭, 2007)。由於學習者能透過互評機制反思自己的作品是否也達到課程要求的標準，因此能從自我反思與同儕回饋中獲得調整自我認知結構的機會，進而精緻化個人的學習。

本研究旨在建置一符合動畫學科教學之整合式數位學習系統，在導入同儕互評機制後，學習者之自我調節策略(self-regulation strategies)與社會比較(social comparison)心理等因素，對2D 動畫專題學習成效之影響，同時並探討學習者對同儕互評機制之態度為何。整體而言，本研究目的如下：

- (1).建置符合動畫學科教學需要及具同儕互評機制之整合式數位學習系統。
- (2).進行教學實驗以探討自我調節策略與社會比較差異之變項，對學習者在同儕互評學習系統中之學習成效的影響。
- (3).探討學習者對同儕互評機制之態度為何。

2.文獻探討

本研究主要之學習情境為導入同儕互評機制之動畫教學數位學習系統，學理基礎聚焦於數位學習之自我調節策略與社會學習理論之社會比較觀點。此外，系統建置另以 Topping(1998)之同儕互評理念與實施方法為核心基礎，進行如下之文獻探討。

2.1.數位學習與自我調節策略

數位學習是一種以學習理論為基礎，強調運用資訊與傳播科技破除學習時間與空間的限制，並以學習者為核心的教學情境設計。研究指出數位學習能透過主動學習、互動與互創學習、模擬學習及累積學習等面向，建構一個提升學習動機與學習成效的數位情境(孫春在、林珊如, 2007)。然而，由於數位學習模式中學習責任多數需落實在學習者自身身上，因此探討學習者如何在網路環境中自主學習之自我調節學習(self-regulated learning)理念，便成為近年來相當熱門的研究議題。

Zimmerman(1998)認為自我調整是一種讓學習者能自我激發、活化認知活動及學習行為，並因此能有系統的達到既定目標的過程。換言之，自我調整能使學習者在面臨新的問題時，自我察覺對所面臨問題的掌控情形、自我反省與自我批判所採用的策略是否適當、並自我監控以有效率的方式來達成所設定的學習目標。Schunk & Zimmerman(1998)則認為自我調整學習應包含自我觀察、自我監控、自我判斷及自我反應等程序。自我觀察是指學習者能對自己的學習表現進行初步的反省與覺知；自我監控則是有計畫的記錄與瞭解自己的行為表現；自我判斷則是依據所設定的目標評估自己的表現是否達成預定進度；而自我反應則是對自我判斷的評估結果，決定採取的後續行動。由此探討可知，在數位學習的環境中，學習者是否能充份掌握學習資源，瞭解自身的學習進度並適當安排自己的學習目標等與自我調節策略有關之因素，將扮演舉足輕重的角色。

2.2.社會學習理念與社會比較心理

社會學習理論認為：人類行為的改變，受學習者不斷進行內在化(internal)認知處理與外在化(external)社會處理之交互作用所引導。Bandura(1986)之社會認知理論(social cognitive theory)則進一步闡述學習者可透過楷模學習(modeling)的方式，藉由觀察同儕學習者的學習歷程與結果，從而內化成屬於自己的知識與經驗。楷模學習係指觀察者在選定一個或多個楷模對象後，在想法、信念、策略或行動上模仿他們，並藉此獲得技能、信念、態度或行為的一種學習方式(Bandura, 1986)。

Schunk(1998)認為同儕楷模(peer modeling)是楷模學習理念中最常應用於促進群體認知學習的方法。然而，學習情境中所崇尚的主流價值，會影響學習者評價成功或失敗的標準。例如：班級風氣強調分數的排名，則社會比較容易成為學習者衡量同儕楷模的努力目標；反之，若班級風氣強調合作互助的話，則團體成長會變成群體知識獲得和學習的共同方向(陳嘉成, 1999)。本研究即從此觀點切入，期能藉由教學實驗探討社會比較心理對同儕互評機制下的學習成效之影響。

2.2. 同儕互評機制的理論基礎與實施方法

Topping(1998)認為同儕互評是指學習程度與知識背景相近的一群學習者，彼此扮演評量者與被評量者之角色，針對彼此作品予以評價的歷程。由於同儕學習者間之認知結構與口語溝通模式較為接近，因此在評價的過程中能提供比專家楷模(教師示範)更貼近學習者程度的觀點。許多實證研究證實，同儕互評機制能促進學習者之認知成長、自我察覺、自我反思及學習成效(劉旨峰, 2002；孫春在、林珊如, 2007)。

事實上，由於資訊與網路科技的成熟發展，使得同儕互評此種評量機制更為普及。因為網路同儕互評模式能克服實施時間與空間的限制，讓學生能在課堂與課餘時間中充份的互動，若再搭配討論區與 E-Mail 等工具的使用，的確能突破傳統同儕互評效率不彰之缺失(Jonassen, 2000)。Falchikov(1995)與 Liu, Lin 及 Yuan(2002)等人的研究均指出：網路同儕互評能改善大多數傳統同儕互評在實施上的缺點，進而獲致更好的學習成效。此外，Liu (2002)等人在國內所進行的研究結果亦建議，網路同儕互評可以採取匿名的實施方式，較不會造成東方學生因同儕壓力所造成的評分不公，亦可提高同儕互評的信度與效度。

吳昆庭(2007)整理有關同儕互評的研究中歸納：多數同儕互評的研究僅探討實施方式與學習成效方面的議題，較少探討在同儕互評的過程中何種因素會影響同儕互評的結果。因此，本研究將研究核心聚焦於數位學習理念之自我調節策略與社會學習理論之社會比較觀點，以實驗研究探討同儕互評過程中學習者個人的認知調整與社會比較心理，是否影響其在網路情境中之動畫學習成效，期能提供同儕互評相關研究之參考。

3. 系統環境與教學實施

3.1. 研究對象與研究設計

本研究旨在建置一符合動畫學科教學的整合式數位學習系統，在導入同儕互評機制後，學習者之自我調節策略與社會比較心理等因素對動畫學習成效之影響。本研究以準實驗研究法，進行為期 18 週之專題與實驗教學，實驗對象為北區某科技大學兩班選修「2D 創意動畫概論」之大二學生共 93 名學生，扣除未全程參與及無效樣本後，實際參與人數為 88 位(男 37、女 51)。參與者在學期之初即依研究工具調查結果區分為高、低自我調節策略組以及高、低社會比較心理組，再以二因子變異數分析探討自我調節與社會比較對採用同儕互評機制之數位學習成效之影響。此外，本研究於實驗課程結束後另以態度問卷方法探討學習者對同儕互評機制之有用性的看法為何。

3.2. 研究工具

本研究之自我調節策略量表為參考 Pintrich 和 Degroot(1990)所編製的學習動機與學習策略量表(motivated strategies for learning questionnaire, 簡稱 MSLQ)，從中選用與本研究目的相符之認知調整策略分量表，該量表原始信度為.943 (Cronbach's $\alpha = .943$)，量表效度為專家效

度。MSLQ 採用 Likert 五點量表之作答形式，由「完全不符合」到「完全符合」的連續選項，分別給於 1、2、3、4、5 分，填答分數愈高代表在各題項上有較佳的表現。

此外，本研究另參考林梅琴、黃佩娟(2000)所編訂之「技職院校學生學習狀況及學習困境之調查問卷」，選取其中有關同儕學習之社會比較心理部份題目，修改成適合本研究之問卷調查工具。原始問卷之信度為.859 (Cronbach's α = .859)，量表效度為專家效度，作答形式及給分方法亦為 Likert 五點量表模式。本研究修正後問卷施測結果之信度為.863 (Cronbach's α = .863)。此外，為瞭解學習者對同儕互評機制之態度為何，本研究另自編同儕互評之有用性態度問卷(6 題)，亦為 Likert 五點量表模式施測，施測結果之信度為.813 (Cronbach's α = .813)。

表 1 本研究使用之問卷工具施測結果之信度摘要表

研究工具	題數	Chronbach's α
自我調節策略(MSLQ)	6	.872
社會比較與同儕楷模	6	.863
同儕互評之有用性(態度問卷)	6	.813

3.3.系統環境與架構

本研究教學網站之建置以 Apache、MySQL 與 PHP 等開放式套件架構而成，內容管理則採用內容管理(Content Management System, CMS)套件 Joomla，來管理學習者出版的多媒體動畫內容，課程設計與各週次教學重點如表 2 所示。課程理念是以營造學習、分享、參與與競爭的同儕互評為目標，因此網站建置方向以提供重複學習、刺激創意表現、增加學習樂趣與獲得適度肯定為原則(如圖 1 所示)。此外，學習者在本研究教學實驗階段必須完成三項課堂作業、期末則需完成符合完整功能需求之動畫作品(詳圖 2)。

至於同儕互評方面，學習者並非依據個人主觀進行評分，而是由教師考量動畫學習進度與作業規畫，提供完整的課程要求項目及配分評分表(如圖 3)，以 5 顆星(五等第)的評等方式，要求學習者觀摩他人的作品並以無記名方式互評。

表 2 本研究之課程設計安排

課程單元	單元教學主題	教學設計與課程內容安排
開課前	融合學生需求	於開課前召集修習過本課程之同學，針對規劃的課程內容共同討論提出建議，並著手收集資料建立單元範例以配合建置數位輔助教材
第 3 週	形狀補間動畫	如開場動畫、動畫遮罩
第 5 週	移動補間動畫	如文字特效、物體旋轉(橫幅廣告創作)
第 7 週	逐格動畫	如人物眨眼、狗搖尾巴
第 9 週	期中評量	加強評量與重點輔助教材之關聯性，以提高學生課後練習與考前複習，達到輔助教材充分利用的目的
第 10 週	公佈期中評量結果	以無紙化的方式公佈個人期中評量結果，主要的目的，除了長期推動無紙化的概念外，並藉以說明尊重個人隱私的重要
第 10 週	聲音與視訊	人物與對嘴
第 12 週	動畫製作流程	完整動畫範例實作
第 14-18 週	期末專題	公佈期末專題進度管控與評分機制，並由老師按每位同學之配合情形與以評分(期末專題創作/校外競賽參與)
第 15-17 週	導引線與遮罩	如行星運行、放大鏡、水波、探照燈
第 18 週	期末考與互評	期末考試(導引線與遮罩)，建構期末展示平台，期末專題互評



圖 1 本研究所建構具同儕互評機制之 2D 動畫數位學習環境

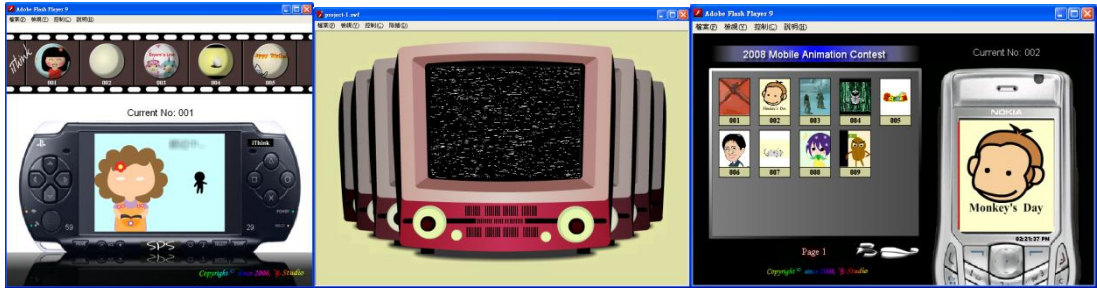


圖 2 本研究教學實驗階段學生必須完成之動畫作品

D962_grade.xls [新增欄位] - Microsoft Excel					
D4					
1. 依你個人的標準，就每部作品評分(0-10 分等第)，也請自評。					
2. 評分時請客觀、公正，一應是現有的規定互評者，作品以 0 分計算。					
3. 評分後請儲存檔案，不用刪除直接上報！IP：4444_專題評分表。					
4. 請確認你的上傳檔案為已評分完成的檔案。					
5. 位參與評分、隨便評分或上傳空檔案者，互評分數以 0 分計算。					
97 學年度第一學期 日四技藝二甲 期末專題 評分表					
序	主題	說明	等第	分數	確認
1	流水號				
2	D962_001	「輕」聲細語從己做起		0	v
3	D962_002	Don't drink on the MRT		0	v
4	D962_003	Merry and Xmas*		0	v
5	D962_004	人人心中都有一個博愛座		0	v
6	D962_005	小心、愛心、心主張		0	v
7	D962_006	小心香蕉		0	v
8	D962_007	不要以為看版是假的		0	v
9	D962_008	少年哪！哇係阿伯呀！		0	v
10	D962_009	心文化-要站穩		0	v
11	D962_010	心主張II		0	v
12	D962_011	手機「心」使用		0	v

圖 3 本研究之動畫專題作業五等第同儕互評評分表

4.結果與討論

本研究歷經 18 週之專題學習活動，所有參與者在期限內均依課程要求完成個人動畫作

品，並透過系統互評機制完成對其它同學的評分。實驗變項各分組之同儕互評得分及學期總成績得分之描述統計結果如表 3 所示。

表 3 自我調節策略與社會比較心理對同儕互評學習成效之平均數摘要

自我調節策略	社會比較心理	同儕互評成績		學期總成績		人數
		平均數	標準差	平均數	標準差	
高自我調節	高社會比較	66.27	8.207	82.86	7.266	22
	低社會比較	69.45	7.123	85.09	7.615	22
	全體	67.86	7.763	83.98	7.441	44
低自我調節	高社會比較	64.82	13.609	80.36	8.335	22
	低社會比較	67.05	11.846	84.32	7.713	22
	全體	65.93	12.659	82.59	8.250	44
全體	高社會比較	65.55	11.130	81.61	7.830	44
	低社會比較	68.25	9.736	84.95	7.576	44
	全體	66.90	10.485	83.28	7.842	88

從表 3 之平均數與變異數之統計顯示，在 2x2 的實驗分組中，高自我調節組在具同儕互評機制的學習成效表現上均優於低自我調節組；然而，值得注意的是低社會比較組在同儕互評與學期成績之得分表現上均優於高社會比較組。

4.1. 自我調節策略與社會比較心理對同儕互評學習成效之交互作用並不顯著

接著，本研究分別以同儕互評成績及學期總成績為依變項，以自我調節及社會比較為自變項，進行二因子變異數分析之交互作用檢定。在同儕互評成績方面，由於 Levene 變異數的同質性考驗中，F 值為 1.508，p 為 .218，未達顯著水準，符合變異數同質性的假設，經二因子變異數分析後交互作用結果如表 4 所示。另一方面，以學期總成績為依變項之分析中，Levene 的同質性考驗 F 值為 .222，p 為 .881，亦未達顯著水準符合同質性的假設，其交互作用統計結果如表 5 所示。

表 4 自我調節策略與社會比較心理對同儕互評學習成效之二因子變異數分析摘要
依變項：同儕互評成績

變異來源	型 III 平方和	自由度	均方	F	p	效果值
自我調節	82.102	1	82.102	.740	.392	.136
社會比較	160.920	1	160.920	1.451	.232	.222
自我調節x社會比較	5.011	1	5.011	.045	.832	.055
誤差	9316.045	84	110.905			

註：R 平方 = .26 (調整後 R 平方 = -.009)

表 5 自我調節策略與社會比較心理對同儕互評學習成效之二因子變異數分析摘要
依變項：學期總成績

變異來源	型 III 平方和	自由度	均方	F	p	效果值
自我調節	42.284	1	42.284	.705	.403	.132
社會比較	245.557	1	245.557	4.097	.046*	.516
自我調節x社會比較	27.284	1	27.284	.455	.502	.102
誤差	5034.773	84	59.938			

註：R 平方 = .059 (調整後 R 平方 = .025)

從表 4 及表 5 之統計得知：在 2x2 的二因子變異數分析中，自我調節與社會比較兩變項間的交互作用並不顯著。

4.2. 低社會比較心理之學習者其學期總成績顯著較高

從表 5 之統計分析可知，雖然自我調節與社會比較變項間的交互作用並不顯著，然「社會比較」之主要效果檢定已達顯著水準($F = 4.097$, $p = .046 < .05$)。可見高、低不同社會比較心理之參與者，在學期總成績之表現上存在顯著差異，經比較其平均數得知，低社會比較心理之學習者其學期總成績顯著較高社會比較心理者為佳。

表 6 高、低社會比較心理對學期總成績之主效果平均數比較

社會比較心理	平均數	標準差	95%信心水準	
			下界	上界
高社會比較	81.614	1.167	79.293	83.935
低社會比較	84.955	1.167	82.634	87.276

此研究結果與研究者所預期之結果不甚相符，一般推論，社會比較心理較高之學習者相對會投入較多的心力在作業上，似乎應獲得較高之成績。本研究之低社會比較組學習者在同儕評分上並未顯著獲得高分，反而在學期總成績上表現得較為傑出，此現象相當有趣。研究者推論或許是因為學期總成績額外採計競賽、筆試及出席率等相關因素，遂造成低社會比較心理之學習者表現較佳，然仍需進一步深入探討。

4.3. 學習者對同儕互評機制均抱持相當正面的態度

本研究另以態度問卷調查方法，探討學習者對採用同儕互評機制之數位學習系統是否有助於動畫學習之看法。各題項之統計結果如表 7 所示。

表 7 本研究參與者對同儕互評機制之有用性看法

問 卷 題 項	平均數	標準差
1.同儕間的觀摩與分享有助於創意的激發與展現？	4.13	.640
2.網站中所營造的觀摩分享氣氛，能激發我的創意展現？	4.07	.603
3.課程中參與每位同學作品的評分，能增加我學習的樂趣？	4.02	.742
4.課程中同儕互評的機制，我覺得是公平且合理的？	3.82	.653
5.課程中同儕互評，會讓自己更投入作品的創作？	3.91	.560
6.我在創作的努力能獲得同儕的肯定？	3.76	.678
問 卷 總 計	3.951	.421

從表 7 可知，學習者對於本研究之數位學習系統導入同儕互評機制是否有助於動畫學習之看法均持相當正面的態度。六個題項之總平均數達 3.951，且所有題項之平均數均大於五等第量表之中數 3。其中，第四題及第六題詢及有關是否認同互評機制之公平性時，相對於其它題項得分略低，似乎與前述 ANOVA 分析結果互為呼應。

本研究旨在瞭解技職院校 2D 動畫課程之數位學習系統在導入同儕互評機制後，學習者之自我調節策略與社會比較心理對數位學習成效的影響。研究結果發現：學習者自我調節策略的高低對本研究學習成效之影響並不顯著，而社會比較心理則對學習成效產生微妙而有趣的作用。未來研究將擴大樣本空間，並改善研究設計與問卷工具，期能進一步探討自我調節策略與社會比較心理對同儕互評機制的影響。

致謝

本研究相關教學實驗及使用之學習輔具乃是由行政院國家科學委員會經費補助，計畫編號：NSC 97-2511-S-147-001。

參考文獻

- 王明潭（2007）。運用相互教學法融入動畫設計對國小五年級學生英語閱讀理解及動機之研究，國立屏東教育大學教育科技研究所碩士論文。
- 李宜芬（2006）。電腦動畫輔助教學之適切性探究—2D 與 3D 電腦動畫輔助教學指標之建構，國立臺北教育大學國民教育學系碩士論文。
- 林梅琴、黃佩娟（2000）。專科學生學習低成就的成因及學習困境之研究--以德明商專為例。德明學報，16，373-396。
- 吳昆庭（2007）。高層次提問與同儕互評對網路學習者反思之影響，國立中山大學資訊管理研究所碩士論文。
- 馬紀楨（2008）。3D 動畫應用於國小四年級自然領域之教學成效-以月相概念為例，國立臺東大學教育學系碩士論文。
- 孫春在、林珊如（2007）。網路合作學習：數位時代的互動學習環境、教學與評量。台北，心理。
- 陳伶伶（2004）。多媒體動畫融入國小六年級教學之學習成效研究--以分數之加、減、乘法為例，國立臺南大學數學教育學系碩士論文。
- 陳嘉成（1999）。成就目標、動機氣候、自我歷程與自我調整策略、持續學習動機和數學成就之關係。國立政治大學教育研究所博士論文。
- 黃幸雯（2007）。以圖文大師動畫語音軟體探討中度智能障礙學生功能性字彙識字學習成效之研究，臺北市立教育大學特殊教育學系碩士論文。
- 梁雪娟（2005）。空間能力與動畫學習對學習成效影響之研究，國立彰化師範大學商業教育學系碩士論文。
- 劉旨峰（2002）。網路同儕評量之研究，國立交通大學資訊科學所博士論文。
- 劉嘉慶（2005）。晶體模型融入教學、電腦動畫教學與傳統式教學法對高中學生晶體構造學習成就影響之研究，國立高雄師範大學化學系碩士論文。
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Falchikov, N. (1995). Peer feedback marking: developing peer assessment. *Innovation in Education and Training International*, 32, 175-187.
- Jonassen, D. H. (2000). *Computers as mindtools for schools: Engaging Critical Thinking*. (2nd Ed.). Merrill, New Jersey: PrenticeHall.
- Liu, E. Z. F., Lin, S. S. J., & Yuan, S. M. (2002), Alternatives to instructor assessment: A case study of comparing self and peer assessment with instructor assessment under networked innovative assessment procedures, *International Journal of Instructional Media*, 29(4), 1-10.
- Schunk, D. H. (1998). Peer modeling. In K. J. Topping, & S. W. Ehly, (Eds.), *Peer-assisted learning* (pp. 185-202). Mahwah, NJ: Lawrence Erlbaum Associates.

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.

Zimmerman, B. J. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional model. In D.H. Schunk & B.J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 1-19). New York: Guilford.

以貝氏網路為基礎進行資訊科技融入國中數學課程與評量之實施研究---以一元一次多項式單元為例

A Study of Integrating Information Technology into Math Course and Measurement in

Junior High School Based on Bayesian Networks at One-variable Linear Polynomial Unit

陳威聖、林士勛*、施淑娟**、劉育隆***

亞洲大學資訊工程學系碩士在職專班

國立臺中教育大學教育測驗統計研究所*

國立台中教育大學數學教育學系**

亞洲大學資訊工程學系***

【摘要】 本研究目的以國中「一元一次多項式」單元，建立一套以貝氏網路為基礎的電腦適性診斷測驗，並研發數位教材之教學內容，而後進行教學成效評估。首先，將受試者分成實驗組及對照組進行團班教學。其次，將所研發之教材內容提供實驗組教師教學，並讓學生做電腦適性診斷測驗。最後，進行補救教學後再讓學生做一次電腦適性診斷測驗。以兩次施測後的成果評估資訊融入教學之教學成效，研究結果顯示使用數位指導教材的實驗組不管在教學或補救成效均顯著優於傳統教學的對照組。

【關鍵詞】 貝氏網路、資訊科技融入教學、一元一次多項式、評量

Abstract: The study is to develop the knowledge structure of computerized adaptive test based on Bayesian Networks by the unit of one-variable linear polynomial in junior high school and build digital teaching materials for the evaluation of the teaching effects. First, divide the subjects into experimental group and control group to conduct teaching. Second, the teacher of experimental group use the materials developed by the researchers to conduct teaching activities, and makes all students take the first computerized adaptive test. At last, have them take the second test after finishing the remedial teaching. The results show that the experimental group is much better than control group.

Keywords: Bayesian Network, information technology, one-variable linear polynomial, measurement

1.前言

1.1. 研究動機與背景

皮亞傑的認知發展理論研究指出：「具體運思期在7歲至11歲，特徵為分類、對應關係與數的具體運思活動，而形式運思期在11歲至成年，特徵為假設—演譯及抽象思考」，而國中生開始學習代數的年齡在12歲左右，由具體運思期之算術導向形式運思期之代數的重要關鍵，所以一元一次式的學習對國中生來說極具重要性與挑戰性，如何改進教學是一值得深入探討的課題。

在資訊科技發達的社會中，學校教育能善用科技改善教學環境，並透過資訊將教學結構化，使學生的學習更具成效，先前相關研究亦指出資訊化教學能達到事半功倍(李淑娟, 2006)。此外，如果教學能夠針對學生學習的盲點進行補救，對學生的學習更會有顯著的助益。因此，完整的教學應包含教學、評量以及補救教學，所以本研究將三者以一套數位教

學模式之研發，並評估其成效。而此套數位教學模式中將結合知識結構與貝氏網路為基礎之電腦化適性診斷測驗系統(楊智為、劉育隆、楊晉民、曾彥鈞, 2006)作為評量工具，配合資訊科技將教材數位化，以達到創新教材，量身定做的適性學習模式。

基於以上背景分析，研究者擬以貝氏網路為基礎，進行資訊科技融入國中一元一次式的數學課程與評量的研究。

1.2. 研究目的

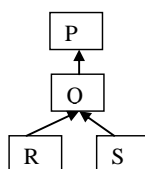
本研究目的如下：首先，比較數位教學模式與傳統教學模式之教學成效；其次，比較數位教學模式與傳統教學模式之補救教學成效；最後，探討數位教學模式對不同能力學生學習成效的影響。

2. 文獻探討

2.1. 以貝氏網路為基礎之電腦化適性診斷測驗

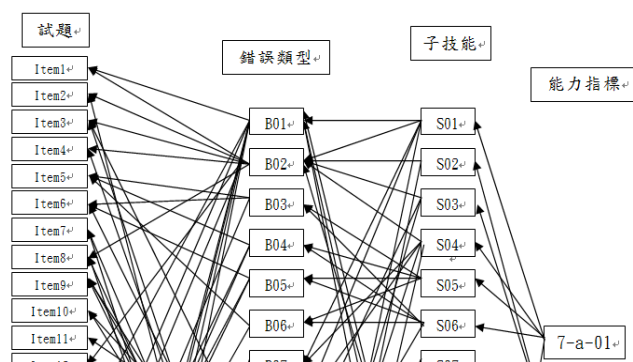
本研究在資訊融入測驗的設計上，是以貝氏網路 (Bayesian network) 為基礎的電腦化適性診斷測驗。測驗系統是以 Airasian & Bart (1973) 的順序理論 (ordering theory, OT) 找出貝氏網路中的知識結構，作為選題的策略，再利用貝氏網路機率更新的技術提供錯誤類型診斷的模式，簡要說明理論如下：

一、試題結構理論：郭伯臣、謝友振、張峻豪、蔡坤穎(2005)的研究結果顯示，使用順序理論(OT)結構之適性測驗選題策略，所需訓練樣本題數較少而且可節省施測題數，優於 IRS 與 Diagnosys。所以本研究採用 Airasian & Bart (1973) 的順序理論 (OT) 來估計試題結構，並用於適性診斷測驗流程之建立。在 OT 理論中，試題與試題的聯合與邊界機率(joint and marginal probabilities)估計得到如圖像 1 的試題結構圖，即可轉化為概念結構圖，如 $R \rightarrow Q$, $S \rightarrow Q$ 代表試題 R 及試題 S 所測量的概念是試題 Q 測量概念之下位概念，代表學生若答對 Q 時，也會答對 R 及 S。因此當學生答對 Q 時，則 R 及 S 就不須再測驗，因此可達到適性省題的目的。



圖像 1 OT 理論之試題上下位關係圖

二、貝氏網路：貝氏網路包含兩個部份，包含節點 (node) 和連結 (link)。而本研究中，節點分為四種，包含能力指標 (competence Indicators)、子技能 (skill)、錯誤類型 (bug) 和試題 (item)。而節點之間利用有向邊且非循環的連結，連結之有無代表節點間的關係是否為相依或獨立的條件之情形，有連結表示兩節點是條件相依。我們以圖像 2 的一元一次式貝氏網路部份摘錄進行說明，稱 S01 為 B01、B02 的父節點，而 B01、B02 則為 S01 的子節點。



圖像 2 一元一次式貝氏網路圖（部份摘錄）

三、適性測驗：本研究之適性測驗係以試題結構理論及貝氏網路為基礎的電腦化測驗，而電腦能依受試者的作答反應控制題目數量及施測的時間，並能保持一定範圍的診斷正確率。

2.2. 一元一次式試題分析

本研究試題做為診斷測驗及補救教學的依據，而試題的錯誤類型係由相關文獻，及研究者教學經驗所編制而成，如表 1：

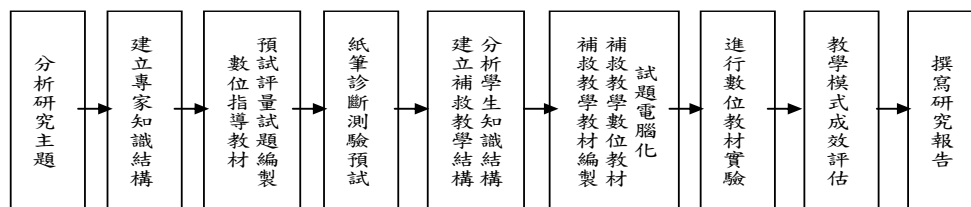
表 1 相關文獻與部分錯誤類型之對照表(摘錄)

代號	錯誤類型	相關文獻
B01	缺少數字與文字符號間的運算法則的概念	王如敏(2004)
B11	分配律使用不當(忘記分配後項)	王如敏(2004)
B12	做一元一次多項式的化簡時，增加文字符號	蕭宇欽(2007)
B13	做一元一次多項式的化簡時，遺漏文字符號	蕭宇欽(2007)
B16	缺少有理數運算的概念	研究者教學經驗

3. 研究方法與步驟

3.1. 研究流程

本研究根據國中一年級「一元一次式」單元所包含能力指標及專家知識結構，編製數位指導教材，並依該單元之錯誤類型編製預試評量試題，再經由紙筆測驗，分析學生知識結構並修正試題，建立電腦化適性診斷測驗。然後，依照學生與專家知識結構，編製補救教學教材。最後，將實驗組與對照組進行教學實驗，評估教學成效。整個研究流程如圖像 3 所示：



圖像3 研究流程圖

3.2. 研究工具

本研究教學模式研發與後續教學實驗需使用之研究工具的編製說明如下：

3.2.1. 紙筆診斷測驗 本研究之紙筆診斷測驗依據「電腦化適性診斷測驗之研究」(陳怡如、吳慧珉、黃碧雲，2004)建立專家知識結構圖，並據此做為紙筆診斷測驗的命題，其中，每一題之誘答選項皆據學生常見錯誤類型加以設計。紙筆診斷測驗共有2份複本測驗，每份試題為35題。紙筆診斷測驗係以8年級學生為施測對象，有效樣本339人，資料處理後，經由統計軟體

SPSS信度分析，所測得之Cronbach's α 值為0.933，顯示本工具有良好的內部一致性。難度範圍在0.191到0.327；鑑別度範圍在0.289到0.654，平均鑑別度為0.395，表示此份試題有很好鑑別度。

3.2.2. 電腦化適性診斷測驗 本研究採用以知識結構與貝氏網路為基礎之電腦化適性診斷測驗(Bayesian network based adaptive test, 簡稱BNAT)作為分析工具，利用試題結構作為選題策略，再利用貝氏網路進行推論，應用於診斷學生錯誤類型，在節省題數、施測時間及預測精準度方面都有最佳的表現(楊智為、劉育隆、楊晉民、曾彥鈞，2006)。

3.2.3. 數位指導教材 先以專家知識結構的節點進行單元教學重點編製，形成以教師為基礎之教材，包含數位教學媒體、教學手冊及單元講義。再整合學生試題與專家知識結構，設計學生加油手冊及補救教學動畫等教材，最後，編製「學生學習回饋單」，主要在瞭解學生對教材使用的滿意度及學習態度的調查。各項數位指導教材簡述如下：

一、教學手冊是提供電腦適性教學使用。手冊中提供詳細教材使用說明、補救教學重點對照表、詳細教案、學生講義詳解和學生加油手冊詳解，以方便任課教師使用。本單元由16個學習重點組成，每一個重點包含重點名稱、觀念導入、重點整理、布題、與自我練習。

二、學生講義是供學生上課和練習用。在單元中的每個子技能為一個教學重點，而每個重點中，教師將觀念導入後，再透過例題進行解題教學，進而澄清觀念。爾後，學生可利用練習題自我挑戰。整本講義的內容包括重點整理、例題、例題詳解及練習題。

三、教學多媒體是供教師進行教學使用，不同於傳統黑板的教學方式，教學多媒體配合各學習重點，將資訊融入教學，呈現生動畫面與詳細解題步驟，較易引起學生學習興趣。

四、學生加油手冊配合單元概念及錯誤類型編寫，供學生進行補救教學。每一個重點練習皆提供練習題，而在手冊最後附上綜合進階題，供提早完成補救的學生進行加深加廣的練習。

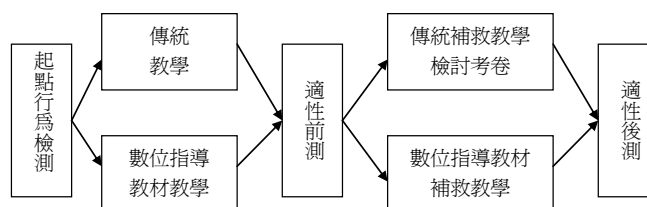
五、補救教學多媒體是供學生進行線上補救教學用。不同於紙本的方式，補救教學多媒體配合各學習重點，利用簡報軟體進行編製，每個教學重點利用範例引導學生進行學習，使學生更易進行概念澄清，幫助學生再學習。

3.3. 研究對象

本紙筆測驗採隨機簡單抽樣，施測對象為自97學年度國中二年級學生，其中北部有3個班級，中部有5個班級，南部有4個班級，有效樣本共計339人。而教學實驗對象以南投縣立偏遠地區的2所國中一年級學生進行教學實驗，實驗的班級數為4個班級，其中對照組有2個班級，共50位學生；另1組為實驗組，亦有2個班，共52位學生。

3.4. 實驗設計

實驗組採數位指導教材教法，對照組採傳統教學教法，實驗流程如圖像4，說明如下簡述。



圖像4 實驗流程圖

一、起點行為檢測：本研究依據教育部(2003)數學學習領域國中一年級能力指標內容編製起點行為檢測試卷，目的在瞭解受試者起點行為能力。

二、單元教學(210分鐘)：同一單元教學中，實驗二種不同教學方法，實驗組以數位指導教材進行教學；對照組由傳統教材進行教學。

三、適性前測(45 分鐘)：兩組受試者在單元教學完畢後，以 BNAT 系統進行適性前測，測驗完畢後由系統產生學習診斷報告書，畫面如圖像 5。



圖像 5 BNAT 系統測驗畫面暨學生個人診斷報告圖

四、補救教學(45 分鐘)：在適性前測後，實驗組依據系統報告診斷書之補救重點，進行補救教學；對照組依據考卷的題目，進行全班補救教學。

五、適性後測(45 分鐘)：兩組受試者在補救教學完畢後，以 BNAT 系統進行適性後測，測驗完畢後由系統產生學習診斷報告書，畫面如圖像 5。

六、實驗變因設計：本實驗教師皆為具豐富教學經驗之數學教師，實驗變因設計如表 2。

表2 實驗變因設計表

組別	實驗變因－教學方法	補救方法	實驗控制
實驗組	數位指導教材	系統報告診斷書	相同的九年一貫能力指標
對照組	傳統紙本教材	測驗試卷檢討	

4. 研究結果

4.1 實驗組與對照組之教學成效與補救教學成效

本節主要在探討使用相同單元的前提下，比較實驗組與對照組的起點行為檢測和適性前測間之教學成效，與適性前測和適性後測間的補救教學成效。

一、教學成效分析：以起點行為檢測的成績為共變量，組別為固定因子，適性前測成績為依變數，顯著水準設為 0.05，分析結果如表 3 所示，實驗組與對照組學習成效具有顯著差異 ($P=0.000<0.05$)。由表 4 可知實驗組進步分數優於對照組，即實驗組之教學成效優於對照組。

表3 共變數分析摘要表

依變數	來源	自由度	F檢定	顯著性
適性前測	起點測	1	36.560	.000
	組別	1	16.389	.000
	誤差	99		
	總和	102		

表4 實驗組與對照組之平均數比較表

分組	平均數		進步	個數
	起點行為檢測	適性前測		
實驗組	50.0	62.4	12.4	52
對照組	49.3	51.7	2.4	50

二、補救教學成效分析：以適性前測成績為共變量，組別為固定因子，適性後測成績為依變數，顯著水準設為 0.05，分析結果如表 5 所示，實驗組與對照組學習成效具有顯著差異 ($P=.006<0.05$)。由表 6 可知實驗組進步分數優於對照組，即實驗組之補救成效優於對照組。

表5 共變數分析摘要表

依變數	來源	自由度	F檢定	顯著性
適性後測	前測	1	297.501	.000
	組別	1	7.787	.006
	誤差	99		
	總和	102		

表6 實驗組與對照組之平均數比較表

分組	平均數		進步	個數
	適性前測	適性後測		
實驗組	62.4	66.0	3.6	52
對照組	51.7	53.0	1.3	50

4.2 數位教學模式之教學成效與補救教學成效

本節主要在探討實驗組在不同能力組別進行教學前後成績及補救教學前後成績之比較。

一、數位教學模式對不同能力學生教學成效分析：將實驗組受試者依起點測驗的成績分為三組，前 33%當作高分組，後 33%當作低分組，其餘為中分組，來比較實驗組在不同能力組別進行教學前後成績之比較，比較結果如表 8。由表 8 顯示各能力組學生適性前測的平均分數皆比起點測的平均分數進步，而在中低能力分組由 t 檢定結果顯示起點測驗與適性前測具有顯著差異，尤以低分組學生進步達 20 分以上，表現最佳。實驗數據顯示數位教學模式在教學成效上，對於中低分組學生的學習能達到顯著有效的助益。

表8 受試實驗組學生起點、前測t檢定

分組	平均數	標準差	t	自由度	顯著性
高分組（起點-前測）	-3.705	10.269	-1.488	16	.156
中分組（起點-前測）	-12.062	16.778	-2.876	15	.012
低分組（起點-前測）	-20.684	18.009	-5.006	18	.000

二、數位教學模式對不同能力學生補救教學成效分析：將實驗組的受試者依適性前測的成績分為三組，前 33%當作高分組，後 33%當作低分組，其餘為中分組，來比較實驗組在不同能力組別進行補救教學前後的成績比較，結果如表 9。由表 9 顯示各能力組學生適性後測的平均分數皆比適性前測的平均分數進步，而在中能力分組由 t 檢定結果顯示適性前測與適性後測具有顯著差異，以中分組學生進步達 5 分以上，表現較佳。實驗數據顯示數位教學模式在補救教學成效上，對於中分組學生的學習具有顯著的助益。

表9 受試實驗組學生前測、後測t檢定

分組	平均數	標準差	t	自由度	顯著性
實驗組高分組（前測-後測）	-2.705	8.052	-1.385	16	.185
實驗組中分組（前測-後測）	-5.750	6.698	-3.434	15	.004
實驗組低分組（前測-後測）	-2.526	8.126	-1.355	18	.192

三、綜合以上分析，因為高分組學生的成績較高，進步空間有限，故數據顯示不管在教學成效或補救教學成效皆較為不顯著；中分組學生的教學成效或補救教學成效皆為顯著，進步的分數至少都在 5 分以上；低分組學生的教學成效則較為顯著，進步成績達 20 分以上，但在補救教學上可能受限於學生理解能力或四則運算能力，導致補救教學成效較不具顯著。

5. 結論與建議

5.1 結論

本研究以結合知識結構與貝氏網路理論之電腦化適性診斷測驗為評量工具，發展數學一元一次式數位教學模式。經教材開發、教學實驗與結果分析，得到以下結論：

第一，針對數位教學模式與傳統教學模式之教學成效而言，實驗組的成績顯著優於對照組的成績，顯示在以貝氏網路為基礎，使用數位指導教材及電腦化適性診斷測驗的教學成效較佳。

第二，針對數位教學模式與傳統教學模式之補救教學成效而言，實驗組的成績顯著優於對照組的成績，顯示在以貝氏網路為基礎，使用數位指導教材及電腦化適性診斷測驗的補救教學成效較佳。

第三，就能力分組表現而言，在兩者教學模式的比較，顯示以貝氏網路為基礎，使用數位指導教材及電腦化適性診斷測驗的教學成效，中低能力組學生上有顯著的助益；而在補救教學成效上，對中能力組學生有顯著的助益。

5.2 建議

本研究根據研究設計與結果提出下列幾點建議：

第一，數位指導教材確實能有效提升學生學習的興趣與成效，且可有效提昇教學效率，建議教師可多善用電腦多媒體編製教材進行數位化教學。

第二，根據研究結果，數位教學模式在「一元一次式」單元的實施成效顯著，確實可提供教學及補救教學的實施。建議未來可嘗試將此模式進一步推廣至數位平台上，提供學校或補習教育的使用。

第三，未來可以進一步探討此模式在數學領域其他單元內容實施之成效，以擴大應用層面，增加系統的實用性。

5.3. 致謝

感謝國立臺中教育大學「團班教學和個別指導之教材與評量以及其相關行政管理系統的建置研究」整合型計畫。

參考文獻

- 李淑娟(2006)。符合SCORM2004與試題結構理論之電腦化適性測驗系統及動畫補救教學模組-以國小數學領域五年級能力指標幾何為例。亞洲大學資訊工程學系在職進修教學碩士論文，未出版，台中縣。
- 何政翰(2005)。國小數學領域電腦適性化測驗系統之建製。國立台中師範學院數學教育學系碩士班碩士論文，未出版，台中市。
- 施淑娟(2006)。應用貝氏網路進行國小五年級「小數」單元學習診斷之研究。國立台灣師範大學教育心理與輔導研究所博士論文，未出版，台北市。
- 郭伯臣、謝友振、張峻豪、蔡坤穎(2005年5月6-7日)。以結構理論為基礎之適性測驗與適性補救教學線上系統。台灣數位學習發展研討會，國立台灣師範大學。
- 陳怡如、吳慧珉、黃碧雲(2004)。電腦化適性診斷測驗之研究。《測驗統計年刊》，12輯，61-61。
- 教育部(2003)。《國民中小學九年一貫課程綱要數學學習領域》。台北市：教育部。
- 楊智為、劉育隆、楊晉民、曾彥鈞(2006)。結合試題順序理論與貝氏網路之電腦適性測驗演算法之探究。第十一屆人工智慧與應用研討會，高雄應用大學。

- 王如敏(2004)。國二學生解一元一次方程式錯誤類型分析研究。國立高雄師範大學數學系碩士學位論文，未出版，高雄市。
- 謝和秀(2001)。國一學生文字符號概念及代數文字題之解題研究。國立高雄師範大學數學系教學碩士學位論文，未出版，高雄市。
- 邱濱文(2004)。模糊理論在學習診斷資訊系統之應用-以國小學童解未知數解題程序的錯誤類型為例。國立新竹師範學院進修暨推廣部教師在職進修數理研究所數理教育碩士班數學組碩士學位論文，未出版，新竹市。
- 蕭宇欽(2007)。高雄地區七年級學生解一元一次方程式迷思概念之分析研究。國立高雄師範大學數學系教學碩士學位論文，未出版，高雄市。
- Airasian, P. W, & Bart, W. M. (1973). Ordering theory: A new and useful measurement model, *Educational Technology*, 56-60.
- Appleby, J., Samuels, P., & Treasure-Jones, T. (1997). Dianosys- A Knowledge-based Diagnostic Test of Basic Mathematical Skills, *Computers and Education*, 28(2), 113-131.
- Stuart Russell & Peter Norvig (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall.
- Takeya(1991).New item structure theorem. Tokyo: Waseda University.
- Lee, J. "Diagnosis of bugs in multi-column subtraction using Bayesian networks,". Unpublished Ph.D. Thesis , Columbia University.2003.
- Vomlel, J. "Building adaptive tests using Bayesian networks," *Kybernetika*—Volume40 (2004) ,Number 3,p.333-348

遊戲互動操作式題型與劇情導向式評量題型之比較

—以國中理化「溫度與熱」為例

梁家銘、曾憲雄*、翁瑞鋒**

國立交通大學理學院網路學習學程

mypearl.n196g@g2.nctu.edu.tw

國立交通大學資訊工程學系/亞洲大學資訊工程學系*

sstseng@cs.nctu.edu.tw

國立交通大學資訊工程學系**

roy@cis.nctu.edu.tw

【摘要】本研究主要是探討遊戲互動操作式題型與劇情導向式評量題型做評量面向比較。本研究所設計之題型，主要是藉由遊戲情境的互動設計，以操作為導向，並透過 Modified Multi-stage Graph 模組達到評量問題解決能力的目的。本研究使用準實驗研究法，研究對象為國中理化老師共 5 人，針對遊戲互動操作式題型與六大學習網中物理科的劇情導向式評量題型做比較。研究結果顯示，教師認為互動操作式題型可以提高參與評量的意願，經由討論分析，遊戲互動操作式題型比傳統紙筆評量在技能及情意面向上有顯著的差異，而遊戲互動操作式題型亦比劇情導向式評量題型在技能面向上有顯著差異。

【關鍵詞】遊戲互動操作式題型、劇情導向式評量題型、實作評量

Abstract: This paper compared several assessment properties of the game-based interactive test item and scenario-based test item. The Modified Multi-stage Graph model is applied to show the different interactive operations between two types of test items. In the experiment, five high school Physics teachers are involved. The understandability and effectiveness of two types of items are evaluated. The result shows that the game-based interactive test item has better understandability and effectiveness.

Keywords: game-based interactive test item, scenario-based test item, performance-based assessment

1.前言

國中自然領域中包含了許多知識性的概念，如力學、熱傳導、光學等，這些知識在教育部六大學習網中有相當多的資料可供參考，若「評量內容」仍是以傳統紙筆測驗的選擇題的方式，較難評斷學生是否會將知識性的概念運用在解決問題上。

問題解決能力不容易以傳統選擇題測驗方式加以評量，而實作評量（Performance Assessment）則具有此功能，唯實作評量在時間、評分標準上有實施上的困難（Wolfe & Miller, 1997），所以在一般課堂上操作不易。因實作評量費時且評量方式不易制定，故在執行上有相當的困難度，若能設計出具有實作評量的效果又能省時，則能提供教師另一個評量選擇。

賴忠良與黃天佑（民 95）研究發現遊戲式評量可以藉由遊戲的趣味性提高學生參與評量的意願，讓學生在遊戲的互動過程中可以得到樂趣。本研究希望讓學生可以從遊戲的操作過程了解自己是否能應用指標能力來解決遊戲情境中出現的問題，並比較遊戲互動操作式題型、傳統紙筆測驗題型及劇情導向式題型之題意描述與測驗評量之效果。

參考認知師徒制裡的理論，老師為了要評量學生的認知程度，因此要設計一連串的實做問題來給學生解決，但是實際問題情境在編輯時，如何對應到背後的評量 model 是很難的。本研究希望以真實情境設計遊戲互動操作式題型，並導入 MMG 的概念，將評量模式模組化，

以改善實作評量的缺點，並達到評量問題解決能力的目標。學習情境的設計，主要是透過改良式的多階段圖表模型(Modified Multi-stage Graph, 簡稱 MMG), 規劃遊戲式學習平台裡各種動作與環境、動作與動作間之相互因果順序與限制的關係。本研究之學習內容是以六大學習網中物理課程之相關數位教材，利用 Scratch 做為評量設計的平台。透過準研究設計，探討遊戲互動操作式題型、傳統紙筆測驗題型及劇情導向式題型之題意描述與測驗評量效果之比較。

研究結果發現遊戲互動操作式題型能將所學的知識概念應用在題目情境中，且因屬操作式的平台，所以在情境及技能應用上有較佳的表現；而六大學習網的劇情導向式評量題型大多是以選擇題的方式呈現，題目在情境上的流暢性較佳；而傳統紙筆測驗因屬文字描述，除知識層面較易發揮效用外，情境及技能的應用皆難以呈現。

2.文獻探討

2.1 評量的重要

評量是教學過程中非常重要的一環，「評量（考試）引導教學」，更是中外皆有的現象。雖然教學、評量也是廣義課程的一部份，但真正決定或引導課程改革成效的，教學、評量卻是重要的影響因素之一。是以在談課程改革的同時，決不能將改革的力量，停留在課程設計本身的表面層次，而是要將改革植基在實際的教學活動，並深化在評量的運用上，方能帶動課程改革的效果（教育部，2001）

2.2 實作評量

評量可以檢示學習者的學習成果，讓教學者藉此採取補救措施或改變教學策略，以提昇教學效果，唯因教學現場在時間、空間及社會風氣的影響，評量內容偏向以傳統紙筆測驗為主，但傳統評量的試題容易出現缺乏彈性、內容窄化、且只評量到片段的知識，評量過程易被忽略，在此提出另一可行的評量稱為「實作評量」（蔡清田，2000）。

實作評量是一種非紙筆測驗的評量類型，評量內容是以學生實際完成一件特定任務或工作，並針對其表現而給予評分。

表格 1 實作評量依測驗情境真實度的高低分為五種類型：（Gronlund,1993）

類型	例子
1.紙筆表現（paper-and-pencil performance）	完成一份實驗報告
2.辨認測驗（identification test）	辨認物理實驗中的可變變因及應變變因
3.結構化表現測驗（structured performance test）	依說明的步驟操作一個實驗
4.模擬表現（stimulated performance）	使用電腦模擬物理實驗，利用模擬系統，將實驗工具佈置好，並進行模擬實驗
5.工作樣本（work sample）	在實驗室中進行實驗

2.3 數位遊戲式學習

Hogle（1996）認為遊戲在教育上有多項優點。（一）可刺激動機及興趣：遊戲元件的好奇與期望可提高學習者的動機與興趣。學習者面臨遊戲的挑戰時，能願意不斷的嘗試，以獲得成就感。（二）提昇記憶能力：遊戲比傳統學習更可強化曾學習過的知識記憶。（三）提

供練習及回饋：針對不瞭解的課程內容可以讓學習者反覆的操作，藉此以澄清迷思概念，並獲得即時的回饋。（四）提昇高層次能力：成功的遊戲經驗可提昇學習者的批判性思考及問題解決能力。

2.4 認知師徒制的學習理論

Collins (1988) 認為學習者長期處在某個真實的情境中，跟隨一位或數位教師，如同過去徒弟跟隨師傅一樣，從旁深入地觀察、模仿師傅的處理問題的經驗與行為外，學習者也經由觀察得到處理問題的經驗，了解完整的情境脈絡，進而發展出新的知識與技能。在自然的情境，學習者會經歷許多不同的狀況，逐漸熟練地發現困難與問題，並應用所學的各式知識來尋找線索、測試自己的假設，以解決所遭遇的問題。而遊戲互動操作式評量即是參考此理論，依學生階段式回答模式來進行評量。

3. 研究方法

本研究使用準實驗研究法，研究對象為新竹市某國中理化老師共 5 人，請 5 位教師針對本研究設計之遊戲互動操作式題型與六大學習網中物理科的劇情導向式評量題型做比較分析，並針對平台的遊戲性、概念分析做討論。

本研究所用之工具如下所列：

3.1 遊戲互動操作式題型

(1).以 Scratch 為設計平台，將能力指標轉化成操作型測驗內容，並以 MMG 模組分析學生解題過程，依據操作結果給予評語。評量的內容採取互動式遊戲設計，如圖 1、圖 2 所示。

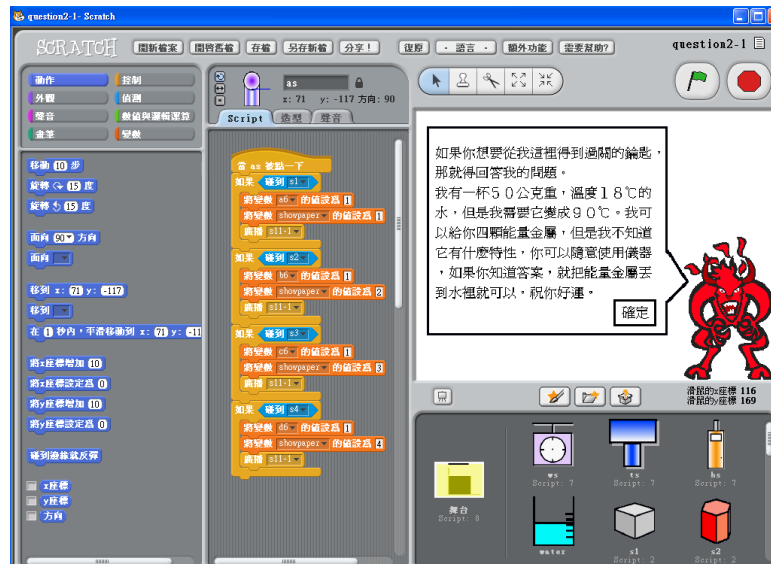


圖 1 Scratch 設計畫面圖

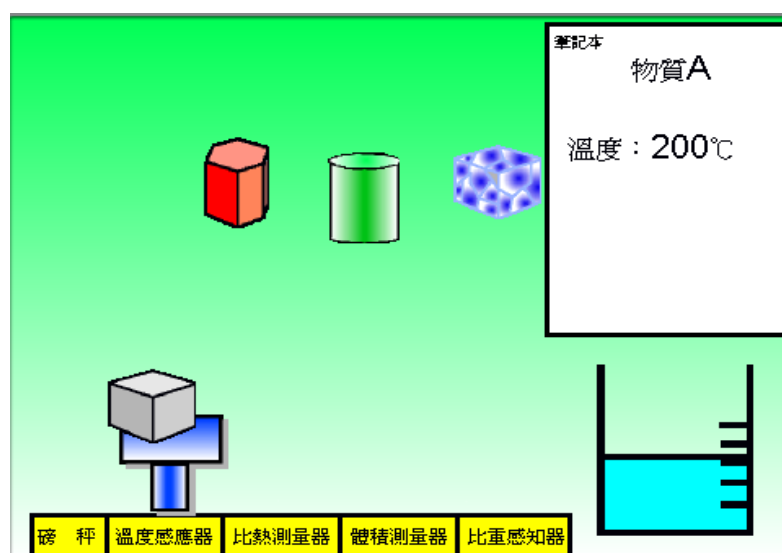


圖 2 遊戲互動操作式題型執行畫面

(2).遊戲情境內容設計：以火場逃生作為評量腳本主題，不同的關卡提供各種工具做選擇，操作不同的工具做為解題策略，預先設定有幾種組合模式是可以過關的

(3).遊戲互動操作式題型之內部測驗流程設計工具（MMG）

在遊戲互動操作式題型之測驗平台內，各科學習情境的設計，主要是透過改良式的多階段圖表模型(Modified Multi-stage Graph, 簡稱 MMG)，規劃遊戲式學習平台裡各種動作與環境、動作與動作間之相互因果順序與限制的關係。在遊戲的平台後端，期望能系統化地觀察學生在學習環境內，問題探究的歷程上，但在一般的網頁式學習環境內，通常較不容易紀錄與了解學生的學習行為模式，但若利用遊戲背後的階段式任務設計，可以解讀學生採取的學習策略。在此我們將互動式學習平台設計為 MMG，其中每個階段由多個節點組成，以節點表示動作，節點和節點之間由有向箭頭，表示階段間的動作順序，以及節點間產生的連帶影響或因果關係，這是一個預先設定好的架構，可用來設計每個階段的環境和動作。MMG 圖例如下：

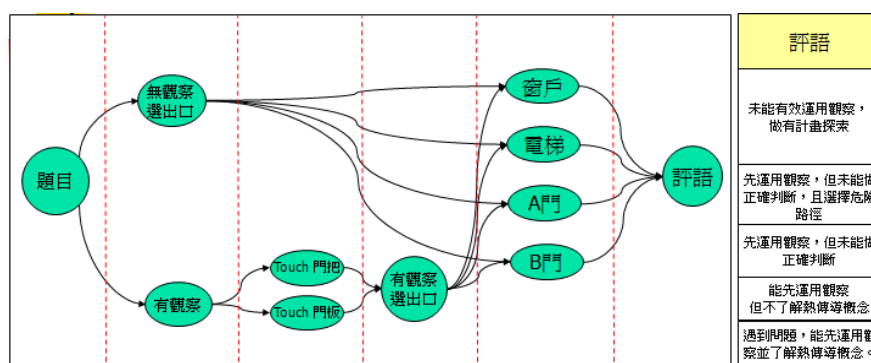


圖 3 MMG 評量模組圖

將不同任務種類需要哪些操作，透過模組化的 MMG 來組合出來（所以先把 MMG 設計成一顆顆可以組合的積木），然後將元件對應上每個動作，以及動作後元件的狀態要有何變化，將以上工作搭配過場動畫，設計一個腳本來描述，之後希望可以透過 wiki 的介面來輔助提供多人共同設計。

(4). 相關指標如下所列：

表格 2 問題解決相關能力指標

過程技能能力指標-歸納、研判與推斷

1-4-4-1 藉由資料、情境傳來的訊息，形成可試驗的假設。

科學與技術認知-交互作用的認識

2-4-5-6 認識聲音、光的性質，探討波動現象及人對訊息的感受。

2-4-5-7 觀察力的作用與傳動現象，察覺力能引發轉動、移動的效果。以及探討流體受力傳動的情形。

思考智能-推論思考

6-4-2-2 依現有理論，運用演繹推理，推斷應發生的事。

設計與製作

8-4-0-4 設計解決問題的步驟

3.2 理化科傳統紙筆測驗

測驗範圍為二年級理化-溫度與熱，試題為是非題及單選題共 30 題，測驗內容以符合九年一貫自然與生活科技課程綱要的能力指標為依據。例題：一杯質量為 200g、溫度為 60°C 的水放在空氣中降溫，若不考慮水蒸發時質量之微小變化，當水溫降到 40°C 時，約散失多少卡的热量？(A) 16000 卡 (B) 8000 卡 (C) 4000 卡 (D) 2000 卡。(國民中學學生基本學力測驗推動工作委員會，97)

3.3 六大學習網的劇情導向式評量題型

六大學習網的劇情導向式評量題型主要是將傳統紙筆測驗的試題數位化後，再以故事情境將試題串連起來，以增加趣味性，而在試題數位化後，可使效率提高，節省大量的人力及時間成本。遊戲畫面如圖 4 所示。



圖 4 六大學習網的劇情導向式評量題型

4. 研究結果與討論

4.1 質化研究

本節主要透過訪談方式，整理歸納遊戲互動操作式題型、傳統紙筆測驗題型及劇情導向式題型之題意描述與測驗評量效果之比較如下表。

評量種類	題意描述			測驗評量效果比較
	知識	情意	技能	
遊戲互動操作式題型	中	高	高	試題的設計由學科概念而來，重視學生解決問題的歷程，強調將知識應用在問題解決能力的評量；情境設計與真實生活產生關聯，避免學習本身與生活無關；遊戲性的闖關設計，能提昇學生參與評量的意願
傳統紙筆測驗	高	低	低	紙筆測驗在編輯試題、施測、閱卷及評分過程中花費較高的成本，紙筆測驗與實作評量不同，它無法以操作的方式施測，故在問題解決能力評量方面較弱，且故事式的文字描述會造成題目冗長，產生閱讀困難的問題。
劇情導向式題型	中	高	低	將傳統紙筆測驗的試題數位化後，再以故事情境將試題串連起來，以增加趣味性，且在試題數位化後，可使效率提高，節省大量的人力及時間成本。但存在著和紙筆測驗相同的問題，即在問題解決能力評量方面較弱，且有無法略過施測試題的問題。

4.2 使用者問卷結果量化分析與討論

使用者問卷主要是針對受試者在使用遊戲互動式操作題型後，對評量概念內容、系統操作是否上手、使用意願等之看法做探討。參加受試者為 5 位教師，本量表係採用 Likert 式五分量表，受試者分別由「非常同意」、「同意」、「無意見」、「不同意」、「非常不同意」五個選項以評定本問卷，填答者依其對不同敘述之不同贊成程度以作勾選。問卷結果如下：

問 卷 題 目	非常同意	同意	無意見	不同意	非常不同意
1 我覺得這一個遊戲評量的內容活潑生動？	20%	60%	20%	0%	0%
2.我覺得這一個遊戲評量很容易操作？	0%	20%	60%	20%	0%

3.我喜歡這種用遊戲來測驗物理科目的評量方式？	40%	60%	0%	0%	0%
4.如果可以，下次我願意再使用這一個遊戲評量進行測驗？	20%	80%	0%	0%	0%
5.我覺得在遊戲的過程中，我可以應用學校課程學到的知識？	80%	20%	0%	0%	0%
6.我覺得像這樣以遊戲的方式來測驗，比較不會有壓力？	0%	60%	20%	20%	0%
7.在遊戲中，我很清楚該用哪些知識才能順利過關？	80%	20%	0%	0%	0%
8.我希望以後在學校各科課程中，能夠讓我們使用這一個遊戲評量？	20%	60%	20%	0%	0%
9.我很希望能和同學討論及分享過關的技巧和秘訣？	20%	20%	60%	0%	0%
10.我覺得如果能和同學比賽過關，整個評量過程會更有趣？	60%	40%	0%	0%	0%
11.在遊戲過程中，我很努力解決問題以便能順利過關？	60%	40%	0%	0%	0%
12.整體而言，我覺得這一個遊戲評量對我的學習很有幫助？	20%	60%	20%	0%	0%

因為目前研究結果僅做前導研究，以做為平台修正分析及後續之研究參考，參與討論教師認為遊戲互動操作式題型能將所學的知識概念應用在題目情境中，且因屬操作式的平台，所以在技能應用上有較佳的表現，而六大學習網的劇情導向式評量題型大多是以選擇題的方式呈現，題目在畫面及情境上的流暢性較佳，另外，部分教師認為，本研究所設計的互動式遊戲在操作上有較高的難度，推論應該是此平台屬較新穎且非線性的評量方式，在操作不熟悉的情況下，容易出現不知道下一步要做什麼的問題，未來在版本修正上會增列「操作說明」及「問題解說」等選項，以便讓初學者更易上手，減少評量誤差。

5.結論與建議

「資訊融入教學」是教育部近幾年推動的主要方針之一，所以各種資訊化的教學法因應而生，但目前常見使用的評量還是以紙筆測驗為主，本研究所設計之遊戲互動操作式題型則提供了另一種不同的選擇，除了可寓教於樂外，更可以讓學生在遊戲情境中操作解題過程，將所學過的知識應用運用在解決問題上。

本研究進一步將會根據參與研究教師所以提供之建議，做系統面的修改及概念融入評量的研究，除此之外，本研究之相關平台及製作的模組將會建構在 wiki 介面，提供教師共同研究創作的工具，以期能讓評量能有更多面向。

參考文獻

教育部(2001)。成績評量準則分析。

國民中學學生基本學力測驗推動工作委員會(2008)。自然科基本學力測驗試題。

蔡清田(2000)。教育行動研究。台北：五南。

賴忠良、黃天佑(民 95 年 9 月)。遊戲式網路評量-網路尋寶系統實作。郭譽申(主持人)，E-Learning。第十七屆物件導向技術與應用研討會，長庚大學。

- Collins, A., Brown, J. S., & Newman, S. E.(1988), Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Cognition and instruction: Issues and agendas*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gronlund, N. E. (1993). *How to make achievement tests and measurements* (5th ed.). NeedhamHeights, MA: Allyn and Bacon.
- Jan G. Hogle(1996),*Considering Games as Cognitive Tools: In Search of Effective "Edutainment"*, University of Georgia Department of Instructional Technology.
- Wolfe, E.W., & Miller, T.R.(1997). Barriers to the implementation of the portfolio assessment in secondary education. *Applied Measurement in Educational*, 10(3), 235-251.

「分數的乘法」數學單元之建構反應題電腦化測驗

The Computer-Based Test for Constructed-Response Items of “Fraction Multiplication”.

吳任婕、鄭涵*、施淑娟**、郭伯臣***

臺中教育大學教育測驗統計研究所

臺中教育大學教育測驗統計研究所*

臺中教育大學數學教育系**

臺中教育大學教育測驗統計研究所***

【摘要】本研究嘗試開發數學領域「分數的乘法」之建構反應題電腦化測驗，藉由電腦紀錄學生詳細的解題過程，期望得到學生更多元之迷思概念，並透過專家檢視受試者之解題過程，給予錯誤類型。依據「分數的乘法」數學單元之知識節點命題，設計適合使用電腦化方式的試題，並初步建立試題之錯誤類型。經過實際施測實驗後，本研究發現藉由電腦紀錄學生建構反應題電腦化測驗之解題過程，可以得到學生更多元的錯誤類型，以幫助教師掌握學生的迷思概念及學習狀況。

【關鍵詞】建構反應題、電腦化測驗

Abstract: This research attempt to development the computer-based test system for constructed-response items of “fraction multiplication”. We expect to obtain more misconception of students from problem solving process by computer. And inspect the problem solving process with experts, gives the wrong type. To design the examination based on knowledge of “Fraction Multiplication”, then choices the suit examination with computerization way, and establishes mistake type for the all questions. After the experimentation, we uses the identification way with computer, establishes a set of coding rule on the system, compared the identification by experts and computer distinction. The study result gives: We can obtain more misconception of students from problem solving process by computer.

Keywords: Constructed-Response Item, computer-based test.

1. 研究動機與目的

本研究著力於開發數學領域「分數的乘法」之建構反應題，結合電腦化測驗系統，達到自動化計分與診斷之功能，降低教師閱卷負擔。將「分數的乘法」數學單元之試題電腦化，並藉由解題過程，幫助教師快速找出學生學習盲點。此外，在推動教育改革之際，期望藉由評量之革新來提升教師教學的品質和學生學習的成就。一些大型入學測驗和證照考試，除了選擇題外，增加建構反應題或實作測驗部分。目前國內數學領域之電腦化測驗大多以選擇題或填充題形式呈現，只能紀錄受試者答案，無法詳細判斷學生是猜對答案或同一答案之不同解題歷程。故本研究擬開發電腦化建構反應題（constructed-response）題型，除了能記錄學生所選之答案外，亦可詳細記錄學生之解題與列式歷程、解題時間，及解題所使用之工具（如繪圖題、操作題）等，及精準診斷學生之迷思概念及學習狀態。

基於此研究動機，本研究目的如下：

- (1) 編製數學領域「分數的乘法」之建構反應題電腦化測驗。
- (2) 探討學生在建構反應題電腦化測驗之「分數的乘法」錯誤類型。

2. 文獻探討

為了達到本研究之目的，在相關文獻部分，針對「分數的乘法」教材分析、電腦化測驗兩部分來探討。

2.1. 「分數的乘法」教材分析

「分數」是國小學童最容易產生學習困難的主題之一（呂玉琴，1991；陳靜姿，1999）。因此，本研究選擇此題材進行研究，企圖透過建構反應題之電腦化測驗設計，能更進一步了解學生完整的分數學習表現，本研究單元的範圍係指依據教育部 92 年發佈之「國民中小學九年一貫課程綱要」中數學學習領域第三階段能力指標，所選取之康軒出版社六年級「分數的乘法」單元。此單元需培養的能力指標整理如下表 1。

表 1 「分數的乘法」單元的能力指標

編號	說明
N-3-4	在具體情境中，解決分數乘以分數的問題，進而形成分數倍的概念。
CN-13	補充分數乘法及除法的算則。
CN-14	加強分數的兩步驟四則混合計算。
C-R-2	能察覺數學與其他領域之間有所連結。
C-R-3	能瞭解其他領域中所用到的數學知識與方法。
C-S-4	能運用解題的各種方法：分類、歸納、演繹、推理、推論、類比、分析、變形、一般化、特殊化、模型化、系統化、監控等。
C-S-5	瞭解一數學問題可有不同的解法，並能嘗試不同的解法。
C-C-3	能用一般語言與數學語言說明情境與問題。
C-C-5	用數學語言呈現解題的過程。
C-C-6	用一般語言及數學語言說明解題的過程。
C-C-7	用回應情境、設想特例、估計或不同角度等方式說明或反駁解答的合理性。
C-C-8	能尊重他人解決數學問題的多元想法。

過去已有許多研究透過紙筆測驗及質性研究的方式提出學生在分數乘法之各種解題類型，尤其是學生常見的錯誤類型或迷思概念，提出與本單元相關之錯誤類型研究，其整理如下表 2。

表 2 「分數的乘法」單元相關之錯誤類型研究

錯誤類型	相關研究
林榮煌（2006）	➤ 誤用約分概念： 1. 分母相約分。2. 整數和分母相約分。 3. 整數和分子相約分。
	➤ 誤用擴分概念：整數同時乘上分母和分子。
	➤ 用分數除法運算：乘數轉成倒數。
	➤ 分母相同時，分子相乘，分母不乘。
	➤ 基本概念錯誤
	➤ 以直接法進行運算，忽略是否為同分母。 分母加（減）分母，分子加（減）分子
湯錦雲（2002）	➤ 用加法算則計算：
	1. 分子和整數相加。2. 分母相同時，分子相加。

	<ul style="list-style-type: none"> ➤ 通分後，分子為大數減去小數。 ➤ 借位的錯誤：從整數所借的 1，直接在分子加 10。
陸雅林（2007）	<ul style="list-style-type: none"> ➤ 與加法混淆使用： <ol style="list-style-type: none"> 1. 分母相乘，分子相加。 2. 分母相加，分子相乘。
Lankford（1972）、Edwards（1983）、Painter（1989）、Ashlock（1990）、Souviney（1989）	<p>求出公分母後放在分母，而分子為原分子相加。</p>

2.2. 電腦化測驗

將試題電腦化，可以給學生及時回饋，具有計分快速之優點。藉由電腦能儲存大量訊息之優勢，記錄學生相關作答歷程，可以獲得更豐富知學生作答訊息以更精確的推論學生之相關能力。NetPASS 是一套由 Cisco Networking Academy Program (CNAP) 研發的系統，用於電腦網路設計、結構和偵測錯誤的評量及訓練。為了評量學生是否可以讓網路運作，所以評量的設計必須在深度及廣度上都必須超越其他的線上測驗。此系統可供學生練習電腦網路技巧並從學生作答反應得到回饋，亦可由學生作答反應得到之證據，並可估計及儲存學生多次作答反應之技能及能力。NetPASS 所建立的模擬環境是以完善的心理學為基礎用以支援學習，當學生第一次登入時，可於頁面上選擇範圍（設計、結構或偵測錯誤）及不同的難度等級（簡單、中級、困難）；在學生作答完成後，NetPASS 針對學生的錯誤予以回饋（Behrens, Mislevy, Bauer, Williamson, & Levy, 2004）。因此，本研究將參考 NetPASS 來建置電腦化測驗系統。

3. 研究方法

3.1. 研究流程

蒐集「分數的乘法」單元相關文獻，並分析本單元之能力指標；依照本單元相關錯誤類型文獻及參考實際教學上學生容易具有之迷思概念來建立試題，並開發「分數的乘法」單元之建構反應題電腦化測驗。再者，進行教學實驗，讓學生實際上線施測，透過系統記錄學生的解題過程。專家由系統所呈現之學生解題過程，進行錯誤類型編碼。整個研究流程如圖1所示：

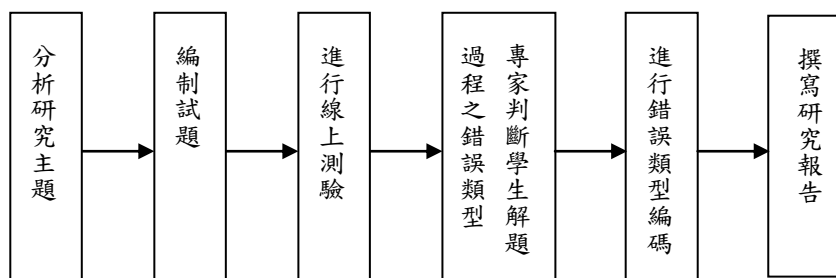


圖1 研究流程圖

3.2 研究對象

施測對象為九十七學年度台中縣共 3 所國小之六年級學生，從中抽取 6 個班級，有效樣本 158 人。

3.3 編製之試題

建構反應題題型之建置流程是先依據知識節點命題，選出每一個實驗單元適合做成建構反應題題型的概念，將這些概念進行分析與命題，命題時以開放式題型為主，以引導學生產出各式各樣之解題歷程與使用工具。再者，建構反應題題型在測驗時，除以電腦紀錄學生解題歷程外，也會隨機挑選學生做錄影，以這些方式去記錄學生所有可能的作答歷程。

本研究編製試題時，依據相關研究之錯誤類型及教學者之教學經驗，表 3 為本研究之錯誤類型分類。

表3 本研究之錯誤類型分類

B01	沒有化成最簡分數。	B02	誤用約分概念。
B03	誤用擴分概念：整數同時乘上分母和分子。	B04	用加法算則計算。
B05	分配率和乘法結合，未以（）標示。	B06	帶分數和假分數轉換的錯誤。
B07	用分數除法運算：乘數轉成倒數。	B08	分母相同時，分子相乘，分母不乘。
B09	與加法混淆使用。	B10	基本概念錯誤。
B11	以直接法進行運算，忽略是否為同分母。	B12	求出公分母後放在分母，而分子為原分子相加。
B13	通分後，分子為大數減去小數。	B14	借位的錯誤。
B15	只完成一部份的計算。例如：整數只跟分母計算。	B16	無法由乘數與 1 的大小關係推論被乘數與積的大小關係。
B17	對於題意的理解錯誤。	B18	沒有先乘除後加減。
B19	沒有使用（）。		

探討學生解題歷程可能包含幾種情形：

- (1) 列式正確，計算正確，答案正確；
- (2) 列式正確，計算錯誤，答案錯誤；
- (3) 列式錯誤，答案正確；
- (4) 列式錯誤，答案錯誤。

如以選擇題型施測，(1) 和 (3) 都是答對，但其實 (3) 可能是猜測而答對應該算答錯，而 (2) (4) 雖然是答案錯誤，但 (2) 應該可部分給分。如以建構反應題題型施測，則可區分此四種情形，避免學生猜題。本研究編製之測驗的試題共 30 題，若學生測驗之作答時間以一小時為限，考慮學生在建構反應題需耗費較長作答時間，且分數的乘法計算也較複雜，經評估後，挑選出須計算的 4 題試題作為建構反應題，故試題分布為 4 題建構反應題及 26 題選擇題。表 4 為選擇題編製之試題範例。

表 4 「分數的乘法」單元之試題範例

編碼	1-2, 2-2, 3-2 帶分數×帶分數應用題（答案化成最簡分數）
題目	() 1. 一箱木瓜重 $12\frac{8}{15}$ 公斤，一箱哈密瓜的重量是木瓜的 $2\frac{3}{4}$ 倍，一箱哈密瓜重多少公斤？（答案需化成最簡分數）

① $24\frac{2}{5}$ ② $9\frac{3}{5}$ ③ $34\frac{14}{30}$ ④ $34\frac{7}{15}$				
選項	選項 1	選項 2	選項 3	選項 4
反應 類型	B10 基本概念錯誤。 $12\frac{8}{15} \times 2\frac{3}{4}$ $= 12 \times 2 + \frac{8}{15} \times \frac{3}{4}$ $= 24\frac{2}{5}$	B06 帶分數和假分數 轉換的錯誤。 $12\frac{8}{15} \times 2\frac{3}{4}$ $= \frac{96}{15} \times \frac{6}{4} = 9\frac{3}{5}$	B01 沒有化成最簡 分數。 $12\frac{8}{15} \times 2\frac{3}{4}$ $= \frac{188}{15} \times \frac{11}{4} = \frac{1034}{30}$ $= 34\frac{14}{30}$	◎

在上述試題的選項設計中，除了正確答案外，其餘選項皆包含一種錯誤類型，同一種錯誤類型，可能有不同之解題歷程，不同解題歷程代表不同之迷思概念。學生即使選了正確的答案，也不代表真正答對。選項 1 是乘法的基本概念錯誤，學生若選擇 1，則表示學生可能犯了「整數乘以整數，分數承以分數」的錯誤類型；而學生若選擇選項 2 表示學生在乘法帶分數轉換為假分數的概念錯誤，可能犯了「忽略分母，整數乘以分子」的錯誤類型；而若學生選擇選項 3，則表示學生概念都清楚，但未化成最簡分數；而選項 4 為正確答案。此題中，學生若選擇選項 3，其實學生大致上是可以解決這個問題的，只是忽略題目應化成最簡分數的要求，雖然答案錯誤，但應可以部份給分。

在「分數的乘法」單元中，學生可能有很多種不同的解題過程，以選項 2 為例，當學生具有「帶分數和假分數轉換的錯誤」之迷思概念時，除「分母不變，整數與分子相乘」外，還有「分子不變，分母與整數相乘」之錯誤類型。若以建構反應題進行測驗，不僅可觀察學生的作答情形，也可以給學生一些實質上分數的回饋，選擇選項 3 的學生可得到部分分數；而題目的選項均設定為題目中的兩個數字相乘，但在實際狀況中，也許學生會有其他不同的作答反應，若學生在選項中未看到與自己算出來的相符答案，學生可能用猜的，這樣就不是教師所想要的錯誤類型了，若能利用建構反應題，將可以更仔細的知道學生之迷思概念及學習狀況。

從這個例子可以看出建構反應題題型能記錄學生更多解題訊息，提供更多證據推論學生的能力。但是，編製建構反應題題型需要比較耗費較多的時間與人力成本，其流程也較一般紙筆測驗或選擇題題型的電腦化測驗繁瑣，下面簡述建構反應題電腦化測驗之建置流程（圖 2）與系統介面說明。

3.3.1 建構反應題題型介面說明

若欲使用電腦化測驗評量「分數的乘法」的應用問題時，選擇題的形式是較常見的，選擇題選項雖有經過設計，但選項也有其對應之迷思概念，而學生選擇錯誤選項有兩種情況，可能是犯此種錯誤類型，也可能是用猜的；且選擇題無法記錄學生之解答過程。若使用建構反應題題型，學生的解題過程及最後答案是否正確，皆可藉由電腦直接記錄，提供更多訊息，作為判讀學生迷思概念的參考，以下為「分數的乘法」適用於建構反應題型之介面操作說明，如圖 3 所示。「分數的乘法」例題解題歷程，如圖 4 所示。



圖3 「分數的乘法」試題介面

(1) 工具說明 I：分數顯示狀態

輸入：點選後輸入的分數則會顯示在左側紀錄欄內。

算式重整：左側紀錄欄若無顯示輸入的值，可點選算式重整，可將紀錄欄顯示更新。

清空整數：將輸入的整數清除。

清空分母：將輸入的分母清除。

清空分子：將輸入的分子清除。

(2) 工具說明 II：符號區

+ **-** **x** **=**：輸入後在左側紀錄欄內顯示運算符號。

(**)**：輸入後在左側紀錄欄內顯示左括號、右括號。

算式重新整理：左側紀錄欄若無顯示輸入的值，可點選算式重整，可將紀錄欄顯示更新。

下一行算式：左側紀錄欄一行算式紀錄完成後，點選下一行算式可換行。

清除算式：將左側紀錄欄內的紀錄全部清空。



圖4 「分數的乘法」解題歷程例題

(3) 系統畫面說明

學生使用視窗右方的工具列輸入答案，學生的作答歷程呈現在試題下方，系統會記錄學生解題歷程於資料庫中（表 5）。其中，「frac」是代表分數，「\\」是代表前面的數字連著後面的分數，「{x}{y}」是代表分數的分子與分母，前者為分子，後者為分母；舉例來說，

「12\\frac{8}{15}」代表為 $12\frac{8}{15}$ ，「\\frac{15}{1}」代表為 $\frac{15}{1}$ 。

表 5 資料庫所紀錄之學生解題歷程

姓名	編號	資料庫所紀錄之學生解題歷程
徐慧軒	06468460212	12\\frac{8}{15}* 1\\frac{1}{4}
黃薇庭	06468460213	12\\frac{8}{15}* 1\\frac{1}{4} = 19\\frac{1}{2}
林鈺傑	06468460214	12\\frac{8}{15}* 1\\frac{1}{4} = \\frac{15}{1}
陳宜卉	06468460215	12\\frac{8}{15}* 1\\frac{1}{4} = 15\\frac{}{}
易庭莉	06468460216	12\\frac{8}{15}* 1\\frac{1}{4} = 2\\frac{8}{3}

4. 研究結果

4.1 學生解題過程之錯誤類型分析

經過實際測驗後，透過專家判別將學生解題過程之錯誤類型進行分析，以圖4例題為例，全體學生在此題共產生9個錯誤類型，其數量統計表如表6，與選擇題型相較之下，更可清楚學生之迷思概念。從解題過程，可以清楚得知學生學習上之迷思概念，學生的最後答案已不是唯一判斷的依據，透過建構反應題，可讓教師除了知道學生答錯，可更清楚知道學生錯在哪裡，針對全班所犯較多之錯誤類型進行補強。以圖4例題為例，此題有66位學生答對，20位學生空白，同時具備兩個錯誤類型以上的學生大約5位；其中，B20和B21是在原先預設之錯誤類型中未出現的，錯誤類型B20為學生觀念正確，但在計算上有迷思概念而不能答對，但因學生只寫下列式及答案，未多列出詳細過程，以致於研究上無法細分學生於哪種計算錯誤之錯誤類型，統整列於錯誤類型B20。錯誤類型B21為學生將此題以加法計算，但在分數的加法

上，又有出現錯誤計算，例如： $12\frac{8}{15} + 1\frac{1}{4} = \frac{21}{31}$ ，其中， $21 = 12 + 8 + 1$ ， $31 = 12 + 15 + 4$ 。

表 6 第 27 題之全部學生解題錯誤類型數量統計表

編號	錯誤類型	學生犯錯誤類型之數量
B01	沒有化成最簡分數	2
B04	用加法算則計算	1
B06	帶分數和假分數轉換的錯誤	3
B09	與加法混淆使用	5
B11	以直接法進行運算，忽略是否為同分母。	4
B15	只完成一部份的計算。例如：整數只跟分母計算	17
B17	對於題意的理解錯誤	18
B20	列式正確，計算錯誤	23
B21	加法概念錯誤	4

5 結論與建議

5.1 結果

透過上述的研究，建構反應題電腦化測驗除了可以清楚記錄學生之解題過程，也可以讓教師清楚學生之迷思概念，不僅可得知學生更多元之錯誤類型，亦可以降低教師閱卷的時間，並可以提供給教師補救教學的依據。與傳統紙筆測驗及選擇題相較，建構反應題能較精確得到學生之迷思概念；學生在做建構反應題時，因電腦化測驗的方式對學生來說較新穎，學生會覺得以這樣的方式測驗數學是有趣的，在測驗過程中，也許可以降低學生對數學的恐懼感，進而提升學生想要學習數學之興趣。

5.2 建議

建構反應題電腦化測驗之建置需耗費時間與人力，且學生在測驗建構題時需要比較長的測驗時間及熟悉系統介面，所以，除建立系統外，可再配合適性測驗診斷系統及貝氏網路達到省題及精準度之功效。另外，判別錯誤類型之機制，可以利用特徵萃取及資料探勘方式(Data Mining)發展，以達到更高之辨識率。

參考文獻

- 呂玉琴(1991)。分數概念：文獻探討。台北師院學報，4，573-606。
- 林榮煌(2006)。國小六年級學童分數乘除概念與運算錯誤類型之研究。國立台中教育大學教育測驗統計研究所碩士論文，未出版，台中市。
- 湯錦雲(2002)。國小五年級學童分數概念與運算錯誤類型之研究。屏東師範學院數理教育研究所碩士論文。
- 陳靜姿(1999)。國小四年級兒童等值分數瞭解之初探。論文發表於第十五屆科學教育學術研討會暨第二屆科學教育學會年會。國立彰化師範大學。
- Ashlock, R.B.(1990).Error patterns in computation: *A semi-programmed pproach(5thed)*.Columbus, Ohio:merril.
- Behrens J. T., Mislevy R. J., Bauer M, Williamson D. M., and Levy R (2004). Introduction to Evidence Centered Design and Lessons Learned From Its Application in a Global E-learning Program. *The International Journal of Testing*, 4(4), 295-301.
- Lankford, L.K. (1972). Final Report : Some computational Strategies of seventh grade pupils, Charlottesville, VA: University of Virginia. (ERIC Document Reproduction Service No. ED 069 496).
- Painter, R.R(1989). A comparison of the procedural error patters, scores, and other variables, of select ed groups of university and eight-grade students in Mississippi on a test involving arithmetic operations on fraction.(Doctoral dissertation, University of Southem Mississippi, 1988).
- Souviney, R.J. (1989). *Learning to teach mathematics*. Columbus, OH: Merrill. (CC: QA135.5.S68)

高校教学质量监控综合平台的研究与设计

The Research and Design of College Teaching Quality Monitoring Platform

刘英群、韩锡斌、杨娟、张明

清华大学教育技术研究所

{liu-yq, hanxb, juan-yang, zhangmm}@tsinghua.edu.cn

【摘要】 在中国大陆，高校教学质量监控体系方面的研究与实践在不断深入，但在具体实施环节，现有的质量监控信息化系统却无法提供全面的支持。为此，本文提出一个教学质量监控平台，该平台的设计基于教学质量监控体系的研究理论和高校实际需求，能够与 THEOL 教务管理信息系统无缝整合，并支持多种监控模式和监控方法。本文主要介绍了该平台的总体架构，及各子系统的设计思想和功能结构。

【关键词】 质量监控体系、教学事故、教学评价、教学督导

Abstract: Although the research and practice of teaching quality monitoring system has developed rapidly since 1990s in Mainland China, no existing network-based monitoring system can provide a total solution for its application. Thus, this paper proposes a teaching quality monitoring platform. It is designed based on the theory of teaching quality monitoring system and the actually requirements of colleges. It can be integrated with the THEOL educational administration system and supports several monitoring modes and methods. This paper mainly introduces the architecture of the platform and the design of the four subsystems.

Keywords: teaching quality monitoring system, teaching accident, teaching evaluation, teaching supervision

1. 引言

高等教育质量保障是现今高等教育研究领域的热点问题之一，而高校本科教学质量监控则是高等教育质量保障的重要组成部分，是确保高校教育教学质量和人才培养质量的关键。因此各高校对教学质量监控都给予了极大的重视，并成立专门的机构来负责教学质量监控的研究、建设和实施。但是，目前普遍存在的问题是各高校监控制度和组织机构的建设相对完善，而信息化建设稍显滞后。随着教育信息化进程的不断推进，传统教育教学管理进一步向数字化、网络化发展，基于网络的教务管理信息系统得以普遍应用，为教学质量监控的信息化带来了契机。因此，引入信息化手段以提高教学质量监控的效率、范围和质量已势在必行。

2. 教学质量监控体系信息化相关研究及问题分析

所谓高校本科教学质量监控体系，是指高校在一定的教育教学理念和教学目标的指导下，采用一定的手段和方法，对本科教学的全过程进行评价、监督和诊断，对学校的本科人才培养活动实行体制化、结构化、持续化的监控，并且将各个环节、各个部门的活动与职能合理组织起来而形成的一个权限，任务明确又互相协调、互相促进的，能够切实保障和提高本科教学质量的质量管理体系（司俊峰，2007）。教学质量监控体系由监控与评价的对象——客体、监控与评价的实施者和组织者——主体、评价方法与指标、调整机制等要素构成，包括目标的确定、各主要教学环节质量标准的建立、信息的收集整理与分析（统计与测量）、评估、信息反馈、调控等环节（曹洪欣，2007）。教学质量监控客体，包括决定教学质量、

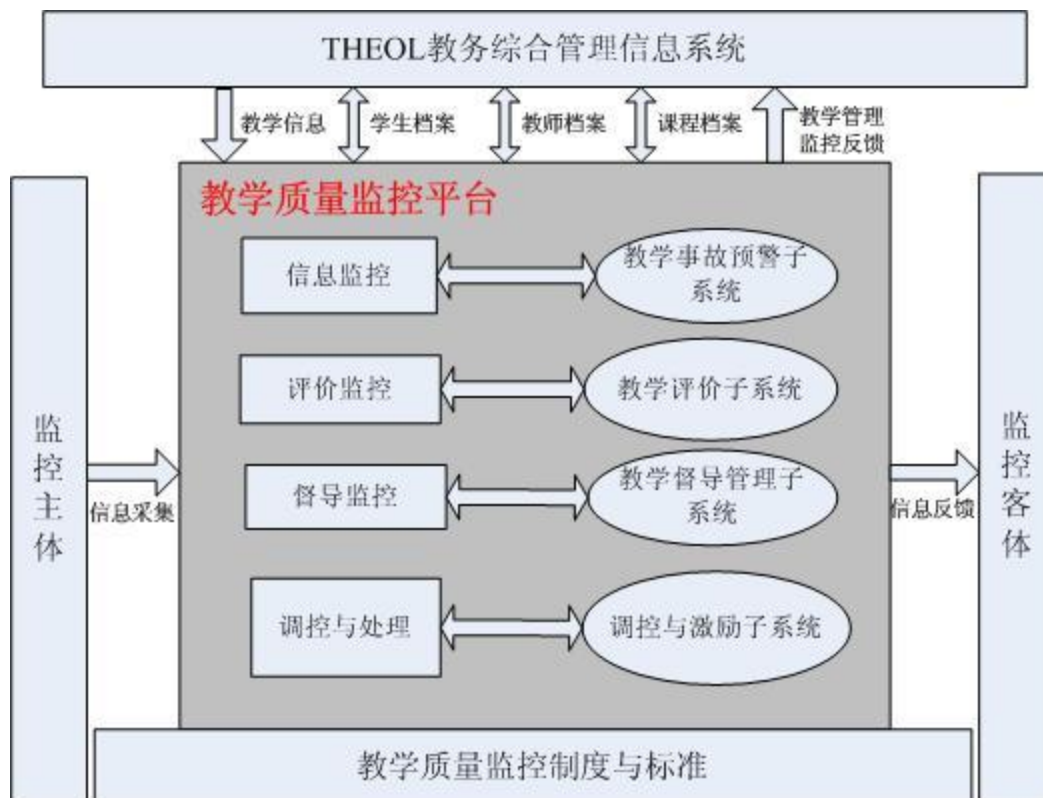
影响教学质量的所有因素，即教学过程所涉及的所有硬件环境与软件要素。教学质量监控的主体就是教学质量监控的实施者和组织者，分为外部监控主体和内部监控主体，其中内部监控主体是由学校、院/系、教研室构成的三级监控组织，包括高校内部的行政职能部门、广大教师员工及学生等，是监控体系中的主要力量（陈建湘和谢再莲，2004）。根据监控方法的不同，可分为信息监控、评价监控和督导监控。根据监控模式的不同，分为集中监控、分散监控和分级监控。

在中国大陆，自 20 世纪 90 年代开始的教学质量监控体系理论研究及实践已经初见成效，但是在具体实施过程中，监控信息化方面还存在诸多问题。有些高校仍然采用传统的信息采集方式，工作量大，效率低。有些高校虽然应用了信息化手段，但是只局限于某个或某几个环节，没有从监控体系的角度统筹规划，如引入教学评价系统（张忠玉、迟国彬和陈品，2006）实现评价监控，引入教学督导系统（何勇、裴晶莹、罗圣国和熊楚强，2007）实现督导监控等。这样会导致监控体系的各个环节出现脱节，使得督导的评价结果无法与教学评价的结果相整合，评价结果也很难起到反馈和调控的作用。此外，在有些高校，同时运行着教务综合管理信息系统和教学评价或督导系统，但是两者却未能有效的整合在一起。

为解决上述问题，我们研究和设计了一个高校教学质量监控平台。该平台以教学质量监控体系的研究理论为指导，结合各高校实际需求，支持多种监控模式和监控方法，能够较好的支持高校教学质量监控体系的实践。

3. 教学质量监控平台总体架构

在高校教学质量监控体系中，监控目标、组织机构、监控制度和监控标准是监控实施的基础，也是监控平台运行的保障。在此基础上，平台涵盖了监控体系中信息收集、诊断、评估、督导、调控和信息反馈等各个环节。以下是平台的总体架构图：



图像 1 平台总体架构图

教学质量监控平台既可以独立运行，也可以和 THEOL 教务管理信息系统（马勃民、程建钢和韩锡斌，2004）无缝整合。平台能够从教务系统中获取课程、教师、学生及教学管

理信息，并将监控结果反馈给教务系统。作为另一种信息采集的方式，平台通过动态的角色设置和灵活的权限控制支持不同级别、不同类型的监控主体在线输入监控信息。监控客体分为系统客体和自定义客体两种，系统客体包括教师、课程、学生、教材、教学设施等，自定义客体可以由用户依照实际需求来设置，平台中的监控结果可以直接反馈给相应的客体。由于监控主体和客体可灵活设定，因此可以较好的支持集中监控、分散监控和分级监控三种不同的监控模式。平台提供教学事故子系统、教学评价子系统（杨红、韩锡斌、程建刚和陈刚，2005）和教学督导管理子系统以分别支持信息监控、评价监控和督导监控。此外，调控与激励子系统实现了对监控结果的调控和处理功能。

总之，平台在架构上具有以下特点：

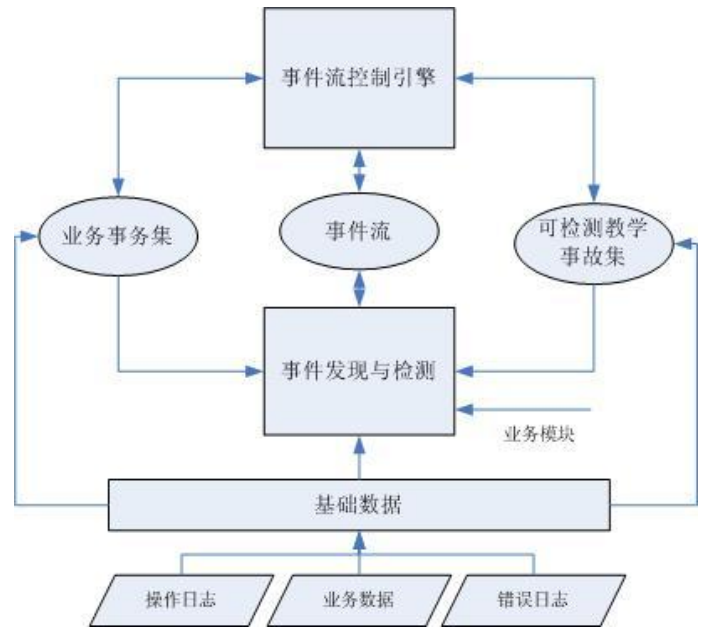
- (1) 支持与教务管理信息系统无缝整合。
- (2) 支持监控体系中信息收集、诊断、评估、督导、调控和信息反馈等各个环节。
- (3) 监控主体和客体可灵活设置，从而支持多级监控和多种监控模式；
- (4) 同时支持信息监控、评价监控和督导监控三种监控方法。

4. 平台子系统的设计

4.1. 教学事故预警子系统

随着教务管理信息系统应用的不断深入，学校的实际教学运行对系统的依赖程度越来越大。因此一旦出现系统使用不当或操作失误，出现管理漏洞，就会对教学过程局部或整体在教学秩序或教学质量方面产生重大影响，严重的甚至会造成教学事故。一般而言，这种漏洞可以通过信息监控的手段协助管理人员及时发现问题，从而在一定程度上减少这种事故的发生。教学事故预警子系统基于 THEOL 本科生教务管理信息系统，能够自动检测系统的异常操作，针对某个业务目标实现面向主题的数据一致性和冲突检查，并通过业务级事务控制实现进度控制和提醒，支持教学事故的分析、诊断和预警。

教学事故的检测可以分为两个步骤：（1）匹配触发条件；（2）启动检测程序，检测数据状态。为了减少预警系统与原业务系统之间的耦合度，我们借鉴事件驱动的架构（EDA）设计思想，提出了一个基于事件流的教学事故预警系统设计框架，如下图所示：



图像 2 教学事故预警系统框架图

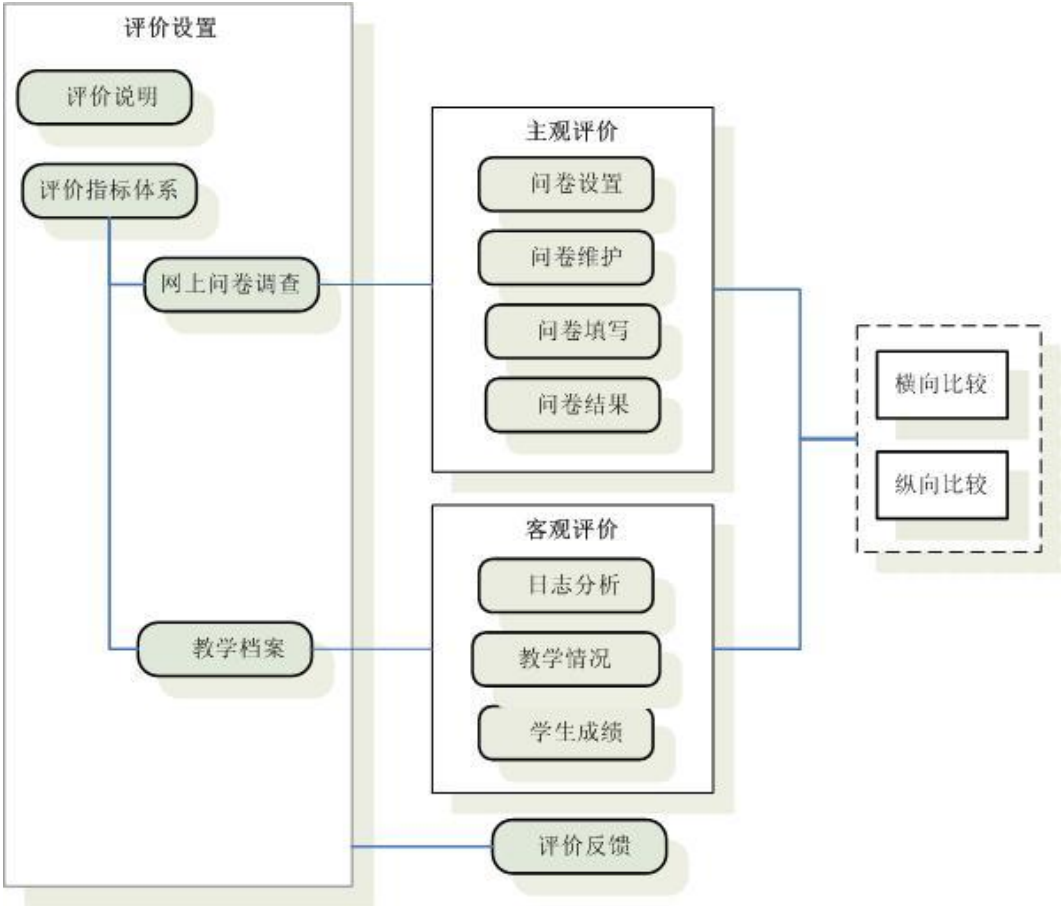
在该框架中，事件发现与监测部件和事件流控制引擎部件是其中的两个核心部件。事件发现与监测部件根据事故的触发条件来检测、接收、过滤和创建事件，依照时间顺序加入到事件流中，并保证事件流中的事件都是有效事件。事件主要来源于三个方面：

- (1) 基础数据：包括操作日志、业务数据和错误日志等。此时事件发现与检测部件是以主动模式，通过数据分析来采集事件。
- (2) 业务功能模块：由用户操作来触发某个事件。此时事件发现与监测部件是以被动模式，接收来自业务功能模块所产生的事件。
- (3) 由事件发现与检测部件根据触发条件自动创建的事件。

事件加入到事件流之后，事件流控制引擎将匹配事故集和事务集，来决定启动哪个事故的检测程序，并维护事件的状态及事故的检测状态。

4.2. 教学评价子系统

在教学评价子系统中，评价客体可以是任课教师、学生、课程、教材等教学过程中的任何软硬件要素，评价主体可以是教师、学生、专家等。根据不同的评价主体，系统可以预置不同的评价导向和指标体系。评价主体也可以查看相关教学档案和教学管理信息，从而做出综合评价。系统的结构框图如图 2 所示：



图像 3 教学评价系统结构框图

(1) 评价设置

评价设置包括评价说明和评价指标。评价说明是一份可公开的评价方案，使得评价透明化，利于监督。评价说明一般应包括评价目的、评价者（包括他们所占的权重）、评价对象、评价内容、评价指标体系、分析方法。评价指标由实施评价的组织者设计，规定必备的评价

内容项目，同时允许评价客体选择或新建具有自己特色的、能体现自己教学特点和风格的评价内容项目。

(2) 主观评价

指采用调查问卷的形式进行非量化的评价。系统首先根据目前先进的评价理念，预置一整套基本的问卷指标体系，用户基于开放性的指标构建体系，可以根据时间、目的、不同学制、不同学科专业、不同教学方式进行调整，建立起适合自己院校特点的评价系统。这样，管理员编制各种评价量表并通过问卷来收集非量化信息，将量化评价与非量化评价有机结合，他人评价与自我评价有机结合，以实现评价客体的全方位评价。

(3) 客观评价

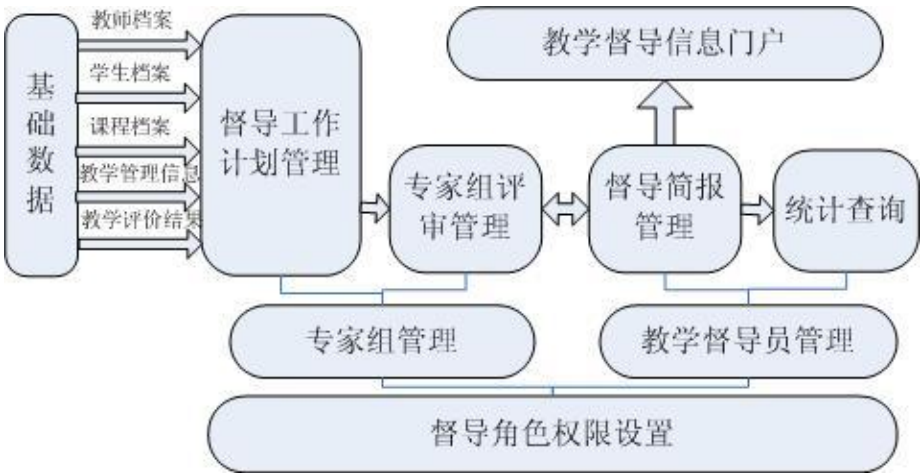
客观评价信息的获取主要来源于教务管理信息系统中收集的教师授课信息、学生成绩信息、选课信息、教学设施的使用情况、教务管理人员的操作日志等。系统会根据指标体系对来自教务系统的原始数据进行统计和处理，并得出评价结果。客观评价是对主观评价的有效补充，能够在一定程度上减轻主观评价中噪音数据对评价结果的影响。

不管是主观评价信息还是客观评价信息，都可以在一定范围内进行横向和纵向比较。同时，评价反馈可随时进行，突出发挥评价在教学进行过程中的控制作用。此外，必须与自我评价相结合，突出在工作过程中的自我认识、自我评价、自我调控和自我完善。

4.3. 教学督导管理子系统

高校内的教学督导工作是按照国家有关教育的方针、政策、法规，对学校教学活动全过程及公共管理进行经常性的定时或随机检查、监督、评价与指导，为学校决策部门提供改进教学及其管理的依据和参谋意见，从而实现其激励功能，保证教学质量的不断提高。教学督导的基本内容是督教、督学和督管，其监控主体是教学督导员，监控客体是教师、学生和管理人员。教学督导员一般由具有丰富教学经验、先进教育教学理念和深广学科专业知识的退休老教师担当（李志平和李晓，2008）。

教学督导管理子系统为教学督导员的教学督导工作提供了基本的信息采集、存档、查询和统计功能，以提高工作效率、增加透明性、扩大影响。以下是教学督导子系统的功能结构图：



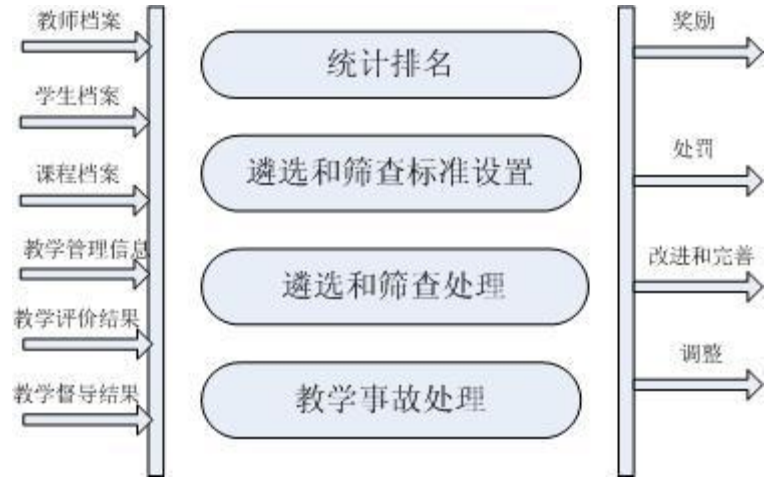
图像 4 教学督导子系统功能结构图

首先由督导专家或教学督导员制定督导工作计划，工作计划的制定主要依据一些基础数据，包括教师档案、学生档案、课程档案、教学管理信息及教学评价结果等。督导工作计划需要经过专家组评审，评审通过后生成听课表。教学督导员实施监督后，将督导结果录入系统，形成督导简报，并发布到督导信息门户上。督导专家和教学督导员都可以通过统计查询

模块查询本学期及历年的督导结果。系统管理员可以对督导专家和教学督导员进行管理，并通过角色权限设置模块控制不同的角色管理权限。

4.4. 调控与激励子系统

调控与激励子系统的设计目标是对质量监控的结果进行统计和处理，使监控结果真正反作用于高校教学活动，从而有针对性的改进教学和管理工作，提高教学质量。在实际应用中，围绕教学活动的计划、执行和结果三个主要环节，可以实现前馈控制、过程控制和反馈控制（司俊峰，2007）。在调控与激励子系统中，提供了丰富的针对教学评价结果和督导结果的统计与排名功能。针对不同的评价客体 and 教学环节可以自定义遴选和筛查标准，结合教务管理基础数据，对教学评价结果和督导结果进行审查，甄别典型，发现问题。遴选和筛查的结果也可以作为学校职称评定、评奖评优、改进教学过程的依据。调控与激励子系统的功能结构如下图所示：



图像 5 调控与激励子系统框架图

首先，由管理员根据具体的筛查目标自定义筛查指标体系，并设定不同的筛查结果所对应的结论。使用该指标体系，在遴选和筛查处理模块中对原始数据进行统计处理，获取量化结果，并计算出筛查结论。此外，调控与激励子系统也包含了对教学事故的处理。教学事故处理是当教学事故发生后，按照学校对不同级别教学事故的处理规定，对相关责任人做出的处罚措施。本模块将记录教学事故的详细信息和处理结果信息。

5. 总结与展望

教学质量监控平台以高校教学质量监控体系的总体结构为研究出发点，将体系中的各个环节、不同的监控模式和监控方法有机整合在一起，为高校教学质量监控的运行和实施提供了必要的信息化处理平台。该平台具有教学事故预警、教学评价、教学督导和调控与激励四个子系统，提供可扩展的定制策略和灵活的权限控制机制，满足不同高校对教学质量监控的信息化需求。

目前，该平台还处于研究和设计阶段，其中的部分子系统应用实施。在接下来的工作中，将进一步细化平台功能，以保证平台的开发和应用。

参考文献

李志平，李晓（2008）。本科大学教学督导的理论基础和实证研究。《黑龙江高教研究》，2008 年第 5 期，84-86

- 曹洪欣（2007）。基于 B/S 的教学督导评估系统的设计与实现研究。硕士论文，2007
- 司俊峰（2007）。普通离校本科教学质量内部监控体系研究。硕士论文，2007
- 何勇，裴晶莹，罗圣国，熊楚强（2007）。研究型大学本科教育教学督导工作信息化管理初探。北京市高等教育学会 2007 年学术年会论文集，672-677，2007
- 张忠玉，迟国彬，陈品德（2006）。基于网络的课堂教学质量监控系统的设计与实现[J]。《电化教育研究》，2006，(07)
- 陈建湘，谢再莲（2005）。试论高校内部教学质量监控主体体系、内容体系和运行机制的构建[J]。《教育与职业》，2004，(15)
- 杨红，韩锡斌，程建钢，陈刚（2005）。基于网络的教师教学评价系统[J]。《教育信息化》，2005，(04)
- 蒋荣萍（2005）。高校教学评价技术的发展研究。硕士论文，2005
- 马勃民 程建钢 韩锡斌（2004）。组件式综合教务管理系统的研究与构建，GCCCE2004

電腦語音報讀系統對閱讀障礙考生的影響

The influence of computerized read-aloud testing accommodation on dyslexic student

黃怡馨、曾芬蘭、林明志、許福元、宋曜廷*
國立台灣師範大學 心理與教育測驗研究發展中心
國立台灣師範大學 教育心理與輔導學系*

【摘要】 本研究欲探討電腦語音報讀系統對閱讀障礙考生的影響。透過標準化測驗、自行研發系統及自編研究工具，對曾參與 971 基測的 18 位閱讀障礙考生進行準實驗研究。研究結果證實：1. 電腦語音報讀系統相較於紙筆測驗，讓閱讀障礙考生在成績上有顯著的進步；2. 電腦語音報讀系統相較於真人報讀，讓閱讀障礙考生在成績上有顯著的進步；3. 無報讀功能之電腦作答方式相較於紙筆測驗，對閱讀障礙考生的成績沒有影響；4. 閱讀障礙考生普遍認同報讀的重要性，且越認同的受試者，使用電腦語音報讀系統的成績比紙筆測驗的成績進步越多；5. 閱讀障礙考生普遍支持電腦語音報讀系統，且越支持的受試者，使用該系統的成績比真人報讀的成績進步越多。

【關鍵詞】 閱讀障礙、真人報讀、電腦語音報讀系統

Abstract: This study specifically explores the influence of computerized read aloud system on dyslexic students. Two groups of dyslexic students ($n=18$) participated. How the subjects performed with computerized read aloud system and thought of the system were investigated. One standardized Chinese test was conducted to measure their academic ability variations. Two questionnaires were used to examine their opinions regarding different testing accommodations. In addition, tracing record of computer system, observers, and semi-structured interviews provided further information. The results revealed that participants' grades were raised with computerized read aloud system compared to paper and pen testing as well as human read aloud accommodation. In summary, the findings are interpreted as followed. Firstly, the subjects performed better with computerized read aloud system than with paper and pen. Secondly, the subjects performed better with computerized read aloud system than with human read aloud. Thirdly, without read aloud function, computerized testing didn't make any difference compared to paper and pen testing. Fourthly, the more supportive toward computerized read aloud system, the more progress the subjects can make based on the same test. And finally, all subjects give high credits toward easy manipulation and clear voice of computerized read aloud system.

Keywords: Dyslexic student, human read aloud, testing accommodation, computerized read aloud system

1. 緒論

1.1. 研究動機與目的

國內外許多文獻已經證實，部分閱讀障礙學生在閱讀過程中，會遭遇識字及理解上的困難。此外，每一個閱讀障礙學生的個別化差異很大，在閱讀時所遭遇的困難也不同。找出最適合閱讀障礙學生的測驗調整方式，是他們能否在測驗時發揮實力的關鍵。

目前真人報讀是國中基測對閱讀障礙學生最常提供的服務，而真人報讀有以下限制：一、每個老師報讀內容不一致；二、報讀老師可能會有念錯或不知道如何報讀的狀況；三、學生無法選擇作答的題序，也無法隨時檢查或自主安排每題作答時間；四、老師不經意的動作或

語調可能影響考生作答反應。因此，電腦語音報讀系統的研發動機是要達到報讀內容一致及作答適性化的目標。

本研究期望透過準實驗研究設計來檢驗以下假設：第一、向閱讀障礙學生讀出文字內容對他們的閱讀理解是有幫助的，換句話說，相較於傳統紙筆測驗，提供報讀可以讓他們更能發揮實力，獲得較高的測驗分數；第二、電腦語音報讀系統所提供的個別化作答方式，可以讓閱讀障礙學生在考試過程更適性；第三、閱讀障礙學生本身能感受到電腦語音報讀系統對他們有幫助且給予支持。簡言之，本研究的目的是欲瞭解電腦語音報讀系統是否能讓閱讀障礙學生在考試時更容易理解題目以及是否讓考試過程更適性，而研究結果期能提供往後試務或教學上之應用。

2. 文獻探討

2.1. 語音報讀在特殊生測驗調整上的成效

McDonnel、McLaughlin 和 Morison (1997) 指出測驗調整為促進身心障礙學生參與測驗的重要方法之一，透過改變測驗實施或受試者反應方式來補償或改善因障礙使測驗分數失真的影響。美國國家教育統計中心 (National Center for Educational Statistics, NCES) 對調整 (accommodation) 的定義是指「變更測驗呈現方式或受試者反應方式，包括呈現型式、反應方式、測驗情境、時間或安排上各種的變更。這些變更基本上不會改變水準、內容或表現的標準」(Mastergeorge & Miyoshi, 1999)。而 Tindal et al. (1998) 認為適合的測驗調整應該只對特殊生有益而對一般生沒有影響。他指出測驗中所做的調整，是用以確認所獲得的測驗結果不是因為學生的障礙所導致，而是能真正的測量出該測驗所欲評量的內涵 (Tindal, 1998)。簡言之，測驗調整主要目的是要給有特殊需求的學生一個平等的機會，在不受障礙影響下展現出他們真實的知識和技能 (Elliot, Thurlow, Ysseldyke, & Erickson, 1997)。

回顧近年的文獻，Chiu and Pearson (1999) 發現最常被使用的測驗調整方式為「延長考試時間」，而改變考試場景及改變測驗作答方式為最不常被使用的測驗調整方式。值得注意的是，因為語音報讀服務往往併同延長測驗時間一起實施，所以要單獨推論其中一個測驗調整是否有效並不容易。在本研究設計中，因為只單就語音報讀的影響進行探討，所以並沒有在施測時延長測驗時間。

語音報讀是閱讀障礙學生參加測驗時經常使用的測驗調整方式。雖然有些研究已經證實語音報讀的測驗調整方式可以改善閱讀障礙學生的考試表現 (Calhoon, Fuchs, & Hamlett, 2000; Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000; Meloy, Deville, & Frisbie, 2002)，但也有研究顯示語音報讀並不一定對所有的閱讀障礙學生都有益 (Helwig et al., 2002)。

雖然語音報讀被視為閱讀障礙學生參加考試時最需要的服務之一，但該如何實施語音報讀，仍然有許多爭議。目前常見的採用方式有三：真人報讀、錄音機統一播放語音、及電腦語音報讀。

Landau et al. (2003) 認為真人報讀有三大缺失以至於影響學生的表現。第一、真人報讀的品質不一致，有時候會有發音錯誤或唸錯的情形發生；第二、考生不願意或是不能夠要求報讀老師重唸某些考題部分；第三、因為報讀老師的語調或是不經意的意見或評論，可能會影響考生的作答反應或是注意力。此外，有相關研究證實，不管使用真人報讀、錄音帶或是影片播放語音，都會或多或少引導學生作答，並且影響學生本身的作答速度而產生負面影響 (Hollenbeck, Rozek-Tedesco, Tindal, & Glasgow, 2000)。為了達到考試作答個別化的目標並克服上述真人報讀的缺失，本研究採用電腦語音報讀系統進行施測。

此外，根據美國教育部 (U.S. Department of Education, 2001) 對測驗調整的進一步說明：測驗調整應該是根據個別學生的需求來選定，而且要與學生所接受的指導與教學一致。所以，真正有效的測驗調整方式，是要找出學生的個別困難所在，並根據每個學生及每個教室所進

行的教學差異而改變，所以，我們可以利用電腦語音報讀提高教學與測驗間的相關性以及達到個別化目標。換言之，在個別化設計的電腦語音報讀協助下，較易真實評量教學上的成效。例如可以透過提供施測獨立性或自行調整作答速度來符合個別化差異，更可以設計與教學內容有更多的交互作用的評量工具等，以上皆是真人報讀無法達到的目標。相關研究亦發現，在教學過程中，電腦語音報讀也有正面的影響，例如因為可以讓他們自行調整閱讀速度，因此增加學生在閱讀文章的獨立性(Dawson, Venn, & Gunter, 2000; Hebert & Murdock, 1994; McCullough, 1995)。

2.2. 語音報讀之爭議

雖然語音報讀被視為是幫助閱讀障礙學生最有效的方法，而且長期被廣泛地使用在需要較多識字與理解能力的閱讀測驗上。然而，有些人認為語音報讀並不適用在閱讀測驗，因為「讀」與「聽」是兩種截然不同的輸入管道，會改變測驗本身欲評量的能力。不過，仍有一些教育專家認為語音報讀方式是可以接受的，因為閱讀測驗的目的主要是測量理解能力並非閱讀能力(Elliott, Ysseldyke, Thurlow, & Erickson, 1997)。

然而，語音報讀這項測驗調整方式對不同族群考生的影響仍舊是不一致且難以解釋的(Helwig, Rozek-Tedesco, & Tindal, 2002)。有些研究認為這種方式有助於特殊生在考試上的表現(Fuchs et al., 2000; Tindal, Heath et al., 1998)，但也有其他研究結果顯示這樣的測驗調整對所有考生（不管有無障礙）都是有益的(Harker & Feldt, 1993; Meloy et al., 2002)，而這樣的研究結果並不符合測驗調整只應有利於特殊生的定義。

此外，造成語音報讀研究結果不一致的因素還有年級、考試型態和考生障礙類型及障礙內涵等(Fuchs et al., 2000; Helwig et al., 1999)。所以，研究人員也呼籲，在使用語音報讀等測驗調整時，對學生個別化差異要更加注意(Elliott et al., 2001; Helwig et al., 1999; Tindal, Heath et al., 1998)，才能夠發揮實質的效益。

3. 研究方法

3.1. 研究設計

本研究採「準實驗設計」(quasi-experimental design)來瞭解電腦語音報讀系統對閱讀障礙考生的影響為何。研究者讓閱讀障礙受試者參與兩次考試，前測與後測的施測材料同為95年第一次國中基測（以下簡稱951基測）國文試題，前測與後測的施測時間同為70分鐘，實驗組跟對照組的前測都是採紙筆測驗。然而，唯一不同的操作是：實驗組在後測採電腦語音報讀，而對照組採電腦作答，但是沒有報讀功能。研究者從97年第一次國中基測（以下簡稱971基測）接受真人報讀服務的184位特殊考生中挑選國文成績較高且量尺分數相似的閱讀障礙考生18位，隨機分成實驗組與對照組接受前測與後測。前測與後測的施測距離時間以不超過3天為原則，因為研究者擔心天數若相隔太久，會有其他變因介入（例如受試者複習考試內容等）。

3.2. 研究樣本

本研究的受試對象是學習障礙類別的閱讀障礙學生，必須持有鑑定文件及具備申請報讀資格。本研究目前的受試者皆來自曾接受971基測真人報讀服務的184位特殊考生名單，所以受試者的年齡都為15歲。依據這些特殊考生國文科正式測驗成績結果發現，約有40位考生成績落在13-23分，5位考生成績落在30分以上，其他考生成績都低於12分。為了避免受試者因為國文程度較低而大量猜題，進而影響到本研究結果，我們將受試者鎖定在國文成績為量尺分數14~23，答對題數為21~28（題本共48題）的特殊考生。符合資格且可以參與施測的個案共18位，並將這18位考生隨機分為實驗組及對照組。

3.3. 研究工具

（一）學科能力測驗工具：951基測國文科（詳見相關網站之歷屆基測試題）

(二) 自編問卷：研究者自編兩份問卷，皆由實驗組填答。分別為問卷 QI（紙筆測驗與電腦報讀的比較）與問卷 QII（真人報讀與電腦報讀的比較）。問卷答案選項分為非常不同意～非常同意四個程度，非常不同意為 1 分，不同意為 2 分、同意為 3 分、非常同意為 4 分

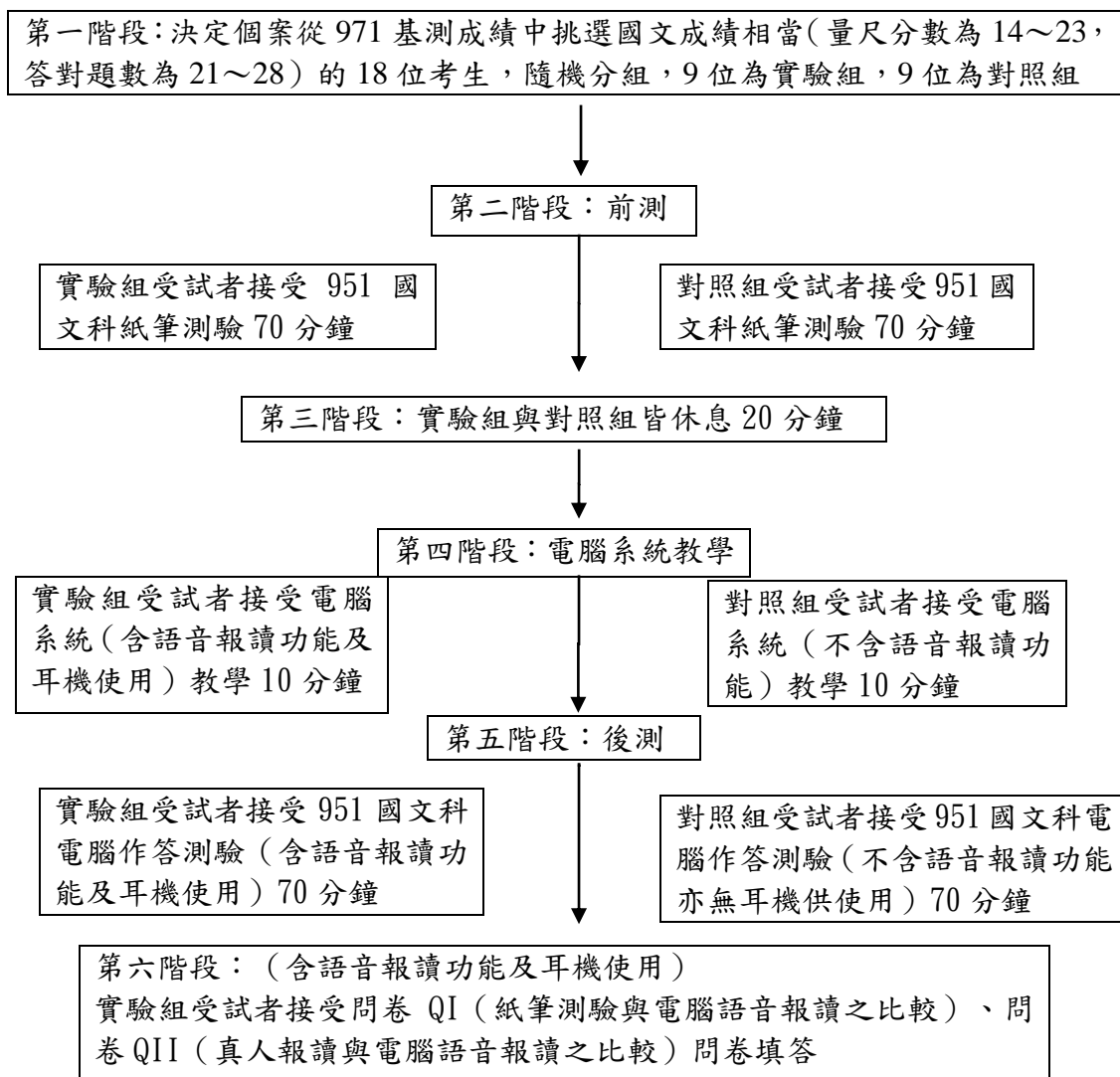
(三) 自行研發之報讀相關系統：

(1) 電腦語音報讀系統（分為學生端及施測者端兩種系統畫面：學生端所看到的系統又分為登入畫面及施測畫面，施測畫面有單題形式及題組形式）。

(2) 電腦追蹤記錄：受試者作答時會有追蹤系統紀錄所有作答反應，並於施測完畢後轉出相關資料，藉以瞭解受試者作答期間的：1. 檢查頻率、2. 作答題序、3. 每題作答時間、4. 全部作答時間、5. 聽取情形（重聽或跳過之題目與答案）。

3.4. 研究程序

研究流程分為六個階段，各階段說明如圖一。



圖一 研究程序圖

此研究主要是想探討電腦語音報讀系統對閱讀障礙考生的影響，所以測驗成績與問卷得分的相關分析只應用在實驗組。當初設計對照組的後測作答方式（電腦作答無報讀）目的是為了消除外界對可能有的練習效應之疑慮。也就是說，若是只有實驗組參與受試，而實驗組在後測的顯著進步可能會被質疑是練習效應的結果而非電腦語音報讀系統的影響（因為同樣一份測驗施測兩次）。現在加入對照組，同樣的進行前測與後測，前測完全與實驗組一樣，後測則是與實驗組使用的施測材料一樣，但是，電腦語音報讀被拿掉了，單純用電腦作答。這

樣一來，若是實驗組後測成績比前測成績明顯進步，但對照組卻沒有，我們可以推論實驗組的顯著差異是因為電腦語音報讀系統所造成，亦即電腦語音報讀系統對閱讀障礙考生有幫助。

4. 研究結果與討論

4.1. 量化統計

因為本研究個案樣本數很少，只有 18 位，所以採用 nonparametric 分析法進行研究分析。分析內容主要針對測驗成績及問卷得分，測驗成績（包括前測、後測、基測）表示個案的答對題數，而問卷分數計算則是非常不同意=1 分、不同意=2 分、同意=3 分、非常同意=4 分。如果是負向題，則會在學生作答完畢後先把分數轉換。所以問卷 QI 得到的分數結果解釋為：分數越高，表示越肯定報讀的重要性（與傳統紙筆測驗相比）；問卷 QII 的分數解釋為：分數越高，越支持電腦語音報讀系統（與真人報讀相比）。各項統計分析結果說明如下：

首先分析實驗組與對照組在前測及 971 基測答對題數的比較。若實驗組與對照組為隨機分配，兩組分別在前測與 971 基測的答對題數應該無顯著差異。研究者利用 Mann-Whitney U-test 來驗證，結果顯示實驗組與對照組在前測及 971 基測的成績表現沒有顯著差異如表一，因此推論實驗組與對照組能力相當，為隨機分配。

表一 實驗組與對照組前測成績差異

比較成績	組別	N	Mean	Sd.	Z	P
前測	實驗組	9	22.00	2.95	-0.981	0.327
	對照組	9	20.33	3.31		
971 基測成績	實驗組	9	24.00	2.59	-.224	0.823
	對照組	9	23.77	2.48		

接著同樣利用 Mann-Whitney U-test 來檢驗實驗組與對照組在接受後測時，所使用之施測材料一樣，唯有施測系統不同的情況下，兩組答對題數是否有顯著差異。由表二可以看出兩組的後測成績表現有顯著差異， p 達 0.013。

表二 實驗組與對照組後測成績差異

組別	比較成績	情境	N	Mean	Sd.	Z	P
實驗組	後測成績	電腦語音報讀	9	26.33	4.33	-2.481	0.013*
對照組	後測成績	電腦作答(無報讀)	9	20.77	3.70		

* $P < .05$

欲比較在不同的施測系統介入後，同樣一份測驗材料，實驗組與對照組的受試者在前測與後測的答對題數上有無差異，我們利用 Wilcoxon Signed Ranks Test 分別作驗證。表三可以看出施測系統不同的確造成實驗組前、後測表現上有顯著差異， $p=0.008$ ；表四則呈現對照組在施測系統改變後，前測與後測成績並沒有顯著差異， $p=0.102$ 。

表三 實驗組受試者在不同施測系統下前測與後測成績表現差異

組別	比較成績	情境	N	Mean	Sd.	Z	P
實驗組	前測成績	紙筆測驗無報讀	9	22.00	2.95	-2.670	0.008**
實驗組	後測成績	電腦語音報讀	9	26.33	4.33		

** $P < .01$

表四 對照組受試者在不同施測系統下前測與後測成績表現差異

組別	比較成績	情境	N	Mean	Sd.	Z	P
對照組	前測成績	紙筆測驗無報讀	9	20.33	3.31	-1.633	0.102
對照組	後測成績	電腦作答(無報讀)	9	20.77	3.70		

因為實驗組每位受試者都參與了 971 基測國文科真人報讀，也接受電腦語音報讀系統 951 基測國文科施測，兩份測驗題目欲測試的能力類似且題量相同（48 題），而且 951 與 971 基測國文科平均難度接近，分別為-0.40 及-0.53，加上受試者是短時間內接受這兩次測驗，能力上應該沒有太大的變化。因此，為了進一步比較真人報讀與電腦語音報讀的差異，我們利用 Wilcoxon Signed Ranks Test 來分析實驗組在後測與 971 基測的答對題數是否有顯著差異。分析結果顯示實驗組在採用真人報讀的基測答對題數與電腦語音報讀的後測答對題數有顯著差異， $p=0.024$ 如表五，所以我們可以合理推測電腦語音報讀更能夠讓閱讀障礙受試者發揮本身實力。

表五 實驗組在後測成績與 971 基測成績之差異

組別	比較成績	情境	N	Mean	Sd.	Z	P
實驗組	後測成績	電腦語音報讀	9	26.33	4.33	-2.257	0.024*
實驗組	971 基測成績	真人報讀	9	24.00	2.59		

* $P < .05$

研究者設計兩份問卷讓實驗組填答，分別為問卷 QI：紙筆測驗與電腦語音報讀系統之比較，以及問卷 QII：真人報讀與電腦語音報讀系統之比較。因為問卷 QI 的內涵主要比較無報讀時與有報讀時，受試者對題目的理解差異，所以若問卷得分越高，表示受試者越需要報讀服務。因此本研究假設受試者若是在問卷 QI 的分數越高，在後測採用電腦報讀的進步分數（即後測答對題數減去前測答對題數）也會越多。為了驗證此假設，研究者分析問卷 QI 得分與前測、後測進步成績相關性。由 Spearman rank correlation 統計結果我們可以發現，實驗組前測、後測進步成績與問卷 QI 得分（分數越高表示越肯定報讀重要性）成高度相關 $r=0.987$ ， $p=0.000$ 如表六。

另外，問卷 QII 主要內涵為二：一、受試者對 971 基測真人報讀的看法；二、受試者對電腦語音報讀的意見。受試者若是在問卷 QII 的得分越高，即表示他們越支持電腦語音報讀，因此我們假設若受試者在問卷 QII 得分越高〔表示越支持電腦語音報讀〕，其後測表現應該比 971 基測進步越多。表七為實驗組後測與 971 基測的進步分數與其問卷 QII 得分的 Spearman 相關，由結果顯示，實驗組 971 基測、後測進步成績與問卷 QII 得分高度相關 $r=0.774$ ， $p=0.014$ 。

表六 實驗組前測與後測的進（退）步成績與問卷 QI 的相關

實驗組	N	Mean	SD	Correlation Coefficient	Sig
前測與後測的進步成績	9	2.38	2.52	0.987**	0.000**
問卷 QI 得分	9	57.00	5.36	0.987**	0.000**

** $P < .01$

表七 實驗組 971 基測與後測的進（退）步成績與問卷 QII 的相關

實驗組	N	Mean	SD	Correlation	Sig
-----	---	------	----	-------------	-----

				<i>Coefficient</i>	
971 基測與後測的進步成績	9	2.38	2.52	0.774*	0.014*
問卷 QII 得分	9	60.11	5.44	0.774*	0.014*

* $P < .05$

4.2. 電腦語音報讀追蹤觀察記錄

研究者將實驗組的九位受試者編碼 A1~A9。表八為電腦追蹤記錄整理，呈現實驗組每一位受試者從開始測驗到結束的作答情況，因為轉出之原始資料非常繁瑣，研究者加以整理後分為五個觀察向度加以說明。

表八 電腦追蹤記錄整理一覽表

觀察向度	受試者表現	說明
(1) 檢查頻率	除了 A7，其他每一個受試者都有使用重聽的功能	重聽的頻率由高到低為： 1. 題目內容 2. 選文內容 3. 答案選項
(2) 作答題序	整體而言，大部分受試者都習慣從第一題開始作答到最後一題	除了： 1. A5 在考試過程有 2 次回去重聽檢查並改答案 2. A3 有跳過一個題組（3 題）沒作答，等全部寫完再回頭作答
(3) 每題作答時間	每題作答時間每個受試者都不一定，因為每一個受試者會重聽自己想聽的部分	題目文字短的，作答速度較快；題目文字越長，受試者會花較多時間聽取及重聽。但是整體而言，平均每題作答時間仍比紙筆測驗時快。
(4) 全部作答時間	每一個受試者作答完畢後都有檢查是否每一個選項都有填答	按下送出作答內容的時間平均比紙筆測驗快 10 分鐘
(5) 聽取狀況	每一個受試者都有跳過某些部分沒有聽的狀況出現	發現： 1. 大部分受試者在答案選項在 4 個字以下時會跳過不聽，例如甲乙丙丁的組合答案選項 2. 有 3 個受試者重聽題組選文 2-3 次 3. 大部分受試者會重聽文字較長的題目，對於需要排序或找主旨的題目，許多受試者都聽三遍以上

4.4. 討論

茲就前述統計分析結果進行下列討論：

(1) 測驗成績分析

研究者將 18 位受試者隨機分為實驗組與對照組，實驗組採取電腦報讀系統施測。研究者擔心，實驗組與對照組雖為隨機分組，但是可能兩組之間本來就存在組間差異，所以先分析兩組在 971 基測及前測成績的表現差異，發現都沒有顯著差異，顯示兩組的能力沒有顯著差異。因此我們可以推論實驗組及對照組若在後續的表現有顯著差異的情況發生，可能是兩組間不同實驗處理的影響。

首先本研究結果顯示，在測驗題目、測驗場景都相同，只有施測系統不一樣的狀況下，實驗組的後測成績比前測成績有顯著進步；相對的，對照組的前、後測答對題數並沒有差異。我們可以推論電腦語音報讀的確對閱讀障礙學生有正面且顯著的影響。

同樣的，在兩次施測題目及施測場景都相同而只有施測系統改變的情況下，實驗組在後測電腦報讀的表現上明顯較前測紙筆測驗有顯著進步，所以我們可以說電腦語音報讀相對於傳統紙筆測驗來說，對閱讀障礙學生是有顯著且正面的影響。相對的，對照組受試者兩次施測表現並沒有顯著差異，顯示在沒有語音報讀的狀況下，傳統紙筆測驗方式與電腦作答方式對閱讀障礙學生來說並沒有顯著影響。

此外，這些受試者都參與過 971 基測的真人報讀服務，亦即一個教室配置一位報讀老師，現場念題目給學生聽。根據學生描述，一個教室可能有 10-25 位學生不等。這樣的方式受到學習障礙協會的批評，亦受到很多學習障礙學生、家長的詬病。他們認為報讀老師素質不齊，而且一個教室學生那麼多，報讀老師無法照顧個別考生的需求，再加上每一個報讀考場所配置的報讀老師都不一樣，所以每一個教室裡的老師的報讀方式與品質都不一樣，對學習障礙學生來說並不公平。為了要瞭解電腦語音報讀系統與真人報讀對閱讀障礙學生所造成的影響差異，研究者比較了實驗組 971 基測成績與後測成績。分析結果顯示，採用電腦語音報讀的實驗組在後測的成績表現顯著優於 971 基測的成績，但由於兩次測驗並不是完全相同，為了推論電腦語音報讀效果的確比真人報讀佳，仍需要進一步的研究確認。

(2) 問卷得分說明

透過問卷，研究者想證實以下研究假設：實驗組中進步分數越多的受試者，對電腦語音報讀系統越支持。因為研究者已經將負向題分數調整過了，所以問卷得分越高即表示受試者越支持電腦語音報讀系統。研究結果顯示實驗組的前測與後測的進（退）步分數與問卷 QI：紙筆測驗與電腦語音報讀系統之比較的問卷得分達高度相關 0.987， $p=0.000$ ，所以我們可以推論越支持使用電腦語音報讀且在傳統紙筆測驗感到挫折的受試者，在使用電腦語音報讀系統時更能發揮本身實力。

另外，研究結果表示實驗組的 971 基測與後測進（退）步分數與問卷 QII 得分有高相關 0.774， $p=0.014$ 。所以我們可以推論，越支持電腦語音報讀且無法適應真人報讀限制的受試者，在電腦語音報讀系統下接受施測更能發揮本身實力。

5. 結論與建議

5.1. 結論

閱讀障礙學生有其生理上的限制及教育上的需求。現行真人報讀服務的確有一些無法克服的缺失，因此，如何改善報讀服務並研發適合閱讀障礙學生的輔具，是特殊教育該努力的方向。本研究之研究結果顯示以下幾點：

（一）電腦語音報讀系統與傳統紙筆測驗相較而言，的確對閱讀障礙學生有顯著的幫助，因為閱讀障礙學生在電腦報讀系統下答對題數明顯增加。

（二）電腦語音報讀系統比起真人報讀，對閱讀障礙學生有顯著的成效，閱讀障礙學生的答對題數也明顯增加，雖然此結果仍須進一步研究。

（三）電腦作答（無報讀）相較於傳統紙筆測驗對閱讀障礙學生的影響並沒有顯著差。（四）閱讀障礙學生普遍來說對電腦語音報讀系統抱持非常正面且支持的意見。

- (五) 閱讀障礙學生普遍較支持電腦語音報讀，而對真人報讀都有其不滿意之處。
- (六) 由相關性分析來看，對電腦語音報讀系統越支持的受試者，其前測與後測之間的進步分數越高。
- (七) 由相關性分析來看，對電腦語音報讀系統越支持的受試者，其後測成績比 971 基測成績的進步分數越高。
- (八) 電腦語音清晰度及接受度都廣泛受到受試者肯定，而且受試者對於所研發的電腦語音報讀系統的操作方式也都表示很簡單沒有問題。
- (九) 研究者建議電腦語音報讀系統可以推廣到教學現場，讓閱讀障礙學生在學習過程中能克服障礙達到最佳學習效果。畢竟要每科教材、每篇文章都有人報讀給閱讀障礙學生聽是不可能的事，但是有了電腦語音報讀系統，即可以克服人力資源上的瓶頸，讓閱讀障礙學生學習過程更無礙，也可以做到讓教學與評量結合。

5.2. 建議

根據本研究結論，對未來研究提出以下建議：

- (一) 閱讀障礙個案數量較少且大多學業成就不佳，所以他們參與施測時，學業成就很可能影響了真正的施測結果。研究者認為要真正瞭解電腦語音報讀系統的影響，應該使用適合且對個案來說應該具有鑑別度的評量工具。因此，研究者建議日後可以自行發展難易度適合特殊生的測驗，或是尋找學業程度較高、能力較好的特殊生來施測。若是條件允許，個案數量要再增加。
- (二) 研究者認為電腦語音報讀系統應該要進一步推廣到教學現場。如何在教學現場將這個系統發揮的最好？授課老師、學生、家長對這個系統的學習功能接受度如何？都值得深入研究與瞭解。

參考文獻

- Calhoon, M. B., Fuchs, L., & Hamlett, C. L. (2000). Effects of computer-based test accommodations on mathematics performance assessments for secondary students with learning disabilities. *Learning Disability Quarterly*, 23(4), 271-282.
- Chiu, C.W.T., & Pearson, P.D. (1999). Synthesizing the effects of test accommodations for special education and limited English proficient students. Paper presented at the 1999 National conference on Large Scale Assessment, Snowbird, UT.
- Dawson, L., Venn, M. L., & Gunter, P. L. (2000). The effects of teacher versus computer reading models. *Behavioral Disorders*, 25(2), 105-113.
- Elliot, J., Thurlow, M., Ysseldyke, J., & Erickson, R. (1997). Providing Assessment Accommodations for Students with Disabilities in State and District Assessment. (NCEO Policy Directions/Issue 7). Minneapolis, MN: National Center on Educational Outcomes. Retrieved August 1, 2008 from <http://www.coled.umn.edu/NCEO/>.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., and Karns, K. M. (2000). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review*, 29(1), 65-85.
- Harker, J. K., & Feldt, L. S. (1993). A comparison of achievement test performance of nondisabled students under silent reading and reading plus listening modes of administration. *Applied Measurement in Education*, 6(4), 307-320.

- Hebert, B. M., & Murdock, J. Y. (1994). Comparing three computer-aided instruction output modes to teach vocabulary words to students with learning disabilities. *Learning Disabilities Research & Practice*, 9(3), 136–141.
- Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *Journal of Educational Research*, 93(2), 113–125.
- Helwig, R., Rozek-Tedesco, M. A., and Tindal, G. (2002). An oral versus a standard administration of a large-scale mathematics test. *Journal of Special Education*, 36(1), 39-47.
- Hollenbeck, K., Rozek-Tedesco, M. A., Tindal, G. and Glasgow, A. (2000). An Exploratory Study of Student-Paced versus Teacher-Paced Accommodations for Large-Scale Math Tests. *Journal of Special Education Technology*. 15(2), 27-36.
- Landau, S., Russell, M., Gourgey, K., Erin, J., & Cowan, J. (2003). Use of the Talking Tactile Tablet in mathematics testing. *Journal of Visual Impairment and Blindness*, 97(2), 85–96.
- Mastergeorge, A. & Miyoshi, J. (1999). Accommodations for Students with Disabilities. Retrieved May 5, 2008 from <http://www.cse.ucla.edu/reports/tech508.pdf>
- McCullough, C. S. (1995). Using computer technology to monitor student progress and remediate reading problems. *School Psychology Review*, 24(3), 426–439.
- McDonnel, L. M., McLaughlin, M. J., & Morison, P. (1997). Educating one and all: Students with disabilities and standards-based reform. Washington DC: National Academy Press. (ERIC Document Reproduction Service No. ED409677)
- Meloy, L. L., Deville, C., and Frisbie, D. A. (2002). The effect of a read-aloud accommodation on test scores of students with and without a learning disability in reading. *Remedial and Special Education*, 23(4), 248-255.
- Schulte, A. G., Elliott, S. N., and Kratochwill, T. R. (2001). Experimental analysis of the effects of testing accommodations on students' standardized achievement test scores. *School Psychology Review*, 30(4), 527-547.
- Tindal, G., Heath, B., Hollenbeck, K., Almond, P., and Harniss, M. (1998). Accommodating students with disabilities on large-scale tests: An experimental study. *Exceptional Children*, 64(4), 439-450.
- Tindal, G. (1998). Models for understanding task comparability in accommodated testing. Retrieved May 9, 2008, from <http://www.coled.umn.edu/nceo/Accommodations/SEAEXEC.html>

特殊教育啟智班 Moodle 學習檔案評量系統之開發與可行性評估

Development and Evaluation of a Special Education Learning E-Portfolio System

徐國樑、劉漢欽*

苗栗縣頭份國小，台灣

國立嘉義大學數位學習設計與管理學系，台灣*

【摘要】 對智能障礙者而言，紙筆測驗往往不適合且不公平，相反地學習檔案評量能針對每一位特殊學生需求，多方面獲取學生學習資訊，然國內針對特殊教育啟智班設計之學習檔案評量系統仍十分有限。本研究宗旨在探討具效果的啟智班學習檔案評量系統應具備之功能架構。並據之，整合 Moodle 課程管理系統，開發啟智班學習檔案評量系統。邀請二位特殊教育暨檔案評量學者、二位教育科技學者、四位啟智班教師、四位啟智班學生家長，以訪問調查方式，探討系統之可行性及其接受度。初步可行性評估結果證實本系統之可用性及接受度，研究者提供許多改善系統功能建議，以供未來系統設計與相關研究之參考。

【關鍵詞】 特殊教育、學習檔案評量、電子檔案、Moodle

Abstract: Although, pencil and paper assessments have been used in the traditional classrooms for years, they are not suitable for assessing the achievement of students with Intellectual disabilities. It is important to integrate information from various resources to present the learning outcome of students with Intellectual disabilities. To achieve the abovementioned purpose, learning portfolios are commonly adapted to present the learning outcome of each individual with Intellectual disabilities. However, only limited number of teachers implemented learning portfolio as a means for assessment because it can be time-consuming building portfolios for students with special needs. This study explored mandatory features of an e-portfolio system for teachers to effectively construct electronic portfolios, especially for students with Intellectual disabilities. Further, a special education e-portfolio system was developed according to the explored features by modifying the open source MoodleTM course management system. Two special education specialists, two educational technology experts, four special education teachers, and four parents of Intellectual disabled students were interviewed and surveyed for their responses toward the system. The results showed that the e-portfolio system was accepted by the interviewees. Suggestions and comments for improving the e-portfolio system and for future research were provided in this paper.

Keywords: Special education, learning portfolio, electronic portfolio, Moodle

1. 前言

傳統評量著重以分數、量化方式描述學生學習結果（張美玉，2000）。這種量化後的測驗分數常常只能提供有限的訊息，難以測量學生真正能力，也無法說明學生問題所在，更難以描述在一段時間內學生學習改變情形（鄒慧英，2000）。對於身心障礙特殊兒童或不同文化背景學生來說，使用文字符號作答之傳統評量，往往不公平且難以適切（邱上真，1994）。鑑於傳統評量之不足，另一股以學習者為中心，著重個別學習特質差異、兼具學習歷程改變與最終結果的評量方式～學習檔案評量（portfolio assessment），近年來受世界各國重視，逐漸地被應用在教育領域。文獻指出學習檔案評量有助於身心障礙特殊兒童之學習，及對特殊教育工作人員執行教學計劃有所助益，譬如：自我決策能力（self-determination）提升、監督學習結果、改善個別化教育計劃編寫及執行時的缺失等（Agran, Blanchard, Wehmeyer & Hughes, 2002; Demchak & Greenfield, 2000; Ezell, Klein & Ezell-Powell, 1999; Salend, 1998）。

學習檔案係指有計畫性地依循特定目的或學習目標，收集和記錄足以佐證學習者在特定領域上之學習事蹟，包括各類評量結果、學習作品、學習省思（reflection）、互動回饋等（Arter & Spandel, 1992; Stiggins, 1997）。「學習檔案評量」就是依據標準評析檔案夾內的學習事蹟，以了解學生在特定領域的學習成效（吳清山、林天祐，1997）。隨著時間經過，學習檔案內之資料量亦逐年累加，傳統學習檔案評量面臨著資料保存、資料檢索、資料分享和即時更新等困擾，這些缺點及不便大部份都能夠藉由網路化學習檔案評量方式獲得解決（李新民，2000）。

藉由國內逐年增加之網路化學習檔案評量實證研究，獲知結合網際網路，實行學習檔案評量，在教育領域已逐漸受到重視，且具相當程度之普及地位（李玟儀，2005；侯錦輝，2004；陳凱貞，2004；曾江合，2003；楊政坤，2004）。這些以補救教學、英語學習或資訊學習為主之實證研究，除了開發學習檔案評量系統外，同時亦研發課程學習管理平台，如此即可引用學習者在課程學習管理平台中所有的學習訊息於學生學習檔案。然而，其發展功能良莠不齊，大部份未學習物件標準化，或是沒有針對進度落後的學生進行個別式輔導與對話等功能，就課程學習管理方面難與專門之課程學習管理系統 Moodle（Modular Object-Oriented Dynamic Learning Environment）抗衡，再加上針對特定使用者發展學習環境，因此難以應用在國民教育階段特殊教育啟智班。而國內針對國民教育階段特殊教育啟智班設計之學習檔案評量系統及實證研究，卻極為有限。

2. 研究目的

本研究宗旨在探討具效果的啟智班學習檔案評量系統應具備之功能架構。並據之，整合 Moodle 課程管理系統，實作適合啟智班之學習檔案評量系統（Special Education Electronic Portfolio，簡稱 SEEP），利用訪問調查方式評估系統之可行性，冀有助於身心障礙啟智班特殊教育工作人員執行教學計劃，及智能障礙兒童學習歷程之描述和改善。

3. 研究方法

依據文獻探討擬定初步系統功能架構，據之建立系統雛型，與特殊教育暨檔案評量專家、教育科技專家、啟智班教師、家長，利用研究者自編訪問大綱進行系統可行性評估，並依據評估結果進行系統程式及系統功能架構之修正工作。

3.1. 研究工具

3.1.1. SEEP 系統功能架構 為（1）個別化教育計劃學習目標系統（2）鏈結參照（3）評分指引（4）其他評量（5）學習事蹟。以下說明各功能用途、內容及支持納入系統架構之文獻：

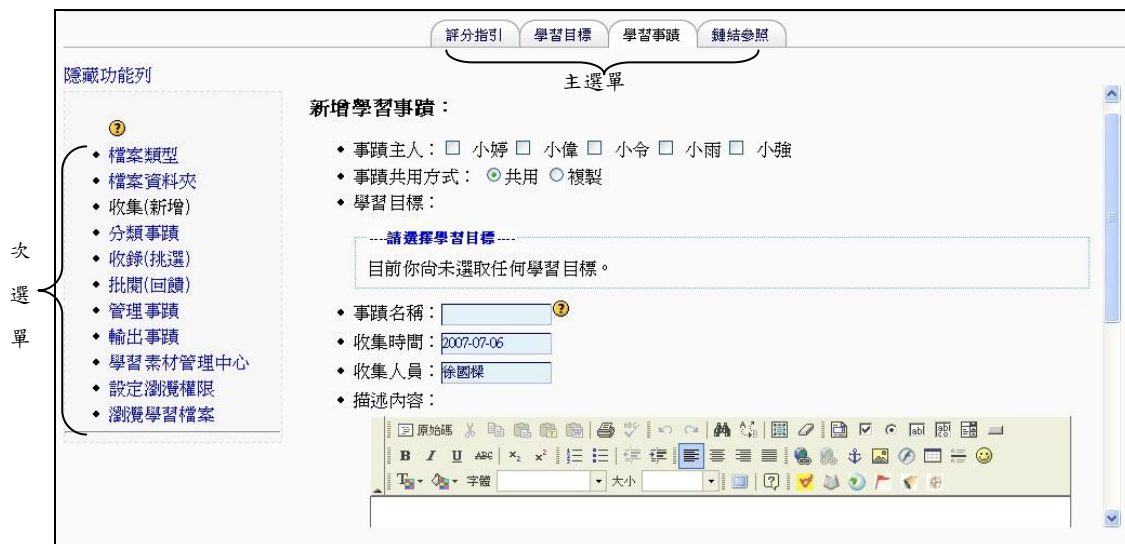
表 1 SEEP 系統功能架構

功能名稱	說明
（一）	A. 用途：
個別化教育計劃學習目標系統	協助教師針對特殊學生學習需求，擬定個別化教育計畫學習長短程目標。並據此學習目標，為學習檔案收集學習事蹟之依據。
	B. 內容：a.編訂、修改 IEP 學習目標。b.輸出學習目標成 doc 或 html。
	C. 支持本立論之文獻（supporting）：
	1. 學習檔案係指依循特定目的或學習目標，收集和記錄足以佐證學習者在特定領域上之學習事蹟，包括各類評量結果、學習作品（artifacts）、學習省思（reflection）、互動回饋等（Arter & Spandel, 1992; Stiggins, 1997）。

2. 每一位特殊學生的教育需求差異大，教學者需以學生學習需求提供量身訂製之教學措施（特教育法施行細則，2003）。	
鏈結參照	<p>(二) A. 用途： 簡化學習檔案內複雜之學習事蹟、學習目標、評量結果三者關係。</p> <p>B. 內容：a. IEP 學習目標能與學習檔案內之學習事蹟或評分指引鏈結。 b. 條列化 IEP 學習目標、事蹟與評分指引三者關係。</p> <p>C. 支持本立論之文獻（supporting）： 1. 鏈結參照能讓學習事蹟進行有意義組織及串聯（李新民，2000）。 2. 鏈結參照能協助瀏覽者找尋學習事蹟時，免於承受傳統檔案翻箱倒櫃之苦，也是電子檔案異於傳統檔案之處（Barrett, 2001）。</p>
	<p>(三) A. 用途： 「評分指引」為評量人員評定學生學習檔案等級或分數的依據。</p> <p>B. 內容：a. 能擬定評分等級（能描述表現標準及範例）及評分項目。 b. 評分指引要能彈性地變更對應之評分項目及評分等級。</p> <p>C. 支持本立論之文獻（supporting）： 1. 評分指引可作為與家長溝通之工具，並塑造評分之客觀性（鄭美紅，2003）。 2. 評分指引可能明確地告訴相關人員（如：評量人員、學生、家長）分數背後代表的意義、學習者目前之學習狀況（Airasian, 2001）。</p>
學習事蹟 管理工具	<p>(四) A. 用途： 協助教師於學習檔案之執行期和管理期，進行學習事蹟之收集、挑選合宜事蹟、針對學習事蹟進行省思、展示等。</p> <p>B. 內容： a. 收集文字、圖、影音視訊等學習事蹟。b. 依使用者需求對學習事蹟分類。 c. 為學習檔案挑選學習事蹟。c. 將事蹟輸出成 doc、html 格式或 blog 網頁。 d. 能對學習事蹟進行批閱與回饋。e. 隱藏事蹟回饋協助使用者進行省思。f. 能查詢學習事蹟。g. 能編輯、刪除學習事蹟。</p> <p>C. 支持本立論之文獻（supporting）： 學習檔案的製作過程，與學習事蹟建立及管理密不可分： 1. Danielson 與 Abrutyn（1997/2000）認為讓學生參與挑選學習事蹟的評價過程，有助於學生省查過去學習的優劣所得、產生自我省思。 2. 透過對自己學習事蹟的評價以及其他相關人員與學生的互動回饋過程，有助於規劃未來的學習設定目標（Paris & Ayres, 1999）。 3. Shaklee、Barbour、Ambrose 與 Hansford（1997）認為實施檔案評量最困難的地方，在於如何有效率地管理和提取眾多學習事蹟中有用的訊息。</p>
	<p>(五) A. 用途： 將特殊教育班教師評量學生學習成效常用之評量工具（觀察、訪談等），收納於系統架構，做為學習檔案內佐證學生學習成效之事蹟。</p> <p>B. 內容：a. 評定量表。b. 檢核表。c. 軼事紀錄表。</p> <p>C. 支持本立論之文獻（supporting）： 1. 學習檔案內容可以包含一些足以反應學生學習成效之測驗、觀察紀錄、訪談內容等評量結果（吳毓瑩，2005）。 2. 觀察（observation）、訪談（interviews）、測驗（tests）是評量活動中最常見使用的形式，也常常被用於蒐集特殊需求兒童之重要資料上（Wolery &</p>

接前頁

3.1.2.系統內容 本系統可區分為前端及管理後端。前述六大部分為系統後端管理部份，而學習事蹟部份可以輸出成動態網頁與 Html、Word 型態等前端呈現，藉此達到下列諸項多重用途之特色：（1）將評量檔案轉化為展示性檔案：讓未具後端管理身份者瀏覽釋出之學生學習檔案，快速方便地展示學生學習進度及個人能力。（2）將評量檔案化為多重型態 Blog：動態網頁能以啟智班全班 Blog、學生個人 Blog。（3）輸出學習檔案能與一般套套軟體結合進行後續編輯：藉由輸出網頁 Html 格式及微軟 Word 格式，與一般電腦套裝軟體進行單向資料交換工作，讓使用者能在使用系統同時，又能享用一般電腦套裝軟體之功能。



圖像 1 本系統後端管理首頁

3.1.3 訪談大綱內容 特殊教育暨檔案評量專家之訪談著重於本系統提供之功能項目，是否符合學習檔案評量之特性、是否符合特殊教育教師評量之特性。教育科技專家重於系統設計架構、系統界面設計部份進行評鑑。啟智班老師訪談著重系統功能之重要性、符合使用者需求等部份。學生家長訪談著重系統內容呈現方式、對於了解學生學習狀況和親師溝通方面。

4. 訪談結果分析（系統可行性評估）

經上述各方人員評估後，歸納整理如下：

（1）受訪專家表示，本系統符合學習檔案評量特性～多元化、利於溝通、有助省思活動進行：

受訪專家覺得本系統接納文字、圖像、聲音、視訊等學習事蹟。這種多元化之接納方式，讓本系統所建立之學習檔案能收錄學生各式各樣之學習紀錄，例如：學習單、觀察紀錄、學習活動等（訪評鑑-E3-1-1，訪評鑑-E1-2-2）。本系統提供多方不同身份者登入提供意見之功能，或家長親自到學校瀏覽本系統，或與教師面對面溝通討論學生學習情形之替代方式進行，足見本系統符合學習檔案利於溝通之特性（訪評鑑-E2-1-6）。此外，不同身份的使用者能針對學習紀錄進行交互討論、隱藏回饋，此有助於使用者進行省思活動（訪評鑑-E2-1-16）。

（2）受訪專家及啟智班教師對本系統功能之設計規劃持肯定、符合使用需求之態度：

A. 製作學習檔案同時亦完成個別化教育計畫：

特殊教育法規定教師需為啟智班學生擬定個別化教育計畫，藉由本系統製作學生學習檔案，亦同步完成編擬個別化學習目標（訪評鑑-T1-3-2、訪評鑑-T4-3-2、訪評鑑-T3-3-14）。

B. 利於資料整合，無需花費時間於資料的轉換或瀏覽設備之切換工作：

除了學生學習單、學習作品是紙本外，其他的事蹟大部分已數位化，若進行傳統檔案評量，教師需花費時間和精神於文件的轉換或瀏覽設備之切換工作（訪應用-T1-1-30、訪應用-T2-2-12、訪應用-T3-1-19、訪應用-T4-4-30）。

C. 藉本系統學習目標和學習結果間之對應，有助於教師撰寫學習評量結果，進行教學決定：

傳統學生檔案評量，較無法標示每一個學習目標與哪些學習事蹟有關係，亦無學習事蹟和學習目標間之對應查詢，是故在綜合相關的學習事蹟，撰寫學習目標的學習結果時會花費較多時間和精神。而本系統透過滑鼠點按動作，就能呈現學習事蹟和學習目標間對應關係，系統設計規劃良好（訪應用-T1-1-20、訪應用-T2-1-23、訪評鑑-E2-1-3，訪評鑑-E4-2-12）。

受訪教師 T1（訪應用-T1-1-28、訪應用-T1-1-29）表示：

連結學習目標和學習事蹟在做傳統檔案時太花時間，最終並沒有法駐記…所以最後在寫評量結果時會很花時間去比對哪些事蹟和學習目標的關係，再撰寫學習結果

而受訪教師 T2（訪應用-T2-1-29）表示：

若於作品上註明和哪些學習目標有關係，可能會破壞作品的原貌，比較不希望這樣去做，所以最後在評量學習結果時，會花費不少的精神和時間，去比對和書寫學生學習結果

（3）藉本系統內容，有助於家長清楚快速地理解教師所要表達之意思：

藉由本系統視訊或相片的呈現，在親師溝通上，有助於家長清楚正確地，理解老師表達的意思，無需不斷地詢問、費心地猜測或一時記不清楚再次看系統內容會比較了解（訪反應-P3-1-4、訪反應-P4-1-6）。

家長 P4（訪反應-P4-1-6）表示：「老師利用視訊與我討論孩子學習過程，讓我更清楚老師給學生提供了哪些口語或肢體，協助程度如何等。」

（4）藉本系統內容，有助於家長真實地勾勒出學生在校學習情形：

藉由學生學習視訊之播放，有助於家長勾勒出學生在校行為規矩，對部份受訪家長而言，其可信度高於紙筆觀察紀錄（訪反應-P2-1-8、訪反應-P2-1-3）。

5. 結論與建議

本研究發展之特殊教育啟智班 Moodle 學習檔案評量系統，在各方人員評估中顯示本系統不僅能協助啟智班教師進行學生學習檔案評量，亦符合特殊教育評量精神，能與個別化教育計畫學習目標系統結合，結構性呈現學習目標、學習事蹟關係，有助於教師撰寫學習評量結果，並據之進行教學決定。此外，藉由本系統家長能清楚快速地理解教師所要表達之意思，真實地了解學生在校學習情形，縮短親師溝通距離。未來將進一步推廣應用於更多啟智班。

綜合研究者省思整理如下，以供後續系統更新或未來類似系統之開發相關人員參考：

- （1）身心障礙包括很多障礙類別，因此建議加入不同障礙類別的學習目標，如：視覺障礙肢體障礙等，讓更多不同障礙類別的特殊教育班級使用。
- （2）學習目標系統能與市面上現有學習目標管理系統進行資料交換。
- （3）建議增加 IEP 法律規定項目，讓本系統變成 IEP 系統暨檔案評量系統。
- （4）系統具備形成性的評量在裡面，因此建議加入提示功能，定期提醒教師接下來的行程。

參考文獻：

- 吳清山、林天祐（1997）。卷宗評量。《教育資料與研究》，15，69。
- 李玟儀（2005）。建置教學歷程檔案系統之研究—以中文補救教學為例。未出版博士論文，育達商業技術學院資訊管理所，台北。
- 李新民（2000）。電子卷宗在教育上的應用。《教育資料與研究》，35，92-99。
- 邱上真（1994）。落實全面身心障礙教育—回應黃委員榮村所提教育改革研討會會議報告。《教改通訊》，1，40-45。

- 侯錦輝 (2004)。個別化學生學習歷程檔案的評量系統-以竹苗區某國小中文補救教學。未出版碩士論文，育達商業技術學院資訊管理所，台北。
- 張美玉 (2000)。歷程檔案評量的理念與實施。《科學教育》，231，58-63。
- 教育部 (2003)。《特殊教育法施行細則》。教育部臺參字第 0920117583A 號令修正發布第十三條條文。
- 陳凱貞 (2004)。實施國小英語電子化檔案評量研究-以一國小三年級班級為例。未出版碩士論文，台北師範學院兒童英語教育研究所，台北。
- 曾江合 (2003)。網路化個人及團體學習歷程檔案系統之建置與評估。未出版碩士論文，交通大學資訊科學系，新竹。
- 楊政坤 (2004)。國小英語學習檔案評量網路系統之設計與實施成效。未出版碩士論文，台北師範學院兒童英語教育研究所，台北。
- 鄒慧英 (2000)。國小寫作檔案評量應用之探討。《初等教育學報(台南師範學院)》，13，141-181。
- 鄭美紅 (2003)。設計多元化科學習作。香港：香港教育學院，未出版手稿。
- Agran, M., Blanchard C., Wehmeyer, M., & Hughes, C. (2002). Increasing the problem-solving skills of students with developmental disabilities participating in general education. *Remedial and Special Education*, 23(5), 279-288.
- Airasian, P. W. (2001). *Classroom assessment concepts and applications*. Boston: McGraw-Hill.
- Arter, J. A., & Spandel, V. (1992). NCME instructional module: using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, 11(1), 36-44.
- Barrett, H. C. (2001). *Electronic portfolios*. Retrieved August 10, 2006 from the World Wide Web: <http://electronicportfolios.com/portfolios/encyclopediaentry.htm>
- Danielson, C., & Abrutyn, L. (2000)。檔案教學(鄭英耀、蔡佩玲譯)。臺北市：心理。(原著出版於 1997)。
- Demchak, M., & Greenfield, R. G. (2000). A transition portfolio for Jeff, a student with multiple disabilities. *TEACHING Exceptional Children*, 32(6), 44-49.
- Ezell, D., Klein, C. E., & Ezell-Powell, S. (1999). Empowering students with mental Retardation through Portfolio Assessment: A tool for fostering self-determination skills. *Education and Training in Mental Retardation and Developmental Disabilities*, 34(4), 453-463.
- Salend, S. J. (1998). Using portfolios to assess student performance. *Teaching Exceptional Children*, 31(2), 36-44.
- Shaklee, B. D., Barbour, N. E., Ambrose, R., & Hansford, S. J. (1997). *Designing and using portfolios*. Boston: Allyn and Bacon.
- Stiggins, R. J. (1997). *Student-Centered Classroom Assessment*. New Jersey: Upper Saddle River.
- Wolery, M., & Wilbers, J. S. (2005)。學前融合教育方案(鐘梅菁、吳金花譯)。台北：華騰文化股份有限公司(原著出版於 1994)。

應用 S-P 表分析理論輔助國中學生資訊搜尋能力診斷與適性化教學之探究

A Study of Applying the Student-Problem Table Theory to Support Diagnosis and Adaptive

Learning on the Information Search Literacy of Junior High School Students

王曉璿、劉宜欣*、柯立芳*

國立台中教育大學數位內容科技系

國立台中教育大學數位內容科技系碩士班*

hswang@mail.ntcu.edu.tw、sindy_liu1234@yahoo.com.tw、solmi.solmi@msa.hinet.net

【摘要】本論文提出一套應用「Student-Problem Table」分析理論的學習診斷系統來輔助國中學生之診斷與適性化教學。學生透過此學習診斷系統，將本身資訊搜尋的能力記錄於此系統中，教師則可以利用 S-P 表分析理論將學生分類為多個學習特徵群組，每組學習特徵皆反應出學生不同的學習困難，教師在網路資訊搜尋的課程中，可以針對不同群組的學生，以不同的教學方式，提供不同的教學策略來進行補救教學。藉此協助不同資訊能力之學生提升其網路資訊搜尋技能，激發學習潛能，以達適性化教學之效。

【關鍵詞】 S-P 表分析理論、資訊搜尋、學習診斷、適性化教學

Abstract: This article proposed a diagnosis system of learning results based on the 「Student-Problem Table」 theory for supporting adaptive learning. The students can use the diagnosis system of learning results to trace their information searching results and the teacher can utilize the S-P Table to group students based on their information searching ability. According to this analysis of the S-P Table, we can apply it in remedial instruction and independent assistance for individual student groups. This research integrates the S-P table theory as the implement of diagnosis system for students in order to analyze the information searching ability after the learning, and the information could be taken as a reference guideline by which the teacher can start the remedial instruction for his/her student.

Keywords: Student-Problem Table, Web Information Searching, diagnosis of learning results, adaptive learning

1.前言

1.1.研究背景

隨著資訊科技的蓬勃發展，國中學生的網路資訊搜尋能力也日趨重要。Rouet(2003)曾指出，當資訊網路科技在教育體系中普遍利用時，學習者為了找出課程相關資料，其資訊搜尋能力也將更加重要。

目前國中小階段的課程中，常見各科教師利用學習單的方式，要求學生上網找尋相關資料，藉此增加學生資訊搜尋能力。然而網路資訊搜尋與傳統紙本資料查找大不相同。雖然網路環境讓資訊隨手得，但對國中小階段的學生來說，要在這浩瀚的網路資訊中找到正確的資訊，實非易事。歐陽閭(2002)指出學生在網路資訊搜尋上常見的問題，包括對搜尋任務未能深入分析，以致關鍵字的使用不當，未能找到最好的資訊，或擷取到不正確的資訊。

然而每位學生面對資訊搜尋的困難不盡相同，不同地區、族群及家庭資訊設備之國中學生，在資訊科技使用及網際網路使用、網路素養及搜尋技能等方面，都有顯著的差異，可見學生資訊能力的落差，也是造成資訊搜尋結果差異的主要原因之一（黃威堯，2004）。因此在

同一班級中，各個學生的資訊能力不盡相同，如何診斷學生個別的資訊搜尋困難，而教師應利用診斷的結果，來檢視和獲致學生的學習成果及教師本身的教學成效等相關訊息，並利用這些訊息來幫助學生克服學習困難、激發學習潛能，以及協助教師調整教學並提供適性的教學活動，藉此來協助不同資訊能力之學生提升其網路資訊搜尋技能為本論文的研究動機。

1.2. 研究目的

本研究旨在發展一套能診斷學生資訊搜尋困難的系統，藉此來協助不同資訊能力之學生提升其網路資訊搜尋技能。此診斷系統利用 S-P 表分析理論 (Takahiro Sato, 1975) 為基礎，針對學生的作答反應資料來診斷學生在資訊搜尋過程中所遇到的困難為何，利用統計方法進而分析學生是因為誤用關鍵字或是搜尋引擎使用上的問題等，其目的在獲得每個學生的網路資訊搜尋技能診斷資料，藉此讓教師能針對不同資訊能力之學生提供適性化教學。本研究期協助教師改進資訊教學，並有效提升不同資訊能力學生之資訊搜尋能力落差。

1.3. 研究範圍與限制

要使用本系統需具備基本資料搜尋能力與具備語文能力，以確保教學活動、系統操作能夠順利進行，而本研究之對象皆為國中一年級之學生，對於網路之瞭解皆有一定之水準，故有關係統應用及成效的評估，不宜過度加以引伸。

2. 文獻探討

2.1. 網路資訊搜尋

Kuhlthau(1991)定義資訊搜尋過程是「使用者特定的問題或主題，為了去擴展自我的陳述知識，從相關資訊中發現含義後的建構活動」。在網路搜尋引擎和 WWW 技術進步和發展下，使得資訊搜尋的方式產生了極大的改變，越來越多的學生利用網路資訊搜尋方式來尋找資訊。

在陳雪菱(2003)的研究中發現青少年網際網路的應用活動中以搜尋資料排名第一，可見網路資訊搜尋在現今學生學習活動中扮演著重要的角色。但是對於國中小學生來說，因為沒有完善的資訊操作技能，亦或受限於對搜尋主題的不了解，常常對於資訊搜尋是一片茫然不知如何著手搜尋所需資訊。胡嘉璽(2000)曾表示，找資料最大的問題在於不懂得分析資料，便直接開始尋找。另外學童的父母處於較低的社經地位時，將導致其資訊資源分配的缺乏，造成資訊搜尋上的障礙，文化及語言的障礙亦是影響的因素(張瀚文, 2000)，由此可見每個學生所面臨的資訊搜尋困難會因為家庭環境、語文能力、資訊能力等不同因素而有所落差。

2.2. 網路資訊搜尋問題診斷

網路資訊搜尋是一種複合能力的表現，要評量學生的資訊搜尋能力並不容易。過去的研究中所使用的評量方式，不外乎以螢幕擷取軟體來紀錄學童在不同任務的網路搜尋過程，分析他們使用的搜尋策略。或是以訪談來蒐集資料，主要探討學童之網路搜尋資訊之動機、檢索問題等。由此可見網路資訊搜尋的評量方式不外乎使用攝影機或電腦軟體紀錄、訪問研究對象，但是評量過程卻曠日費時。所以在本研究中，先以基本的資訊搜尋題目來測試學生在試題上的作答反應，再利用 S-P 表來分析學生資訊搜尋困難，使得教師能利用簡單的問答系統得知學生的學習困難，以節省教師和學生的時間。

S-P 表(Student-Problem chart)分析法，為日本學者佐藤隆博(Takahiro Sato)於 1970 年代所創，該方法分析每位學生及每個試題的作答反應組型，嘗試以幾個指標化數據作為診斷或判讀該反應組型是否為不尋常或異常的一種測驗分析方法(余民寧, 1997)。所謂的 S-P 表分析，針對每位學生與每個試題的作答反應組型所產生的注意係數，與整份測驗卷的差異係數做分

析。注意係數與差異係數這二項指標都是用來協助教師診斷學生表現、試題的品質以及教學成果的重要資訊之一，也可將這些指標作為改進教學、編製教材與輔導學生之參考。

2.3. 適性化補救教學

補救教學能夠使學生的學習獲得改善，並且減少學習挫折。但每個人的特質及能力不同，學習過程也有所差異，唯有針對每個人不同需求進行的補救教學才有意義。

然而，在過去資訊搜尋相關的研究中，許意苹(2007)運用 Wh-Questions 問題分類及摘要策略教學法來改善關鍵字運用的技巧缺乏，並提升網路資訊搜尋的正確性及縮短搜尋資訊的時間。程璟滋(2006)在班級上實施腦力激盪搜尋策略，學童學得聯想關鍵字的方法，使其於網路資料查詢時，能夠針對題目聯想出多個適合的關鍵字。林郁璇(2005)發展引導式搜尋介面，進一步探討引導式搜尋介面，是否能有效地提升學童搜尋之正確率、效率。在這些研究中可以發現，研究者針對學生在資料搜尋時所面對不同搜尋的困難，分別提出不同的解決策略，確實能夠更有效率的提升資料搜尋成效。但所使用的研究方法往往是全班採用同一種教學策略，顯少針對學生個別的問題給予不同的補救教學，而本研究中則是先對班級中的學生進行資訊搜尋能力的診斷和分析，再利用分析的結果來針對不同資訊搜尋能力的學生給予不同的補救教學，這也是本研究與前人之研究最大的不同處。

3. 研究方法

本研究以台中縣某國中一年級學生為研究樣本，從中選取一個班級共 24 人作為研究對象，學生具備有基本的電腦操作及語文能力，也有上網搜尋資料的經驗。圖 1 為本研究的概要流程圖。在本研究中，學生需在線上資訊搜尋測驗系統中進行問答式的搜尋測驗，且學生在作答題目時必須填寫其搜尋關鍵字和搜尋到之解答。研究者利用線上資訊搜尋測驗系統收集學生的測驗資料後，再使用 S-P 表理論來分析個別學生資訊搜尋能力並將學生依據 S-P 表分析法分類為多個群組。之後再討論針對不同群組的學生以不同的教學方式來進行補救教學。

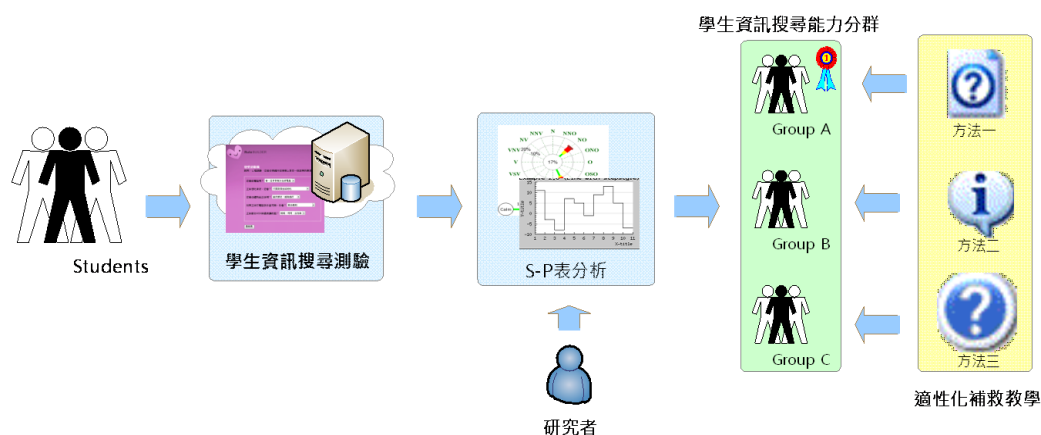


圖 1 研究方法流程圖

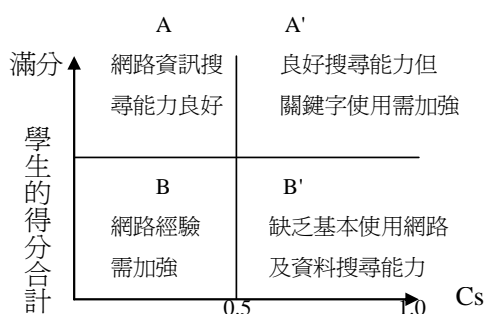
3.1. 線上資訊搜尋測驗系統

本系統是利用 PHP + MySQL 所建置而成的。測驗題目設計則分為三大類型，分別為測驗學生的網路經驗、選擇關鍵字的能力、摘要的能力。學生對網路認識的深淺包括對搜尋引擎的認識、上網的經驗多寡等。例如：下列何者不是搜尋引擎？(1) Yahoo (2) Google (3) PC Home (4) Blog。此類題型主要是用來判斷學生對於網路的認識和使用上是否熟練。測驗學生選擇關鍵字的能力的題型主要是以符合國一學生程度的生活常識為出題方向，學生在線上系統中看完題目後必須自行選擇最佳的關鍵字。例如：台灣縱貫鐵路在什麼時期完成；學生的答案是選擇該問題最適當的搜尋關鍵字，例如：(1)縱貫鐵路 (2)台灣+縱貫鐵路 (3)台灣+縱貫鐵路+

完成 (4)台灣縱貫鐵路在什麼時期完成。此題型設計主要是為了解學生對於題目本身是否能擇出最適合的關鍵字。而摘要的能力主要是測驗當學生搜尋到一連串的資料後，能否從資料中擷取出正確的答案，例如學生利用所選擇的關鍵字進行網路搜尋後找到一篇網路資料，文章裡有以下敘述「縱貫線的興建，始於清領時代末期，時任台灣巡撫劉銘傳的規劃，…。1896年，日本開始統治台灣之後，立即將縱貫鐵路的興建列為首要的施政計畫…。」。受測學生必須在文章內容中找出正確答案。而此測驗系統會將學生的答案記錄和作答時間等相關資訊上傳至資料庫中，做為後續 S-P 表學習診斷分析的資料。

3.2. S-P 表分析

S-P 表分析是結合認知理論和測驗理論所發展出來的認知診斷評量技術，將學生在試題的作答反應予以圖形化的測驗分析法。余民寧(2002)根據佐藤博士提出的判斷標準，將學生的學習特徵分為六種類型，分別為：學習穩定型學生、粗心大意型學生、努力不足型學生、缺乏充分型學生、學力不足型學生、學習異常型學生。然而本研究將學習特徵重新定義為四種類型(圖 2)，分別為：網路資訊搜尋能力良好(A)、具有良好搜尋能力但關鍵字使用需加強(A')，具有基本搜尋能力但使用網路經驗需加強(B)、缺乏基本使用網路及資料搜尋能力(B')。在 A 區學生的學習良好，網路搜尋工具使用熟悉度高，對於關鍵字的選擇也比較熟練。A'區學生的學習穩定，對於關鍵字選擇並不熟悉。B 區學生的學習不穩定，對於網路使用經驗不足，但已有基本網路搜尋概念。B'區學生學習不穩定，對網路搜尋工具熟悉度不夠。在善用 S-P 表於資訊搜尋測驗的過程中，S-P 表成為有用的教學輔助工具，建立活化有效的教學管理系統，不僅可了解學生的學習成就，亦能找到診斷學生學習困難之線索，以提供個別化教學設計與實施的改進之依據，一方面可增進教學效率；另一方面可以作為補救教學的參考。



注意係數與得分的對照

圖 2 學生診斷分析圖

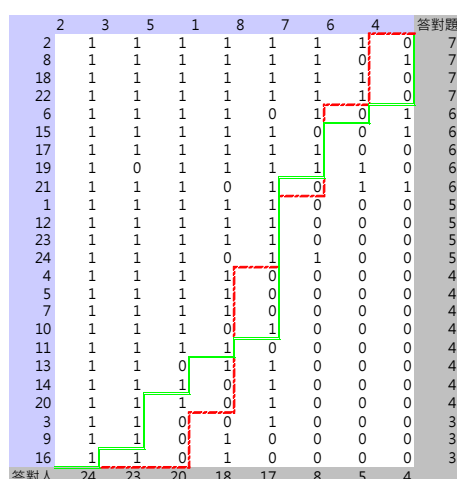
4.結果與討論

4.1. S-P 表編製與討論

在本次研究過程中，利用一節電腦課的時間施測，總共有八題題目。利用線上資訊搜尋測驗系統搜集了24位學生的作答結果，經評分後以S-P表分析方法制作出S曲線和P曲線之圖形。S曲線和P曲線之圖形詳細制作過程如下，利用Excel將原始得分矩陣資料依照每位學生得分總分高低，由上往下依序排列，如有總分相同時，亦可照座號大小順序排列，接下來按照試題答對人數多寡，由左(答對人數最多之試題排在最左邊)到右依序排列，遇有相同得分之試題，亦可依題號之大小順序排列，最後，根據每位學生答對的題數(簡稱總分)，從左向右數出與總分相同之試題個數，並在右邊畫上一條分界線，由高分往低分畫出每位學生總分所對應的分界線，再將這些分界線的下方利用直線連接，則會形成一階梯狀之曲線，此曲線即稱為「S

曲線」，同理，依據每道試題的答對人數，從上往下數出與答對人數相同之學生個數，並在其下邊畫上一條分界線，由左端往右端分別畫出每道試題之答對人數所對應的分界線，則會形成一階梯狀之曲線，此曲線即稱為「P曲線」，繪製結果如圖3所示(紅色線是S曲線，綠色線是P曲線)。

圖3中顯示本次測驗結果，學生答對最多題為7題，最少為3題，S曲線分隔了學生答對與答錯的區域，此曲線以左的範圍，代表學生大都答對了試題，可以看出學生學習成就達成的程度。而P曲線是用來區分試題答對與答錯人數的分界線。題目2為大多數學生皆答對，題目4為大多數學生皆答錯，可以看出班級學生達成與未達成教學目標的程度。排列在S-P表左上方者，代表能力較好之學生與較簡單之試題，大多數是被期望答對的試題，所以，這個區域應該出現大多數的1。相反的，在S-P表右下方者，應該出現大多數的0。



(差異係數:0.336)

圖3 S曲線和P曲線圖

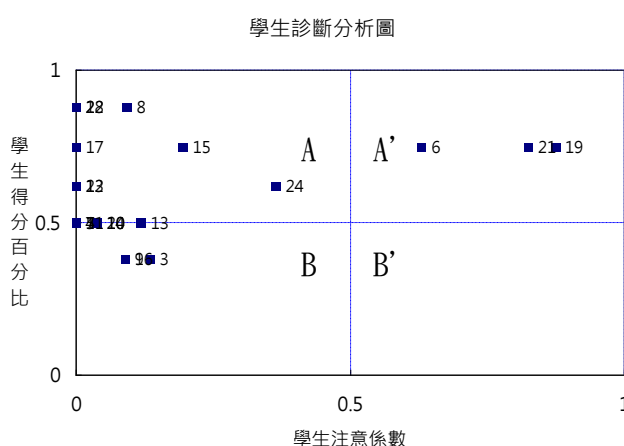


圖4 學生資訊搜尋能力分類圖

4.2. 學習診斷分析結果

S-P表解讀上不具有視覺化的效果也不容易將學生分群來比較，因此可以將S-P表利用公式算出學生的注意係數(caution index)，注意係數是S-P表針對「個別」學生與試題所使用的另一類係數，可分為學生注意係數(CS)與問題注意係數(CP)兩種。主要是用來作為判斷學生或試題在反應組型中是否有異常現象的指標，教師可利用這些指標瞭解學生問題所在。在圖4的學生資訊搜尋能力分類圖中，使用S-P表理論公式可把圖3的結果轉換成如圖4的四個象限，即可學生的測驗資料快速的分為四大群組以了解各個學生的學習狀態。我們依照只四種群組討論分析其每種群組所適合的個別化教學設計。

- (A)網路資訊搜尋能力良好：表示該生具備有良好的網路資訊搜尋能力，能選用較佳的關鍵字做資料搜尋，也能在搜尋結果中找出所需要資訊，在補救教學的實施會以開放性問題做為練習，加強學生將網路上的資源摘要成一篇報告的能力。
- (A')具有良好資訊搜尋能力但關鍵字使用需加強：具有良好資訊搜尋能力，但關鍵字的使用會混淆不清，常會造成關鍵字的誤用，在補救教學方面可以利用概念圖的方式提升學生對搜尋任務的定義，加強關鍵字使用的正確性。
- (B)具有基本資訊搜尋能力但使用網路經驗需加強：雖然具備有基本的資訊搜尋能力，但是較缺乏搜尋策略，常將整句問題當成關鍵字搜尋，對於找到的資訊也無法正確找出答案，在補救教學方面可以採用問題分類及摘要策略教學法，改善關鍵字運用的技巧。
- (B')缺乏基本使用網路及資料搜尋能力：學習特徵落在該群的學生表示對於資訊搜尋能力低落，缺乏使用搜尋引擎的能力，因此補救教學措施包括一連串的資訊搜尋加強課程。

5. 結論與未來發展

本研究以 S-P 圖表分析各個學生資訊搜尋能力的結果，以了解個別學生的學習行為、全班學習成效及整體測驗的品質。以 S-P 表為學習評量的分析工具，而教師應利用診斷的結果，來檢視和獲得學生的學習成果及教師本身的教學成效等相關訊息，並利用這些訊息來幫助學生克服學習困難、激發學習潛能，以及協助教師調整教學並提供適性的教學活動，藉此來協助不同資訊能力之學生提升其網路資訊搜尋技能。本研究未來計劃再加入 S-P 表理論中試題診斷分析，來了解試題是否能真實反應出學生的資訊搜尋能力。而未來也將增加適性化教學系統，針對學生資訊搜尋能力的弱點，做更有系統的補救教學。

參考文獻

- 余民寧(1997)。有意義的學習——概念構圖之研究。商鼎文化出版社。
- 余民寧(2002)。教育測驗與評量—成就測驗與教學評量。台北：心理出版社。
- 林郁璇(2005)。引導式搜尋介面創發之研究—以國小四年級學童資訊尋求模式為基礎。國立台中教育大學教學科技研究所碩士論文。
- 胡嘉璽(2000)。網路資料搜尋攻略。台北：學貫行銷。
- 張瀚文(2000)。從小學生之資訊需求與資訊尋求行為探討學校資訊網路與系統設計。台北市立圖書館訊，17(3)，43-56。
- 許意苹(2007)。國小高年級學生網路資訊搜尋課程發展與成效之研究。淡江大學教育科技學系碩士在職專班碩士論文。
- 黃威堯(2004)。國中生數位落差現況探討之研究。高雄師範大學工業科技教育學系碩士論文。
- 程璟滋(2006)。腦力激盪搜尋策略對國小學童網路資料蒐集能力之影響。國立嘉義大學教育科技研究所碩士論文。
- 歐陽閻(2002)。學生網路資訊檢索能力之評鑑研究。論文發表於第六屆全球華人電腦教育應用大會暨全國教育資訊劃論壇 2002 年國際學術研討會，北京市。
- 陳雪菱(2003)。國高中生網路資訊尋求行為影響因素之研究—以台北市為例。元智大學資訊傳播學系碩士論文。
- Kuhlthau, C.C.(1988).Developing a Model of Library Search Process: Cognitive and Affective Aspects. *RQ*, 28, 232-242.
- Route,J.-F.(2003).What was I looking for? The influence of task specificity and prior knowledge on students' search strategies in hypertext. *Interacting whit computers*, 15, 409-428.
- T. Sato, "Introduction to student-problem curve theory analysis and evaluation," Tokyo:Meiji Tosho, 1975.

英文寫作自我效能量表之發展

Development of the English Writing Self-Efficacy Inventory

張菟真、廖長彥*、辜玉旻

國立中央大學學習與教學研究所

altheawcc@gmail.com, kuyumin@gmail.com

國立中央大學網路與學習科技研究所*

calvin@cl.ncu.edu.tw

【摘要】隨著線上寫作平台的興起與寫作趨勢的改變，探討科技環境對於寫作歷程與表現的研究也日趨多元。本研究以 Bandura 的自我效能理論為基礎，以發展符應現今大學生學習環境的英文寫作自我效能量表為主要目標。研究樣本為台灣某國立大學生共計 113 名，經過探索性因素分析共得 20 題及 5 個因素，分別為「寫作態度」、「寫作任務」、「寫作歷程」、「寫作技巧」及「寫作回饋」，總解釋變異量為 49.98%，並以驗證性因素分析與效標關聯效度進行「英文寫作自我效能」的構念效度驗證。背景變項分析則發現大學生的性別與年級別在英文寫作自我效能上沒有顯著差異。

【關鍵詞】 英文寫作自我效能、問卷發展

Abstract: With the increasing application of web-based technologies to support the writing process, it has become necessary to ask how the process of writing is integrated with the features of web-based environment. The purpose of this study was to develop a writing self-efficacy inventory (WSEI) for university students' online English writing. In the pilot study, the item analysis and the factor analysis on the WSEI was employed to test its validity on 113 college students from a university located in the north part of Taiwan. The final version of inventory contains 20 items including five subscales, including writing attitude, writing task, writing process, writing skill and writing feedback. The cumulative percentage of variance accounted for by these five factors was 49.98%. The confirmatory factor analysis and the Criterion-Related Validity are used to test the validity of WSEI. The analysis on background variables indicated that there were no significant differences on the sex and grade variables.

Keywords: Writing Self-efficacy, Questionnaire Development

1. 前言

過去將寫作視為是一件作品的產生，強調寫作形式的練習與改正，包括段落結構、標點符號等規範，以這樣的成果導向的方式去展現、評量學童原有作文能力。然而，現在的觀點則是將寫作視為一種過程，強調在寫作過程中增加學生的寫作知識，讓學生學習如何寫作，以提昇寫作能力（Klein, 2000; Nelson, 2001; Tynjälä, Mason, & Lonka, 2001）。而隨著網路時代的來臨，Web技術的成熟，寫作得以擴展到網路，線上寫作環境逐漸成熟，如：Blog, Wiki (Ebner, Holzinger, & Maurer, 2007)、Social writing platforms, 如：Facebook, Myspace等 (Alexander, 2006)等，透過科技協助，除了保持傳統寫作的優點，還能與許多人分享、並透過文章快速地與許多人溝通，同時能便捷地與相關資訊連結。

在寫作的研究進展中，寫作與自我效能的關係是主軸之一（Tynjälä, et al., 2001）。先前的研究指出寫作表現與寫作自我效能是互有高相關，而且寫作自我效能是寫作表現的良好預測指標(Meier, McCarthy, & Schmeck, 1984; McCarthy, Meier, & Rinderer, 1985)。在寫作與自我效能的相關研究（彭琪媛，民 96; Pajares, Hartley, & Valiante, 2001）中將寫作的自我效能聚焦

於「寫作技巧」，然而寫作所需要重視應還包括「對於寫作成果優劣的自我歸因」，以及「對於寫作任務的認知」。因此本研究試圖發展適合的寫作自我效能量表，讓學習者了解自身自我效能與寫作能力的關係，也將能讓系統發展者設計出輔助寫作的環境，藉以改善與補足過去研究的限制。

2. 文獻探討

2.1 自我效能

自我效能起源於自 Bandura 的社會學習理論，他認為人類的行為不僅受到結果的影響，而且受到人的認知所形成對於結果期望的先行因素影響(Meier et al., 1984)。根據 Bandura (1986) 的定義「自我效能是指一個人判斷其執行一連串行動的能力，強調不僅一個人所擁有的技術能力，而是這個人運用這些技能達成任務的能力判斷」。此定義明確的陳述「技能成分」與「運用技能達成任務的能力」是截然不同的，也就是說個人相信自己具有達成目標行為表現的能力(Bandura, 1977)。此外，個人所經驗的行動結果會被視為一種關於後續表現的自我回饋，所以個人成功與否的經驗會影響其後續的行動表現。因此，寫作自我效能即是個人對自己完成寫作任務的一種能力判斷，也是影響寫作行為最主要的自我規範機制之一，亦即當面對寫作任務時，個人是否主動地全力以赴，取決於個人對寫作任務性質的認知及個人實力的評估。

2.2 寫作與自我效能的關係

近年來，人們對自我效能在作文過程中的影響做了一些卓有成效的實證探討，包括年齡發展趨勢、性別與種族差異、學習策略等面向(司繼偉，2000；Meier et al., 1984)。由於自我效能在已有技能的整合和應用之間發揮直接的中介作用，因此寫作自我效能對學生作文品質的影響應該會隨著他們學會使用寫作的認知技能和行為技能而趨於明顯。Shell 等人考察了大學生的自我效能，結果預期和寫作成績的關係具有中等的相關，以回歸分析發現，自我效能能夠解釋寫作成績的顯著差異，且此自我評價觀念的作用獨立於實際寫作策略的使用。Pajares 及 Valiante (1997) 的研究結果顯示女生在寫作自我效能比男生高，並且感到寫作更為有用。

目前大多數的研究者偏向採用在短時期內設計不同實驗條件比較學生的自我效能對其作文成績的影響程度。而利用量表進行寫作自我效能的研究並不多(Pajares et al., 2000; 彭琪媛，民 96)，這類的量表又多以重視寫作技巧。而本研究以量表作為寫作自我效能的資料收集方式，並且增加寫作自我效能的向度。

3. 研究方法

3.1. 研究工具

本研究依據概念架構及參考 Pajares 等人(2001)及彭琪媛(2007)的問卷題目編擬 40 個項目及 5 題社會讚許性問題，共計 45 題。涵蓋下列五個向度：(1)個人能力信念、(2)個人歸因信念、(3)寫作成果信念、(4)對寫作的認知信念、(5)寫作態度信念。此初探性量表設計以網路填答的方式，採用李克特式六點量表由受試者自行評量自身對於英文寫作的心理經驗與行為狀況。本研究亦設計研究對象背景問卷，以了解研究對象之基本資料以及英文學習的情況。

3.2. 研究對象

刪除無效問卷後，所收集之問卷對象共 113 位學生，分別來自各學院各年級，平均年齡為 20.7 歲(標準差為 1.67 歲)。以性別區分，女性較多共計 67 名(佔 59.3%)，男性共計 46 名(佔 40.7%)。以學院別區分，以管理學院 48 名(佔 42.5%)為多，次者為理學院共計 25 名(佔 22.1%)，地科學院 13 名(佔 11.5%)、文學院 10 名(佔 8.8%)、工學院 9 名(佔 8%)、資訊電機學院 8 名(佔 7.1%)。以年級區分，大三共計 35 名為多(佔 31%)，次者為大二、大四分別為 27 名(佔 23.9%)、大一最少共 24 名(佔 21.2%)，無研究生。

3.3. 資料分析

本研究分析採用項目分析，以作為篩選或修改個別題項的依據；再以因素分析，檢驗量表的構念效度。本研究的因素分析分為兩個部分，首先進行探索性因素分析，藉以探索量表的因素結構，進一步進行驗證性因素分析，藉以確認量表的因素結構。

4. 研究結果

4.1. 項目分析

項目分析主要目的是在針對預試題目進行適切性的評估。除了遺漏狀況的檢視之外，並進行描述統計檢驗（包括平均數、標準差、偏態係數）、極端組比較等。

4.1.1. 描述統計檢測 本量表為六點量表，中間值為 3.5；各項目的平均數為 2.83~5.26；標準差為 0.77~1.40。檢驗結果發現，項目平均數明顯正偏離，而標準差均十分理想，顯示題目的基本描述統計特徵良好。

4.1.2. 極端組比較 將全體樣本依量表總分的前後 27%極端區分為高、低分組，結果顯示只有 6 題無法鑑別高低分者。其餘題目皆達 .001 顯著水準，顯示這些的題目皆有良好的鑑別度。

4.2. 探索性因素分析 以 113 位預試樣本所進行的主軸因子法，經斜交轉軸後，取特徵值大於 1，經將因素負荷過低或不穩定的題目刪除後，僅餘 5 個因素，共計 20 題，結果請見表 1。五個因素的總解釋變異量為 49.98%。

表 1 寫作自我效能量表的因素矩陣

題目	因素				
	1	2	3	4	5
22.我會主動和同學或朋友交換閱讀作品。	0.785				
45.我主動讓同學或朋友閱讀我的文章。	0.743				
29.我會選擇有挑戰性的題目來寫作。	0.687				
38.不論身在何種學習環境裡，我都會把握練習用英語寫作。	0.562				
14.即使別人評斷我的文章寫不好，我還是會對寫作過程中學到其他東西而高興。	0.538				
12.我能冷靜地面對各種批評，因為我信賴自己的寫作能力。	0.453				
23.我覺得寫作是一種重要的溝通能力。		0.917			
21.寫作是生活中重要的學習能力。		0.755			
28.我認為寫作本身就是一項問題解決的過程。		0.503			
42.我認為寫作只是書寫技巧的練習。		-0.355			
15.我認為寫作是一項很複雜的認知歷程，包括形成概念、組織概念及統整想法等。			0.616		
10.我認為文章的好壞不能用分數來代表。			0.541		
11.我認為完成寫作就像是完成一項藝術作品。			0.492		
9.對我來說寫作並非是一件簡單的事情。			0.393		
39.寫作時，我有把握可以寫出我的想法。				0.687	
43.我覺得寫作時我寫什麼內容我就會學到什麼內容。				0.537	
1.進行英文寫作時，我能正確地使用各種基本的英語寫作技巧。(例如：拼字、字詞、文法、時態、標點等)				0.419	
37.我會主動請教老師或同學有關寫作的問題。					0.624

4.當我受到讚美或鼓勵時，我會在下次的寫作表現上更好。	0.512
31.我認為如果努力多練習，我的寫作能力會增進。	0.477
解釋的變異量%	29.49 7.12 5.82 4.61 2.95
累積的解釋變異量%	29.49 36.60 42.42 47.03 49.98

經由因素分析得出五個因素之後，依據各因素內部的題目意義將各因素加以命名如下：(1)「寫作態度」的信念：指學生能主動積極進行英文寫作任務的信念，例如「我會主動讓同學或朋友閱讀我的文章」、「不論身在何種學習環境裡，我都會把握練習用英語寫作」；(2)「寫作任務」的信念：學生對於英文寫作這項任務的看法，例如「寫作是生活中重要的學習能力」、「我認為寫作本身就是一項問題解決的過程」；(3)「寫作歷程」的信念：學生如何看待英文寫作的歷程，如「我認為寫作是一項很複雜的認知歷程，包括形成概念、組織概念及統整想法等」、「我認為文章的好壞不能用分數來代表」；(4)「寫作技巧」的信念：學生對於英文寫作時的技巧運用之信念，如「寫作時，我有把握可以寫出我的想法」、「進行英文寫作時，我能正確地使用各種基本的英語寫作技巧。」；(5)「寫作回饋」的信念：學生如何因應其他人如何給予回饋的信念，如「我會主動請教老師或同學有關寫作的問題」、「當我受到讚美或鼓勵時，我會在下次的寫作表現上更好」。

4.3.驗證性因素分析

在預試階段，量表的因素結構經過探索性因素分析，發現了五個因素，再以此樣本進行驗證性因素分析，利用 AMOS7.0 軟體，以最大概似法 (Maximum likelihood) 進行參數估計。模式適配分析結果： $\chi^2=342.18$, $p<.001$, $df=160$, $\chi^2/df=2.14$, $RMSEA=.101$, $CFI=.751$, $TLI=.673$ ，顯示理論模型與觀察分數適配度未盡理想。

4.4.效標關聯效度分析

以學英文的年級、現在讀英文的時間、寫英文作文的時間與 5 題的社會讚許性問題為本研究的效標變項，分析這些效標與量表中各因素的相關，其結果如下表 2。

表 2 英文寫作自我效能量表與效標變項的相關

因素名稱	學英文 的年級	讀英文 的時間	作文的 時間	s_1.有時為 寫英文 了討好別 人，我會說 違心之論	s_2.失敗之 後我總是立 刻爬起重新 出發	s_3.我總是 以客觀的立 場來看待事 情。	s_4.對於不 懂的事，我 會裝得很懂 得樣子	S_5.我總是 以正面的態 度面對別人 批評
F1 寫作態度	.04	.30**	.39**	.00	.32**	.10	.03	.35**
F2 寫作任務	-.02	-.03	.03	.02	.22**	.03	-.07	.06
F3 寫作歷程	.22*	.06	-.03	.02	.17	.08	-.18	.33**
F4 寫作技巧	-.15	.00	.12	-.32**	.21**	.03	-.03	.00
F5 寫作回饋	-.14	.32**	.17	-.10	.18	.07	.04	.15

* $p<.05$ ** $P<.01$

因素一寫作態度與讀英文的時間 ($r=.30$)、寫英文作文的時間 ($r=.39$) 社會讚許性問題 2 ($r=.32$) 跟問題 5 ($r=.35$) 的相關達顯著 ($p<.01$)，意即寫作態度越好者，其讀英文的時間、寫英文作文時間則越高；因素二寫作任務僅與社會讚許性問題 2 ($r=.22$) 的相關達顯著 ($p<.01$)，意即無論學生如何看待寫作這件事情，都與英文的效標無關；因素三寫作歷程與學英文的年級 ($r=.22$, $p<.05$) 及社會讚許性問題 5 ($r=.33$, $p<.01$) 的相關達顯著，即當學生越小學習英文，對於英文寫作較持結果論的態度，而年紀越大學習英文，則較重視英文寫作的歷程學習；因素四寫作技巧僅與社會讚許性問題 1 ($r=-.32$) 及問題

2 ($r=.21$) 的相關達顯著 ($p<.01$)；因素五寫作回饋與讀英文時間 ($r=.32$) 的相關達顯著 ($p<.01$)，即讀英文時間越長者，其越重視他人給予的寫作回饋。

4.5. 背景變項分析

4.5.1. 性別變項 性別在五個因素的差異檢定，係以多變量變異數分析來檢驗。經 Wilk's λ 考驗未達顯著水準 (Wilk's $\lambda=.97$, $p>.05$)，表示不同性別的大學生在英文寫作自我效能的整體層面上沒有差異。

4.5.2. 學院別 大學生的不同學院別在整體多變量變異數考驗上未達顯著差異 (Wilk's $\lambda=.70$, $p>.05$)，表示不同學院別的大學生在英文寫作自我效能的整體層面上沒有明顯的差異。針對各因素進行不同學院別差異檢定時發現，只有「寫作態度」($F=3.74$, $p<.01$)及「寫作技巧」($F=3.46$, $p<.01$)達顯著，經事後檢定，均為文學院與理學院兩者差異達到顯著，即文學院在寫作態度與寫作技巧上面的得分高於理學院。其餘的三個面向並沒有顯示學院別的差異。

4.5.3. 年級別 大學生的不同年級別在整體多變量變異數考驗上未達顯著差異 (Wilk's $\lambda=.90$, $p>.05$)，表示不同年級別的大學生在英文寫作自我效能的整體層面上沒有明顯的差異。針對各因素進行不同學院別差異檢定時發現，這五個面向亦無顯示年級別的差異。

5. 結論與限制

5.1. 研究結論

5.1.1. 英文寫作自我效能的因素結構 本研究目的探討大學生的英文寫作自我效能，編製了一份「英文寫作自我效能」量表，經過項目分析、因素分析共得到 20 題及五個因素，分別為「寫作態度」、「寫作任務」、「寫作歷程」、「寫作技巧」與「寫作回饋」。以三個效標變項及 5 題的社會讚許性問題進行效度評估，結果指出這五個因素和英文寫作自我效能的效標關聯程度相當分散且關係較低。此外，本研究亦進行英文寫作自我效能的驗證性因素分析之驗證，以評估此測量工具的因素結構是否恰當，結果顯示假設此模型的適配度未盡理想，表示英文寫作自我效能量表的五個測量向度需要再進行修正。

5.1.2. 英文寫作自我效能可以作為大學生自我效能能力判斷的工具 從背景變項的分析發現，沒有顯示性別或是年級的差異。經事後比較「寫作態度」及「寫作技巧」此二面向顯示文學院學生從事英文寫作時，其寫作態度與寫作技巧會影響英文寫作的知覺信念高於理學院學生。

本研究的結果發現，大學生在從事英文寫作時，對於本身英文寫作能力以及對英文寫作的知覺信念，可以藉由評量工具來測得個別差異。然而，本研究的量表仍有修正的空間，因為重新審視第一次量表設計的狀況，許多問題具有兩個面向的意含，進行因素分析時，其因素負荷量便會分散到不同因素，因此題目刪題的狀況也就比較多。而且根據驗證性因素分析的結果顯示，需要更進一步進行模式修正。

5.2. 研究限制

本研究在樣本的蒐集上，礙於人力與人際網絡的限制，僅能透過網路的傳遞讓研究參與者自願填寫問卷，無法以隨機抽樣的方式來選取參與者，雖在樣本蒐集過程中，盡可能的將樣本予以分散，避免過度集中。然而本研究對象均為國立大學的學生，因此樣本取樣上可能無法顧及其他學校類型的大學生，例如私立學校、科技大學等，因此抽樣問題為本研究的限制。

參考文獻

- 司繼偉 (2000)。寫作自我效能感的研究進展。《心理學動態》，8(3)，50-55。
- 彭琪媛 (2007)。網路日記對寫作表現及寫作自我效能影響之研究-以朝陽科技大學學生為例。朝陽科技大學應用外語研究所碩士論文，未出版，台中縣。
- Alexander, B. (2006). Web 2.0: A new wave of innovation for teaching and learning? *Educause Review* 41, 33-44.

- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs: Prentice Hall.
- Bandura A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Ebner, M., Holzinger, A., & Maurer, H. (2007). Web 2.0 Technology: Future Interfaces for Technology Enhanced Learning? *Lecture Notes in Computer Science 2007*, 4556: pp. 559–568.
- Klein, P. D. (2000). Elementary students' strategies for writing-to-learn in science. *Cognition and Instruction*, 18(3), 317-348.
- McCarthy, P., Meier, S., & Rinderer, R. (1985). Self-efficacy and writing: A different view of self-evaluation. *College Composition and Communication*, 36(4), 465-471.
- Meier, S., McCarthy, P. R., & Schmeck, R. R. (1984). Validity of self-efficacy as a predictor of writing performance. *Cognitive Therapy and Research*, 8(2), 107-120.
- Nelson, N. (2001). Writing to learn: One theory, two rationales. In P. Tynjälä, L. Mason, K. Lonka (Eds.), *Writing as a learning tool* (pp. 23-46). Norwell, MA: Kluwer Academic Publishers.
- Pajares, F., Hartley, J., & Valiante, G. (2001). Response format in writing self-efficacy assessment: Greater discrimination increases prediction. *Measurement and Evaluation in Counseling and Development*, 33(4), 214-221.
- Pajares, F., & Valiante, G. (1997). Influence of writing self-efficacy beliefs on the writing performance of upper elementary students. *Journal of Educational Research*, 90, 353-360.
- Tynjälä, P., Mason, L., & Lonka, K. (2001). Writing as a learning tool: An introduction. In P. Tynjälä, L. Mason, K. Lonka (Eds.), *Writing as a learning tool* (pp. 7-22). Norwell, MA: Kluwer Academic Publishers.

高校数字校园建设水平评价指标体系的研究

An Evaluation Index System for Digital Campus of Higher Education Institutions

韩锡斌、刘英群、张明、周潜

清华大学教育技术研究所

邮件信箱: hanxb@tsinghua.edu.cn

【摘要】 高校数字校园建设水平的评价指标体系是实施高校数字校园建设水平评价的基础。本文通过文献分析和高校调研的方法, 确定了数字校园建设水平评价的目的、评价对象的内涵和组成部分、评价指标构成, 并提出了评价指标量化计算的方法。

【关键词】 评价指标体系、高校数字校园、高校信息化、指标量化

Abstract: An evaluation index system is the basis of assessing digital campus of higher education institutions. With the research method of document analysis and questionnaire investigation in Chinese mainland universities, this paper describes the purpose of the evaluation and the meaning of digital campus, and presents the evaluation indexes and the method to calculate the scores of every index.

Keywords: evaluation index, digital campus of higher education institutions, IT used in higher education institutions, quantitative index

1. 前言

教育信息化是教育发展历史上最为深刻的变革之一。高校数字校园建设与应用是教育信息化的重要组成部分。中国大陆高校信息化建设进展很快, 但普遍存在着资金大量投入而成效不佳的问题, 制约了高校信息化的发展。

作为教育主管部门需要采取有效措施促进高校数字校园的建设, 提高校园网的应用水平和使用效率。近年来各级教育主管部门在资金投入、相关政策与法规、技术标准等方面均有大的举措, 取得了明显的成效, 但尚未对高校数字校园建设的水平进行过全面系统的评价。要实施评价, 制定一套科学而可行的指标体系是关键, 本文将探讨高校数字校园建设水平的评价指标体系。

2. 相关研究分析

随着信息技术的飞速发展, 信息化的浪潮席卷全球, 国内外的专家学者和从业人员从各个角度对于国家信息化、地区信息化、城市信息化、企业信息化、政府信息化和教育信息化等展开了相关研究和实践, 并且在各个领域的信息化评价领域开展工作, 产生了许多研究成果。

2.1. 宏观层次的信息化测评研究

上世纪六十年代美国的 Machlup 提出了 GDP 比重法 (Machlup, 1962), Porat 和 Rubin 提出了就业结构分析法 (Mach, Porat & Rubin, 1977), 他们通过信息产业对国民经济的贡献来测度国家或地区信息化发展水平。日本学者小松崎清提出了信息化指数法 (小松崎清, 1965), 从邮电、广播、电视、新闻出版等行业中, 选出 11 个要素共 4 组指标, 即信息量 (Q)、信息装备率 (E)、通信主体水平 (P) 和信息系数 (V), 由此计算出一个反映社会信息化程度的总体指标——信息化指数。这些宏观层次的测评研究虽然只能用作宏观层面的信息化测量, 不能直接运用于高校信息化评价, 但是却为信息化水平的测算提供了重要手段, 尤其

是信息化指数测度方法的测量结果是无量纲的相对量，即在确立某一基准点的前提下的相对比较值，具有较好的可比性。

2.2 微观层次的信息化评价研究

与国家和地区层面的信息化评价不同，各类组织，如企业、政府和高校等，信息化评价与其特殊的行业性质紧密相关，都更强调信息技术的“投入/产出”的信息化效率分析，以及管理对信息化价值实现的作用，因此其评估体系具有更大的复杂性。2002年10月9日，中国国家信息化评测中心（NIEC，National Information Evaluation Center）正式推出中国第一个面向效益的信息化指标体系——《中国企业信息化指标体系》（国家信息化评测中心，2002）。国际上开展电子政务评价的咨询企业主要有 Accentrue 和 Gartner 咨询公司。

2.3 高等院校信息化评价研究综述

（1）高校信息化调查研究

高校信息化调查有针对性地调查高等教育信息化发展过程中存在的热点问题，并根据调查结果对政府相关教育行政部门政策制定者和高校提供信息化发展的建议。国际上开展的高校信息化调查主要包括 CCP（the Campus Computing Project）、EDUCAUSE 的 CDS（Core Data Survey）和 CIS（Current Issues Survey）。中国大陆高校信息化调查研究主要是 2002 年和 2004 年北京大学教育学院教育技术系进行的中国高等教育信息化发展调查研究（赵国栋，2005）。

（2）高校信息化评价指标体系的研究

此研究思路提出高校信息化评价指标体系，对高校信息化的整体状况给出定性或定量的测度，从而达到对参与评价的各个高校进行评比的目的。但是，目前此类方案尚无得到广泛认同的评价指标体系。这类思路的典型研究成果有：美国的 STaR（School Technology and Readiness）评估量表（CEO Forum on Education and Technology, <http://www.ceoforum.org>）。EDUCAUSE 制定的 “The Student Guide to Evaluating Information Technology on Campus”（EDUCAUSE Website, <http://www.educause.edu>）。《关于我国高等学校信息化评价指标体系构建的理论探讨》提出的评价指标体系（安宝生和衷克定，2004）。上海复旦大学研究制定了《上海市高校信息化评价指标体系及参考权重》（张成洪，2006）。

3. 研究方法

（1）确定高校数字校园调研评价指标

分析高校数字校园的内涵以及影响高校信息化的相关非技术因素的分析，以期考察高校信息化不同于其他行业信息化的特殊性，得出高校数字校园的组成部分，从而初步提出一级指标项。分析已有的高校信息化测评的相关文献，归纳出国内外已有的 21 套高校信息化的评价指标体系。通过对比分析这 21 套高校信息化评价指标体系中指标，本着尽可能全方位反映数字校园具体内涵的原则，选择了 110 个调研指标项供调研参考。

（2）高校数字校园调研

1) 高校实地访问调查：根据上述调研指标项制作了高校信息化现状访谈问卷，选定高校后由专人实地调研了 24 家高校的信息化数据。调查过程中采集数据的同时还要记录和反馈指标框架中存在的问题，反馈结果对指标体系进行修正。

2) 网络普查：由于教学是高校的中心工作，以往的指标体系中对教学的评价研究较少，因此，本课题针对教学信息化，采用网络普查、教师访谈和问卷调研方法进行深入研究。对中国大陆 742 所本科院校的网络教学平台拥有情况以及基本的使用情况进行了网络普查。

3) 教师问卷调查: 问卷调查的方式是到不同高校或利用研讨会的机会进行面对面的问卷填写, 确保问卷的质量。此次调研涉及院校 20 所, 其中本科院校 17 所, 高职高专类院校 3 所。共计回收有效问卷 313 份。同时又进行了网上问卷调查, 回收教师有效问卷 124 份, 学生有效 2550 问卷份。

(3) 高校数字校园指标体系的确定: 根据调研结果, 确定高校数字校园指标体系中的层级关系和各指标的权重。根据可操作原则, 提出指标的量化计算办法。

4. 分析讨论

综合文献分析结果, 对评价指标体系的研究需要阐明四个问题: 评价的目的、评价对象的内涵和组成部分、评价指标构成和评价指标的量化计算。

(1) 高校数字校园评价的目的

所谓“评价的目的”就是评价结果给谁用于何种目的(安宝生和袁克定, 2004)。

高校信息化评价的目的大致可以分为三类: 第一类是由研究机构针对大学进行调研, 提出研究报告, 阐明信息化的现状、存在的问题和发展趋势, 其结果可以供主管政府部门、高校和研究者参考。第二类是由政府主管部门主导, 对学校信息化水平进行评估, 目的是检测各学校信息化发展状况, 引导学校信息化建设的方向, 为政策学校信息化政策制定提供信息依据。第三类是由研究机构面向学生提供各大学信息化的应用水平, 为学生选择大学提供参考。

本文的研究是“北京市教育科学十一五规划课题”项目的一个部分, 其主要目的就是为北京市教育主管部门提供政策建议。因此本文的研究属于第二类。

(2) 高校数字校园的内涵与组成部分

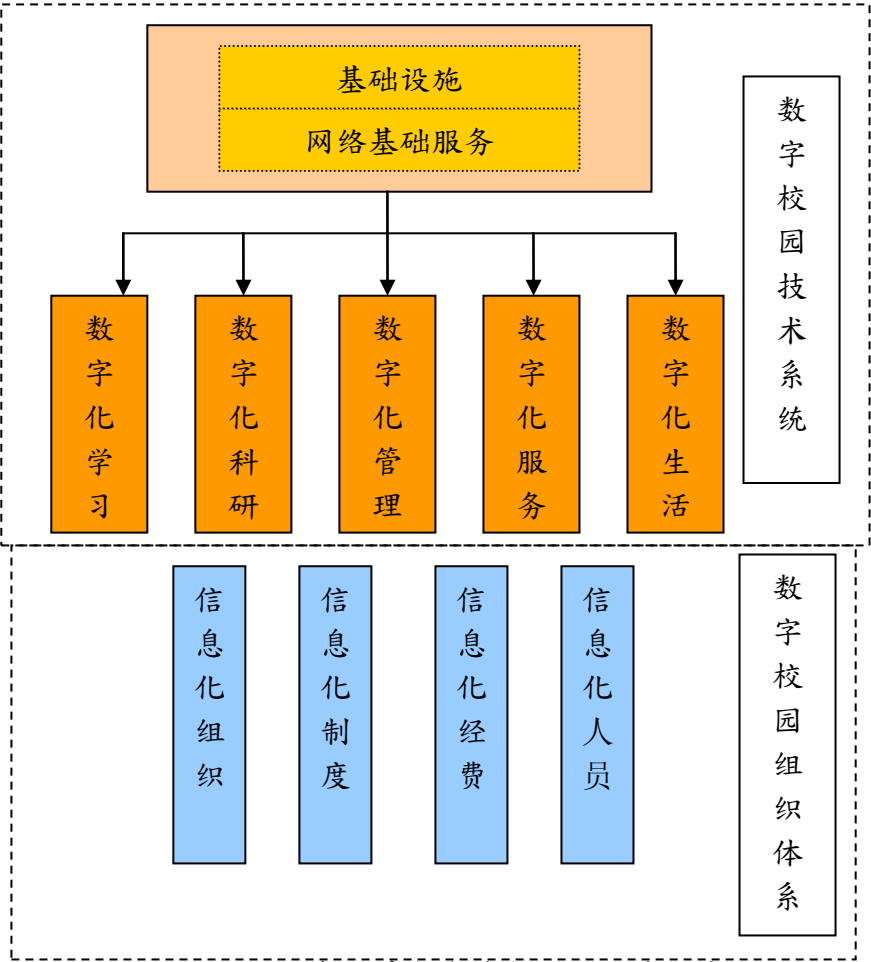
高校数字校园是指利用多媒体、网络技术, 将学校的主要信息资源数字化, 并实现数字化的信息管理方式和沟通传播方式, 从而形成高度信息化的人才培养环境。(全国高校教育技术协作委员会数字校园建设委员会, 2004)。据此得出, 在基础设施建设之上, 高校数字校园建设的重点应该包括五个主要方面: 数字化学习(e-Learning)、数字化科研(e-Research)、数字化管理(e-Management)、数字化公共服务(e-Service)和数字化社区生活(e-Life), 这五个方面是高校数字校园面向应用的五个主要组成部分, 被称为数字校园的 5E 架构(韩锡斌等, 2005)。

一项研究(韩锡斌等, 2005)运用知识管理的思想和方法对数字校园的 5E 架构进行了进一步修正, 认为基于知识管理的数字校园的 5E 架构中, 高校的组织体系建设是整个数字校园建设的有机组成部分。在该架构中除了技术系统之外, 又包含了一个与之并行的组织体系。数字校园组织体系建设涉及很多方面, 已有的许多关于高校信息化评价研究中, 都已考虑了这些因素。

综合以上研究, 高校数字校园包括技术系统和组织体系两大部分(如图像 1 所示), 其中技术系统包括基础设施建设、网络基础服务和五个方面的应用, 即数字化学习、数字化科研、数字化管理、数字化公共服务和数字化社区生活; 组织体系分为四个方面: 信信息化组织、信息化制度、信息化经费和信息化人员。

(3) 高校数字校园评价指标构成

根据高校数字校园建设内容的组成部分(如图像 1 所示)来构建高校数字校园的评价指标体系。参考已有的高校数字校园评价的研究成果, 共设计了 110 个供调研的指标项, 这些指标项尽可能全方位反映数字校园建设的具体内涵。



图像 1 高校数字校园的组成部分

通过在调研过程中对各个高校的实际访谈和分析结果,本课题得出了高校数字校园评价指标体系的最终结果,如表格 1 所示,总共 7 个一级指标项、 15 个二级指标项,44 个三级指标项。

表格 1 最终的高校数字校园评价指标体系

一级指标项	一级指标项分值	二级指标项	二级指标项分值	三级指标项数目
1. 信息化基础设施	100 分	1.1 信息化设施建设水平	50 分	5
		1.2 校园网管理水平	50	5
2. 网络基础服务与公共服务	90 分	2.1 网络基础服务水平	30 分	3
		2.2 图书馆数字化服务水平	30 分	3
		2.3 应用集成水平	30 分	4
		3.1 教学资源的建设和应用	30 分	3
3. 数字化学习	80 分	3.2 教学过程信息化支持水平	20 分	2
		3.3 教学支持服务水平	30 分	3
		3.1 科研管理与活动信息化的水平	40 分	1
4. 数字化科研	40 分			
5. 数字化管理	40 分	4.1 管理信息化的水平	40 分	1
6. 数字化生活	20 分	6.1 生活服务信息化水平	20 分	2

7. 数字校园组织体系	130 分	7.1 信息化组织	30 分	3
		7.2 信息化制度	40 分	4
		7.3 信息化经费	30 分	3
		7.4 信息化人员	30 分	3

在上述指标项的确定时，遵循以下五个原则：

- 导向性：从注重基础设施建设转向注重数字校园主要领域的信息化建设，从注重硬件建设转向注重应用软件建设，从注重系统构建转向注重系统的深入应用，从注重技术系统的建设转向技术系统与组织体系建设并重，从注重信息化面向教职员工服务转向面向学生服务；
- 简约性：尽量选取较少的指标反映较全面的情况，所选指标要具有一定的综合性，不同指标之间的相关性尽可能小；
- 可操作性：所选取的指标含义明确，便于从高校相关部门进行数据采集，同时便于量化计算；
- 有效性：所选取的指标的量化指数应该有一定的区分度，便于测量不同高校信息化的水平差异；
- 可延续：所设计的指标体系的框架（一级指标项）可随时间相对稳定，而二级指标尤其是三级项可以在内容上调整和拓展。

（5）高校数字校园评价指标量化计算方法

为了便于高校进行自测和横向对比分析，需要将高校数字校园评价指标体系进行量化测量计算。

日本学者小松崎清提出了信息化指数法（小松崎清，1965），其测度结果是无量纲的相对量，即在确立某一基准点的前提下的相对比较值，具有较好的可比性。本文也采用该思路，给每个三级指标设定能够进行量化测量的考察点，获得考察数据，再通过指标加权获得无量纲化总指数，从而反映高校数字校园建设与应用的状况。量化的方法有以下两种：

第一类指标：能够通过量化的观测值直接获取考察数据，判定该考察数据在哪个取值区间，再将其所在取值区间的指标分值作为该指标的无量纲指数。取值区间的划分依据目前的公认经验值和调研结果。

第二类指标：不能通过量化的观测值直接获取考察数据，需要将构建的系统、采取的措施或提供的服务等通过选项来表示，每个选项赋予一定的分值，选择的选项越多（或越少）则该指标项的得分就越多（或越少），从而将定性描述的指标转化为可以定量测度的无量纲分数。每个选项的得分也是依据调研结果和关于该选项重要性的经验判断。

高校数字校园指标总分为 500 分，各指标项的分值如表格 1 所示。

5. 结论

本文对高校数字校园的评价指标体系、评价方案和网上评价的软件系统原型三个方面进行了研究，研究结论如下：

- （1）提出了高校数字校园建设水平评价指标的框架，该框架涵盖了高校信息化建设的五个主要方面：数字化学习、数字化科研、数字化管理、数字化公共服务和数字化社区生活，同时考虑了四类非技术因素：信息化组织、信息化制度、信息化经费、信息化人员。
- （2）在上述框架基础上，提出了测定高校数字校园建设水平的评价指标体系，该指标体系 7 个一级指标项、 15 个二级指标项，44 个三级指标项。上述指标项都是通过在校内实际调研和访谈的基础上确定的，具有较强的可操作性。
- （3）访谈与问卷调查结果表明：数字校园建设中对于作为学校中心工作——教学的支持远远不够，不仅体现在教师课程教学中信息技术手段应用不足，而且也反映在校园网对学生

基于网络的学习支持不够，因此在进行评价指标和权重设定时突出了信息技术对课程教学支持及其效果体现的重要性。

参考文献

- 安宝生，衷克定（2004）．关于我国·高等学校信息化评价指标体系构建的理论探讨。《中国电化教育》。2004，（4）：24~27。
- 国家信息化评测中心(2002)。中国企业信息化指标体系。
http://ciq.com.cn/tx_jbzhb.htm。
- 韩锡斌，杨娟，陈刚（2005）。基于知识管理的大学数字校园的概念、架构和策略。《中国远程教育》。2005，（8）：39~42。
- 全国高校教育技术协作委员会数字校园建设委员会（2004）。建设数字化学习环境的规划方案（讨论第四稿）。2004 高校教育技术协作委员会数字校园会议。
- 小松崎清(日)(1994)。《信息化与经济发展》。北京：社会科学文献出版社。
- 赵国栋(2005)。关于中国高等教育信息化发展状况的调查与分析。《中国远程教育》。2005，（8）：44~48。
- 张成洪（2006）。《高校信息化的规划与评价》。上海：复旦大学出版社，164~166。
- Machlup, F. (1962). The Production and Distribution of Knowledge in the United States. New Jersey: Princeton University Press, 1962. 107~108.
- Porat, M. U. & Rubin M. (1977). The Information Economy: Development and Measurement. Washington, D.C.: Government Printing Office, 1977. 424~501.

基於連續機率比檢定模式之電腦化自然發音精熟測驗遊戲的設計與發展

The design and development of the Sequential Probability Ratio Test based

computerized mastery testing for Phonics game

林逸農、王詩婷*、林政廷**、楊接期***

國立中央大學網路學習科技所

輔仁大學心理所*

國立中央大學網路學習科技所**

國立中央大學網路學習科技所***

【摘要】 隨著台灣對於國際化的重視，英語學習開始向下延伸成為小學生的學習內容。英語學習的基礎在於發音的正確性，因此精熟正確發音與否成為學習的重點。而連續機率比檢定模式提供精熟與否的參考。本研究以自然發音為學習內容設計樂趣化的數位學習教材與評量，並結合測驗理論發展成可做診斷功能的系統。診斷為精熟者，表示學習者對於學習內容的瞭解程度較好；未精熟者，且可列為補救教學。本研究所開發的測驗遊戲仍在初始階段，希冀能讓學習者在使用時，提供自然發音的學習精熟情況，收集更多學習表現，提供適性的輔助。

【關鍵詞】 連續機率比檢定模式、精熟測驗、樂趣化學習、自然發音遊戲

***Abstract:** With attaching great importance to the internationalization in Taiwan, English learning begins at the elementary. Pupils should take English class. The foundation of English learning is the accuracy of pronunciation, and therefore learner's mastery of the accuracy of pronunciation is very important. And the Sequential Probability Ratio Test (SPRT) offers a reference standard of mastery. The practicality of the propose methodology is demonstrated through an digital game-based learning with learning contents of phonics. The development of game combines with testing theory which can diagnose the learning performance. If the result is "expert (mastery)", it means that learners' understanding of learning content is very well. If the result is "beginner (non-mastery)", the result indicates that English remedial teaching improves underachievers' English performances. Such research is still in its infancy, but we hope that the digital game-based learning can offer learner the level of mastery for Phonics and collect more learning behaviors to offer suitable assistance.*

Keywords: Sequential Probability Ratio Test, mastery testing, digital game-based learning, Phonics game

1. 研究背景與動機

英語學習的推動一直是台灣教育工作重視的部分。目前教育部已經策劃從國小三年級開始進行英語學習課程(教育部，2006)由此可知英語學習的重要性。對於英語學習者而言，認知發音正確與否是英語學習很基礎的階段(Bus & van IJzendoorn, 1999)，而發音技巧的習得可有助於學習者在音韻覺知(phonemic awareness)上有更充分的遷移應用(Ehri et al., 2001)。而網際網路的普遍和多媒體技術的進步，許多教材也在網路上呈現，並與遊戲互相結合，發展為數位式的教育遊戲教材。

在遊戲教材的設計上，除了重視學科內容的學習外，也必須重視評量的部分。許多遊戲教材，經由遊戲得分來表示學習者的能力表現，但是少有遊戲教材討論到，分數要多少才是適當的。從教學和評量的角度來看，標準參照提供教學上一個重要的依據(吳裕益，1986)，及格與精熟與否都是評量上需要討論的重點。本研究由學習精熟的角度來探討自然發音的學

習表現，以精熟與否的觀點作為學習表現的依據。採用連續機率比檢定(Sequential Probability Ratio Test; SPRT)作為精熟與否的參照標準(Wald,1947)，以遊戲方式結合自然發音的學習內容，讓學習者在遊戲中測驗，並診斷學習單元的精熟與否。

本研究目的，主要開發以連續機率比檢定(SPRT)為基礎的測驗，透過遊戲方式整合自然發音的學習內容，發展自然發音精熟測驗遊戲。希冀學習者經由遊戲過程能有學習的評量，並且瞭解自己對於學習內容的精熟情況。

2.文獻探討

2.1. 連續機率比檢定模式

連續機率比檢定(SPRT)，是Wald根據機率模式所推導出的決策模式(Wald,1947)。Ferguson(1969)將其應用於教育測驗的領域。近期部分研究開始發展電腦化精熟測驗，做為精熟與否的決策模式(李茂能，2001；Ferguson, 1969)。測驗時採隨機選題，根據受試者每次的答題反應，估算能力值，並判定受試者是否達到精熟(mastery)或未精熟(non-mastery)水準。SPRT採用「連續答對或答錯題目的可能性機率」 λ 值，其公式為 $\lambda = (P_m/P_{nm})^{\sum r} (1-P_m/1-P_{nm})^{\sum w}$ ，(其中 P_m 為預設精熟者最低答對率； P_{nm} 為預設未精熟者最高答對率； r 為對答情形， w 為錯答情況。)，來與事先設定之精熟決斷點(Upper Decision Point, UDP)，其公式為 $UDP = (1-\beta)/\alpha$ 與未精熟決斷點(Lower Decision Point; LDP)作比較，其公式為 $LDP = \beta/(1-\alpha)$ 。 λ 值若 $\geq UDP$ 為「精熟者」；若 $\leq LDP$ 為「未精熟者」。而 α 、 β 表示誤差容許度； α 表示型 I 錯誤(Type I error)， β 表示型 II 錯誤(Type II error)。研究中發展的診斷精熟測驗的參數設計以Frick(1986)的論點，當 $P_m=.85$ ， $P_{nm}=.6$ ，且 $\alpha=\beta=.025$ 時，SPRT 模式與傳統精熟測驗結果的一致性高達0.84至0.98。

2.2. 精熟診斷測驗

由教學的觀點來看，精熟診斷測驗是教師為了檢驗學生的學習成果，在教學前事先訂定精熟標準，在教學告一段落後，對學生進行以教材內容範圍的測驗，並判定學生是否達到預定的精熟標準(余民寧，1997)。精熟診斷測驗的目的，旨在瞭解學生是否熟練教材內容及是否達到教師預期的成就水準。

2.3. 樂趣化學習

數位遊戲式學習(Digital Game-Based Learning)又稱為樂趣化學習，主要透過遊戲結合了學習內容，亦可藉由遊戲中的 Flow 過程(Csikszentmihalyi, 1990)，讓被動學習成為主動的參與(Gee, 2003)。近年來其他學者的研究也確認了數位遊戲式學習可帶來正面的教育意義(Kirriemuir & McFarlane 2004; Prensky, 2003;)。Squire(2003)認為，以電玩遊戲的方式進行學習，隨著遊戲的挑戰性、幻想性、好奇心和回饋性等豐富的機制，可以增進學生學習的知識。有學者認為遊戲可以幫助學習(Amory, Naicker, Vincent, & Adams, 1999)或是讓學習趣味化(Malone, 1983)，在這個過程中，可以增加學生玩中學(learning by playing)的學習機會。(Kirriemuir & McFarlane, 2004)。由網路搜尋發現非常多自然發音學習教材與樂趣化學習的整合教材。但結合精熟診斷的測驗概念較少，因此本研究以 SPRT 機率模型來估算學習者的學習精熟表現，讓學習表現除了原始分數外另有精熟分數可以參考。

2.4. 自然發音

在自然發音中，字母與讀音之間的規律性佔有相當大的比例(Crystal, 1987)，而學習自然發音法可以提供學習者豐富的語音線索而有助識字的學習(Adams, 1990)。在自然發音法中，音韻覺知的習得是被所探討的重點，音韻覺知是音素分割(phonemic segmentation)的能力，指將語言設想為一連串的音素區塊(phonemic segments)，由單字和音節中找出這些音素區塊的能力。Treiman 與 Baron(1981)認為音素分割的能力和閱讀上使用字母-聲音的規則的相關程度最高，可見自然發音的學習對於字母-聲音的重要性。音韻覺知能力和閱讀之間具有可預測的關係，而音韻的教學可改善閱讀的表現(O' Connor & Jenkins, 1995)，因此自然發音法的學習是需要被重視的。

3.遊戲式自然發音學習系統的開發

本研究發展的斷系統主要包含兩個部分，第一部份是互動式多媒體自然發音練習教材部分，第二部分是診斷測驗系統部分

3.1.自然發音練習教材

自然發音的學習內容上，以同字異音的教材為學習內容。採用 ou 發[au]和[u]的單字做為學習內容，每個發音提供 8 個英語單字練習，共 16 個單字。此部分以遊戲方式進行，提供兩個關卡(圖 1)。學習者登入後可點選欄位上的半隱藏圖片聽取單字發音，點選下方的英文字按鈕也會出現單字的發音，找到相同的發音，把文字拖曳到欄位上，便會顯現圖片，點選該圖片可聆聽其單字發音。這兩個關卡提供 300 秒給學習者練習，倒數計時的設計讓學習者能有時間的挑戰來完成遊戲。遊戲進行中，學習者所點選、移動物件和時間點皆被系統紀錄做為後續的遊戲行為紀錄。



圖 1 自然發音練習教材學習關卡

3.2.診斷測驗系統

本研究以 SPRT 來發展精熟診斷的測驗系統，試題編製以上述 16 單字做測驗內容。測驗時先讓學生以拖曳方式完成英語單字和中文的配對(圖 2 左)，正確後出現發音按鈕(圖 2 右)，聆聽選出錯誤的發音單字。依照 16 個單字編製 24 題的測驗題，每題包含兩個[au]和兩個[u]，其中一個音是錯的，將[au]唸成[u]，或是[u]唸成[au]。



圖 2 診斷測驗系統

本研究設計的診斷測驗系統流程如圖 3。在教材練習後，開始進行測驗。系統隨機產生題目，當受試者點選答案後，開始進行 λ 的估算，以 UDP 和 LDP 的閾值決定是否繼續測驗和精熟與否。

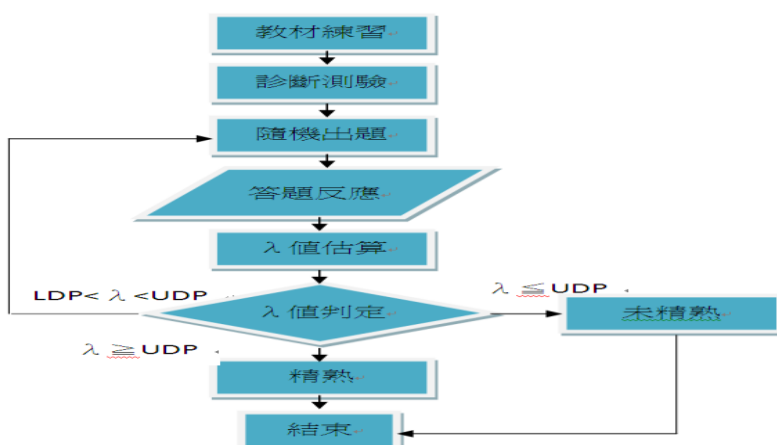


圖 3 診斷系統流程圖

施測結束後，系統會報導結果，以提供受試者知道測驗結果(圖 4)，包含精熟與否、答題數目和正確完成數目。



圖 4 測驗結果

4. 實驗設計與結果初探

本研究進行了初探性實驗，讓學習者練習教材後，進行診斷測驗，並觀察受試者在實驗進行中的反應。

4.1. 教材練習的紀錄分析

表 1 分為[au]和[u]兩部分的教材，分別紀錄完成時間、文字發音(點選英語文字產生發音)、圖片發音(點選隱藏圖片產生發音)和發音複習(完成文字與圖片組合的發音)。從表 1 的數據可知，[au]完成的時間比[u]久，可能是在第一部分的練習上花了許多時間，剩下的時間有限想

盡快完成。此外可以發現幾乎[au]和[u]兩部分的文字發音居多，表示大部分的學童以認識單字來完成練習，不是以聲音配對來找出答案。在發音複習上，學習者練習的次數很少，可能是時間限制所致，無法有充分複習聆聽單字聲音的機會。

表 1 教材練習的紀錄表

代號	[au]完成 時間(秒)	[au]文字 發音(次)	[au]圖片 發音(次)	[au]發音 複習(次)	[u]完成 時間(秒)	[u]文字 發音(次)	[u]圖片 發音(次)	[u]發音 複習(次)
50301	114.342	10	0	0	24.192	11	0	0
50302	209.415	23	3	0	76.16	13	10	0
50308	174.013	26	0	1				
50317	126.861	10	0	0	64.165	25	0	0
50322	243.128	21	0	23	106.303	82	0	8
50324	200.203	33	1	8	86.703	34	0	5
50325	149.021	27	0	1	67.747	42	0	0

4.2. 診斷測驗結果

本研究的初步結果(表 2)，經學習診斷後，發現僅有 4 位學習者精熟。精熟者從答對率來觀察最高是 100%，最低是 86.36%。這和傳統採用百分等級的方式不同，考量學習後是否精熟，其答對率的標準較高。由未精熟者來看，最高答對率是 75%，最低是 57.89%。傳統的觀點認為 70%左右的成績應是中上等能力，但在精熟測驗是列為不精熟，因此精熟測驗在處理學習者是否對於教材內容的熟悉性提供較嚴格的標準。從時間反應來看，可以發現精熟程度和完成時間並非成正比，有的學習者有精熟能力，反應時間也快，可列為優等學習者。有學習者花許多時間，但也有精熟程度，可能是謹慎學習者。

表 2 診斷測驗分析表

代號	答題完成 最少時間(秒)	答題完成 最多時間(秒)	答題完成 平均時間(秒)	總答 題數	答對 題數	答對 率	是否 精熟
5030201	7.03	25.47	11.27	24	17	70.83	否
5030801	7.03	31.56	12.08	19	11	57.89	否
5032201	6.09	32.35	13.09	24	18	75.00	否
5032401	6.87	16.72	10.33	10	10	100.00	是
5032402	5.16	19.28	18.47	22	19	86.36	是
5032501	5.81	22.93	9.69	24	16	66.67	否
5032502	5.20	50.87	15.91	11	11	100.00	是
5032901	8.13	43.91	16.63	11	11	100.00	是

5. 討論與結論

本研究主要以 SPRT 的理論發展診斷測驗，學習和評量內容加入了遊戲的成分，讓學習者在進行中，減少測驗的排斥和增加友善的操作介面，讓學習與測驗上能有更多的親和性。本研究所收集的樣本雖少，卻能反映精熟診斷測驗的精神。精熟診斷測驗強調精熟與否，對於學習精熟的學習者，面對相同難度不同題目有能力解題，是精熟的表現，精熟者在測驗的作答情況是反應出對題目的最低猜測。本研究希冀在題目上能收集大樣本資料來估算題目難度，選用難度相當的試題編製，以提升測驗的信度和效度。

參考文獻

一、中文部分

余民寧(民86)。教育測驗與評量。臺北市：心理出版社。

李茂能(2001)。電腦化適性測驗的過去、現在與未來。台灣教育，604，52-61。

李源煌(民84)。效標參照測驗之題目難度分配型態對循序漸進機率比率決策模式(SPRT)判別學生學習精熟或未精熟之影響。測驗年刊，42，415-430。

吳裕益(1986)。設定標準參照測驗通過標準問題之探討。測驗年刊，32，163-175。

教育部(2006)。國民中小學九年一貫課程綱要—語文學習領域(英語)。2009年3月1日，取自 <http://www.edu.tw>

二、英文部分

Adams, M. J. (1990). *Beginning to Read--Thinking and Learning about print*. MA:MIT.

Amory, A. Naicker, K. Vincent, J. Adams, C. (1999). The use of computer games as an educational tool: Identification of appropriate game types and game elements. *British Journal of Educational Technology*, 30, 311-322.

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91, 403-414.

Crystal, David.(1987).*The Cambridge encyclopedia of language*. Cambridge, England: Cambridge University.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

Ehri, L. C., Nunes, S. R., Willows, D. A., Schuster, B. V., Yaghoub-Zadeh, Ferguson, R. (1969). *Computer-assisted criterion-referenced measurement* (Report WP-41). Pittsburgh: Pittsburgh University, Learning Research and Development Center.

Frick, T. W. (1986). *An investigation of the validity of the sequential probability ratio test for mastery decision in criterion-referenced testing*. Paper presented at the Annual Meeting of the American Educational Research Association.

Gee, J. P. (2003). What Video Games Have to Teach Us About Learning and Literacy? *ACM Computers in Entertainment*, 1(1), 1-4.

Kirriemuir, J. & McFarlane, A. (2004). *Literature Review in Games and Learning*. A Report of NESTA Futurelab. http://www.nestafuturelab.org/research/reviews/08_01.htm

Malone, TW (1983). Guidelines for designing educational computer programs. *Childhood Education*, 59, 241- 247.

O'Connor, R. E., & Jenkins, J. R. (1995). Improving the generalization of sounds/symbol knowledge: Teaching spelling to kindergarten children with disabilities. *The Journal of Special Education*, 29(3), 255-275.

Prensky, M. (2003). Digital game-based learning. *Computer in Entertainment (CIE)*, 1(1), 21-21. *Reading Research Quarterly*, 36, 250-287.

Squire, K. (2003). Design principles of next- generation gaming for education. *Educational Technology*, 43(5), 17-23.

Treiman, R., & Baron, J. (1981). *Segmental analysis ability: Development and relation to reading ability*. In G. E. MacKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice*(vol. 3). New York: Academic Press.

Wald, A. (1947). *Sequential analysis*. New York: Wiley

A Folksonomy Aggregation Concept Tagging Scheme for Collaborative Test Sheet Design

Ching-Yi Liao (廖經益), Shian-Shyong Tseng(曾憲雄), Jui-Feng Weng(翁瑞鋒),
Jun-Ming Su(蘇俊銘)

Department of Computer Science National Chiao Tung University, ROC

sstseng@cis.nctu.edu.tw

Abstract: The idea of Concept tagging for item becomes more and more important for adaptive learning while interpreting the testing result of the student. In particular, while conducting a collaborative test sheet design, the item designed from different teachers may have various concept annotations for the item, synonyms (multiple tags for the same concept), homonymy (same tag used with different meaning), and polysemy (same tag with multiple related meanings) are likely to arise, and the meta noise may cause some incorrect interpretation and the low quality of interpretation. In our work, we propose a Folksonomy Aggregation Concept Tagging Scheme (FACTs) to aggregate the collaborative concept annotations from different teachers. The experimental result shows that the FACTs can helpfully provide the amendment of collectively concept annotation automatically, and can provide the recommendation for the teachers on demand.

Keywords: Folksonomy, concept annotation, IRT, data mining, learning performance.

1. Introduction

In the adaptive learning, if the annotated concepts can be properly tagged on each item, the assessment of the learning performance becomes easier. The testing result can be interpreted by the teacher according tagged concept. Therefore, the concept annotation for the item is important for learning diagnosis and adaptive learning [3][4][5] while conducting adaptive learning for the student. Currently, the available items may be developed separately by different source such as instructors, experts, textbooks, educational publishers, etc. While merging the items to form a test sheet, the tagging problems such as synonyms (multiple tags for the same concept), homonymy (same tag used with different meaning), and polysemy (same tag with multiple related meanings) are likely to arise when items' tags are provided from different sources. In the folksonomy tagging environment, we propose a concept ontology approach to reduce the complexity of concept tagging, but the following resulting issues of concept tagging still need to be solved:

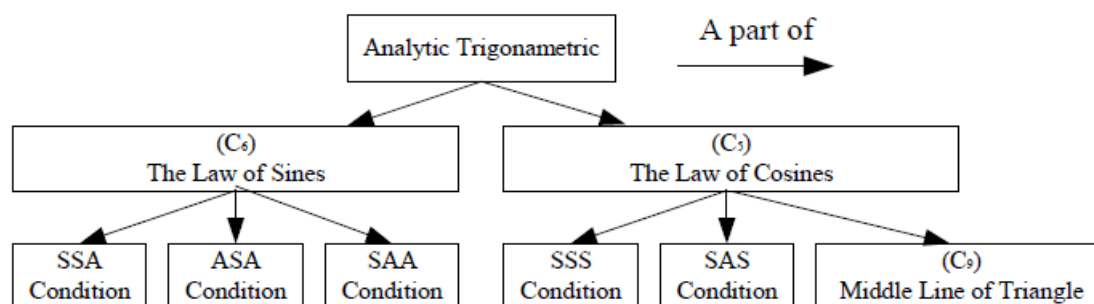


Figure 1. The domain concept ontology of the “Trigonometric” in high school mathematics.

- i) Concept granularity: The concept tagged with different granularity level may have difference interpretation for the testing result. For example, if an “Analytic Trigonometric” test item is tagged with the concept annotation “The law of sines”, it’s concept granularity is different from tagging “SSA condition” (The law of sine applied in SSA condition in the triangle) as the concept annotation refer to the domain concept ontology shown in Figure 1.

ii) Tagging relationship: The logic relationship of concepts tagged includes the logical “AND” and “NOT”. When merging the two concept annotations directly without consensus, it is impossible to interpret the testing result of the student what concept is well or poor learned according to the tagged concepts. Since the tagged concepts should be entirely well learned for the student to be able to solve the problem, but it fails when there are several solutions for solving the item.

In the collaborative item design environment, the tagging concept may have various concept granularities and discordant tagging relationship incompletely. In order to have a consentaneous tagging, the criteria of “Well Annotating” should be defined. We define that if the tagged concept which can really represent the concept learning performance of the student can be tested by the item, it is called well annotated. Specifically, for an item which is well annotated, the correct answer of the item implies the well learning of the annotated concept, and the well learning of the annotated concept implies the item is to be answered correctly. Thus, the emendation of the concept annotation can be held by the addition of proper annotation and the cancellation of improper annotation based on the specification mentioned above.

Before checking the sufficient and necessary relationship of the item and the annotated concept, the learning performance of concept is needed and the item difficulty and the discrimination should be considered to avoid the bias during learning performance estimation. Secondly, the specific relations among item testing result and annotated concept learning performance need to be found out to evaluate the appropriateness of the concept tagging. Thus we proposed the Folksonomy Aggregation Concept Tagging Scheme (FACTs) to have consentaneous tagging.

The FACTs applies Item Response Theory as the scoring function to evaluate the learning performance with the consideration of the item’s difficulty and the item’s discrimination. We further apply data mining technique to discover the frequent relations between item testing result and tagged concept’s learning performance. The discovered association rule is interpreted by the heuristic pattern rules we proposed to fine tune the concept annotation.

In the experiments, we make convenient for the teachers to tag the concepts in a more flexible tagging relationship, the effectiveness of FACTs is verified by the mathematics testing result of a senior high school in Taiwan. As a Conclusion, the FACTs can fine tune the concept tags automatically and thus provide an environment for teachers in collaborative test sheet design without the interruption of specific limitation of tagging. Furthermore, FACTs may give recommendation on user demand, such as a feedback for the teacher to check whether the designed item is fit with the testing concept object of two-way specification table.

2. Related Works

In order to have more significant interpretation of the testing result, it is convenient to have the item tagged with tested concepts, which provides the aid of finding the performance of concept learning and provides adaptive learning accordingly[3][4][5]. It may be easy for the teachers to form a test sheet by selecting the items for the multi-source and tagging the items with the tested concept collaboratively. But then a great deal of overlapping and incomplete information may occur while merging the collaborative tag directly.

The key issues in collaborative tagging include inconsistency and ambiguity. The simple and directly method is the use of a formal taxonomy or ontology derived through data mining [1][9]. However, the ontology methodology still produced a relatively high error rate by a searching query

test [12]. Statistical and pattern methodologies is well working in general Internet indexing and searching, such as the Google PageRank [10] or Amazon's collaborative filtering system [7]. However, these approaches are relatively indirect toward reducing inconsistency and ambiguity. Social network approaches such as building various ontologies from tags on the basis of concepts and communities [9] to improve the understanding of tag behaviors, although the concept-oriented ontologies conform much more closely to a formal structure, but the methods don't directly address the issues of inconsistency and ambiguity. Visualization such as Cloudalicious illustrates the tag clouds or folksonomies as they develop over time [11]. However, as with the social network methods, these approaches can greatly improve user behavior but do not necessarily reduce tag inconsistency and ambiguity.

While developing the designed items with concept annotation, the item can be obtained by different source and tagging collaboratively based on a given domain concept ontology as a visualization social networking tagging. Even though, tagging an item in a collaborative environment may be not easy to have the same concept granularity. Furthermore, while merging the items directly with different tagging concept may cause various tagging relationship. Thus FACTs is proposed to emend the merging concept annotation automatically to have a consensus tagging formation for an item and give the recommendation on user's demand.

3. Folksonomy Aggregation Concept Tagging Scheme

In the collaborative item design environment, consentaneous annotation is needed while merging the item tags. We believe that if the tagged concept can be tested by the item and really response the concept learning performance of the student, then the item is said to be a well concept annotating item. Specifically, our heuristic is based on the emendation are satisfying the sufficient and necessary condition of "Well concept annotating". Canceling and appending the concept annotation if the items' testing result and the tagged concepts' learning performance do not satisfy the condition mentioned above.

The FACTs we proposed is shown in Figure 2. We assume that the item is designed and is tagged according to the given domain concept ontology collaboratively. First of all, the Testing Result Table (TRT) and the Item Concept Relationship Table (ICRT) is given, the IRT-based Concept Learning Performance (IRT-CLP) is built based on the Item Response Theory (IRT), and is fuzzified into symbolic data for Association Rule Data Mining process. The mined association rule is treated as a relationship pattern, which can be interpreted by the heuristic rules based on the sufficient and necessary condition of the item testing result and the annotated concept learning performance. Thus the concept annotation can be fine tuned automatically, and the recommendation of amending the concept annotation can be provided on demand.

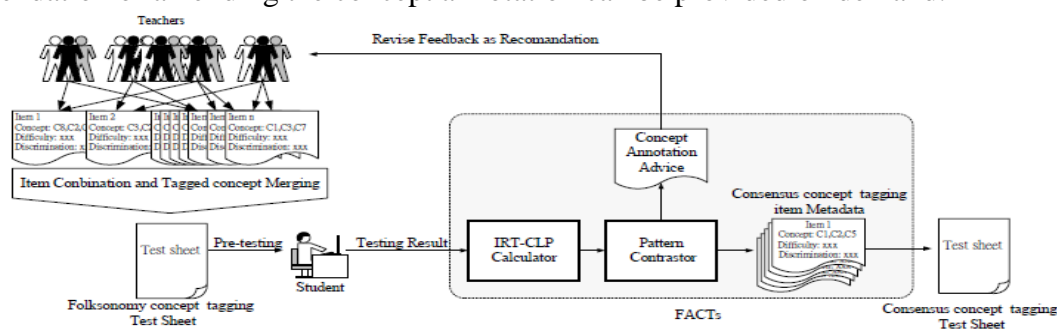


Figure 2. The Folksonomy Aggregation Concept Tagging Scheme (FACTs)

4. IRT-based Concept Learning Performance Calculator

Since the learning performance based on the testing result can be affected by the item difficulty, item discrimination and the student's ability, the item parameters and student ability are estimated based upon the testing result by Joint Maximum Likelihood Estimation [8], and the weighted score of each item is different based on the difficulty P_j , the discrimination D_j and the ability x_j with the Testing Result Table (TRT). The item weight scoring function is based on the Two-Parameter Logistic Model of IRT [2], and Weighted Item Score Table (WIST) related to each item and student can be obtained [6].

In the collaborative environment, each teacher has their own concept tagging. According to the item concept annotation of each teacher, the Item-Concept Relationship Table ICRT is built. While merging several ICRT from different teachers, the Merge Item-Concept Relationship Table MICRT is defined as the maximum union of each ICRT. Figure 3 shows an example of MICRT merged from ICRT of three teachers'.

Figure 3. Merge Item-Concept Relation Table (ICRT)

Concept Test item	ICRT of Teacher A				ICRT of Teacher B				ICRT of Teacher B			
	C ₁	C ₂	C ₃	C ₄	C ₁	C ₂	C ₃	C ₄	C ₁	C ₂	C ₃	C ₄
T ₁	0.1	0.2	0.4	0	0.2	0.3	0	0	0	0.3	0.5	0
T ₂	0.3	0	0.3	0	0.2	0	0	0	0	0	0.4	0
T ₃	0	0	0.7	0	0	0	0.5	0	0	0	1	0
T ₄	0.1	0.3	0.4	0	0.1	0.1	0.5	0	0.2	0.3	0.5	0

Merge by
Maximum

Concept Test item	MICRT of the three teacher			
	C ₁	C ₂	C ₃	C ₄
T ₁	0.2	0.3	0.5	0
T ₂	0.6	0	0.4	0
T ₃	0	0	1	0
T ₄	0.3	0.3	0	0.4

While having the WIST and ICRT, we can further define the Concept Performance Index Table (CPIT) of each item relative to each student. Meanwhile, the IRT-based Concept Learning Performance (IRT-CLP) is calculated according to the CPIT. In order to present the high and low learning performances of each concept of the student, the threshold of the membership functions can be defined by the educational expert to fuzzify IRT-CLP, and the concept performance degree $C_{i.L}$ (poor performance) and $C_{i.H}$ (well performance) are then obtained for each student [6].

5. Aggregation of Concept Tagging

• Unfitness Association Pattern Mining Tagged Concept and Item

After the fuzzification of IRT-CLP, the concept performance " $C_{k.H}$ " and " $C_{k.L}$ " of the students and the testing result " $T_{i.H}$ " (correct answered) and " $T_{i.L}$ " (wrong answered) of the items are formed as the input transactions for the association rule data mining, and are used to discover the unfitness pattern of the tagged concept and item. With the input transaction data, the association patterns of tagged concept and item can be obtained by satisfying the threshold value of "Support" and the "Confidence" value of the association rule data mining.

• Heuristic rules of proper and improper tagging

1) Heuristic of addition of proper annotation:

Generally, the well learning of the tagged concept of the item will result in correct answering the item. Else if the well learning of the annotated concept of the item has incorrectly answered, some of the key concepts are probably missing to be annotated to the item. Thus, if there is a pattern " $T_1.L \rightarrow C_1.H, C_2.H, C_3.H$ ", it means that some concepts are missing or the annotated concepts are too general (coarse concept granularity). To fine tune the annotation of item T_1 , other possible concepts related to T_1 or the sub-concepts referred to the domain ontology can be added.

2) Heuristic of cancellation of improper annotation:

If the concept is not tagged in the relationships of logical AND, it implies that some of the annotated concepts may not be tested in the testing, thus the item may have some redundant concepts annotated. Actually, the learning performance of a redundant annotated concept may not affect the answering of the item; thus if the item with low learning performance tagged concepts is correctly answered, it may imply that such concept is improper tagging. For example, if pattern " $T_2.H \rightarrow C_1.L, C_2.H, C_3.H$ " exist, it implies that the concept C_1 might be redundant for test item T_2 . Then the tag of concept C_1 is removed to fine tune the concept annotation.

6. Experiment

There are 21 mathematic teachers, who are divided into four groups participated in this experiment. Each group is expected to design 4~5 items with concept tagging collaboratively. Therefore, 16 items can be chosen to form a test sheet. The pre-test is undertaken for 782 students at K-11 grade in a senior high school. The assessment with 16 items including 9 concepts with average test score 67.04. Four groups of item concept tag are merged into MIRCT.

We utilize the freeware of Calibration Software for the 3 Parameter Logistic IRT Model [6] to estimate the 16 items' discrimination, difficulty and 782 students' ability. We apply the association rule data mining with two stages, Pattern Matching and Objective Tag Mining. In the Stage I, Pattern Matching is adopted with the minimum support 0.2 and minimum confidence 0.1, and in the Stage II, Objective Tag Mining is adopted with the minimum support 0.2 and the minimum confidence of 0.7. Three specific patterns have been found in Stage I Pattern Matching, and eight association rules are found in Stage II Objective Tag Mining. According to the matched patterns, the concept annotations of the items T_6 , T_7 and T_{16} are amended by adding the miss-annotated concept. Item T_6 is amended by adding the miss-annotated concept C_6 , and the item T_7 is amended by adding the miss-annotated concepts C_3 , C_4 , C_6 and C_8 , and the item T_{16} is amended by adding the miss-annotated concepts C_3 , C_4 and C_6 . The amendment of concept annotation is shown in Table 1.

Table 1. The result amended by the Intelligent Concept Annotation Scheme.

Test Item	Original tagged Concept	untagged Concepts	Aggregation Tagged Concept
T_6	C_3, C_4	C_6	C_3, C_4, C_6
T_7	C_5	C_3, C_4, C_6, C_8	C_3, C_4, C_5, C_6, C_8
T_{16}	C_5	C_3, C_4, C_6	C_3, C_4, C_5, C_6

The aggregation effect of the FACTs is certified by comparing the original concept tagging and the consentaneous tagging result of the items. The satisfying response of 21 teachers expressed in typical five-level Likert scale item is shown in Table 2.

Table 2. The concept tagging aggregation satisfying response of the FACTs

1. Do you agree the concept tagging aggregation of the FACTs ?					
Satisfying response	1	2	3	4	5
Numbers of teacher	0	0	3	14	4
Percentage of the teacher	0%	0%	14.3%	66.7%	19.0%

Note: 1: Strongly disagree 2: Disagree 3: Neither agree or disagree 4: Agree 5: Strongly Agree

7. Conclusion

In our experiment, the test sheet is constructed with item design and concept tagging collaboratively by 21 high school mathematic teachers with teaching experiments more than ten years. The efficiency of FACTs has been identified by the participated teachers. The FACTs we proposed expands the expressive power of tagging concept for the item, and the inconsistent concept granularity level can be fine tuned. We are looking forward to having an item bank management based on user generated item collecting and sharing in Web 2.0 services.

Reference

- [1] S. Golder and B.A. Huberman (2006), "Usage Patterns of Collaborative Tagging Systems," Journal of Information Science, 32(2), pp. 198-208.
- [2] R.K. Hambleton and S.H. Waminathan (1985), Item Response Theory, Kluwer-Nijhoff Publishing, Boston, Massachusetts.
- [3] G.J. Hwang, C.L. Hsiao, and C.R. Tseng (2003), "A Computer-Assisted Approach to Diagnosing Student Learning Problem in Science Course", Journal of Information Science & Engineering, 19(2), pp. 229-248.
- [4] C.L. Hsiao, G.J. Hwang, and C.R. Tseng (2001), "An interactive concept relationship construction assisted system for learning diagnosis," The 5th Global Chinese Conference on Computers in Education, pp. 925-932.
- [5] C.S. Hsu, S.F. Tu, and G.J. Hwang, "A Concept Inheritance Method for Learning Diagnosis of a Network-based Testing and Assessment System", Proceedings of The 7th International Conference on Computer-Assisted Instructions, 1998, pp. 602-609.
- [6] Ching-Yi Liao, Shian-Shyong Tseng and Jui-Feng Weng (2008), "Building an Intelligent Concept Annotation Scheme for Adaptive Testing", The 4th Taiwan E-learning Forum, pp.22.
- [7] B. Linden, B. Smith, and J. York (2003), "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, 7(1), pp. 76-80.
- [8] Rudner, Lawrence M. (2005). PARAM-3PL Calibration Software for the 3 Parameter Logistic IRT Model (freeware). Available: <http://edres.org/irt/param>.
- [9] P. Mika, "Ontologies Are Us: A Unified Model of Social Networks and Semantics", Proceedings of 4th International Semantic Web Conference. (ISWC), Springer, 2005pp. 522-536.
- [10] I. Rogers (2002), "The Google PageRank Algorithm and How it Works", <http://www.iprcom.com/papers/pagerank/>.
- [11] T. Russell, "Cloudalicious: Folksonomy Over Time", Proceedings of 6th ACM/IEEE-CS Joint Conference Digital libraries, ACM Press, 2006, pp. 364-364.
- [12] P. Schmitz, "Inducing Ontology from Flickr Tags", Proceedings of International World Wide Web Conference Steering Committee (IW3C2), 2006.

鷹架式的模型追蹤於 VB 程式設計理解層級之評量

Scaffolding-based Model Tracing for VB Programming Comprehension Assessment

胡莉玲、曾憲雄*、翁瑞鋒**

亞洲大學資訊工程學系/大同技術學院資訊管理系

jasmine@ms2.ttc.edu.tw

國立交通大學資訊工程學系/亞洲大學資訊工程學系*

sstseng@cs.nctu.edu.tw

國立交通大學資訊工程學系**

roy@cis.nctu.edu.tw

【摘要】 本研究提出一個鷹架式的模型追蹤於 VB 程式設計理解層級評量(Scaffolding-based Model Tracing for VB Programming Comprehension Assessment)之機制，根據事先定義好的 VB 程式設計理解層級之本體論，將收集而來的 VB 經典程式做上註記，將內隱的程式間之關係外顯化，為診斷出學習者理解層級的能力，而精心設計成不同類型的填空題型，使得每個不同的填空題型背後皆含有相對應的 VB 程式語言的概念，並搭配以範例學習之技巧，期待找出學生可能的迷思概念，並給予適當的鷹架學習程式，以幫助學習者學習。

【關鍵詞】 模型追蹤、鷹架學習、本體論、迷思概念、以範例學習

Abstract: This paper proposed scaffolding-based model tracing for VB programming comprehension assessment mechanism. The assessment system is according to predefined VB-Programming Comprehension Ontology (VB-PCO). For implicit relation of the programs to explicit, we annotate the VB classic programs that from the experts or the teachers. In order to diagnose learner's the VB comprehension level capability, we design a lot of kinds of cloze testing. The cloze testing all can mapping to the concepts of the VB programming, and match the skill of learning by example to find the learner's possible misconceptions, then give the appropriate scaffolding to assist learner learning.

Keywords: Model Tracing, Scaffolding Learning, Ontology, Misconception, Learning by Example

1. 前言

在計算機科學的教學科目中，程式設計課程的學習是很重要的，其主要目的是要教導學生學習程式語言，並學會利用程式語言來解決問題的一種訓練。通常，程式設計學習的能力可以分為知識層級(Knowledge Level)之敘述語法(Syntax)的學習、理解層級(Comprehension Level)之程式語意(Semantic)的理解與應用層級(Application Level)之程式的編寫(Coding)。

讓學生擁有 VB 程式設計理解層級的能力，是可以透過臨摹(Imitate)的方式來學習的。如何讓學生透過臨摹的方式來學習程式語言呢？首先，可以從專家或有多多年教授 VB 程式設計經驗的老師處詢問或收集一些 VB 的經典程式範例題型，讓學生藉由瞭解這些經典程式範例題型，臨摹其解題方法與技巧，來幫助學生學習解題的技巧。

縱然以此方式來學習是很自然的，然而對於相似題目的程式，仍然會有部份學生無法瞭解。例如：瞭解選擇排序法的解題方式，對插入排序法的解題仍可能存在迷思概念(Misconception)。根據我們的觀察，對於學習成效較好的學生，是因為他們可以瞭解到內嵌(Embedded)在程式中的知識架構，也就是說，他們可以很容易的將這些知識架構，對應到自己認知上較高階層的語用概念(Pragmatic Concept)。

所謂程式設計上的語用概念，指的是相同的敘述在不同的程式中，有著不同的用法與其意義的。例如：宣告變數中，變數的用法卻是不盡相同的，它可能是做為初值的設定、邊界的判斷或終止執行的判斷用。我們認為題目相似的程式，其程式的知識架構是相近的，然而其敘述的語用概念卻是不相同的，而這些語用概念大部份都是屬於認知結構的。

因此，我們認為如果可以將這些原本隱含在程式背後的語用概念給顯現出來，那麼情境認知理論(Situated Cognitive Theory)中的鷹架學習(Scaffolding Learning)應可以用來支援學生在認知架構上的學習。也就是說，當學生在程式設計的解題過程中，有認知上的迷思概念時，若能提供適當的鷹架輔助學習，是可以幫助學生來解除認知上的迷思概念。

在程式設計領域的眾多研究中，已有許多不同型式的學習鷹架被提出，其中最經典的大概是屬於模型追蹤(Model Tracing)的技巧了。所謂模型追蹤，主要是針對特定題目的特定程式，依據大多數學習者的錯誤樣式，來追蹤出學習者的迷思概念。也就是說，使用模型追蹤是可以藉由學習者的學習路徑(Learning Path)，來瞭解學習者的學習過程與行為，進而預測出學習者的學習問題與常犯的錯誤，並給予適當的輔助學習。

雖然模型追蹤可以找出學習者的迷思概念，然而它只能追蹤到單一程式的迷思概念，並無法追蹤到題目間或是程式間的關係與其共同概念。也就是說，傳統的模型追蹤是屬於較局部的概念(Local View)，它只能找出程式本身的語法或語意上的迷思概念，而無法找出屬於整體概念(Global View)的語用迷思概念。

傳統的模型追蹤之所以無法找出語用迷思概念，是因為它無法瞭解到程式與程式間的關係，也因而無法有效的找出相關的語用迷思概念。根據我們的觀察，題目類似的程式，其中是有著許多隱含的關係存在的，例如：選擇排序法與泡沫排序法為兩個題目類似的程式，其目的皆為將一連串原本雜亂無章的資料按其順序排序，其程式結構是相近的，所使用的語用概念也是相近的。也就是說，當學生瞭解了選擇排序法的解題方法與技巧，那麼在解泡沫排序法時，學生應該也很容易可以將兩者間相似的語用概念運用在泡沫排序法的解題上的。因此，我們提出一個想法，希望能運用程式與程式間的關係，來找出學生在學習程式設計時的語用迷思概念，進而給予適當的補救教學活動，幫助學生擁有程式設計理解層級的能力。

至於，如何建置鷹架呢？在程式設計學習領域的研究中，曾有許多學者提出卷宗評量法、除錯評量法等多種不同的評量方式，為了建構適性化的學習鷹架，我們認為學生的學習歷程是需要被考量進來的，因此，在本研究中，我們提出了一個鷹架式的 VB 程式設計理解層級之評量機制(Scaffolding-based VB Programming Comprehension Assessment)架構，依據學生的學習歷程，瞭解其學習狀態而給予適合的鷹架，幫助學生的學習。

2. 文獻探討

由於程式設計的學習技能各有不同，輔助程式設計的教學系統亦有多種型式。傳統的程式偵錯器(Debugger)大概是學生在學習程式設計時最初接觸的系統，偵錯器雖然可以偵測出程式語法上的錯誤，但是它卻沒有解釋錯誤的原因，也無法偵測出語意上的錯誤，因此偵錯器較適合有經驗的程式設計師，不適合程式設計的初學者。對於將學習程式語言以動畫的方式來輔助教學的系統(Crews & Ziegler, 1998)，雖有解釋錯誤原因的設計，但大都只能單向的使用，沒有與學習者互動的介面設計，也沒有對使用者程式設計能力或理解的評量機制。另有一些智慧型教學系統(ITS, Intelligent Tutoring System)，如：GIL(Reiser, 1992)、DISCOVER(Ramadhan, 1997)或 LISP 教學(Anderson & Reiser, 1985; Anderson & Pelletier, 1991; Anderson, Corbett, Koedinger & Pelletier, 1995)等為使用模型追蹤方法的互動式系統，雖有解釋錯誤原因，也有設計與學習者互動之介面，然而卻沒有顯示類似程式與程式間的關係，亦即智慧型教學系統雖能找出語法與語意的錯誤，卻無法追蹤出語用的迷思概念。

綜合以上，傳統的程式設計教學系統並無法有效的提供學習者在理解層級上所需要的鷹架，在本研究中，我們想要提供鷹架式的評量架構，以評量出學習者在理解層級上的迷思概念，給予適當的鷹架學習與補助的學習建議。

3. 鷹架式模型追蹤以範例學習之流程

本研究於前已討論過，學生可由臨摹而習得解題方法與技巧，藉由此觀念，本研究提出了以範例學習(Learning by Example)的想法，並搭配鷹架式模型追蹤，企圖藉以找出學習者在理解層級之迷思概念，並能給予適當的鷹架學習。本研究提出的鷹架式模型追蹤以範例學習之流程。鷹架式模型追蹤以範例學習之流程主要有四個步驟，現分別說明如下：

(1). 經典範例程式的收集(Classical Sample Program Collection)

我們由老師或專家處收集整理一些較經典的程式範例，參照事先定義好的程式設計理解層級之本體論(Programming Comprehension Ontology, PCO)，對程式加以註記分類後，再存回程式庫做為出題之用。

(2). 程式設計鷹架的建置(Programming Scaffolding Construction)

利用事先建置好的程式設計理解層級之本體論與詮釋資料(Metadata)，來建置鷹架學習模組，當學習者被偵測出錯誤時，則依其學習狀態，選擇最適合的鷹架程式提供學習者學習。詮釋資料相似度愈高的程式，則較易被選取為適合的鷹架程式。

(3). 填充測驗之模型追蹤(Cloze Testing Based Model Tracing)

本研究的每個測驗題目都是經過精心設計的，其填空的背後也都有一個相對於 VB 程式語言的概念存在。採用填充測驗(Cloze Test)的理由，主要為填充題型作答時較沒有線索可循，可降低學習者猜測的行為。

(4). 鷹架式的補救學習(Scaffolding-based Remedial Learning)

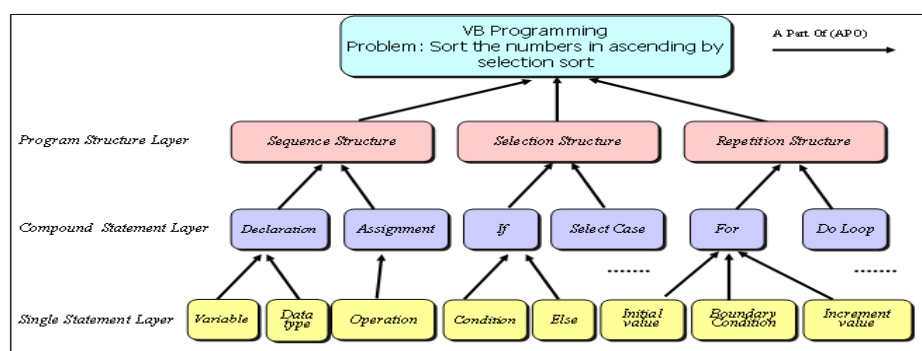
系統會將學習者的答案與事先建置好的正確樣式(Correct Pattern)或錯誤樣式(Buggy Pattern)做對照，以確認學習者之迷思概念，評量結果並可回饋至適性化出題模組，繼續下一題之測驗，亦可將評量結果與迷思概念做對照，依此提供相對應的鷹架學習給學習者，或依學習者的學習狀態給予補救矯正的學習建議。

4. 鷹架式 VB 程式設計理解層級之評量機制

為了診斷學生的迷思概念，我們先建構好的 VB 程式設計之理解本體論(PCO)，做為提供認知上鷹架學習之參考，再根據題目的特性來定義問題的詮釋資料，以描述程式與程式間之關係，做為解釋語用概念之依據。

4.1 程式設計理解層級之本體論(Programming Comprehension Ontology, PCO)

定義此程式設計理解層級之本體論(PCO)是為了要展現 VB 理解層級之程式結構，如圖像 1 所示。



圖像 1 程式設計理解層級之本體論

程式設計理解層級之本體論(PCO)可用三個層次來展現：

- (1). 程式結構層次(Program Structure Layer)：依 Dijkstra 的結構化程式設計的理论，我們將此層次又分為 Concatenation, Selection 與 Repetition。
- (2). 複合敘述層次(Compound Statement Layer)：這個層次依程式結構層次再細分為 Declaration, Assignment, If, Select-Case, For 與 Do-Loop。
- (3). 單一敘述層次(Single Statement Layer)：此層次則再依複合敘述層次再細分，如：Variable, Data Type, Operation, Condition, Else, Initial Value, Boundary Condition, Incremental Value 等。

4.2 詮釋資料(Metadata)的設定

我們以詮釋資料(Metadata)來展現程式與程式間之關係，做為解釋語用概念之依據。此詮釋資料分別以問題型式(Problem Type)、程式化結構概念(Structured Programming Concept)與複合敘述(Compound Statement)等三個屬性來表示程式與程式間之關係，其範例如表格 1 所示。

表格 1 詮釋資料之範例

Metadata Attribute	Value
Problem Type	Selection Sort, Exchange, one dimension array, nested-for
Structured Programming Concept	Nested(Repetition, Repetition), Nested(Repetition, Selection)
Compound Statement	Declaration, Assignment, If, For
Sample Program	
Problem: Sort the numbers in ascending by selection sort.	
10	Dim I, J as Integer
20	Dim s(10), Temp as Single
30	For I = 1 to (UBound(s) - 1)
40	For J = I + 1 To UBound(s)
50	If s(I) > s(J) Then
60	Temp = s(J)
70	s(I) = s(J)
80	s(J) = Temp
90	End If
100	Next J
110	Next I

4.3 鷹架式模型追蹤於VB 程式設計理解層級之評量機制(Scaffolding-based Model Tracing for VB Programming Comprehension Assessment Scheme)

本研究提出一個鷹架式模型追蹤於 VB 程式設計理解層級之評量機制，如圖像 2 所示。

本架構主要有三個模組：

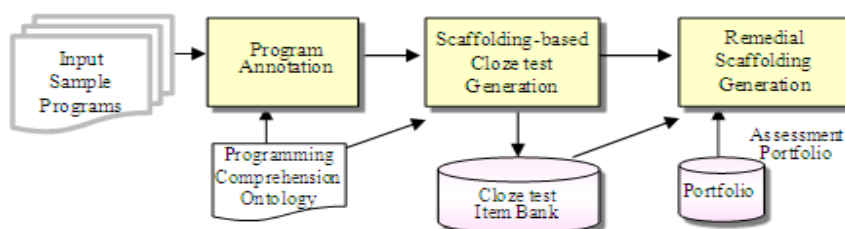
- (1). 程式註記模組(Program Annotation)

首先我們將收集來的範例程式，參照事先定義好的 PCO，加以註記後儲存。

- (2). 鷹架式填充測驗出題模組(Scaffolding-based Cloze Test Generation)

本架構採用的是填充測驗，參考註記後之範例程式與 PCO 後，存回測驗題庫。

- (3). 鷹架學習程式產生模組(Remedial Scaffolding Generation)

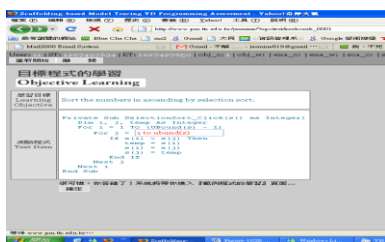


圖像 2 鷹架式模型追蹤於 VB 程式設計理解層級之評量

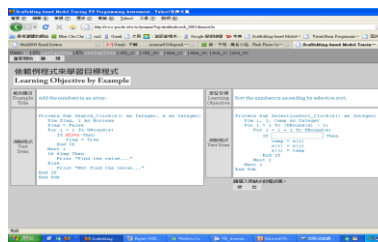
由學習者的答案與學習者的學習歷程，做為產出適當的鷹架學習程式，以供學習者做為補救學習之用。

5. 實驗與討論

為了驗證本研究所提之理論，我們架設了一個 PHP 搭配 MYSQL 的網路平台，實驗對象為 34 名的大專生，其中 18 名學生為專二生，16 名學生為專五生。本研究先對兩組學生實施前測試後，使用本系統學習後，再進行後測試，以比較學習前後之高分群與低分群學生學習之成效。此網路平台的部份實作畫面如圖像 3 與圖像 4 所示。

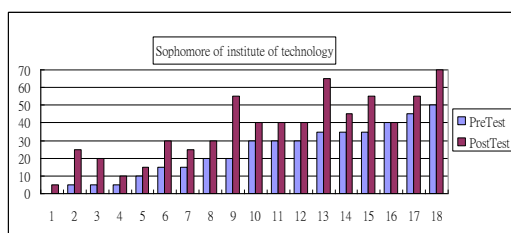


圖像 3 實驗實作畫面一

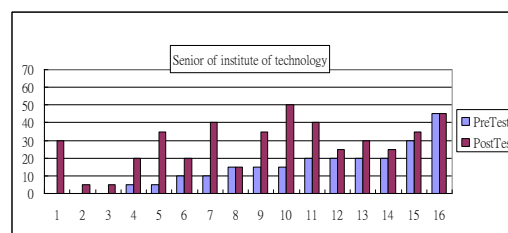


圖像 4 實驗實作畫面二

實驗結果，如圖像 5 與圖像 6 所示。圖像 5 為專二生的實驗結果，圖像 6 為專五生的實驗結果。由實驗結果顯示，學生在使用本教學系統後，成績都明顯的提高。



圖像 5 專二生的實驗結果



圖像 6 專五生的實驗結果

經由實驗，本系統有幾點發現：(1)可以促使學習者去比較並瞭解不同題目之間的關係，(2)引導學習者重新思考書本上的程式範例，(3)許多學習者認為本系統確實可以幫助他們瞭解程式設計的概念。

6. 結論

本研究為了要診斷學習者的迷思概念而設計了程式設計理解層級之本體論，並提出了鷹架式模型追蹤於 VB 程式設計理解層級之評量架構。

未來的研究方向，可以增加更多的測驗題目，與擴增主題項目，也可以分析學生的行為，依學生的特質不同，探勘出他可能的迷思概念，在教學的同時，即可給予適當的加強與建議。

參考文獻

- Anderson, J. R., & Reiser, B. J. (1985). The Lisp Tutor, Byte, 10, 159-75.
- Anderson, J. R., & Pelletier, R. (1991). A Development System for Model-Tracing Tutors, Proceedings of the International Conference of the Learning Sciences, 1-8, Evanston, IL.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned, Journal of the Learning Sciences, 4(2), 167-207.
- Crews Jr., T., & Ziegler, U., (1998). The Flowchart Interpreter for Introductory Programming Courses, Frontiers in Education Conference, 1998. FIE'98 28th Annual.

- Ramadhan, H. A., (1997). Improving the engineering of model tracing based intelligent program diagnosis, IEEE, Proc.-Softw. Eng.
- Reiser, B., (1992). Making process visible scaffolding learning with reasoning-congruent representations, Proceedings of the 2nd conference on Intelligent tutoring system (ITS'92), Montreal.

應用模糊方程式近似值解法之技能評量分析

曾煥雯、吳孟軒*

國立台灣師範大學應用電子科技學系

hwtzeng@ntnu.edu.tw

國立台灣師範大學工業教育學系*

【摘要】 具有豐富經驗之教師，進行技能評量時，思維中存在著可靠評量尺度以及各評量項目間權重分配，這是多年教學經驗所累積出來。然而，實際技能評量經驗顯示，有各種環境因素會影響評量者評分權重，例如節省時間、成品水準過於接近、省麻煩等考量下，經常捨去逐一打細項分數，而直接對某一個子項目做出評量結果，為此我們懷疑權重分配被扭曲的可能性頗高。研究重點於對此議題，發展出由自由度分析法延伸出全新的反模糊法則近似值解法，探究評量者進行整體評量時，思維中是否能秉持原訂權重分配，或有何種程度的變動，並分析逐項評量再加總與直接給出總評之間的偏差現象。

【關鍵詞】 反模糊法則、技能評量、權重分析、近似值解法

Abstract: A expertly teachers of skill assessment always has reliable scale and weighting in his own perception. They are constructed for many years by experiences of practical operation and knowledge acquisition. But, while considers about the loading and inconvenient, expertly teacher sometimes gives out the totally score of whole item's assessment, without deals to every sub-items. We doubt, in this light way of measuring will cause the distortion of original weighting that was set by committee. It is the major purpose of this project, to propose a new Fuzzy Decomposition Method about the approximate solution of fuzzy equation, for the analysis usage. By this way, we could understand whether the weighting distribution between sub-items was changed or not. If changing was happen, how about the range of deviations.

Keywords: Fuzzy Decomposition Method, Skill Assessment, Weighting Analysis, Approximate Solving Method

1. 研究背景

所謂模糊綜合評判，就是要對某一對象進行全面性的評價。例如對一個技能成品的評價就應考慮尺寸、外觀、功能、精密度、色彩等多項因素。在涉及到產品鑑定，資格評審，人才選拔，工作效果等劃分等級事物時，可以用模糊綜合評判，因在這類問題中每個等級之界線是模糊的，而利用模糊綜合評判可以更好的模擬人的思維[1, 4]。而專家思維也可以用反模糊綜合評判來加以逆向分析之。欲經由具豐富學識、經驗的委員，探求其個人對於評鑑細項間權重的分配時，首先請其對各個細項作出評判得出 \underline{R} 矩陣，並對於整個項目作整體性的評估得出 \underline{B} ，再藉由反模糊的技術（即求解模糊方程式），求得其個人對於權重分配向量 \underline{X} ，供為委員會規劃權重的參考[5]。上述求解 $\underline{X} \circ \underline{R} = \underline{B}$ 例中，已知 \underline{R} 與 \underline{B} ，則 \underline{X} 的解存在於一個可能的範圍中，範圍中最大值的通解只有一組，然而其最小值不一定是唯一的，如圖 1 所示。

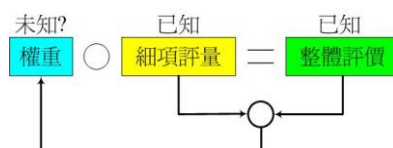


圖 1：反模糊綜合評判

模糊綜合評判方程式 $\underline{X} \circ \underline{R} = \underline{B}$ ，已知 \underline{R} 與 \underline{B} 求 \underline{X} 。

求解最大值及最小值時，它們都是極端的，都是具有單一的自由度，即通解只有一個。最大值解邏輯特性為廣度優先 $Max = \min(\max_1 \max_2 \dots \max_n)$ ，亦即它具有可能性(possibility)的特質；最小值解邏輯特性為深度優先 $Min = \max(\min_1 \min_2 \dots \min_n)$ ，即具有必要性(necessity)的特質。最大值雖是廣度優先邏輯導向，然繼而求取所有可能解的交集(Min)，導致只有一個解答的特性，所以它的通解值是唯一的最大值。最小值雖是保守優先邏輯導向，然而求取所有可能解的聯集(Max)，這個先緊後鬆的特性，導致最小值通解不一定唯一，如圖 2 所示。

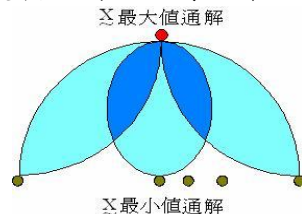


圖 2：反模糊解範圍

現有的求解的過程中有各種的策略先求出若干個可能的解，再以計算後距離最短(最貼近)者為通解，或以系列的過程推導出通解者，要將它轉成計算機的程式時，往往結構顯得繁複，程式的寫作與維護都不容易[1,2,6]。進一步，考量及最大值與最小值具有一定的跨距(span)，於分析時必須時時思考到底以何種特質（廣度或深度）的導向較為適合，為此，研究首先提出以解模糊方程式近似值法則如下，期使求出解答兼具廣度與深度兩者性質，又能夠探討分析出評量者心中權重分佈[2,5]。

2. 近似值解法

依據 Tsukamoto 所提出的方程式 $X_j \wedge R_{ij} = B_j$ 中的 X_j 的可能解有三種狀態，修正如下：

- (1) 若 $R_{ij} > B_j$ 時，則 $\max(X_j) = B_j$ ， $\min(X_j) = B_j$ 。
- (2) 若 $R_{ij} = B_j$ 時，則 $\max(X_j) = 1$ ， $\min(X_j) = B_j$ 。
- (3) 若 $R_{ij} < B_j$ 時，則 $\max(X_j) = B_j$ ， $\min(X_j) = R_{ij}$ 。

以上述的三種狀況為基礎，分別由各元素的最大值與最小值先建立起 M 及 m 兩個判別矩陣。由於最大值的廣度導向，直接取各最大值的最小者，即能得出通解。最小值的判別矩陣需先修正，之後直接判別 m 的最大值，得出的即為最小值通解。整個求解的次序如圖 3 所示。

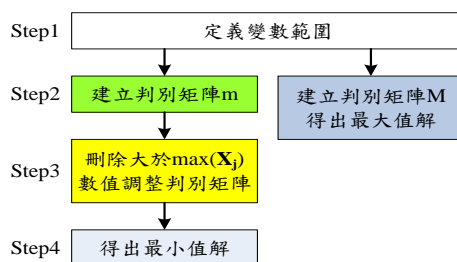


圖 3：近似值解法流程

求解例子： $B = [0.2 \quad 0.4 \quad 0.6 \quad 0.7]$

$$R = \begin{bmatrix} 0.2 & 0.4 & 0.6 & 0.7 \\ 0.2 & 0 & 0.6 & 0.8 \\ 0.1 & 0.4 & 0.6 & 0.7 \\ 0 & 0.3 & 0 & 1 \end{bmatrix} \quad M = \begin{bmatrix} 1 & 1 & 0.2 & 0.2 \\ 1 & 0.4 & 1 & 0.4 \\ 1 & 1 & 1 & 0.6 \\ 1 & 0.7 & 1 & 0.7 \end{bmatrix} \quad m = \begin{bmatrix} 0.2 & 0.2 & 0.1 & 0 \\ 0.4 & 0 & 0.4 & 0.3 \\ 0.6 & 0.6 & 0.6 & 0 \\ 0.7 & 0.7 & 0.7 & 0.7 \end{bmatrix}$$

近似值解法： $MaxV = \min(X_{1j} \quad X_{2j} \quad X_{3j} \quad X_{4j}) = [1 \quad 0.4 \quad 0.2 \quad 0.2]$

求得 $\text{Max}(X_j)$ 後對判別矩陣 m 做調整，每一個元素大於 $\text{Max}(X_j)$ 同列元素值者將被刪除（歸零），再依照每一列數直取最大值就是近似值解法的最小值通解。

$$\text{原先 } m = \begin{bmatrix} 0.2 & 0.2 & 0.1 & 0 \\ 0.4 & 0 & 0.4 & 0.3 \\ 0.6 & 0.6 & 0.6 & 0 \\ 0.7 & 0.7 & 0.7 & 0.7 \end{bmatrix} \quad \text{經調整後 } m' = \begin{bmatrix} 0.2 & 0.2 & 0.1 & 0 \\ 0.4 & 0 & 0 & 0 \\ 0.6 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 \end{bmatrix}$$

近似值解法： $\text{Min}V = \max(X_{1j} \quad X_{2j} \quad X_{3j} \quad X_{4j}) = [0.7 \quad 0.2 \quad 0.1 \quad 0]$

3. 研究目的

當今全球資訊網路相當發達，模糊綜合評判技術已經能夠成功應用於電腦網路上，然而反模糊的技術一直都相當的繁複且運算維度又高，研究報導相當少且都是工程應用為主。為擴展技職研究領域廣度與深度，突破評量探索空間，本研究的重心即在於探討符合下列技術要求的反模糊近似值分析工具[1][3-4]：1. 迅速求得反模糊近似值，2. 具有高相容性能供給不同的瀏覽器來使用，3. 搭配資料庫，程式物件化，4. 計算機的規格要求簡化。

4. 系統架構

本研究的執行所建構的網頁架構如下圖 4 所示，介紹了研究的概要、模糊理論、解模糊方程式理論，以及核心的解模糊方程式的及時分析網頁。網頁的架構主要有專家評量反模糊分析、反模糊數學理論、等相關內容如下圖 5 所示。進入技能評量網頁後，首先是對評量者進行網頁操作介紹，包括成品的內涵與數量，三個評量重點剝線、絞接、及外觀，以及這三個重點細項間的權重設定。

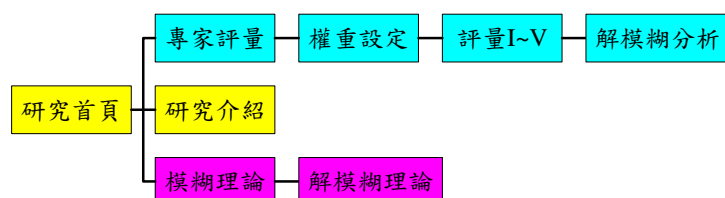


圖 4：本研究網頁主架構



圖 5：網站之系統介紹

5. 技能評量設計

基於下列幾項的考量，本研究決定以單芯線連接之技能評量為例，進行網路分析的測試：

1. 技能成品細部能完整呈現於網頁上。
2. 評量細項明確且項目有限。
3. 專業教師都熟悉的技能。

為了多角度的探究評量者的思維，研究者按照評量細項的特徵，即剝線、絞接、外觀等三個細項，分別製作了五個成品，它們在四個細項的表現上各不相同。這五個絞線連接的技能成品呈現分別如圖 6 所示，表一為研究者對五個成品就三個評量細項的分析。

以單芯電線的絞接為例，在室內配線的競賽或檢定的現場評量中，國內都是直接給出整體的結果，而捨去透過細項評量，再依據評量手冊中的權重來計算整體性的評量結果。這種現象存在已久，探究其原因，經常是源於評量的負荷大加上時間又短，因而簡化評量的實施。

以單芯電線的絞接為例，在室內配線的競賽或檢定的現場評量中，國內都是直接給出整

體的結果，而捨去透過細項評量，再依據評量手冊中的權重來計算整體性的評量結果。這種現象存在已久，探究其原因，經常是源於評量的負荷大加上時間又短，因而簡化評量的實施。

表一：五個成品表現分析

組別	成品	剝線	絞接	外觀	總分
1					
2					
3					
4					
5					

所完成的網頁如圖 7 所示，經由八次完整的評量測試，得到相關的數據如下。在進行評量時，基於心中思維的探討需要，系統首先會要求使用者給出就剝線、絞接、外觀等三個評量細項之個人所認為的權重百分比，然後逐一的呈現研究者所提供的技能成品。



圖 6：五個評量的技能成品



圖 7：技能評量網頁

針對逐一的成品，請使用者就三個評量細項給出由 1 到 10 的等級給予評分，同時對每一個成品由 1 到 10 分，直接給出總分的等級。如此，即是方程式 $\bar{X} \circ \bar{R} = \bar{B}$ 中，已知細項評分 \bar{R} 與 \bar{B} ，而推論探究 \bar{X} 的操作。

表二：各組細項及總分評量結果

組別	成品	剝線	絞接	外觀	總分	組別	成品	剝線	絞接	外觀	總分	組別	成品	剝線	絞接	外觀	總分
一	1	10	9	8	9	四	1	6	6	6	6	七	1	6	7	7	7
	2	7	5	5	5		2	6	5	7	7		2	3	3	3.5	3.2
	3	9	7	8	8		3	5	5	5	5		3	8	7.6	7.4	7.5
	4	10	10	10	10		4	8	8	9	9		4	5	9	8	8.5
	5	9	7	8	8		5	8	6	6	7		5	7	6.2	6.5	6.4
二	1	7	8	6	7.5	五	1	8	7	7	7	八	1	8	7.5	8	7.6
	2	4	4	3	3.5		2	7	5	6	6		2	4	3.5	6	4
	3	7.6	6.5	6	6.2		3	6	7	7	7		3	7.5	7	7.2	6.8
	4	6.8	8	7.6	7.7		4	8	10	9	9		4	7.2	8.4	8	8.2
	5	6	6	5	5.3		5	8	8	8	8		5	6	6.4	6	6.2
三	1	7	8	8	8	六	1	10	10	10	10						
	2	5	3	4	3.5		2	6	6	6	6						
	3	9	7	7	7		3	7	7	7	7						
	4	6	8	8	8		4	8	8	8	8						
	5	6	4	5	4.5		5	9	9	9	9						

表三：各組權重的原始設定

組別	剝線	絞接	外觀	總分
1	3	4	3	1
2	3	5	2	1
3	3	5	2	1
4	1	8	1	1

5	3	4	2	1
6	4	4	1	1
7	3	6	1	1
8	2	6	2	1
平	2	5	1	1

6.研究發現與討論

系統於評量完成後，即時進行反模糊近似值解推論運算，分析評量者思維中的權重範圍，以最大值(Max)及最小值(Min)為上下界。分別呈現如下表四。表中的 Max 值由於模糊數學的運算原理如同採用最大的容許可能（如圖 2 所示），所以最能成為具有價值的參考邊界，至於 Min 依據模糊數學的原理，它如同最極端邊界。

表四：分析的 Max 及 Min 範圍

組別	剝線	絞接	外觀	組別	剝線	絞接	外觀
1	%	%	%	5	%	%	%
Set Val.	30	40	30	Set Val.	35	45	20
Max	22	36	42	Max	27	27	46
$\Delta_{\text{Max-Set}}$	-8	-4	12	$\Delta_{\text{Max-Set}}$	-8	-18	26
Min	25	35	40	Min	30	25	45
組別	剝線	絞接	外觀	組別	剝線	絞接	外觀
2	%	%	%	6	%	%	%
Set Val.	30	50	20	Set Val.	40	45	15
Max	33	33	34	Max	33	33	34
$\Delta_{\text{Max-Set}}$	3	-17	14	$\Delta_{\text{Max-Set}}$	-7	-12	19
Min	35	35	30	Min	33	33	34
組別	剝線	絞接	外觀	組別	剝線	絞接	外觀
3	%	%	%	7	%	%	%
Set Val.	30	50	20	Set Val.	30	60	10
Max	33	33	34	Max	33	33	34
$\Delta_{\text{Max-Set}}$	3	-17	14	$\Delta_{\text{Max-Set}}$	3	-27	24
Min	35	30	35	Min	32	32	36
組別	剝線	絞接	外觀	組別	剝線	絞接	外觀
4	%	%	%	8	%	%	%
Set Val.	10	80	10	Set Val.	20	60	20
Max	33	33	34	Max	43	28	29
$\Delta_{\text{Max-Set}}$	23	-47	24	$\Delta_{\text{Max-Set}}$	23	-32	9
Min	35	30	35	Min	45	25	30

在八個表中，普遍存在權重分析結果與原設定值差異偏大現象，例如，八組評分的絞接選項配分原訂為 50~80%，但利用近似值解法分析出實際分布為 25~35%。

探討背後所代表的意義為，評量者所預先設定的權重分配，與思維中的真正分配不能完全相符合，而且差異愈大表示符合程度愈差。

另就八組原權重設定平均與 Max 平均值做比較，數據呈現如下表五。

表五：權重平均值分析

權重	剝	絞	外
設定	2	5	1
Max	3	3	3
設定 - Max	3	-2	1

進一步分析發現整體而言，評量者所做出的權重設定值與腦海中思維的分佈有一段的差

距存在。以第一組評量結果為例，進一步的深入探討。這一組中，五個導線連接成品評量數據如表二所示。每一個細項的評量結果及總分都是介於 0~10 之間，經過歸 1 的運算後，成為介於 0~1 之間的數值，所以它的模糊綜合評判運算為：

$$[X_1 \quad X_2 \quad X_3] \circ \begin{bmatrix} 1 & 0.7 & 0.9 & 1 & 0.9 \\ 0.9 & 0.5 & 0.7 & 1 & 0.7 \\ 0.8 & 0.5 & 0.8 & 1 & 0.8 \end{bmatrix} = [0.9 \quad 0.5 \quad 0.8 \quad 1 \quad 0.8]$$

反模糊近似值解法運算求解出權重向量中 $[X_1 \quad X_2 \quad X_3]$ 分佈範圍，以一號評量結果為例，求解出來的值如下：

$$Max = [0.5 \quad 0.8 \quad 0.9], \quad Min = [0.5 \quad 0.7 \quad 0.8]$$

逐一對成品的數據，先用原來所設定的權重計算總分，求出結果與原給出的總分評量值間之偏差百分比值，加總後，呈現出按權重設定值與直接給出評量間的差距。再由反模糊近似值解法推論的 Max 取代原設定的權重設定值，求出 Max 向量值與直接給出評量間的差距。將前一個差距與後一個相比較，求出差距縮減量。八組評量結果差距及縮減情形如表六所示。

表六：八組評量的差距及縮減量

原權重	Max 向量之	差距縮
9.5%	3%	6.5%
22%	9.5%	12.5
12.5	22.4%	-9.9
-45.7	-26.7%	19%
-0.6	0%	0.6%
0%	0%	0%
-14.6	-14.5%	0.1%
8%	14%	-6%

由表六中明顯的呈現出，反模糊近似值解法推論能分析出評量者思維中權重最具代表性的權重值，使細項評分經權重分配後能夠接近總分值。其中第三組縮減量呈現-9.9%第八組縮減量呈現-6%，實際原因為評審的總分與加權間較無邏輯性，使運算後造成發散式誤差，為歸 1 時小數點四捨五入，經多層計算所累計的結果。因此再次證明了研究者當初的假設，即藉由反模糊近似值推論，能確實分析出專家思維中的權重分配。

參考文獻

- [1] 吳萬鋒 (1986)。模糊數學與計算機應用。電子工業出版社。
- [2] 曾煥雯(1997)。建立植基於模糊理論之技能評量模式。國立台灣師範大學工業教育研究所博士論文。
- [3] Fu Cheng Tang.(2000). Perturbation techniques for fuzzy matrix equations. Fuzzy Sets and Systems Vol.109, Issue3, page363- page369.
- [4] A.V. Markovskii.(2005). On the relation between equations with max-product composition and the covering problem. Fuzzy Sets and Systems Vol. 153, page261- page273.
- [5] George J. Klir, Bo Yuan(2005). Fuzzy Sets and Fuzzy Logic: Theory and Applications. Pearson Education Taiwan, page153- page176.
- [6] Yan-Kuen Wu, Sy-Ming Guu(2005). Minimizing a linear function under a fuzzy max-min relational equation constraint. Fuzzy Sets and Systems Vol. 150, page147- page162.

微观建构理论指导下协作知识建构的学习绩效评价研究

The Research of Learning Performance of Collaborative Knowledge Building(CKB) based Micro-construct theory

张伟、谢幼如

华南师范大学教育信息技术学院

[摘要] 许多教育学和心理学的专家学者就学习的本质问题逐步达成了共识,即学习的本质是知识建构。但如何对知识建构过程和结果进行绩效评价,如何构建协作知识建构(CKB)中学习绩效评价体系成为一大难题。本文以微观建构理论为指导,以基于概念性问题的课堂协作知识建构的学习绩效评价为例,探索协作知识建构的学习绩效评价体系。

[关键字] 微观建构、学习绩效、协作知识建构

Abstract: many experts and scholars of Pedagogy and psychology come to an agreement that the nature of learning is Knowledge Building. But how to evaluate the process and result of Knowledge Building and how to build a system of evaluation of learning performance in CKB is a difficult problem. In this article, I will explore the system of evaluation of learning performance in CKB based Micro-construct theory.

Keywords: Micro-construct, Learning Performance, Collaborative Knowledge Building

1.引言

教育领域关于“学习”的研究已经生成了诸多结论。但当再回头重新审视学习时,我们不禁要问,学习的本质是什么?随着新教育理念的引入,一些教育学家和心理学家达成了共识:学习的本质是知识建构。在此背景下,谢幼如教授从协作的角度研究“知识建构”,提出了“协作知识建构(CKB)”的五个过程要素:共享、论证、协商、创作和反思,并有广州市七所中小学校展开了实践探索。但在实践过程,如何对协作知识建构过程和结果进行评价是摆在教师和研究者面前的难题。本文以微观建构为导向,对学习绩效的评价进行深入探索,包括评价思路、评价内容、评价标准、评价的方式方法等。

2.微观建构和学习绩效

2.1. 微观建构

建构主义在近30年的发展过程中逐渐分化成宏观建构主义和微观建构主义,宏观建构主义倾向于本体论层面的研究,对实际教学的可操作性指导不强,而且一些建构主义模型多是粗略的描述学习,并没有细致揭示学习的建构过程,也没有形成自己独特的学习评价体系。

微观建构主义是指把建构主义理论从哲学层面降低到教学应用层面,下放到课堂教学的具体问题解决过程中,为实际教学提供具体指导的学习理论,它甚至强调为某个具体知识点的学习设置情境等。微观建构理论能够从实用和务实的角度刻画知识变化的内在机制。

2.2. 学习绩效

目前关于绩效的定义主要有以下四种观点：一种观点认为绩效是结果；第二种观点认为绩效是行为；第三种观点认为绩效是行为加结果；第四种观点认为绩效将个人的素质和潜能也包括在内。本研究采用第三种观点，即绩效是指与组织总体目标及价值追求相一致的行为倾向和业绩成就（梁林梅，2004）。

协作知识建构的学习绩效是指学生进行协作知识建构的学习行为及最终的协作知识建构的学习成就，如小组智慧制品、个人知识获得、高阶思维能力等（石娟，2008）。可以从以下几个方面来理解此定义：（1）学习绩效是行为和结果；（2）这些行为和结果必须是可测量的、看得见的；（3）这些行为和结果必须是与学习目标相一致的。

3. 基于问题解决的课堂协作知识建构

协作知识建构是指个体在特定的组织中相互协作、共同参与某种有目的的活动，最终形成某种观点、思想、方法等智慧制品的过程（谢幼如，2007）。基于问题解决的课堂协作知识建构是指在课堂教学过程中创设问题情境，当问题呈现后，学习者在小组内协作、共同参与某种有目的的活动（如共享、论证、协商、创作和反思等），最终实现知识建构并生成智慧制品的过程。问题又分为概念性问题、解释性问题和设计性问题三类，本文将以基于概念性问题的课堂协作知识建构为例，进行学习绩效评价的研究。概念性问题的协作知识建构主要经历对象呈现、概念形成、概念运用、作品形成、学习评价和总结反思六个环节。

4. 问题解决的课堂协作知识建构的学习绩效评价体系

4.1. 学习绩效评价理念

微观建构理论所倡导的评价理念侧重对知识建构过程的评价，重点是问题解决过程中微观的协作知识建构活动和知识的不同形态变化（丰富化、内化、外化、概念化等）。所以，要对协作知识建构过程中微观的具体的可测量的行为和结果进行评价。

4.2. 学习绩效评价标准

学习绩效评价的标准有两个来源：

第一、依据每一个章节甚至每一个知识点的学习目标。

第二、依据协作知识建构的五个过程要素，即共享、论证、协商、创作、反思。

4.3. 学习绩效评价思路

微观建构理论倡导情境性评价和一般性评价相结合的评价思路，这对基于问题解决的课堂协作知识建构的学习绩效评价同样具有指导意义。

4.3.1. 情境性评价

情境性评价是对短时间内的学习绩效进行细致描述，尤其对学习者的具体认知过程进行揭示。在基于问题解决的协作知识建构过程中，情境性评价侧重在问题情境中的评价，在问题解决过程中进行评价，获得问题分析、问题解决以及作品创作的思维方式和认知建构过程。

4.3.2. 一般性评价

一般性评价是对较长时期内教学活动和学习效果的整体评价，提供与情境性评价互补的资料，主要侧重于获得学习者学习绩效的一般性资料，主要包括学生的阶段性成就，如创作作品、知识掌握、能力提高等。

4.4 学习绩效评价方法

在微观建构理论倡导的评价思路的启示下，基于问题解决的课堂协作知识建构的学习绩效评价方法主要有以下几种。

4.4.1. 口语报告法

口语报告法是由德国心理学家 Duncker 在 1945 年首次提出的，是揭示人的思维活动的有效方法。学生在问题解决过程中出声思考，讲出思考的内容，研究者或教师要记录学生的口语报告。在评价结束之后，对学生的口语报告进行转写、编码和分析。

4.4.2. 直接对话法

直接对话法是针对错误分析技术而提出，即让学生回忆他们在分析、解决问题的过程中哪些操作是冗余的，甚至是错误的，然后学生自己分析出现操作错误的原因。或让学生自主提出存在问题并分析原因，或教师就容易出错环节和问题与学生对话交流。

4.4.3. 计算机模拟评价法

利用计算机模拟人的思维过程、揭示人对知识的认知机制、外化内在的知识建构过程。如学生在分析解决问题时操作步骤或思维活动可以用计算机模拟的方式将其外显，即让学生依据自己的思维过程编写计算机程序，形成产品，或者为解决问题提供一种或几种策略。

4.4.4. 问卷测评

问卷测评主要是获得学习者的一般性评价资料，是在课堂协作知识建构结束之后实施的，问卷侧重于学生在协作知识建构过程中的行为表现，即学生是否出现过这个行为，而对于行为能达到的深度、持续时间和产生作用等方面并未体现。

4.4.5. 学习反思

学习反思侧重于对学习者知识点掌握情况的评价，学生对问题的协作建构过程进行反思：有什么收获、有哪些缺失、如何改进等。

5. 基于概念性问题的课堂协作知识建构的学习绩效评价

在此简单描述利用上述评价思路对广州市东山培正小学郑炜珍老师《平行四边形和梯形》课进行评价，这节课是典型的基于概念性问题的协作知识建构问题。

5.1. 学习绩效评价的内容和标准

5.1.1. 知识点的评价标准

本节课知识点的评价标准即各个知识点的目标层次，如识记平行四边形和梯形的概念、特征；理解平台平行四边形的关系和四边形的易变特性等。

5.1.2. 行为和结果的评价标准

从协作知识建构的过程要素出发，得出协作知识建构过程中行为和结果的评价维度。

表 1 协作知识建构过程中行为和结果的评价（缩减后）

协作知识建构	行为表现	完成结果
共享	1、与同伴分享了对四边形的理解	形成对四边形、平行四边形和梯形的初步认识
	2、与他人分享在点子图上创作的各种四边形	
论证	1、反对同伴的观点，并且提出自己的看法	不同观点的产生
	2、同意同伴的观点，并给出同意的理由	
协商	1、通过讨论，分析了长方形、正方形的关系	得出长方形和正方形的特征关系
	2、通过与他人讨论，得出平行四边形的特点	
创作	1、利用点子图创作了平行四边形和梯形	多个合格的平行四边形和梯形图案
	2、我们用七巧板工具设计了四边形的图案	
反思	1、我们评价、讨论同伴设计的四边形图案	小组的和个人的平行四

5.2. 学习绩效评价方法的选择

针对本节课，采用课堂提问法、问卷测评法、反思法和档案袋法来评价学生的学习绩效。

5.2.1. 课堂提问法：主要是评价学生对知识点的掌握情况。对于概念性问题评价，通过课堂提问法一般就能够达到评价的目的。

5.2.2. 问卷测评法：教师依据表中列出的可能发生的协作知识建构行为和结果制定测评问卷，以此来检测学生的“做了没有”层面上的学习绩效。

5.2.3. 反思法：教师给学生1~2分钟的时间，让学生反思在理解四边形概念、搜索四边形资源、分享观点和作品以及创作作品等过程中的操作、失误和收获等。

5.2.4. 口语报告法：在小组学习过程中，学生说出自己的思考过程，并在学生旁边放置录音设备，记录学生在协作知识建构过程中的口语报告。

5.3. 学习绩效评价的实施结论

利用上述评价方法对本节课进行学习绩效的评价，得出如下结论：

第一、通过教师的课堂提问发现，绝大多数学生基本能掌握知识点。

第二、学生在知识建构过程中都有问卷中所列行为的参与，但参与程度存在差别。参与程度最深的是“与同伴分享了对四边形的理解”，参与程度最浅的是“同意同伴的观点，并给出同意的理由”。

第三、口语报告法没有获得太多数据，分析原因有二：首先、本节课是基于概念性问题的，难度较低，而且学生多数时间在创作平行四边形和梯形，很少有深度的讨论；第二、学生在协作知识建构的过程中利用了网络平台上的培正论坛进行交流。

第四、学生对自己的协作知识建构过程进行反思，如许瀚文同学说“我觉得应该让我们尝试一下画四边形”；谭子俊同学说“我们应该结合平行和垂直一起学习”，等等。通过反思，学生的确发现了在协作知识建构过程中存在的问题。

微观建构理论的引进对学习绩效的评价提供了直接的理论指导，它倡导在情境中评价、在过程中评价，它以务实的态度指导协作知识建构过程的评价，试图通过多种评价方式来外化学生的协作知识建构过程，同时也赋予“评价”新的含义。

参考文献

- 辛自强（2006）。《知识建构研究：从主义到实证》。北京：教育科学出版社。
- 辛自强（2004）。《问题解决与知识建构》。北京：教育科学出版社。
- 董蓓菲（2005）。《新课程视野中的学习评价——评价类型与策略》中国小学语文教学论坛。
- 石娟（2008）。《网络环境下基于问题的协作知识建构的学习绩效研究》。广东：华南师范大学教育信息技术学院。

大專學生數位健康科技知識分享及學習生涯量表之發展

A Digital Scale Development about Knowledge Sharing of Health Technology and Learning

Career for Colleges Students

鄭金謀

輔英科技大學人文與管理學院

so003@mail.fy.edu.tw

【摘要】本研究旨在建構大專學生數位健康科技知識分享及學習生涯量表，以測試學生認知科技對健康影響之知識分享和學習生涯類型。透過抽樣調查，從全台技職校院取得 14 所大專學生有效樣本 1218 份，統計分析發現學生之學習類型可分擴散型、同化型、聚斂型、調適型等四類。生涯興趣類型分為實際型、研究型、藝術型、社會型、企業型、事務型等六類。健康科技知識分享模式涉及計畫行為、創新科技接受等理論，同時利用結構方程模式(SEM)驗證健康科技知識分享模式具有良好的適配度。健康科技知識分享模式涵蓋行為規範、分享態度、分享意圖、行為控制、分享行為等變項，其中行為規範對分享態度、分享態度對分享意圖、分享意圖對分享行為等均有正向關係。提出結論及建議事項，作為大學校院在健康科技知識交流、數位健康通識教學、產學合作、政府擬訂健康科技政策之參考。

【關鍵詞】 數位量表、健康科技、知識分享、學習風格、生涯類型

Abstract: The purpose of this research was to construct a model concerning knowledge sharing of health technology and learning career, and to compile a scale about knowledge sharing of health technology and learning career to understand the relationship between the knowledge sharing of health technology and learning styles for students. Then, via a survey method with cluster and stratified random sampling from 14 technological and vocational colleges in Taiwan, 1,218 of effective questionnaires were returned. The statistical analysis discovered that for the students belonging to any of the four learning styles, that is, diverging, accommodating, converging and assimilating. The career types of students belonged to one of six types—the realistic type, the investigative type, the artistic type, the social type, the enterprising type, and the conventional type. The results showed that the students' knowledge sharing of health technology consisted of behavioral norm, sharing attitude, sharing intention, sharing behavior and behavioral control in terms of the theory of planned behavior. The model of students' knowledge sharing of health technology was tested with good fit via structural equation modeling. The path correlation of model was based on the theory of planned behavior and technology acceptance model, sharing attitude toward sharing intention, behavioral norm toward sharing intention, and sharing intention toward sharing behavior, which all had positive effects. Finally, conclusions and suggestions were proposed as references for further research, such as transactional platform of knowledge sharing in health technology, the digital health liberal teaching, enhancing the cooperation of industries and institutes, and helping the government drawing up a technological policy.

Keywords: digital scale, health technology, knowledge sharing, learning styles, career types

1. 前言

隨著知識經濟時代的來臨，在面對瞬息萬變的競爭環境中，健康科技的知識分享與學習型態成為達成預期目標與創造競爭力的利器。知識分享可為知識經濟的主軸，但分享知識之前，關鍵性的使命乃為學習(learning)。知識必須源自於學習，同時須經過不斷的修改和更新。

透過學習，可不斷提昇組織成員的能力，而透過知識的分享，使得組織能力加速提昇、組織運作日益茁壯，也可更具有彈性以因應快速改變的環境。數位科技為學習帶來便利，進而強化知識藉由分享所創造的加乘效果(Tsai, & Lin, 2001)，學習者一同擁抱快樂的數位學習生活。

學習風格或類型(learning styles)是指學生在變化多端的科技環境從事學習活動時，經由其知覺、記憶、思惟等心理歷程，在外顯行為上表現出帶有個別差異的學習偏好。學習風格包含認知、情意、社會、生理等因素，具有獨特性、穩定性及一致性。了解學習風格可改善教學，對學生生涯發展亦有所幫助。在眾多的學習風格分類理論中，教育界使用最多的是Kolb的學習類型。大專學生生涯興趣類型，譬如Holland所謂人格與職業類型適配之分類，而其知識分享與其學習風格和生涯興趣類型究竟有何對應關係，此為本研究所欲解決的問題。

近年來高等及技職教育鬆綁後，大學校院快速擴張，為取得競爭優勢，不斷創新研發，提出增強學生各種能力之競爭策略(李國義，2004)，以培育科技人才為其重點發展特色。台灣科學技術白皮書—科技發展遠景與策略(2003-2006)，強調邁向知識、創新的科技發展。科技知識是競爭力的關鍵因素。學生本身就是科技創新過程中的重要參與者，會將新的創意帶入教室、實驗室、工廠，以及新一代的企業。研究者熱衷於健康科技的相關研究，希冀了解學生對健康科技知識分享與學習生涯狀況，進而在通識教學上貢獻綿薄之力。

2. 研究目的

建立大專學生健康科技知識分享與學習模型所依據之理論；發展大專學生健康知識分享學習生涯量表；了解大專學生健康科技知識分享與學習生涯類型之現況，並考驗健康科技知識分享與學習生涯結構模式。

3. 文獻評述

3.1 理論探討

本研究主要參酌計畫行為理論(Theory of Planned Behavior, TPB)、科技接受模式(technology acceptance model)、創新擴散理論(innovation diffusive theory)、健康信念模式(health belief model, HBM)、社會認知理論(social cognitive theory)的觀點，來探討大專學生健康科技知識分享與學習生涯類型。

3.2 數位科技健康知識分享與學習生涯類型

數位學習是使用者透過電腦、廣播、錄音帶、網路等數位科技電子資源媒體來進行學習的方式，並由數位內容及教學方法來創造學習經驗，以達成學習目的。依據Kolb之經驗學習理論，學生的學習風格(learning Style)是影響學生學習和知識分享的重要因素。大專學生的學習風格與知識分享有何關連尚待深究。調適型學習風格的大專學生主動驗證，提供個人化營養資訊，採用健康的飲食行為(Cousineau, et al. 2006)。透過問題呈現與同儕評估可進行網路學習或知識分享(Barak, Rafaeli, 2004; Tseng; Chiang; Hsu, 2008)。

依據Holland(1973)認為個人的職業選擇為其人格的反應，職業興趣量表可以反應個人的自我概念、生活目標等人格特質。Holland認為人們會藉由工作的選擇和經驗來表達自己的興趣、價值觀和人格。為便於分析，他將人與工作領域分成了六種特定的類型。Holland(1985)認為人格會影響職業的選擇；一種職業就好比一種生活方式、生活環境，由此可形成多數人對職業的刻板印象。人格類型若符合其工作環境，這種情況即稱為諧和、契合或適配(congruence)(鄭金謀，2008)。學習思考風格模式可預測Holland的六種生涯類型(Zhang, 2007)。數位學習科技健康相關量表如在組織系統之應用(Wang, Wang, Shee, 2007)，而學校健康科技環境從事知識分享與學習風格類型以及生涯類型之研究，尤其衡量使用數位科技頻繁之大專學生健康衝擊之量表，有急欲拓展之領域。

4. 研究方法

4.1 抽樣調查法

採叢集分層隨機抽樣方式，抽樣技職校院學生進行正式量表測試。按北中南東地區各抽取 2-3 所大專學生，每校 60-150 人，預計 1500-1900 人。施測完後，建立大專學生健康科技知識分享與學習常模，以為解釋結果之依據。

4.2 研究對象

抽樣調查人數，輔英、屏東、高雄海洋、南亞、清雲、台北、高雄第一、南台、中台、台灣、雲林、樹德、弘光等科技大學，共計回收有效問卷 1218 人。其中男生 457 人，女生 761 人。五專 447 人，四技 452 人，二技 319 人。醫護健康類 343 人，管理設計類 177 人，環境生命類 195 人，外語幼教類 173 人，建築光電類 168 人，航海水產類 162 人。

4.3 研究工具

研究者從已發展之技術校院學生健康科技素養量表(鄭金謀, 2008)中之知識分享 18 題；加上修訂自 Kolb 所編製之第三版學習風格量表(Learning Style Inventory, version 3, 1999)，12 題選擇題，所得分數散佈在 Kolb 之四種學習風格類型；另酌修自生涯興趣量表(測驗出版社, 1994)中 60 種職業名稱而分屬六種生涯類型，經統整編修而成大專學生健康知識分享學習生涯量表。可為測試大專學生處在科技環境下對健康認知之知識分享程度，同時了解其所屬學習風格類型和生涯興趣類型以及彼此之對應關係。

4.4 研究假設

1.1 行為規範對分享態度有正向關係；1.2 分享態度對分享意圖有正向關係；1.3 行為控制對分享意圖有正向關係；1.4 行為控制對分享行為有正向關係；1.5 分享意圖對分享行為有正向關係；2.1 學習風格類型在健康科技知識分享各變項上有顯著差異；2.2 生涯興趣類型在健康科技知識分享各變項上有顯著差異。

5. 結果與討論

5.1 大專學生健康科技知識分享與學習生涯量表信效度分析

從已發展之技術校院學生健康科技素養量表(鄭金謀, 2008)中之知識分享 18 題，發現每一題項與分量表相關在.589~.764，因素負荷量在.623~.803，均在標準的範圍內，顯示量表所涉內容之構念效度或準確度相當高。

使用 Cronbach's α 評估各結構指標因素的內部一致性，以顯示量表信度的高低，發現健康科技知識分享 18 題 Cronbach's α 為.953，內含行為規範 4 題為.873，分享態度 3 題為.899，分享意圖 4 題為.850，分享行為 4 題為.836，行為控制 3 題為.833。顯示各題項內容一致性(consistency)相當強。Kolb(1999)學習風格量表第三版，量表中 CE、RO、AC、AE 的四個基本構面具有良好的內部效度.82、.73、.83 及.78，透過 SPSS(統計套裝軟體)第 15 版計算內部一致性：具體經驗的 Cronbach's α 值為.72、省思觀察為.71、抽象概念為.77、主動驗證為.74；顯示各題項內容一致性(consistency)可接受。生涯興趣量表(測驗出版社, 1994)各向度之六種類型的折半信度，其係數在.73~.93 之間，各向度之 α 信度係數在.90~.93 之間。聚斂效度與區辨效度考驗之結果，各相關係數之數值呈現合理之聚斂效度與區辨效度，就各類型總分內部相關及因素分析結果觀之，亦具有符合 Holland 理論結構的建構效度。

5.2 大專學生健康科技知識分享模式

此模式所依據之理論，從文獻評述及健康科技相關研究結果，可發現主要的有計畫行為理論、創新科技接受模式、健康信念模式、社會認知理論等。再者，此模式須經修正之後才適用，其適配度均達到良好適配程度。可發現大專學生健康科技知識分享結構模式相當良好，因此進一步以此模式驗證假設成立情形。

表 1 大專學生健康科技知識分享結構模式估計值

因素	因素	估計值	因素	因素	估計值	假設成立與否
知識分享	---> 行為規範	.892	行為規範	---> 分享態度	.595	成立
知識分享	---> 分享態度	.327	分享態度	---> 分享意圖	.421	成立
知識分享	---> 分享意圖	.492	行為控制	---> 分享意圖	.087	不成立
知識分享	---> 行為控制	.803	行為控制	---> 分享行為	.025	不成立
知識分享	---> 分享行為	.536	分享意圖	---> 分享行為	.398	成立

依表 1 發現本研究所建構之大專學生健康科技知識分享模式，利用 SEM 處理驗證假設，健康科技知識分享涵蓋行為規範、分享態度、分享意圖、行為控制、分享行為等要素。而行為規範對分享態度，分享態度對分享意圖，分享意圖對分享行為等均有正向關係，假設 1.1、1.2、1.5 成立，假設 1.3、1.4 未成立，顯示多符合計畫行為理論、科技接受模式。

5.3 健康科技知識分享與學習風格

經變異數分析發現學習風格類型在知識分享各變項無顯著差異(表略)，假設 2.1 未成立。即學習風格類型在健康科技知識分享上不分軒輊，並無因不同類型而有差異。

5.4 健康科技知識分享與生涯興趣

經變異數分析發現生涯興趣類型在知識分享各變項無顯著差異(表略)，假設 2.2 未成立。即生涯興趣類型在健康科技知識分享上不分軒輊，並無因不同類型而有差異。

6. 結論與建議

6.1 結論

利用結構方程模式驗證健康科技知識分享模式具有良好適配度，知識分享模式涵蓋行為規範、分享態度、分享意圖、行為控制、分享行為等變項，其中行為規範對分享態度、分享態度對分享意圖、分享意圖對分享行為等均有正向關係。統計分析發現學生在學習類型分屬擴散型、同化型、聚斂型、調適型四類，其在知識分享上並無顯著差異，假設未獲支持。興趣類型分屬實際型、研究型、藝術型、社會型、企業型、事務型六類，其在知識分享上亦無顯著差異，假設未獲支持。研究發現技職校院學生健康科技知識分享與學習風格之對應關係不顯著，而與生涯興趣之對應關係有部分顯著，至於學習與生涯類型在知識分享未有顯著交互作用。有待透過如質性晤談方式實地了解學生之健康科技知識分享與學習生涯之關係現況。

6.2 建議事項

建立大學校院在健康科技知識交流管道，相關產業機構可提供學術界研究之園地，若能建構產學合作之橋樑投資生產，則樂待其成。大專學生健康科技知識分享與學生學習生涯所依循之理論及量表，可提供相關研究教師發展科技學習理論及建立量表之參考。本研究攸關大專學生健康之科技知識與學習風格類型及生涯興趣類型研究，結果可為政府擬訂健康科技政策，進而實施嵌入教學統整研究(Hinostroza & Mellar, 2001; Yi, Jackson, Park, Probst, 2006)，造福青年身心健康之參考。重新編製學習風格與生涯興趣量表，將學習風格類型改為創意學習風格，量表可測出學生創意學習風格之高中低三級；生涯興趣類型可改為創新生涯興趣，量表可測出創新生涯興趣高中低三級；然後與健康科技知識分享作一相關或預測研究，進而建立知識分享與學習生涯之結構模型。如此可作為提昇創新能力與競爭力之依循。

7. 參考文獻 (略)

基於微軟 OpenXML 的 Office 操作題自動閱卷的實現

An Open Xml-based implement of automatic evaluation for MS Office Operations

陈杰、朱卫东
北京交通大学

【摘要】在電腦教育領域，微軟 Office 操作能力的考試一直都是重要內容，對於這些操作題的手工評判工作是相當繁瑣的，而且很容易誤判、漏判。在 Microsoft Office 2007 中，微軟採用 Open XML 做為其文檔存儲的格式。本文介紹了實現 Office 操作題自動閱卷的一種方法，該方法憑藉 Open XML 文檔格式的諸多優勢，在實踐領域發揮了較好的效果。

【關鍵字】 Open XML、Office 操作題、自動閱卷

Abstract: The evaluation of operating MS office is important for beginners. But the evaluation for this exam is difficult to be made. With the release of Office 2007, Microsoft use Open XML as a new document format. This article introduced a technique to evaluate the exam above automatically. This new technique is based on MS Open XML and has good benefits in practice.

Keywords: Open XML, MS Office Operations, automatic evaluation

1.前言

在 Microsoft Office 2007 版本以前，二進位檔（包括.doc、.dot、.xls 和 .ppt 檔）在存儲和傳輸資料方面發揮了重要作用，但已經無法滿足新的工作環境的挑戰，例如脫離具體 MS Office 環境的資料共用與文檔分析。到了 Microsoft Office 2007 版本，微軟完成了二進位檔格式到 XML 格式的轉換。新的檔格式稱為 Microsoft Office Open XML 格式（Office XML 格式），並將逐步成為國際標準。

在電腦操作能力考試中，對於微軟 Office 的操作考試是一項重要內容，包括 Word、PowerPoint、Excel 等軟體的操作。對於這些操作的手工評判工作是相當繁瑣的，而且很容易誤判、漏判；在 MS Office 2007 中，由於採用了 Open XML 標準，憑藉 XML 文檔格式的諸多優勢，為實現操作題的自動閱卷提供了一種新的方法。

2.微軟 OpenXML 的特性

微軟 Open XML 格式改善了檔和資料管理、資料恢復以及與行業系統的互操作性。它們擴展了以前版本的二進位檔的功能。任何支援 XML 的應用程式都可以訪問和處理採用新檔格式的資料。應用程式不必是 Microsoft Office 系統乃至 Microsoft 產品的一部分。用戶還可以使用標準轉換來提取或重用資料。此外，安全性問題大大減少，因為存儲為 XML 的資訊實質上是純文本。因此，資料可以無障礙地通過企業防火牆。

新的 Office XML 格式引入了很多優點，這些優點不僅有助於開發人員以及他們構建的解決方案，而且還有助於單個用戶和各種規模的組織。對於實現操作題的自動閱卷系統而言，至少具有以下有利的特性。

1、文檔易於處理：使用 Office XML 格式核心的行業標準 XML，簡化了 Microsoft Office 應用程式和企業業務系統之間的資料交換。解決方案可以通過使用能夠處理 XML 的標準工具和技術來改寫 Office 文檔中的資訊或創建文檔，而無需訪問 Office 應用程式。這種新的格式使開發人員可以創建、修改、分析文檔而無需使用 Office 代碼。

2、公開和免版稅：Office XML 格式基於 XML 和 ZIP 技術，所以它們能夠廣泛使用。這些格式和架構的說明書在 Microsoft 的免版稅許可證下發佈和提供，它是公開提供的，可供企業廣泛使用。這為企業和用戶降低了不少成本。

3、可靠性：Office XML 格式設計為比二進位格式更為可靠，因此有助於降低由於損壞或破壞檔而丟失資訊的風險。

4、高效性：Office XML 格式使用 ZIP 和壓縮技術存儲文檔。新格式的顯著優點就是大大減小了檔大小，比類似的二進位文檔減少了多達 75%。這是結合使用 XML 和 ZIP 技術存儲檔的優點之一。由於 XML 是一種基於文本的格式，壓縮性非常好，而且 ZIP 容器支援壓縮內容，因此用戶可以顯著地減小檔大小。這一特性在包含大用戶數量的線上考試系統中將發揮很大的優勢。

5、安全性：在操作題線上考試系統中，往往需要上傳大量的文檔。而 Office XML 格式的開放性將檔轉換為更安全、透明的檔，用戶和開發人員可以放心地共用文件。此外，Office XML 格式包含一種特定用途的格式，這種格式為具有嵌入代碼的檔賦予單獨的副檔名，從而使 IT 人員能夠快速識別包含代碼的檔。

3. Office 操作題自動閱卷實現的原理和方法

3.1 實現的原理

在微軟 Office 系列軟體操作題的考試中，考生會得到一個包括若干段落或表格的原始文檔(例如 demo.doc)，一段包含操作要求的題幹描述；在根據題幹要求完成相應操作之後，考生提交操作後的文檔，由閱卷者手工閱卷或者自動閱卷系統參照相應的評分規則，對此文檔進行流覽、分析，最後給出考試成績。

其一般流程為：

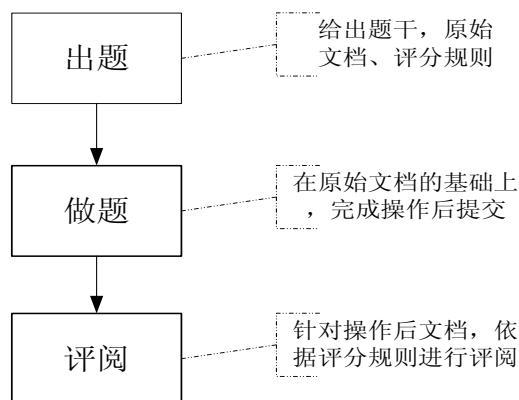


圖 1 操作題考試的一般流程

通過對電腦操作題考試要求和內容的分析，可以歸納出常見的考點集。例如，對於 Word 操作題的考試，考核內容主要是排版和製表，排版包括設置文字屬性、段落屬性、頁面屬性；製表包括設置表格的單格邊框屬性和行、列屬性，通常有如下考點：

表 1：Word 操作題考試常見考點

常見考點			
字體	字型大小	顏色	段落縮進
加粗	插入表格	行/段間距	圖片格式
下劃線	表格邊框網底	頁面分欄	插入文本框
複製粘貼	表格格式	頁眉頁腳	首字下沉
查找替換	插入刪除行列	頁邊距	上標下標

對於上述 WORD 操作題考試中的考點，在 Office2007 的 WORD 文檔（其副檔名為.docx）中都有相應的 XML 節點標籤與之對應。通過建立包含“考點——標籤”對應關係的 Open XML 標籤庫，再結合與每一道考試題對應的評分規則庫，就可以運用 XML 文檔解析器對考生上傳的帶批閱文檔進行掃描、分析，從而實現 WORD 操作題的自動閱卷。而 MS Office 系列中的 PowerPoint、Excel 的操作題自動閱卷的原理與之類似。

3.2 實現的方法

依據上述原理，可以給出操作題自動閱卷的處理流程（以 Word 操作題為例）如下。

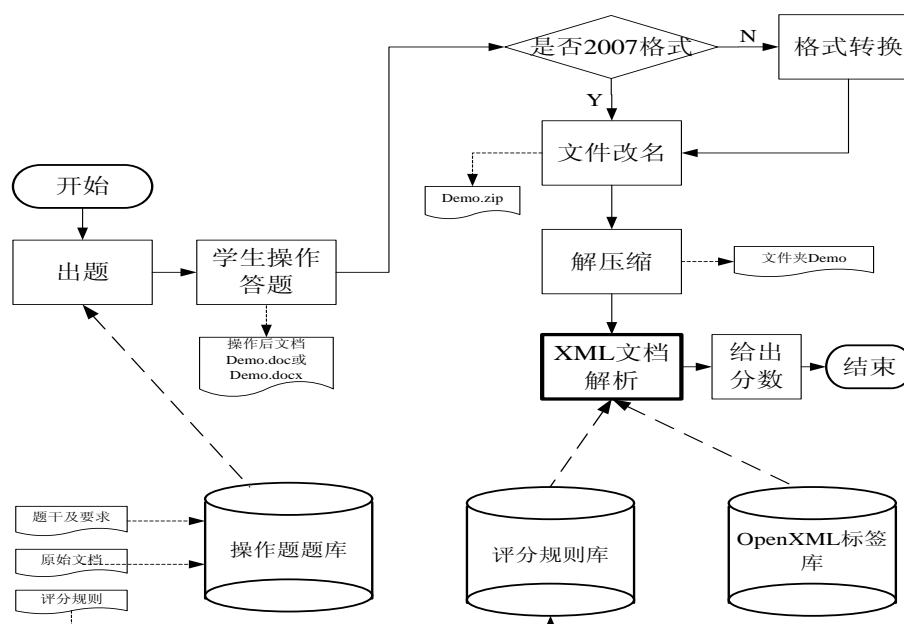


圖 2 操作題自動閱卷處理流程(以 Word 為例)

如圖 2 所示，在整個流程包括了操作題考試的一般流程，即出題、做題、評閱，其中評閱階段是本文關注的重點。評閱階段從考生完成操作提交操作後文檔開始。系統首先要進行文檔預處理，即根據考生所使用的 Office 軟體的版本，決定是否需要對操作後文檔進行格式轉換，得到 2007 版本格式的 Word 文檔，其檔副檔名為*.docx；然後，系統將此.docx 文檔通過檔改名直接改為.zip 文檔，再將其解壓縮，得到一個包含一系列 XML 檔的檔夾，此檔夾包含了所有與考生操作相對應的內容；最後，通過 XML 文檔解析器，依據“評分規則庫”以及“OpenXML 標籤庫”對這一包含 XML 檔的檔夾進行掃描、分析，給出考生得分。

3.3 幾點說明

- 文檔預處理：當考生所使用的 Office 版本是 2007 以下時，需要在預處理階段將其轉化為 2007 版本的格式，使之符合 Open XML 的標準格式，從而才能用 XML 解析器對其進行自動閱卷。這一操作的核心元件由遵循 Open XML 標準的第三方提供；在 Java 平臺中，可以應用 JACOB 實現 Office 文檔版本格式的轉換。
- XML 文檔處理：這是整個自動閱卷的核心模組，其實質是一個基於評分規則庫和 OpenXML 標籤庫的 XML 文檔處理器，可以採用符合 XML 標準的 DOM、SAX API 來實現。這兩者都提供了對 XML 文檔解析的支援，並且並不相互排斥。
- 評分規則庫：評分規則庫是自動閱卷評判的依據之一，與操作題題庫相關聯。在出題階段，除了給出題幹要求及原始文檔，還需要給出相對應的評分規則，包括每一道操作題的若干個考點、各個考點的具體參數值。
- Open XML 標籤庫：XML 處理器對考生提交的文檔進行分析時，根據評分規則庫中的考點，參照 Open XML 標籤庫獲取對應的 XML 標籤特徵，並讀取其參數值，借此判斷考生在某個考點上是否得分。標籤庫的構建，需要基於對微軟 Open XML 文檔格式

的深入細緻分析；需要結合電腦 Office 操作題考試的常見考點（如本文表 1），從中找出常見考點對應的標籤所在的位置以及標籤名稱。需要指出的是，Office 系列的各個軟體對應的文檔雖然都遵循了 Open XML 標準，但在具體的文檔組織、內容上還是存在著一些差異；Word、PowerPoint、Excel 等文檔中各個考點對應的標籤庫有著自己獨立的集合。下表給出了 Word 文檔中 Open XML 標籤庫的示例。

表 2 XML 標籤庫示例

考點名	對應的檔	標籤位置	標籤名
標題樣式	word\document.xml	w:document/w:body/w:p/w:pPr/w:pStyle	w:val
對齊	word\document.xml	w:document/w:body/w:p/w:pPr/ w:jc	w:val
超鏈結	word\document.xml	w:document/w:body/w:p/w:hyperlink	r:id
段落網底	word\document.xml	w:document/w:body/w:p/w:pPr/w:shd	w:fill
著重號	word\document.xml	w:document/w:body/w:p/w:r/w:rPr/w:em	w:val
字型大小	word\document.xml	w:document/w:body/w:p/w:r/w:rPr/w:sz	w:val
字體	word\document.xml	w:document/w:body/w:p/w:r/w:rPr/w:rFonts	w:eastAsia
.....			

4. 應用價值

基於 Open XML 實現的 Office 操作題自動閱卷，具有以下特點：

- 1) 準確率高：通過構建一個與考點精確匹配的 Open XML 的標籤庫，考生的每一項操作都可以找到與之匹配的 XML 標籤，借助一定的模糊評分演算法，可以實現較高的準確率。
- 2) 效率高：借助 XML 在文檔處理中的高效、便捷的特點，與處理以往的二進位檔相比，可以實現較高的執行效率。
- 3) 演算法簡單：DOM、SAX 等元件提供了對 XML 文檔進行處理分析的 API，基於這些 API，對 Office 文檔進行處理的演算法比較簡單，有利於系統的後期升級、維護。

基於上述特點，該方法具有較大的應用價值。既可以將其設計為一個獨立的 Office 操作題閱卷系統，也可以將其嵌入到 E-Learning 中常見的線上考試系統、網路教學平臺中，擴充、優化其功能，實現對 Office 操作題的自動閱卷；例如，在基於 WEB 的“思源網路教學支援平臺”中，就集成了基於該方法的自動閱卷模組，在實踐應用中獲得了較好的效果。

参考文献

- [1] <http://www.microsoft.com/china/msdn/openxml/>, 2009 年 2 月檢索，Open XML 開發中心
- [2] 劉紅梅（2007）。基於 VBA 的 Word 操作題自動閱卷方法的研究與實現。《電腦教育》，2007 年 24 期。
- [3] 張永崗（2007）。基於網路的電腦考試系統的研究。《科技資訊》，2007 年 27 期。

電腦化適性測驗的條件化信度區間表示法

The conditional reliability zone of the computerized adaptive testing

吳佳儒、陳柏熹、周業太*

國立台灣師範大學教育心理與輔導學系

國立台灣師範大學心理與教育測驗研究發展中心*

【摘要】 本研究主要是探討電腦化適性測驗的信度表示方法。電腦化適性測驗的測量精準度是以試題反應理論的訊息量來表示。Raju, Price, Oshima 與 Nering (2007)曾將 IRT 的訊息量概念轉成傳統測驗的信度，並以條件化信度來表示隨著受試者能力而改變的信度值。本研究應用條件化信度發展出電腦化測驗的條件化信度區間表示法。我們以 Chen, Lin, Hsu, & Yen (2007)所發展出來的多元智慧電腦化適性測驗題庫為例來呈現 CRZ，結果發現該題庫在低難度與高難度範圍還需要增補題目，以提升該測驗的適用性。

【關鍵詞】 電腦化適性測驗、試題反應理論、條件化信度、題庫

Abstract: The goal of this research is to develop a representation of the reliability of computerized adaptive testing (CAT). We applied the concept of conditional reliability, suggested by Raju, Price, Oshima & Nering (2007), to represent the conditional reliability zone (CRZ) of the CAT. The CAT items bank for multiple intelligences developed by Chen, Lin, Hsu, & Yen (2007) were used to compute the conditional reliability zone. Results indicated that items for low level and high level abilities need to be added to the items bank.

Keywords: computerized adaptive testing, item response theory, conditional reliability, item bank

研究背景

電腦化適性測驗(computerized adaptive testing; CAT)是一種高效率的測驗形式，它能让受試者只接受約傳統紙筆測驗一半的題數，就能達到與傳統紙筆測驗相同的測量精準度。CAT 的理論基礎是試題反應理論(item response theory; IRT)。在試題反應理論中是以訊息量(information)來表示測量精準度的概念。訊息量有別於傳統的測驗信度(reliability)，而且此訊息量也會隨著受試者的能力不同而變。然而，IRT 與訊息量的概念都較複雜，不容易被理解。因此 Raju, Price, Oshima 與 Nering(2007)將 IRT 的訊息量概念轉成傳統測驗的信度值，並以條件化信度(conditional reliability)來表示隨著受試者能力而改變的信度值。

雖然我們可以用訊息量來表示電腦化適性測驗中的測量精準度，但是每個人所選到的題目都不相同，因此訊息量須等到受試者做完測驗後才能算出來。為了讓測驗使用者在進行測驗之前就能有參考資訊，以作為是否選用該測驗的依據，本研究應用了 Raju, Price, Oshima 與 Nering(2007)所提出的條件化信度，計算整個題庫對不同能力受試者所提供得訊息量，並轉換成該題庫執行電腦化適性測驗的最高條件化信度；再以隨機選題方式模擬算出受試者在執行若干非適性試題後的訊息量，並轉換成該題庫未來執行若干題的電腦化適性測驗之最低條件化信度。將上述兩者間的區間定義為電腦化適性測驗的條件化信度區間(conditional reliability zone; CRZ)。

CRZ 的概念是在受試者執行某測驗題庫的電腦化適性測驗之前，預估的受試者未來可能的信度區間，以供測驗使用者決定是否適合對受試者實施電腦化適性測驗。條件化信度區間也能提供題庫的相關資訊給測驗發展者，以作為增補試題改善題庫的依據。

研究方法

本研究以 Chen, Lin, Hsu, & Yen (2007)所發展出來的多元智慧電腦化適性測驗題庫為例，計算該題庫條件化信度區間，以供改善題庫之參考。該測驗題庫是參考 Gardner(1983)的多元智慧理論所發展出來的。其中肢體動能的電腦化測驗技術尚未開發出來，因此在該測驗中將肢體動能刪除，改用 Goleman(1995)提出的情緒智慧來取代。該測驗最後建置出語文、數學邏輯、音樂節奏、視覺空間、自然觀察、人際互動、自我內省、情緒智慧等八向度的多媒體電腦化測驗題庫。目前共有 200 題，平均每個向度約有 25 題。測量對象為國中與高中(職)學生。該測驗於 2005 年~2006 年間分別對 1261 為國中生與 1076 為高中(職)生學進行預試。本研究使用該測驗預試後所計算出來的多向度試題難易度來計算各題庫對不同能力者之最高條件化信度；並對各向度題庫進行隨機選題，計算該測驗對不同能力者之最低條件化信度。並繪製條件化信度區間圖。

研究結果

研究結果顯示，該多元智慧測驗中各向度題庫的 CRZ 不太相同。以未來若對中等能力者執行 10 題之單向度電腦化適性測驗為例，若以最大概似法(maximum likelihood; ML)進行能力估計，其信度區間約在 0.55~0.8 之間。以語文智慧題庫為例，如圖 1 所示。整體而言，各向度題庫都能對中等能力者提供較高的測驗信度，但是對高能力與低能力者的測驗信度較低。因此未來若要執行電腦化適性測驗，應增補高難度與低難度試題，才能增廣該題庫適用範圍。

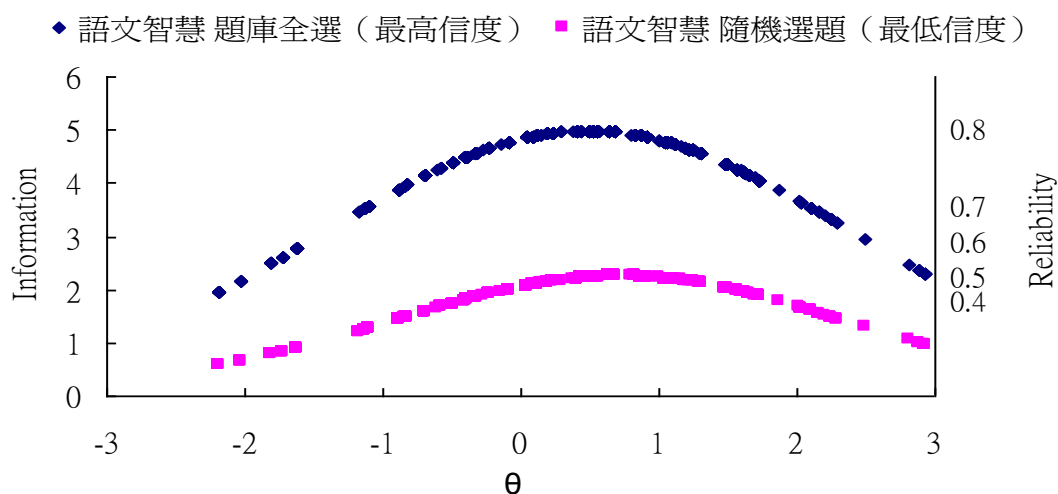


圖 1 電腦化適性測驗的條件化信度區間

參考文獻

- Chen, P.-H, Lin, Y.-L., Hsu, C.-L., & Yen, Y.-C.. (2007). *The Development of Multidimensional Computerized Adaptive Testing for Multiple Intelligence*. Paper presented at the IMPS 2007, the 15th International and 72nd Annual Meeting of the Psychometric Society. Tokyo, Japan.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic books.
- Raju, N. S., Price, L. R., Oshima, T. C., & Nering, M. L. (2007). Standardized Conditional SEM: A Case for Conditional Reliability. *Applied Psychological Measurement*, 31(3), 169-180.
- Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

專題研習中硬支架和軟支架對學生問題解決的影響

The effect of soft and hard scaffolds on students' problem solving in Project-based learning

周玉霞、李芳乐、胡继昌*
香港中文大学资讯科技教育促进中心
香港胡保良局胡忠中学*

【摘要】 本論文對比了課程設計階段的硬支架，和課程實施過程中的軟支架在學生探究專題中對於問題解決過程的影響，選擇了 280 名香港的初中學生，對他們的問題解決過程（問題理解、空間表徵和問題解決方案）進行了研究，結果發現，問題解決過程中教師提供的軟支架起主要作用，教師的教學取向（軟支架）越接近於建構主義學習型，學生的問題解決過程越順利，此時硬支架才能幫助學生在問題理解、問題空間表徵中有更好的表現。但是硬支架對於學生的問題解決方案設計無直接的幫助。

【關鍵字】 專題研習、WebQuest、硬支架、軟支架、問題解決

Abstract: This study was conducted to shed light on the effect of hard scaffolds and soft scaffolds on students' problem solving processes in Project-based learning. 280 secondary students from Hong Kong were chosen to participate in this study and the problem solving process including problem understanding, representation of problem space and problem solving plan were explored. It was found that the key to quality of problem solving lied in the teacher's soft scaffolds. The closer the teacher's teaching orientation is to constructivism, the better students' problem solving. And in this circumstance, the hard scaffolds are helpful for students to understand the problem and represent the problem space, but did not directly affect the problem solving planning.

Keywords: Project-based learning, WebQuest, hard scaffold, soft scaffold, problem solving

1.前言

專題研習 (Project-based learning) 是香港課程改革的四個關鍵項目之一，在各科尤其是通識科的教學中佔據三分之一的課時。過去幾年來專題研習實施中，學生缺乏相關的知識背景和問題解決經驗，而令專題研習往往流於簡單的複製、粘貼，並沒有真正達到提高學生問題解決能力的目的。

在專題研習的早期發展中，Dewey(1938) 就指出在學生的自主學習活動中需要教師的指導，以學生為中心不是放任自流，而是精心設計下的學生探究活動。因此給學生提供支架，可以幫助學生的自主性發揮。在哪個階段給學生提供支架，以及提供怎樣的支架，才能促進學生的探究活動呢？

為了回答這兩個問題，本文通過創設兩種不同支架環境的專題研習（一般的專題研習和 WebQuest），並結合教學過程中教師提供的動態支架，來研究學生在不同階段、不同形式的教學支架中，問題解決過程中的表現。

教學過程中的支架 (scaffolding) 是指教師為學生獨立學習所提供的必要支持 (包括資源、過程等)，使學生在這樣的支持下可以達到最近發展區 (zone of proximal development) (Bransford, Brown, & Cocking, 2000)。

支架理論的發展中，由起初的支持學生學習的策略、工具和指引 (Hannafin, Land, & Oliver, 1999)，近年來又進一步擴展，包括學生可以使用的各種工具和資源，尤其是多媒體網路的支持 (Kao, Lehman,

& Cennamo, 1996)。Brush & Saye (2002) 在課程發展的研究中，進一步提出兩種類型的支架：軟支架 (soft scaffolds) 和硬支架 (hard scaffolds)。所謂軟支架，是指教師或同伴提供的動態的 (dynamic)、情境相依的 (situation-specific) 支架，這樣的支架基於學生的學習回饋。教師需要不斷診斷學生的需要，才能真正提供有效的支架。相比之下，硬支架是靜態的 (static)，是事先根據對學生學習中可能遇到的困難而設計的支持。在教學軟件中經常會嵌入這樣的支架 (Kao, Lehman, & Cennamo, 1996)。本文沿用這樣的分類，在課程設計階段給學生提供不同的支架，對比硬支架和軟支架對學生問題解決的影響。

2. 研究設計

為了進一步深入研究硬支架——在教學設計階段預先根據學生可能的需要而設計的教學支持，以及軟支架——即教師根據學生實際需要而即時提供的教學支持 (Brush & Saye, 2002) 對學生專題研習的影響，本論文創設了兩種不同硬支架環境，並研究在不同硬支架支持下，不同教學取向的軟支架如何影響學生的問題解決。

為了體現硬支架的不同，筆者選擇了兩種典型的代表不同硬支架的專題研習學習環境，一種是一般的專題研習，即教師未事先設計資源和解難步驟，讓學生自己決定解難的過程和查詢相關的資源；另外一種是增加了硬支架（包括專題需要的網路資源、問題解決的方法等）的專題研習——WebQuest 學習環境。後者在專題研習的課程材料中預先設計了學生的解難支架，包括相關的網路資源和“目標——手段”的分階段解難步驟，這些支架都屬於課程實施之前教師預先設計的硬支架。

在軟支架部分設計，參照 Barr & Tagg (1995) 提出的講授取向和學習取向，根據研究者對香港教師的實際教學情況的觀察，提出三種取向的軟支架，即講授取向、學習取向和混合取向。講授取向 (instruction orientation)：教師在教學過程中扮演傳遞資訊的作用，大多數時間學生只是在被動聆聽 (O' Neill & Auken, 2003)。學習取向 (Learning orientation)：教師的角色是一個促進者 (facilitator) 和監控者 (monitor)，目的是幫助和引導學生自主學習 (O' Neill & Auken, 2003)。混合取向 (mixed orientation)，是處於講授取向和學習取向之間的教學取向，指教師在教學過程中既注重講授，也適當組織學習活動，這個過程中教師依然是課堂內的主要講解者，學生既注重基本的知識技能，也有一定空間的探究活動。在設計了不同的教學支架學習環境中，學生的問題解決呈現出怎樣的差異呢？

學生在本論文希望通過研究硬支架和軟支架對學生的問題解決過程的影響，幫助教師更好地設計課程和實施課程，以提高學生專題研習水準。研究問題如下：

1. 硬支架和軟支架對學生問題解決過程有何影響：1.1 學生的問題理解有何不同？1.2 學生的問題空間表徵有何不同？1.3 學生的解難過程策略有何不同？

本研究採用量化研究方法，量化研究理想的抽樣方式是隨機抽象，然而教育研究一般由於種種條件限制，經常會採用方便取樣 (Punch, 2005)。由於時間和專案規模的限制，本研究量化部分的樣本選擇即為方便取樣，以香港沙田馬鞍山某中學的中二中三年級共 10 個班級 280 名學生 (男生 129 人，女生 151 人) 作為研究樣本。

本研究的主要的研究工具是問題解決開放題，研究者和教師共同編制，並進行試測後，得到可接受的信度和效度後才正式施測。由於問題解決過程和學生的學習能力和之前的問題解決技能有關，因此本研究根據學校的分班情況¹，根據學生學力分為又把他們分為兩個學習水準的組別。此外在實驗前進行了學生問題解決的前測，作為協變量，以排除學生問題解決方面的先備條件對此次實驗的影響。

3. 研究發現

教師提供了不同的硬支架和軟支架，它們對學生的問題解決過程有何影響呢？

¹ 学生在小学升中学的时候，根据他们的入学成绩，分为两种水平的学力：低和高。次项研究中，由于使用方便取样，所以未能避免学力这一因素的影响，故纳入分析的变量中。

學生在這兩種不同硬支架的學習環境中（一般的專題研習和 WebQuest），問題解決過程均未有明顯不同，T 檢驗差異不顯著。納入學力、軟支架和性別因素后進行協方差分析，發現教師在課程實施過程中提供的軟支架成為影響問題解決過程策略的主要因素。教師的軟支架越接近於建構主義學習型取向，學生解難過程中的策略也就越接近於專家解難策略。因此教師軟支架對學習環境是否發生作用起著舉足輕重的作用。這與 Bain (1995) 的觀點是一致的。他認為學習環境在對學生發揮影響的時候，離不開教師的詮釋，教師的教學方法影響學生學習過程各種內部因素。具體分析結果如表 1。

表 1. 學生解難策略、學業成績的協方差分析

	比較項目	前測		學力 (A)		軟支架 (B)		硬支架 (C)		性別 (D)		B*C	
		F	p	F	p	F	p	F	p	F	p	F	p
解難過程	問題理解	0.17	0.678	23.63	0.000	1.63	0.199	1.77	0.185	0.57	0.450	4.05	0.045
	空間表徵	7.02	0.009	14.52	0.000	4.38	0.014	2.71	0.101	1.53	0.217	16.96	0.000
	問題方案	0.26	0.609	9.50	0.002	10.16	0.000	0.02	0.877	5.50	0.020		

從表 1 可知，學生學力、教師提供的軟支架影響學生問題解決的兩個重要因素。學力在所有維度上都有主效應；軟支架在問題解決過程中的空間表徵、問題解決方案維度具有主效應。另外，在問題理解和空間表徵維度，軟支架與硬支架有交互作用；此外，性別對問題解決無影響作用，下面分而述之。

3.1. 學力對學生的問題解決過程策略具有主效應

實驗後，控制前測成績的影響，學力高的學生依然在各項問題解決得分方面顯著高於低學力學生，具體分析結果如下，問題理解後測 ($F=23.63$, $p<.001$)、問題空間表徵後測 ($F=14.52$, $p<.001$)、問題解決方案後測 ($F=9.50$, $p<.005$) 均高於學力低的學生。

3.2. 問題理解

硬支架與軟支架對學生的問題理解具有交互作用。在問題理解這個維度，相比硬支架而言，教師提供的軟支架起著主要的影響作用。只有在教師提供的軟支架趨向於建構主義學習型時，WebQuest 中的硬支架才更加有利於學生對問題的深入理解。

3.3. 問題空間表徵

在問題空間表徵這個維度，軟支架和硬支架有交互作用，在提供了硬支架的 WebQuest 學習環境中，軟支架為學習型和混合型取向的班級，學生問題空間表徵成績顯著高於軟支架為講授型取向的班級；當軟支架趨向學習型取向，硬支架才會幫助學生更好地進行問題空間表徵，但是在其他取向的軟支架下，不同硬支架並未有影響。

3.4. 問題解決方案

軟支架對問題方案後測的差異具有主效應。在問題解決方案設計中，之前提供的不同硬支架並未起到明顯作用，而教師提供的軟支架起到了決定性的作用。教師如果能在學生的解難過程中適時提供支持，那麼會幫助學生更好地解決問題，避免只是膚淺地拷貝和粘貼。但是如果教師通過講授來教學生如何一步一步解決問題，並不能幫助學生獲得較有創意的問題解決方案。

另外，性別對問題方案的差異具有主效應，女生比男生的成績更好 ($F=5.50$, $p<.05$)。

4. 結論

學生的問題解決過程受到內外因素的影響，主要的內部因素是學生的學習能力和性別對他們的問題解決過程沒有影響。外部因素則是硬支架和軟支架的組合。

4.1. 學力高，解難易

學力是影響學生的問題解決過程中最主要的內部因素。學力高的學生無論前後測，各項得分都顯著高於學力低的學生。這說明人的智力因素對問題解決這一高層次思維活動有至關重要的決定作用。

4.2. 女生比男生的問題解決方案更具創意

性別對問題方案的差異具有主效應，女生比男生的成績更好。

4.3. 硬支架結合軟支架——在問題理解和空間表徵階段作用更佳

但是硬支架對學生的解難有何作用呢？有賴於實施課程的教師採用何種軟支架來啟動這些硬支架，從而使得學生在解難過程中獲得切實幫助的同時，也有一定的探索和創新。

具體說來，在問題理解這個過程中，混合型的軟支架支持中，學生的成績最佳，其次是學習型和講授型。然而在問題空間表徵這一過程中，硬支架對學生從多種方面分析和表達問題空間，搜尋可能的問題解決途徑具有重要的作用。此時如果教師的軟支架越接近於建構主義學習型，學生的問題空間表徵越豐富多樣。但是當沒有提供硬支架的時候，教師提供的軟支架並沒有太大作用。這說明，學生在問題解決的前兩個階段——問題理解和問題空間表徵階段，硬支架有重要的作用。但是這些硬支架能否達到目的，也要看教師提供何種軟支架來激活這些硬支架的潛在教學意義。

4.4. 學習型取向的軟支架能激發學生創意解難

在問題方案設計過程中，硬支架將不再起到作用。這也說明，硬支架由於相對是靜態的，只能從資源和問題的初期分析過程中幫助學生更好地掌握問題的初始狀況，對問題的可能解決途徑有一些提示，但是在最終解決方案的設計階段，這些支架的作用將逐漸隱去（fading），此時教師根據實際情況提供的軟支架，如評價各方案的優劣，對創新思考的鼓勵，對困難的可能估計及提供優選方案的標準等，都會幫助學生朝著問題解決前進。

總之，問題解決過程中，不同階段需要不同的支架。在問題解決的初期，預先設計適當的資源和問題解決步驟，輔以教師適時而恰當的指導，能幫助學生更好地理解問題和表徵空間。而在問題解決后期，則有賴於教師提供的軟支架，如鼓勵和激發學生創意，容忍學生嘗試和犯錯，鼓勵合作接近於建構主義的教學支持。隻有明確各階段學生的實際需要，才能切實培養和提高學生問題解決水平。

參考文獻

- Bain, J. (1995) Academic's Conceptions of learning and teaching: Implications for University Education. In L. Conrad, and L. Phillips (edition) *Reaching More Students*. Brisbane, Griffith *Institute for Higher Education*, Griffith University.
- Barr, R. B., & Tagg, J. (1995). *Teaching to learning—A new paradigm for undergraduate education*. *Change*, 195.11/12.
- Bransford, J., Brown, A., & Cocking, R. (2000). *How People Learn: Brain, Mind, and Experience & School*. Washington, DC: National Academy Press.
- Brush, T.A. & Saye, J.W. (2002). A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment. *The Journal of Interactive Online Learning*, 1(2), 1-12.
- Dewey, J. (1938). *Experience and education*. Carbondale: Southern Illinois University Press, 1-62.

- Hannafin, M., Land, S., & Oliver, K. (1999). Open learning environments: Foundations, methods, and models. In C. Reigeluth (edition), *Instructional design theories and models*, 2:115-140. Mahway, NJ: Erlbaum.
- Kao, M., Lehman, J., & Cennamo, K. (1996, February). Scaffolding in hypermedia assisted instruction: An example of integration. Paper presented at the annual meeting of the *Association for Educational Communications and Technology*, Indianapolis, IN.
- O' Neill, M. F. & Auken, S. V. (2003). Assessing the impact of a change from a teaching paradigm to a learning paradigm in the introductory course in marketing management. *Journal of Educational Technology Systems*, 32(2&3), 199-210.
- Punch, K. F. (2005). *Introduction to Social Research*. Thousand Oaks, CA: Sage.

導入知識能量地圖於中小學教育評量之研究

Applying Knowledge Map to Design and Develop the Educational Assessment Tool

王宗一、蘇建元、趙偉智*、張駿淵

國立成功大學工程科學系

高雄市左營區屏山國民小學*

1. 研究背景與動機

電腦測驗或線上評量是數位學習領域或資訊教育範疇所不容忽略的重要議題，由於具備立即回饋、個別診斷、易於挑選或組合試題、共享試題題庫等…的優勢，相關研究早已證實其具有和傳統紙筆測驗相同的效力（Clariana & Wallace, 2002），而許多的教學單位或組織便常採取線上評量或電腦測驗來評估學習者的學習成果。雖然有許多發展出的教育評量系統，試圖幫助教師處理教學評量的施測，然而就我們的訪查結果，發覺教師大都仰賴教科書出版商所提供的題庫光碟來進行測驗編制，或藉由網路題庫，找尋題型來進行修改，然而這些未經測試或評估的題目常導致：（1）試題偏重某家出版商的教科書內容，（2）施測內容未符合實際教學目標，（3）沒有提供教師專業判斷或試題篩選機制，（4）題目未對應實際該學的學習概念及，（5）施測範圍下的試題，所對應的學習概念比重難以拿捏。而容易造成教育評量不足以量測學生程度的窘境，尤其，對於教學經驗較少或對該學科領域較生疏的教師而言，評量編制更是一種負擔（Campbell & Evans, 2000）。

為此，本研究提出導入知識能量地圖於教育評量，規劃並建構一個能輔助中小學教師出題的教育評量工具，讓中小學教師能依照設定的目標或學習概念比重來出題，以減少光碟或線上題庫所造成的問題，期能降低目前中小學教師的出題負擔，並提升教育評量的出題效力。

2. 文獻探討

2.1. 中小學的教育評量現況

教育評量主要在收集、解釋證據並讓預期使用者，如教師或學校人員能針對特定目的進行決策（Harlen, 2005），其用意在於支援學習或提供回饋給學生、家長及教師，並確認下個學習步驟；或提供資訊給學校以衡量或評鑑其規準與成效，而評量跟教學息息相關，也稱為教學評量，是教師用以輔助教學，幫助學生學習的重要歷程。不過，有研究認為教學評量的進行佔據了教師超過三分之一的專業時間（Frey, Petersen, Edwards, Pedrotti, & Peyton, 2005），教師要能編製一份具效力的評量內容，必須透過彼此意見交換、經驗分享與交流，才能達成，然而教師平日教授的課程種類繁多，學生數量又多，要編製對學生有意義的評量，實屬不易，容易造成教師自編評量不足以量測學生學習的情形出現（Holler, Gareis, Martin, Clouser, & Mille, 2008），因此，對教師而言，如何提供具體協助教師編製出優良且符合教學需要的評量工具，是值得進一步探討。

2.2. 學習概念與試題關係

概念是建構知識的基本單位（Chang, Chiu, Lin, & Heh, 2003），透過概念間的聯結可使知識的表徵更具真確性。以教育而言，概念也是學校教育內涵的全部（Novak, 1979），在各種學習領域裡，課程單元是由大小不同的學習概念所組合而成，且每個學習概念具有不同的重要程度（張國恩、林水成、陳世旺，1998）。學生若能理解概念，則相關問題便迎刃而解，亦能學習更複雜的知識或概念。在試題的編製上，有學者主張應依循以一道題目對應一具體

學習概念為出題準則，或某個題目會關連到某些學習概念的知識，並對應到某些不同的知識層次（Agarwal, Edwards, & Perez-Quinones, 2006）。而過去也有著重試題與學習概念間的相關研究，如應用試題概念的結構分析來進行學生的學習診斷（Chang, et al., 2003），以試圖瞭解試題背後的隱含概念，或概念與試題間的繼承關係，或根據學生答題情形來發掘學生對概念的精熟程度。因此，試題若能分析出其內含的學習概念關聯性，即可透過試題施測後的分析，來瞭解學習者對所學課程內容或學習概念的認知程度，也可做為教師教學或課程調整的參考。

2.3. 試題分類

由於網路與電腦的普及，教師會採用資訊工具來進行評量編制，使得電子試題也大幅成長，面對龐大且雜亂的電子試題，應設法讓試題能依照學科類別或教學目的進行分類，以方便教師取用。目前電子試題主要仍以文字形式呈現，一般可採行文件分類來就試題進行自動分類。而文件分類是由大量文件資料中所擷取出來的特徵與統計資訊作為該分類依據，將文件分類到最相似的類別。早期常採粹取關鍵字分類進行，後續則有用詞彙權重來進行，例如 TF-IDF（Term Frequency-Inverse Document Frequency）或 VSM（Vector Space Model）演算法，以文件索引向量來計算重要的關鍵字詞，作為評估索引正確與否的權重（Salton & McGill, 1986），因此，就試題內容進行篩選及分類，除了能有效表達試題與重要關鍵詞之間的關聯性外，也能用來區分或判別試題之間的相似度。

2.4. 知識地圖

知識地圖是一種知識表徵方法，能將知識主題的知識結構以視覺或圖形形式表現。知識地圖是由 Dansereau 所提出（Davenport & Prusak, 2000），認為文本、網頁資料或資料庫的資訊，應能藉由地圖形式來描繪與呈現，其主要是提供索引和導引，告知使用者知識所在及如何進行搜尋和擷取，並進一步應用資訊。Zack（2002）認為知識地圖可衡量組織存在哪些知識與知識分佈的指南，需要某項專業知識時，可以圖形樣式表現出組織內的專業人員、知識資產、知識來源或架構、和知識應用的方法，如此使用者將可輕易的將新知識與本身既有的知識關係串聯起來。在國內，有學者就利用知識地圖來分析數位學習領域的學術論文，由系統自動建構概念圖，來了解此領域的重要研究概念與議題（Chen, Wei, Chen, & Kinshuk, 2008）；Castles 等人則透過知識地圖來瞭解學生對學習範圍內的概念理解程度並分析哪些概念對學習者是較為吃力的（Castles, Lohani, & Kachroo, 2008）。

由上得知，教師從光碟或線上題庫來編制試題，仍隱藏許多問題，為此，本研究導入知識能量地圖的想法，讓系統進行自動試題分類與學習概念的粹取，使教師瞭解試題概念的比例分配，同時，教師能藉由設定之施測範圍內不同試題來源的取得，形成學習概念節點的知識地圖，讓教師依照地圖所呈現的概念比重及其關連強度來建構並挑選出合適的試題內容。

3. 系統架構

圖一為所發展之評量工具之系統架構圖，由以下幾個部份構成。

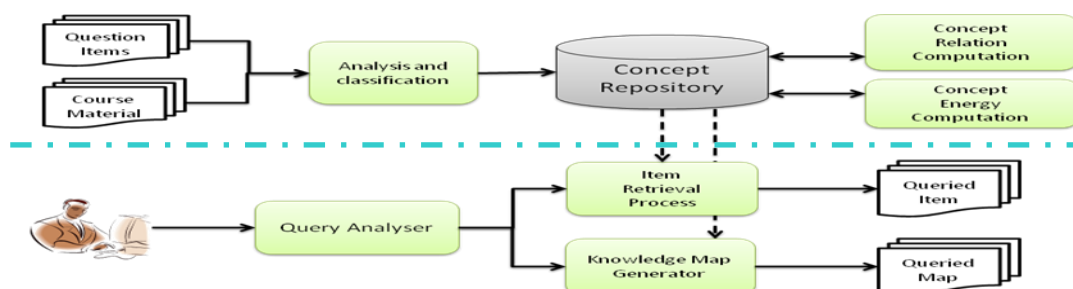


圖 1 所研發之系統架構圖

1. 試題匯入與分析處理：針對教師所選定的電子試卷或評量範圍進行自動分類解析，也就是從所選定的試題數或試卷數中分析並粹取出每個試題所含之學習概念，並將試題進行分類；另外，也提供教師能自行匯入所設計的試題，以增加題目的多樣性。

2. 學習概念產出處理：由於試題與每個學習單元中的學習概念關連密切，我們挖掘隱含在學習單元中的學習概念，並與試卷中的試題來進行學習概念比對，除了依據中研院研發之中文斷詞系統(CKIP)來作為重要特定學習概念詞彙的取得外，也提供教師自行增加或修改重要學習概念詞彙。

3. 計算學習概念關連強度與概念能量：在概念產出處理後，依據教師選定的試題範圍，系統自動建立試題範圍內的學習概念關連強度與概念能量，概念能量代表概念權重，透過此機制運算，教師能清楚看出概念的分佈及試題所對應之每個學習概念的權重及其依存關係。

4. 知識地圖產出處理：負責從教師選定的評量範圍中建構且呈現學習概念的知識地圖，此地圖將幫助教師評估施測範圍內，試題涵蓋的學習概念權重及關係，以方便教師決定或調整試題的類型或數量，更可當作評量學生學習概念熟捻與否的依據。

4. 系統實作

本研究初步建置國小五、六年級「自然與生活科技」領域之課程與試題，首先，教師能經由使用介面先行匯入欲分析或使用之電子試題，或使用系統工具底下所儲存的試題，在介面中選定年級並挑選測驗範圍後（圖2），便能藉由所選定的電子試題數量或試題卷來分析出該範圍內的學習概念能量與其關連強度（圖3），並轉化與呈現具學習概念能量與概念關係強度的知識地圖（圖4），教師便能透過地圖，進一步擇定學習概念下的連結試題數量與型態，或修改與調整所要評估或檢測的學習概念比重。

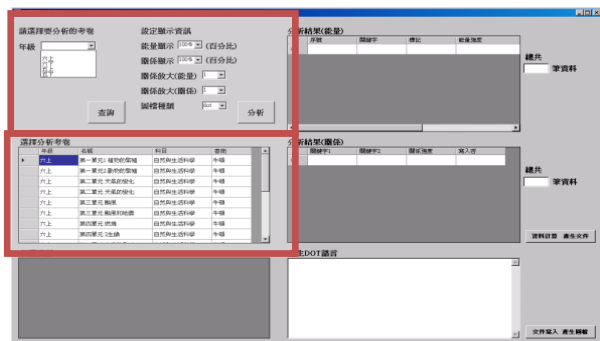


圖 2 選定施測年級與測驗範圍



圖 3 分析出的學習概念能量與概念關係

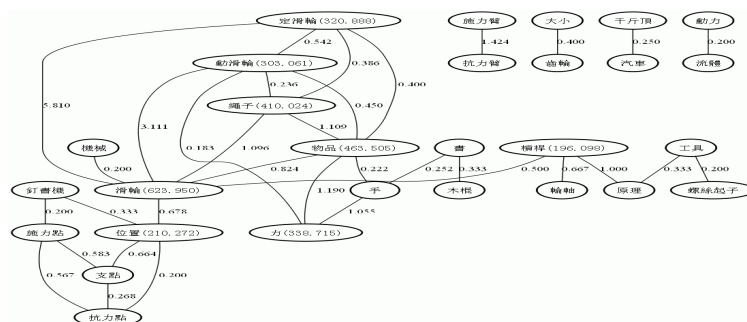


圖 4 轉化出具學習概念能量與概念關係強度的知識地圖

5. 結論與未來方向

本研究結合知識能量地圖並找出施測範圍內的學習概念能量及學習概念關連強度，讓開

發出的評量工具能依照教師設定的課程或施測範圍來動態產生出學習概念節點，並將該範圍內的學習概念與試題進行關連性連結，而以知識能量地圖來表達學習概念間的關係權重。期望改善教師處理評量或試題的方法，使教師能藉著概念間的關係權重、學習概念能量等資訊來取捨試題，進一步幫助教師減輕選題、完成出題的工作。

目前工具尚處於開發階段，待建構完成，我們將收集更多試題進行分析，同時也將調查教師對該工具使用前後的態度及滿意度，作為後續的調整與參考，未來也預計加入更多幫助教師出題或提供教師進行自動選題的機制與功能，使得該工具能更具實用性與效能。

參考文獻

- 張國恩、林水成、陳世旺（1998）。屬性化概念圖的模糊整合，載於 1998 國際電腦輔助教學研討會論文集。
- Agarwal, R., Edwards, S., & Perez-Quinones, M.(2006). Designing an adaptive learning module to teach software testing.
- Campbell, C., & Evans, J.(2000). Investigation of Preservice Teachers' Classroom Assessment Practices During Student Teaching. *Journal of Educational Research*, 93(6), 350-355.
- Castles, R., Lohani, V. K., & Kachroo, P.(2008). Knowledge Maps and Their Application to Student and Faculty Assessment. Paper presented at the 38th ASEE/IEEE Frontiers in Education Conference, Saratoga Springs, NY.
- Chang, J.-C., Chiu, Y.-P., Lin, Y.-Y., & Heh, J.-S.(2003). Learning Diagnosis Process with SCORM Compatible Learning Materials. Paper presented at the WISCS 2003.
- Chen, N.-S., Wei, C.-W., Chen, H.-J., & Kinshuk.(2008). Mining e-Learning domain concept map from academic articles. *Computers & Education*, 50(3), 1009-1021.
- Clariana, R., & Wallace, P.(2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(5), 593-602.
- Davenport, T., & Prusak, L.(2000). Working knowledge: how organizations manage what they know. *Ubiquity*, 1(24).
- Frey, B., Petersen, S., Edwards, L., Pedrotti, J., & Peyton, V.(2005). Item-writing rules: Collective wisdom. *Teaching and Teacher Education*, 21(4), 357-364.
- Harlen, W.(2005). Teachers' summative practices and assessment for learning-tensions and synergies. *Curriculum Journal*, 16(2), 207-223.
- Holler, E. W., Gareis, C. R., Martin, J., Clouser, A., & Mille, S.(2008). Teacher-made assessments: getting them right. *Principal Leadership*, 9(1), 60-64.
- Johnson, C., & Fuller, U.(2006). Is Bloom's taxonomy appropriate for computer science? Paper presented at the Proceedings of the 6th Baltic Sea Conference on Computing Education Research, Koli Calling.
- Novak, J.(1979). The reception learning paradigm. *Journal of Research in Science Teaching*, 16(6), 481-488.
- Salton, G., & McGill, M.(1986). *Introduction to Modern Information Retrieval*: McGraw-Hill, Inc. New York, NY, USA.
- Zack, M.(2002). Developing a Knowledge Strategy. In C. W. Choo & N. Bontis (Eds.), *The Strategic Management of Intellectual Capital and Organizational Knowledge*: Oxford University Press US.

國中社會領域模擬考之線上適性測驗研發

Develop the On-Line Adaptive Testing to Social Studies Mock Exam for Junior High School

陳詩靖*、郭伯臣**、吳任婕**、林文質*

亞洲大學資訊工程學系*

臺中教育大學教育測驗統計研究所**

1. 研究動機與目的

國中基本學力測驗是國中畢業生升學的主要依據，故各國中會針對國三生舉行模擬考，模擬考的題型、考試時間、分數計算方式都以仿照國中基測的方式進行，因此模擬考可視為預測學生國中基測表現的一項重要工具。對於國中社會領域來說，三子科每週授課時數僅一至二節，教師常無暇在教授新進度之餘幫助學生充分複習，若能有一套與國中基測相仿的線上適性測驗，將有助於在基測前有限的準備時間內作最有效率的評量。近年來，有許多相關研究（黃朝恭,1999；蕭顯勝、黃啟彥、游光昭,2007）指出，電腦化適性測驗除可「因材施教」外，更可大幅節省施測試題就能達到與傳統測驗相同的效果。故本研究的主要目的如下：

- （1）以模擬適性測驗探討國中基測社會領域模擬考施行線上適性測驗的成效。
- （2）研發國中基測社會領域模擬考的線上適性測驗及驗證系統成效。

2. 文獻探討

2.1. 國中基測社會領域的試題編製

國中基測試題的研發依據標準化測驗的流程編製。正式施測時，會依據事前公佈的測驗目標，從題庫中抽取試題，組題的原則主要有：試題取材均勻分佈；生活化、統整化素材優先；試題難度以 50-75%通過率為主(林世華,2004)。而根據「國中基測貫徹一綱」說帖（國中基測委員會,2007），編製國中基測題本的藍圖是以各領域的基本能力和內容兩個向度為主，在社會領域方面包含的基本能力主要有：具備社會學科的基礎知識、具備了解社會學科知識內涵的能力、具備批判性思考的能力（國中基測委員會,2000）；內容方面，教育部於民國 94 年公布了「社會學習領域七至九年級基本內容」以降低各版本教材內容差異及作為教科書編寫參考，其內容依地理、歷史及公民三科細分為 55 個細項，本研究在此以 A1~I6 代稱。綜上所述，本研究以社會領域的基本能力、基本內容形成雙向細目表，如表 1 所示。

2.2. 電腦化適性測驗

電腦化適性測驗(Computerized Adaptive Testing, 簡稱 CAT)係指以電腦進行適性測驗，施測時系統會依受試者的能力挑選適合的試題施測，並依作答反應進行能力的估計以挑選出下一試題，直到能力得到精確的估計為止。通常 CAT 可答 1/5 ~ 1/3 即可結束測驗，視題庫中之題目數及所設定的終止標準而定(何榮桂,1999)，故相較於使用傳統紙筆施測的古典測驗，CAT 更能達到測驗「客觀」、「精確」與「效率」的目標(Weiss, 1982)。CAT 有以試題反應理論(item response theory, 簡稱 IRT)為基礎(Wainer, 2000)或以知識或試題結構為基礎(Chang, Liu, & Chen, 1998)兩大類之分，但本研究參照國中基測的模式，故採用以 IRT 為基礎的 CAT。

3. 研究步驟

本研究的目的是在於發展社會領域模擬考的線上適性測驗，故經文獻探討歸納出基測試題的命題原則，再挑選較符合的一回模擬考題作為題庫，再發展成線上適性測驗。步驟概述如下：

(1) 題庫內容分析

本研究的題庫試題取自彰化縣 96 年國中聯合模擬考社會領域一回試題，範圍與國中基測相同。為評估試題內容的分布，以本研究的雙向細目表分析結果如表 1。由表 1 可知，地理、歷史、公民三科題數均為 22 題，符合國中基測委員會（2005）所提的各子學科平均分配之原則。基本內容各細項(A1~I6)也大致維持平均分佈，故其內容代表性不錯；基本能力方面，林秀娟（2007）指出國中基測社會科試題以「了解」層次居多，比例高達半數以上，且近七成為較低層次的認知能力，故本回試題符合此趨勢。

表 1 題庫的雙向細目表

基本內容	七年級(地)					八年級(地)					九年級(地)										
基本能力	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6	B7	C1	C2	C3	C4	C5	C6	C7	C8
基礎知識																					
了解能力		1	2		1	2	2				2		1	1		3	1				2
批判性思考	1	1							1							1					

基本內容	七年級(歷)					八年級(歷)					九年級(歷)						
基本能力	D1	D2	D3	D4	D5	E1	E2	E3	E4	E5	E6	E7	F1	F2	F3	F4	F5
基礎知識	1				1			1			1						
了解能力	1	1	2	1			1		1		1	1		1	1	5	2
批判性思考																	

基本內容	七年級(公)					八年級(公)					九年級(公)						
基本能力	G1	G2	G3	G4	G5	G6	H1	H2	H3	H4	H5	I1	I2	I3	I4	I5	I6
基礎知識											1						
了解能力		2		3	1		1	1	1	2	3	1	2	1			1
批判性思考											2						

(2) 試題的信、效度分析

題庫試題採 SPSS 分析，得到 Cronbach α 係數作為信度分析，結果如表 2。另外，本研究以模擬考量尺分數與受試樣本參加同年度的第一次基測所得量尺分數間的相關程度建立效標關聯效度，由表 2 可見達到 0.891738 的高度相關，顯示該試題對於國中基測有良好預測度。

表 2 模擬考題信、效度

施測人數	Cronbach α 係數	模擬考量尺分數與基測量尺分數 Pearson 相關
424	0.921	0.891738

(3) 試題參數分析

本研究以模擬考的實証資料使用 BILOG-MG 軟體分析試題參數，採 IRT-Rasch 模式，並列出古典測驗理論的指標(答對率、點二系列相關)做為分析參考，所得結果如表 3。

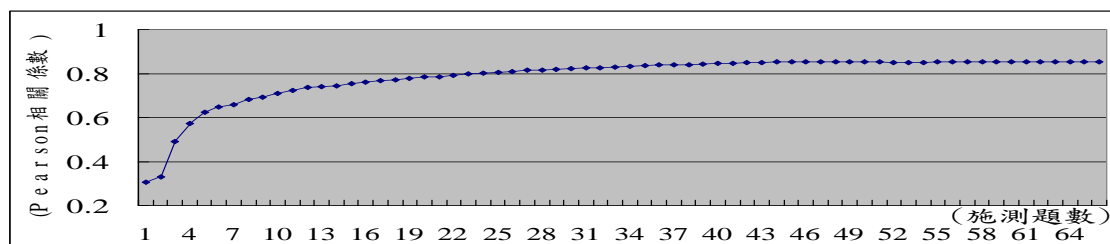
表 3 試題參數分析表

	平均值	標準差	極大值	極小值	集中分布範圍
難度 b 值	-0.68	1.09	1.98	-3.45	-1.77~0.41
通過率(%)	61.62	18.13	95.50	15.80	43%~80%
點二系列相關	0.37	0.12	0.57	0.07	0.25~0.49

由表 3 的試題難度平均值及集中分布範圍(上下一個標準差)可知試題難度符合基測設定的「中間偏易」原則；試題通過率的平均值與集中分布範圍也與基測設定的 50%~70% 通過率的標準相符；試題的點二系列相關平均值 0.37，集中分佈於 0.25~0.49 間，一般說來，測驗中大多數題項的點二系列相關在 0.3 與 0.6 之間為最佳(楊孟麗、譚康榮、黃敏雄, 1993)，故顯示本測驗有不錯的鑑別度。原本點二系列相關極小值為-0.24，但因為該題測驗內容屬於舊教材，現行教科書已不提此內容，且此題答對率僅 15.8%，故除卻此題以外，最小值為 0.077。

(4) 模擬適性測驗

本研究以 manlab 軟體模擬受試者接受適性測驗時，於不同題數時的能力估計值，再利用 SPSS 軟體取其與基測成績的相關，結果如圖像 1，可發現全體的能力估計值與全體基測成績的相關程度於第 24 題時已達 0.8 以上，與測驗完的全體相關 0.891738 很相近。



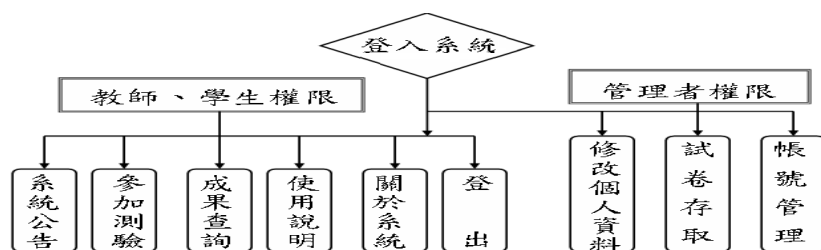
圖像 1 模擬適性測驗圖

4. 適性測驗實作成果

4.1. 系統簡介

(1) 系統功能

本研究使用的系統為台中教育大學教育測驗統計所開發的「電腦適性診斷測驗系統」，其權限與功能如圖像 2 所示。



圖像 2 系統功能圖

(2) 系統介面

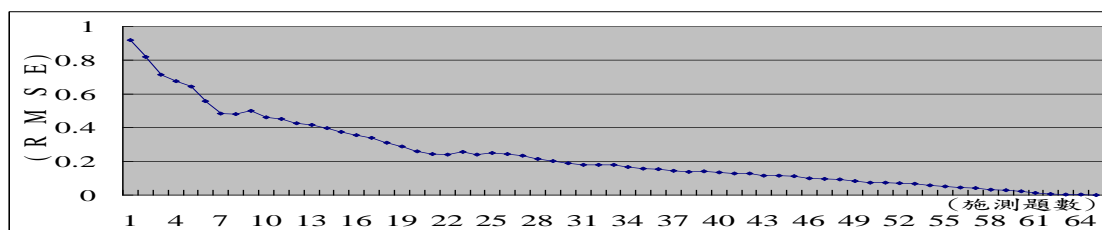
登入後選擇教材單元即可進行測驗，施測完畢後點選「成果查詢」功能，系統會列出受試者歷來施測單元及學生的估計能力值，作為學生自我了解及教師教學補救的參考，如圖像 3。

745060001【王幸祐】 歷來測驗單元			
編號	單元名稱	能力值	列印
1	研發教材 社會第18冊第30單元【基測社會科】-卷1	-0.3105	

圖像 3 施測畫面及測驗結果查詢畫面

4.2. 能力值估計均方根誤差圖

本研究藉由探究適性測驗於不同測驗長度時對受試者能力值估計的情況來驗證系統成效。具體作法是以適性測驗的流程施測完題庫中全部的試題 66 題後，再以均方根誤差(Root Mean Squared Error, 簡稱 RMSE)來分析，如圖像 4 所示。結果當受試者施測到第 30 題時，RMSE 的值達 0.2 以下；施測到第 46 題時，RMSE 的值達 0.1 以下。



圖像 4 能力值估計均方根誤差圖

5. 研究結論

(1) 本研究適性測驗的題庫使用一回模擬考試題建置，內容分布大致符合基測試題出題方向，難度、通過率也符合國中基測試題的基本假定。

(2) 本研究使用模擬考實証資料進行模擬適性測驗，發現測驗長度在 24 題以上時就可達到 0.8 以上的相關，與全體測驗完的相關值 0.891738 很接近，故由此推論，若進行適性測驗僅需作答 34% 的試題就可達到與全測時同樣的測驗效果，故約可節省 63% 的試題。

(3) 本研究透過實際的線上適性測驗，比較在不同測驗長度時對於能力估計的影響，以驗證系統成效。結果發現做到 30 題左右，誤差為 0.2 以下，做到 46 題時誤差在 0.1 以下。

參考文獻

- 何榮桂(1999)。量身訂製的測驗—適性測驗。《測驗與輔導》，157期，3289-3293。
- 林世華(2004)。國中基本學力測驗與聯考實務作法的差異。《飛揚月刊》，14期。2008年11月20日，取自：<http://www.bctest.ntnu.edu.tw/>
- 林秀娟(2007)。《九年一貫社會領域能力指標與國中基本學力測驗社會領域試題分析之研究》。國立台北教育大學社會科教育學系碩士論文，未出版，台北。
- 國民中學學生基本學力測驗推動工作委員會(2000)。「基本學力測驗」社會學科命題方向。《飛揚月刊》，5期。2009年2月10日，取自：<http://www.bctest.ntnu.edu.tw/>
- 國民中學學生基本學力測驗推動工作委員會(2007)。「國中基測貫徹一綱」說帖。《飛揚月刊》，44期。2009年2月10日，取自：<http://www.bctest.ntnu.edu.tw/>
- 黃朝恭(1999)。《國民小學國語科多媒體線上測驗系統建置之相關研究》。國立台中教育大學教育測驗統計研究所碩士論文，未出版，台中。
- 楊孟麗、譚康榮、黃敏雄(1993)。《心理計量報告：TEPS 2001 分析能力測驗，民國九十年第一波高中樣本（高中/高職/五專/國中）》。中央研究院調查研究專題中心『台灣教育長期追蹤資料庫』。
- 蕭顯勝、黃啟彥、游光昭(2007)。網路化科技素養適性測驗系統之建置。《國立台南大學理工研究學報》，40卷1期，1-21。
- Chang, K-E., Liu, S-H., & Chen, S-W. "A testing system for diagnosing misconceptions in DC electric circuits." Computers & Education, 31, 195-210, 1998
- Wainer, H. (Ed.), "Computerized adaptive testing: A primer (2nd ed.). Hillsdale", NJ: Lawrence Erlbaum Publishers., 2000.
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. Applied Psychological Measurement, 6, 473-492.

