

# 線上測驗系統的評分機制及回饋方式對測驗成績、評分效力、測驗系統滿意度之影響研究

楊亨利

國立政治大學資訊管理學系教授  
yanh@mis.nccu.edu.tw

應鳴雄

親民技術學院資訊管理系講師  
queue@seed.net.tw

## 摘要

電腦測驗是資訊教育上的重要議題，許多教育單位及企業組織，利用線上測驗系統來進行學員的學習成效評量。但是填充題及問答題等類型在線上測驗系統的評分實施上仍有許多問題需克服，其評分結果也與傳統紙筆測驗教師評分結果不同，並產生嚴重的測驗評分等化問題。本研究在探討評分機制種類及回饋種類對於測驗成績、評分效力、測驗系統滿意度之影響。為了使線上測驗能具備與傳統測驗相同的評量效力，本研究提出一個智慧型評分機制概念，以期望能解決測驗系統評分效力問題。研究結果顯示評分機制對於測驗成績、評分效力及系統滿意度會產生影響，測驗回饋種類對於系統滿意度也會產生影響。而線上測驗系統若要具備教師紙筆評分的效力，則可在評分機制中提供具有智慧評分能力的機制，另外測驗系統的滿意度也可透過評分機制、回饋種類來提升。

**關鍵字：**線上測驗、智慧評分機制、電腦自我效能、系統滿意、測驗等化

# The Influence of Scoring Mechanism and Feedback of Online Testing System on Testing Score, Scoring Effect, and System Satisfaction

Heng-Li Yang

Professor

Dept. of Management Information System

National Cheng-Chi University

yanh@mis.nccu.edu.tw

Ming-Hsiung Ying

Lecturer

Dept. of Management Information System

Chinmin Institute of Technology

queue@seed.net.tw

## Abstract

The computer-based or online testing has been one of important topics of information education. However, traditional online tests do not provide fill-in-blank items because of grading problems. This research hopes to demonstrate that if properly designed, an online test could achieve the same test effectiveness as manual paper-and-pencil tests.

Furthermore, in order to understand the influence of scoring mechanism and feedback of online testing system on testing score, scoring effect, and system satisfaction, this research built an online test prototype system to provide both traditional and intelligent online test scoring mechanisms, and both KR feedback and comprehensive feedback. 120 students participated in an experiment to compare the differences of testing score, scoring effect, and system satisfaction.

The result indicated that different scoring mechanisms would have a significant effect on test score, scoring effect, and system satisfaction. The different feedback types would have a significant effect on system satisfaction. The prior knowledge and computer self-efficacy would have a significant effect on test scores, but computer self-efficacy would not have a significant effect on system satisfaction. The intelligent scoring mechanism could be equivalent to manual tests.

**Keywords:** Online Test, Intelligent Scoring Mechanism, Test Score Equation, Computer Self-Efficacy, System Satisfaction

## 壹、前言

測驗是利用試題來評量學生成就的一種工具，也是評量學習成果的有效方法。傳統紙筆測驗，無論出題、考試、評分等作業皆需人工處理，考試時間與地點也受到限制，因此測驗效率低落，人工作業錯誤的機會也很高(何榮桂, 1990; 周文正, 1998)。電腦輔助測驗(computer-based testing, CBT)因為具備提高測驗效率、減少測驗時間、降低測量誤差、提供多元豐富的試題呈現方式、額外提供測驗相關資訊等優點(何榮桂, 1990)，因此國內托福考試已採用電腦輔助測驗方式進行，考選部也正積極規劃國家考試採行電腦化測驗作業(考試院, 2003)。

1986年美國心理學會發表電腦基礎測驗在發展、使用及解釋上之指導原則，以確保電腦測驗能具有與傳統紙筆測驗相同的效力(Bugbee, 1996)。因此利用電腦進行測驗時，應注意電腦測驗之效果應該與傳統紙筆測驗相同(何榮桂, 2000; 簡茂發, 1999)。

目前電腦測驗的設計，仍受到傳統紙筆測驗觀念影響，因此考試題目的類型普遍仍以選擇題及文字測驗為主。雖然目前亦有少數系統提供填充題型的測驗，但卻只能做到答案樣式(pattern)比對，而未對答案的詞彙語意進行分析。在填充題測驗中的答案常有許多相同語意的詞彙可互換，因此容易造成電腦測驗評分結果與老師親自評分結果不同，並導致嚴重的測驗評分效力問題。

過去的研究顯示，學習者的先備知識及電腦自我效能程度會影響學習成就的結果(Tam, 1996; Karsten and Roth, 1998; Wang and Newlin, 2002; Cronbach and Snow, 1977; Dochy et al., 1999; 翁榮源, 2003)，因此在衡量線上評分機制對於測驗成績的影響時，有必要納入先備知識及電腦自我效能因素。此外，成功的線上測驗系統，應該與傳統測驗的評分結果等化，並爭取學習者的信賴，甚至學習者能從評分機制及回饋中獲得更多效益。從資訊系統的角度而言，線上測驗系統應提供符合使用者需求與期望之資訊品質服務，以提高使用者的系統滿意度(Pitt et al., 1995; Delone and McLean, 1992; 呂怡緯, 1998; 羅凱揚, 1998)。

綜合上述，本研究將透過實驗方法來探討線上評分機制及測驗回饋對於測驗成績、評分效力及系統滿意度的影響，並同時考量學習者的先備知識及電腦自我效能等變數的影響。

## 貳、文獻探討

### 一、測驗方式的演化

教育目的在於引發學生行為改變，然而教師則經常利用測驗來評量學生學習後的行為改變，因此測驗在各學科中均扮演著重要的角色(林璟豐, 2001)。但是何榮桂等人(1999)認為紙筆測驗在編撰試題、施測、閱卷、評分等過程中，相當耗費人力、物力，而且遠距學習者也未必能參與這種「會考式」的紙筆測驗，因此若透過電腦及網路測驗將能提供適當的支援與輔助。

電腦輔助測驗是將傳統的考試工具及考題轉移到電腦中，讓學生藉由電腦螢幕閱讀考題，利用鍵盤或滑鼠來輸入及點選答案。Alessi 和 Trollip(1991)認為電腦輔助測驗在應用上，具有選擇組合試題、易於產生試題及共享題庫等效用。何榮桂(1990)也認為電腦輔助測驗具提高測驗效率、試題呈現方式多元豐富、減少測量誤差、減少測驗時間及獲得額外測驗資訊等優點。目前許多電腦輔助測驗已利用試題反應理論(item response theory, IRT)來發展電腦化適性測驗系統(computer-adaptive testing, CAT) (Van der Linden and Glas, 2000; Wainer and Dorans, 2000)。而美國的 GRE 測驗及台灣地區托福測驗(TOEFL)也由傳統的紙筆測驗改為電腦化測驗。

近年來，電子化學習及網路遠距教學已成為熱門的議題，以網路測驗所能夠達成的效果而言，它不僅是施測的工具，也是一種教學工具(周文正，1998)。McCormack 及 Jones (1997)認為網路測驗能夠改善評量的程序和方法，因為網路測驗具備有節省時間、即時回饋、減少資源、保存記錄、更加便利等特性。周倩與簡榮宏(1997)的研究則指出，大部份的學生較信任網路測驗的結果，而且認為網路測驗具備效率上的優點，而且也不會影響考試的實力。但是網路測驗也有容易作弊、線上閱讀題目容易疲倦、面對螢幕有礙思考、打字速度影響答題表現等缺點。Mark(1997)及 Bennett et al.(1999)則認為，利用多媒體的方式來呈現更多元的測驗問題，以及利用人工智慧使電腦自動判斷如何去測驗及進行評量，是影響未來網路測驗的二個重要運用技術。

## 二、測驗題型

Gronlund(1998)依據題型的表現方式，將題型歸納成四大類：

- (一) 選擇反應(selected response)：如選擇、是非及配合題等。
- (二) 供應反應(supply response)：如填充、簡答、申論題等。
- (三) 限制性實作(restricted performance)：乃是指高度結構化的任務，例如選擇實驗室設備、測量濕度。
- (四) 延伸性實作(extended performance)：此類型評量需要較多的理解和判斷，如實驗的進行、利用電腦解決問題等。

本研究所指稱的線上測驗系統主要以認知領域為教學目標的測驗為主，題型除了認識型題目(recognition-type items)外，還包含建構型題目(constructed-response items)中的填充題。認識型題目主要包括是非、選擇、及配合題等，建構型題目則包括名詞解釋、填充題、問答題等(鼎茂，2000)。

理論上，網路環境中可以實施任何型態的測驗，但由於現實環境限制及技術上的困難，目前網路測驗題型仍以選擇題、是非題等認識型題目為主。然而，填充及問答題型較少出現在線上測驗系統中，其原因是電腦難以自動辨識詞彙語意及文句結構，因此若能克服這些困難，則對於線上測驗系統的發展將能提供極大的貢獻。

## 三、測驗評分效力

美國心理學會所發表的電腦測驗在發展、使用及解釋上之指導原則，其內容包括電腦所造成的影響效果必須清除、電腦測驗發展者需證明測驗效度、電腦測驗與紙筆測驗的成績等級順序必須相似、平均數散佈情形也需相似(Bugbee，1996)。因此近年來有一些學者便針對電腦適性測驗的試題等化(equation)問題進行研究(Tsai, et al.,2001；Han, et al.,1997)。

林明達(1998)認為使用電腦批閱申論題及填充題時，其答案常有不同的寫法或其他可替代的語意詞彙，因此無法明確的進行評分，也導致電腦評分結果與老師親自評分結果不同。本研究中定義的評分效力，是指線上測驗系統評分結果與紙筆測驗教師親自評分結果近似相等的程度，若是線上評分機制的評分結果能夠與紙筆測驗教師親自評分結果近似相等，那麼則表示此線上測驗系統的評分機制具備極佳的評分效力。

## 四、先備知識及電腦自我效能對於學習成效的影響

先備知識指學生帶進學習過程的知識、技能或能力，它包括長期儲存的知識、知識基礎範疇下的先備知識、經驗知識、背景知識及個人知識等(Jonassen and Gabrowski, 1993)。過去研究指出，學習者對某知識領域所具備的先備知識越高，他們的理解、記憶及認知學習成效就愈好(McCormick and Pressley,1995; Cronbach and Snow, 1977)。此外，Dochy 等人(1999)分析 183 篇先備知識的研究，指出先備知識與學習成果有很強的

相關性，其中 91.5% 的研究顯示先備知識對學習成果有正面的影響，而且先備知識可以用來解釋學習成效 30% 至 60% 的變異。

電腦自我效能是個人對於自己學習電腦與使用電腦能力的主觀判斷，大多數的研究都認為電腦自我效能與學習成績間有正向關係，電腦自我效能越高的人，無論在學習成效、學習成就或工作績效上都會表現的比較好(Tam, 1996; Karsten and Roth, 1998; Igbaria and Iivari, 1995; Compeau and Higgins, 1995; Wang and Newlin, 2002; 黃郁雯, 2003)。

## 五、測驗回饋

回饋(Feedback)是一種提供給個人的資訊，其包括實際績效及被期望績效間的比較資訊。Devedzic(2003)認為回饋應用於績效領域，能增加學習者的學習效果。

洪榮昭(1992)將回饋區分為 KR 回饋(knowledge of result)與 KCR 回饋(knowledge of the correct response)兩種形式。KCR 強調知識的正確回應，並以提供學生正確答案為重點，因此無論學生答對或答錯，此回饋均會提供與此答案知識相關的正確知識概念。KR 則只強調學生對於知識的答題結果，因此 KR 回饋只會提供學生諸如「答對」或「答錯」之訊息，若應用在選擇題型時則具有暗示效果，亦即學生的答案如果是錯的，其他的答案則可能是對的。

Dempsey 等人(1993)，則根據情境及回饋的資訊內容，將回饋分成無回饋、知識結果回饋、正確知識回應回饋、詳盡式回饋(elaborated feedback)、再次嘗試回饋(try-again feedback)。Gilman(1969)的研究發現，接收 KCR 回饋的學生其學習績效顯著優於未接收 KCR 回饋的學生。另外，Dempsey 與 Wager(1988)則依照回饋時間，將回饋區分成立即回饋及延遲回饋。

## 六、資訊系統滿意度

Cyert 與 March(1963)認為資訊系統提供符合使用者需求的服務，則能提高使用者對資訊系統的滿意度。呂怡緯(1998)、羅凱揚(1998)歸納影響資訊系統滿意度的因素包括：正確性、可靠性、時間性、符合使用者需求、對系統的信心、安全性、輸出格式品質、容易使用、支援性、資訊內容品質、容易找到所需資訊等。DeLone 與 McLean(1992)認為要衡量資訊系統效能必須考慮系統品質、資訊品質、資訊使用、使用者滿意度、個人衝擊及組織衝擊等六個構面。

綜合上述，一個成功的線上評分機制，應該與傳統教師親自評分結果等化，線上評分機制的評分結果也應該讓學習者信賴。因此，成功的線上測驗系統除了要從教師觀點來探討評分機制的評分效力外，也需要從使用者滿意度的觀點來評估線上評分機制及測驗回饋是否提供使用者期望的資訊品質服務。

## 七、人工智慧與模糊理論在測驗系統的應用

為了使測驗系統能更具效能，許多研究結合人工智慧及模糊理論等技術，以解決複雜的出題、評分、解釋及回饋等問題(林明達,1998; Marshall, 1993; Sun, 1999; Hwang, 2003; Moundridou,2003)。Devedzic(2003)則更進一步指出「內容導向智慧」(content-oriented intelligence)，將是下一代網站教育系統發展的重要議題。

Zadeh(1965)提出模糊集合論，該理論將傳統集合的特徵函數從 0 和 1 的絕對值，推展為 0 到 1 之間的任意值，此特徵函數又稱為歸屬函數(membership function)。在填充題測驗中，由於不同的填答者所填寫的答案詞彙會因成長背景與習慣而有差異，有些詞彙與標準答案的語意完全相同，有些則是部分相似，因此若使用模糊集合的觀念將能使測驗系統不再只能具有「非對即錯」的二元評分能力。目前國內外對於如何建

立可自行辨識詞彙語意的智慧閱卷系統，並未見有特別的研究。Mark(1997)也認為利用人工智慧使電腦能自動判斷測驗及評量方式，是值得研究的方向。

## 八、目前線上測驗系統及智慧型測驗系統發展

目前國內外對於線上測驗系統或智慧型測驗系統之發展應用相當多，例如美國 GRE 測驗採用電腦適性測驗、美國管理會計師協會 IMA 採用電腦測驗進行會計師認證考試、台灣地區 TOEFL 測驗採用電腦化測驗等。

國內外對於電腦測驗的研究，主要以 Lord(1980)介紹的項目反映理論及 Weiss(1980)介紹的電腦適性測驗為主。國內較為著名的研究包括何榮桂等人(1996)提出採用遠距適性測驗的系統架構與一連串之後續研究，以及游寶達(1998)利用模糊理論及智慧選題法建構 ICAT 系統，以提升對受測者能力評量的精確度和穩定性。

在遠距教學環境中，測驗系統已成為教學系統的重要元件，例如 ClassNet 系統支援教師在線上編寫測驗題目、解答，測驗題型包括是非、選擇及申論，是非及選擇由系統批改，申論則以電子郵件寄給教師批閱(Van Grop and Boysen,1997)。Mallard 系統提供選擇、計算、布林運算等題型，系統會適時進行提示、回饋及計分(Swafford and Brown, 1996)。黃國禎等人(2002)針對測驗題目進行智慧型的線上測驗題型分析與改進研究，並採用題目語意分析來解決相似題目重複出現的問題。

表 1 國內知名線上測驗系統功能比較表

網站名稱	功能	測驗內容領域	測驗類型	其他說明
e-School 電子校園 <a href="http://eschool.com.tw">http://eschool.com.tw</a>	線上測驗、線上考卷製作、線上評分、答案詳解、成績排名、時間計時	兒童部、國中部、高中部、大學、社會大學等各級教育多元內容。	單選題、複選題、填充題	1. 填充題無語意辨識能力 2. 填充題僅可設計一個空格的題目
長春藤線上測驗中心 <a href="http://www.ivytest.com.tw">http://www.ivytest.com.tw</a>	線上測驗、線上考卷製作、線上評分	大學、四技二專、高中、五專、托福、高普考、證券業務、駕照考試等試題測驗	單選題、複選題	
K-bar 考試網 <a href="http://www.k-bar.com.tw">http://www.k-bar.com.tw</a>	線上測驗、線上考卷製作、線上評分、時間計時	升學模擬測驗(大學及高中)、語文能力測驗(中文、日文、托福、GEPT 全民英檢)、高普考、證照等	單選題、複選題、填充題	1. 填充題無語意辨識能力 2. 填充題僅可設計一個空格的題目
狀元考試資訊網站 <a href="http://elec.taivs.tp.edu.tw/kitty">http://elec.taivs.tp.edu.tw/kitty</a>	線上測驗、線上考卷製作、線上評分、記錄及評估題目答題難易度、隨機出題	四技二專、二技、技能檢定、預官、高普考測驗	選擇題	
線上測驗服務中心 <a href="http://www.onlinetest.org">http://www.onlinetest.org</a>	線上測驗、線上評分、成績查詢、時間計時、題目支援圖形內容	學科能力測驗(國文、英文、數學、社會、自然、歷史)	單選題、複選題	
國中生活科技線上測驗系統 <a href="http://home.kghs.kh.edu.tw/olt">http://home.kghs.kh.edu.tw/olt</a>	線上測驗、線上考卷製作、線上評分、成績查詢、時間計時、題目支援圖形內容、試題轉換	國中生活科技領域	單選題、是非題、填充題	1. 填充題僅可設計一個空格的題目 2. 填充題提供相似詞答案處理

## 參、系統評分效力先期研究

為了瞭解國內線上測驗系統提供之功能，本研究針對國內知名線上測驗系統，從功能、測驗內容領域、測驗題型三方面進行比較，結果如表 1。目前國內線上測驗系統以提供選擇題型之測驗為主，而少數可支援填充題型之線上測驗系統，其每個測驗題目幾乎僅能支援一個填充格，而且在填充題的答案分析辨識上，一般測驗系統都採用字元比對方法；換言之，如果標準答案是「雙十節」，受測者輸入「國慶日」、「雙十國慶」則都算答錯，這與紙筆測驗教師評分結果可能不同。然而，表 1 的「國中生活科技線上測驗系統」，其填充題型雖允許老師預先於答案欄中將老師可接受的答案一併輸入在同一個答案欄中，並用半形逗號分開，例如老師可在標準答案中輸入「國慶日,雙十國慶,雙十節」，則當學生輸入「國慶日」、「雙十國慶」、「雙十節」時，

該系統則判定學生答題正確，但這些詞彙卻無法提供給其他相似題目或其他試題重複使用，而且也無法解決相似語意詞彙擴充、相似語意詞彙智慧辨識等問題。

為了瞭解線上測驗與紙筆測驗間的評量等化差異來源，本研究參考 K-bar 考試網及陽明國中官老師課程教學網之測驗系統功能，發展一個一般型線上評分離形系統，此系統包括是非、單選、複選及填充題等題型，而填充題的題目最多可包含 20 個填充格。此外，此離形系統如同一般測驗系統，亦可於線上編輯題庫及線上評分，但填充題之評分方式仍以目前多數測驗系統之字元比對二元評分方式進行。在一般型評分系統發展完成後，本研究針對中部某技術學院資訊管理系修習電子商務課程之 87 位學生，利用平時學習評量時間先進行紙筆測驗，測驗題型包括是非題、單選題、複選題及填充題各 5 題，當所有學生皆完成測驗後，教師請每位同學連至線上離形測驗系統，開啟題目完全相同之電子測驗卷，並將紙筆測驗卷上所寫的答案，鍵入到線上電子測驗卷中。經過事後教師對紙筆評分及測驗系統評分進行比較後發現，學生回答的相同答案卻會在不同的評分方式上出現顯著不同的評分成績，而且線上測驗成績顯著低於紙筆測驗。分析後發現，是非題、單選題、複選題之得分在紙筆測驗及線上測驗之結果並無顯著差異，填充題之得分上在紙筆測驗及線上測驗之結果卻有顯著的分數差異，顯示目前線上測驗系統若支援填充題型測驗，電腦將無法具備與紙筆測驗相同的評分效力。本研究進一步將每一份試卷進行分析比對後發現，填充題型在線上測驗中無法擁有紙筆測驗同樣的評分效力，其主要原因可歸納出以下五點：

- 一、電腦環境：在中文電腦環境中，文字有全半形及大小寫之分，但測驗系統未能辨識。例如答案是半形大寫的「B2B」，但學生可能輸入小寫「b2b」、全形的「B 2 B」等答案之情形。
- 二、可替換之相似詞彙：填充題之答案可能其他可替換之相似語意字彙，但測驗系統未能辨識。例如標準答案為「B2B」，學生可能寫出「B to B」、「Business to Business」、「企業對企業」等相似詞彙。
- 三、填充格的集合順序：某些包含多格填充題目之答案順序是可互換，但測驗系統未能辨識。例如某題目為「電子商務中常見三種的經營模式分別是\_\_\_\_\_、\_\_\_\_\_、\_\_\_\_\_。」此題標準答案若為「B2B、B2C、C2C」，但受測者填寫成「B2C、C2C、B2B」，但測驗系統未能辨識。
- 四、學生中文程度低落：在紙筆測驗中學生若因為某個字不會寫，會有寫錯字、利用同音異字或注音取代的情形，在傳統教師評分時，教師會根據錯別字之情形斟酌給分，但在線上測驗環境中並未提供教師「斟酌」給分的辨識機制。
- 五、隨機性漏字疏忽：某些學生在測驗中因為不經意的疏忽而漏寫部分答案，在紙筆測驗中，教師會考量學生的漏字情形而斟酌給予部分分數，但線上測驗系統則因答案未完全相同，因而評為「答錯」。

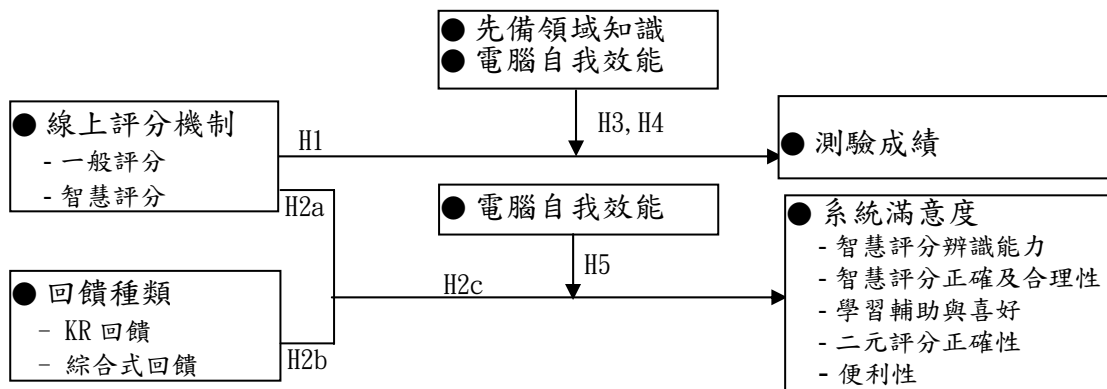


圖 1 研究架構

## 肆、研究方法

為了在課程進行的自然過程中進行變數操控，並能同時進行因果關係探討，所以本研究採用實地實驗法(field experiment)進行研究。

### 一、研究架構

本研究的架構如圖 1。本研究主要探討電子化學習環境中線上測驗評分機制與回饋種類，對於測驗成績、評分效力及系統滿意度的影響。本研究為了避免測驗時間對於研究結果的影響，並對測驗時間進行控制，因此本研究在測驗進行的實驗過程中採取寬裕的測驗時間，以使得所有的受測者皆有足夠時間完成測驗。

### 二、變數設計

本研究主要探討的變數可分成以下三個部分：

(一) 自變數：包括線上評分機制及回饋種類二個變數。

1. 線上評分機制：區分為智慧型評分機制與一般型評分機制。過去文獻指出電腦測驗支援填充題型時，容易產生評分效力問題。本研究初期針對造成評分誤差影響因素的探討中發現，在華文線上測驗環境中，全半形、輸入法、漏字、詞彙語意、答案集合等因素都會造成測驗評分的誤差。本研究中所稱的智慧型評分機制在進行填充題型評分時，會透過智慧評分模組來消彌上述因素所造成的評分誤差。而一般評分機制則採用「非對即錯」的二元評分判斷方式進行所有題型的評分。
2. 回饋種類：本研究歸納洪榮昭(1992)、Dempsey et al.(1993)的回饋類型，在線上測驗系統設計中提供 KR 回饋及綜合式回饋二種類型。KR 回饋是指受測者完成線上測驗後，系統僅給予成績結果，並針對受測者每個題目的回答給予「答對」或「答錯」的簡單回饋訊息。而綜合式回饋不僅在受測者完成線上測驗後，系統會針對受測者填寫的答案給予答對或答錯的回饋，還會給予跟該題目相關的知識輔助、評分解釋及學習建議，綜合式回饋若搭配在具備智慧評分機制的系統中，更能針對漏字、同音異字及詞彙語意相似的評分結果及斟酌扣分部分進行回饋說明。

(二) 依變數：包括線上測驗成績及系統滿意度二個變數。

1. 線上測驗成績：Bostorm(1990)及其他研究，經常會使用測驗成績來當作評量學習成效的指標。本研究所稱的線上測驗成績是指學生經由測驗系統的評分機制進行評分後所獲得之成績。Bugbee(1996)認為不同的評量工具若能具有相同的評分結果，則具有評分效力。因此在本研究之分析中也會以系統評分後的測驗成績與教師親自評分的紙筆測驗成績間之分數差距絕對值，來表示線上評分機制的評分效力，並進一步分析線上評分機制是否能正確的代理出題教師閱卷，並給予正確的測驗成績。
2. 系統滿意度：本研究中的系統滿意度是指學生在使用具有評分機制的測驗系統後，透過線上系統取得測驗後的成績及測驗回饋資訊，最後再根據個人對於測驗系統的功能需求、正確性、時間性、可靠性、系統信心、資訊品質、便利性、支援性及評分合理性等感受，而評估的系統滿意度。本研究參考 DeLone 與 McLean(1992)、呂怡緯(1998)及羅凱揚(1998)對於資訊系統績效及使用使用者滿意度的評估項目，並依照本研究之測驗系統功能特性，發展線上測驗系統之使用者滿意度量表，此問卷以 Likert 五點語意差異尺度來衡量，共有 22 項問題，問項內容請參考本文之附錄。

(三) 干擾變數：包括先備領域知識及電腦自我效能二個變數。

1. 先備領域知識：此部分參考 Jonassen 與 Gabrowski(1993)的定義，將先備領域知識定義為學生在課程學習前已具備的知識、技能與能力。本研究在學習者開始學習領域知識初期，對他們進行學前領域知識能力測驗，每位學習者在測驗中所獲得之成



績，即被視為學習者在進行課程學習前的先備領域知識程度，分數範圍從 0~100 分。本研究所使用之學前領域知識能力測驗題目內容共計 30 題，由於文章篇幅限制，因此無法刊載所有測驗題目內容。

2. 電腦自我效能：參考 Compeau 與 Higgins(1995)及 Igarria 與 Iivari(1995)的定義，本研究的電腦自我效能是指受測學生認知自己在使用電腦及學習電腦上的能力。衡量項目參考 Compeau 與 Higgins (1995)的電腦自我效能量表，此問卷以 Likert 五點語意差異尺度來衡量，共有 10 項問題，詳細問項內容請參考本文之附錄。這 10 項問題的分數總和即為學習者電腦自我效能的量化值，範圍從 10~50 分。

### 三、研究假設

本研究提出以下待驗證的基本假設：

- H1：線上評分機制種類對於學習者的測驗成績有差異。
- H2a：線上評分機制種類對於學習者的系統滿意度有差異。
- H2b：回饋種類對於學習者的系統滿意度有差異。
- H2c：線上評分機制種類對於學習者的系統滿意度因不同的回饋種類而有差異。
- H3：線上評分機制種類對於學習者的測驗成績因不同的先備領域知識程度而有差異。
- H4：線上評分機制種類對於學習者的測驗成績因不同的電腦自我效能程度而有差異。
- H5：線上評分機制種類與回饋種類對於學習者的系統滿意度因電腦自我效能程度而有差異。

### 四、實驗設計

為了檢驗線上評分機制及回饋種類對於測驗成績、評分效力及系統滿意度的影響，本研究依據線上評分機制及回饋種類，設計了 2\*2=4 個實驗組。第 1 組採用智慧型評分機制並搭配 KR 回饋，第 2 組採用一般型評分機制並搭配 KR 回饋，第 3 組採用智慧型評分機制並搭配綜合式回饋，第 4 組採用一般型評分機制並搭配綜合式回饋。

#### (一) 實驗系統

為了模擬真實的網路測驗環境，本研究架設一個三層式主從遠距測驗系統，後端資料庫伺服器採用 Microsoft 公司之 SQL Server 2000，用以執行觸發程序及存放本系統之題庫、成績、語意、申訴記錄等資料。Web 伺服器則使用 Windows 2000 作業系統中的 Internet Information Server 5.0 伺服器軟體，以執行伺服器端的 ASP(Active Server Pages)程式，並藉由 ASP 程式與資料庫進行連結，測驗系統主要使用 ASP 程式語言來發展各項模組功能。

本研究規劃之線上測驗系統架構如圖 2 所示，因研究目的與篇幅限制，無法對系統內相關之演算法及技術方法進行詳細描述，以下僅能針對各元件進行簡要說明：

1. 題庫編修介面：此介面提供試卷及題庫發展者能於線上維護及查詢試卷及題庫內容。
2. 測驗題庫及答案資料庫：儲存測驗系統的考題編號、語幹、選項、答案、題目配分、題目類型、學生資料、考試時間資料、班級資料等。
3. 出題模組：產生線上測驗試題內容及試題順序，每位受測學生所回答之試題內容都是相同的，但試題題號順序及答案選項順序則由電腦亂數產生。
4. 知識擷取介面：提供增修知識庫內容之介面，以持續擴充系統智慧，進而提升線上評分能力，以使系統具有老師的評分判斷特徵與能力。
5. 使用者介面：提供線上測驗、題目解釋、成績查詢及答案申訴之介面。
6. 工作記憶區：儲存評分及模糊判斷的運算過程內容。
7. 語意/同音異字資料庫：存放智慧型評分機制所需要參考比較的相關資料，諸如語意詞彙資料、詞彙間關係值、同音詞庫、全半形對應資料等。

- 8.測驗結果資料庫：存放學生的電子試卷結果，包括參與考試的時間、機器位置、考試歷程、填寫之答案、測驗成績、各題回饋資訊等。每位學生均會產生來自一般型評分機制及智慧型評分機制評分後的二個成績結果。
- 9.申訴處理模組與申訴處理記錄：處理並記錄學生對系統評分及答案質疑的申訴工作。通常教師無法將每個答案的相似語意詞彙全部都事先建立至系統中，因此有可能發生受測者認為自己填答的答案語意是正確的，但系統卻無法辨識出的狀況，此模組便能協助受測者進行申訴。而教師則可利用知識擷取介面來處理學生的申訴問題，若某些相似語意詞彙確實無法被系統正確辨識，教師則可立即進行語意詞彙的擴充，以促使系統之評分效力可不斷提升。

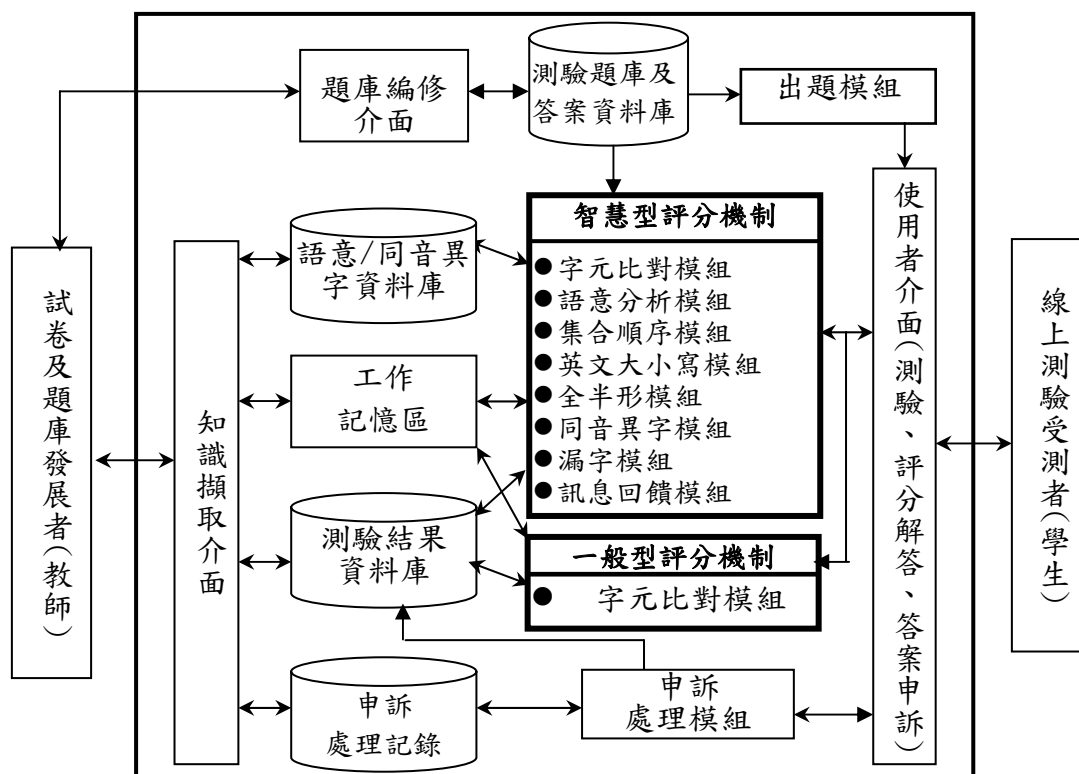


圖2 本研究發展之線上測驗系統

10.智慧型線上評分機制：為了解決填充題型評分時所造成的評量效力問題，有必要發展具備智慧的線上測驗系統來模擬實際教師之評分決策模式。因此本研究發展一個具備題庫編輯、測驗、評分、成績查詢、出題的線上系統，其題型則包含是非題、單擇題、複選題及支援一個以上空格的填充題測驗題型。智慧評分機制為本系統之核心，此機制包含各種分析及計算填充題測驗分數之相關模組。每一個填充格的分數 $S(t_i)$ 是由填答者答案之正確率歸屬函數 $P(t_i)$ 與該填充格之配分 $SI(t_i)$ 決定，即 $S(t_i) = P(t_i) \times SI(t_i)$ 。以下針對各模組功能依照運作順序簡單描述(下文中之變數 $t_i$ 是指測驗題目編號， $AS(t_i)$ 表示受測者所填答之答案， $AT(t_i)$ 表示標準答案)：

- (1)字元比對模組：此模組將 $AS(t_i)$ 與 $AT(t_i)$ 進行字元比對，當所有字元均完全吻合，則受測者所填答案的正確率歸屬函數 $P(t_i) = 1$ 。
- (2)語意分析模組：用以比較 $AS(t_i)$ 與 $AT(t_i)$ 詞彙間是否存在語意關係，此關係若存在，則語意相似係數 $RS(t_i)$ 便決定了正確率歸屬函數 $P(t_i)$ 。
- (3)集合順序模組：用以比較 $AS(t_i)$ 與 $AT(t_i)$ 詞彙間是否存在集合順序關係。例如某填充題的題幹為「電子商務最主要的三種經營模式是

$$P(t) = \begin{cases} 1 & , RS(t) = 1 \\ RS(t) & , 0 < RS(t) < 1 \end{cases}$$

- \_\_\_\_、\_\_\_\_、\_\_\_\_」，假設使用者輸入次序為B2C、C2C、B2B，而測驗的標準答案次序為B2B、B2C、C2C，則本模組會進行集合順序分析，若填答者的答案存在於標準答案集合中，則該填充格之正確率歸屬函數 $P(t_i)=1$ 。
- (4)英文大小寫模組：用以比較AS( $t_i$ )與AT( $t_i$ )詞彙間是否為英文大小寫的關係，此關係若成立，則正確率歸屬函數 $P(t_i)=1$ 。
- (5)全半形模組：用以比較AS( $t_i$ )與AT( $t_i$ )詞彙間是否為文數字全半形關係，此關係若成立，則正確率歸屬函數 $P(t_i)=1$ 。
- (6)同音異字模組：當AS( $t_i$ )  $\neq$  AT( $t_i$ )，且不存在語意關係、大小寫及全半形關係，兩個詞彙間也只有 1 個字不同時，此模組便會啟動。此模組在檢查這個唯一不同的字，是否為受測者使用注音輸入法輸入時因為選字錯誤而造成的同音異字情形，若AS( $t_i$ )與AT( $t_i$ )確實存在此種關係，則正確率歸屬函數 $P(t_i)=1-PW$ 。這裡的PW是教師在系統中設定的同音異字錯誤扣分比例參數，若PW=0.3，則表示同音異字需扣該題 30%的分數，因此正確率歸屬函數 $P(t_i)=0.7$ 。
- (7)漏字模組：此模組用於AS( $t_i$ )  $\neq$  AT( $t_i$ )，且不存在語意關係、大小寫及全半形關係，同時因為受測者的答案少輸入一個字而造成答案遺漏的不完整現象時。若AS( $t_i$ )與AT( $t_i$ )確實存在此關係，則需檢查教師自訂的評分風格漏字扣分比率參數PL。當漏字比例低於 25%時，正確率歸屬函數 $P(t_i)=1-PL$ 。若系統之PL參數設為 0.6，表示漏字比例低於 25%時需扣該題 60%的分數，因此正確率歸屬函數 $P(t_i) = 0.4$ 。至於漏字比例的門檻參數是否為 25%，教師可自行修改。
- (8)訊息回饋模組：提供受測者正確答案回饋，本系統同時採用 KCR 及 KR 之回饋方式。本系統依照受測者組別及其所填寫之答案內容，給予不同的回饋。當受測者的實驗組別支援綜合回饋時，此模組提供測驗成績及模糊評分結果回饋(包含語意關係、大小寫關係、全半形關係、同音異字關係、漏字關係等評分回饋)。
- 11.一般性線上評分機制：僅包含一般線上測驗系統處理填充題之字元比對模組評分方法，此模組將使用者答案與系統內標準答案進行比對，當所有字元均完全吻合，則此題的正確率歸屬函數 $P(t_i)=1$ ，字元若未完全相符，則正確率歸屬函數 $P(t_i)=0$ 。

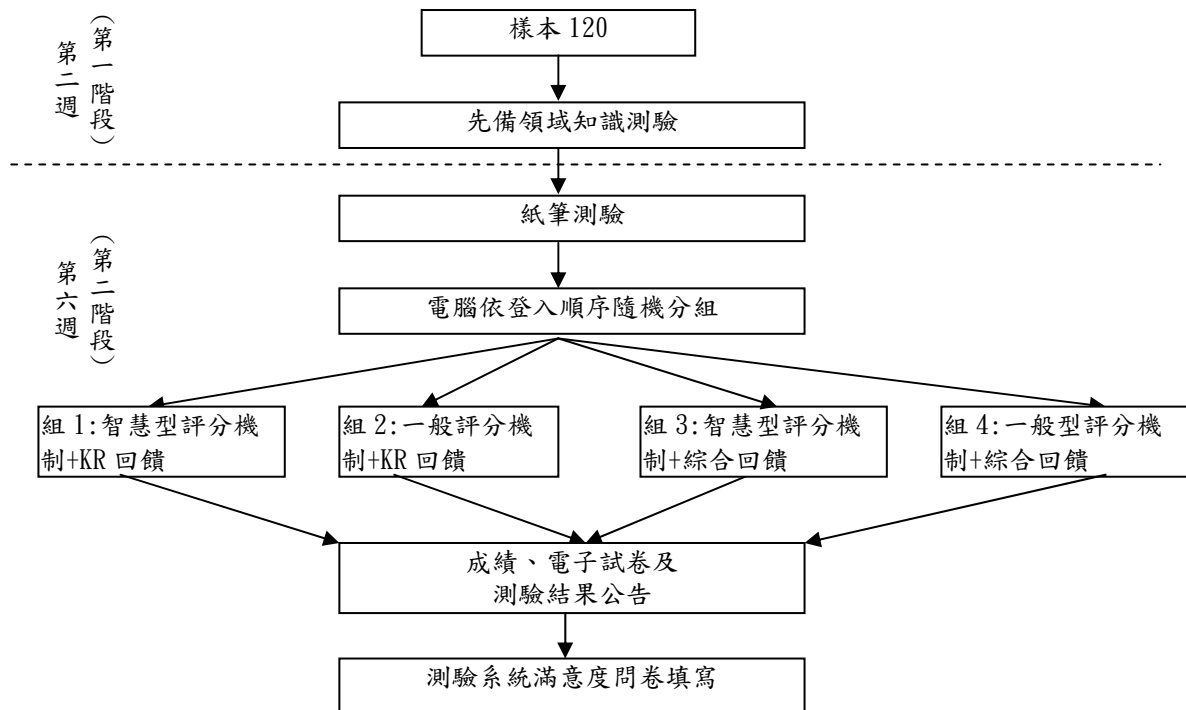


圖 3 本研究之實驗流程圖

## (二) 實驗樣本與實驗流程

本實驗對象以中部某技術學院資管系學生 3 班 120 位修管理資訊系統課程之同學為樣本，採實地實驗法進行線上測驗，並配合問卷進行資料蒐集。在 120 位受測者中男生有 63 位，女生有 57 位。但經過第一階段及第二階段實驗後，由於有 8 位受測者未全程參與實驗或錯誤操作系統導致其實驗資料未能完整取得，後續資料分析將予以扣除，因此全程參與之受測者共有 112 位。而本研究之實驗流程如圖 3 所示。

## 伍、資料分析結果

### 一、樣本基本資料分析

在實驗流程第二階段所實施的測驗系統滿意度問卷中，除了測驗系統滿意度問項題目外，另外還包括受測者的基本資料，基本資料包括性別、電腦使用經驗、網路使用經驗及網路使用頻率。

受測者在性別方面，男生有 58 人(51.8%)，女生有 54 人(48.2%)。在電腦使用經驗方面，使用電腦未達 2 年者有 31.25%，使用 2~4 年者有 39.29%，使用 4 年以上者有 29.46%。在網路使用經驗方面，使用網路未達 2 年者有 20.6%，使用 2~3 年者有 31.3%，使用 3~4 年者有 23.2%，使用 4 年以上者有 25%。在網路使用頻率方面，平均每日上網 1 小時以下者佔 14.3%，平均每日使用 1~3 小時者有 34.8%，平均每日使用 3~5 小時者有 33.9%，平均每日使用 5 小時以上者有 17%。以上資料顯示，受測者的電腦及網路使用經驗與頻率分佈相當均勻。

### 二、信度與效度

根據 Malhotra(1993)的建議，Cronbach  $\alpha$  值只要達到 0.6 以上便可接受該份量表，本研究參考 Compeau 與 Higgins(1995)的電腦自我效能量表，並經二位領域專家針對翻譯成中文的內容及詞句給予意見，加上此量表已經過許多研究的檢驗，因此具備相當高的內容效度。此外，本研究的電腦自我效能量表 Cronbach  $\alpha$  值為 0.8939，顯示具有高度信度。

本研究的系統滿意度量表係參考自文獻，並依照本研究之測驗系統功能特性及專家建議，發展出線上測驗系統之使用者滿意度量表，有其內容效度。為了檢驗滿意度量表的建構效度，本研究採用主成份因素分析，萃取出特徵值大於 1 的五個構念因素，每個項目均以因素負荷量(factor loading)絕對值均大於 0.5 做為因素項目取捨的依據。本研究依據各構念因素所對應的衡量項目加以命名為：智慧評分辨識能力(7 個項目)、智慧評分正確及合理性(4 個項目)、學習輔助與喜好(5 個項目)、二元評分正確性(3 個項目)及便利性(3 個項目)。智慧評分辨識能力因素主要包括評分機制在相似語意字彙、漏字、全半形、同音異字、集合順序等的智慧評分辨識能力問項。智慧評分正確及合理性因素主要包括模糊評分結果的正確性、評分機制與紙筆評分結果的一致性、評分機制斟酌給分及扣分之合理性、填充題型評分結果之正確性等問項。學習輔助與喜好因素則包括對於評分機制輔助學習及個人對於線上測驗系統環境的喜好等問項。二元評分正確性因素則是有關是非題、單選題及複選題等採用二元評分方法的評分正確性問項。便利性因素則包括線上測驗在時間彈性、地點彈性及測驗效率上所產生的便利效果問項。各因素中項目得分加總後，即為各滿意度指標的量化值，這五個滿意度指標的範圍分別是 7~35 分、4~20 分、5~25 分、3~15 分、3~15 分。整體滿意度指標的 Cronbach  $\alpha$  值為 0.93，另外「智慧評分辨識能力」、「智慧評分正確及合理性」、「學習輔助與喜好」、「二元評分正確性」及「便利性」等五項滿意度指標的 Cronbach  $\alpha$  值依序為 0.94、0.87、0.87、0.82 及 0.79，顯示此量表具有高度的信度。

### 三、線上評分機制種類對於學習者測驗成績影響的假設驗證

為了檢驗評分機制對於測驗成績的影響，並排除先備知識及電腦自我效能的干擾，本研究使用單因子共變數進行分析。為了讓分析結果能正確推論，本研究在常態性、獨立性及變異數同質性之檢定上均符合基本假設，另外也對共變數分析之組內迴歸係數同質性(homogeneity of within-class regression coefficient)假設進行檢驗。在評分機制對測驗成績影響的共變數分析中，本研究使用了 Levene 法進行變異數同質性檢定，結果顯示不同的評分機制間無顯著差異， $F(1,110)=0.507$ ， $p=0.478$ 。在線上評分機制\*先備領域知識的組內迴歸係數同質性檢驗部分，其  $F$  值=0.165， $p=0.848$ ，未達 0.05 顯著水準，表示二組迴歸線的斜率相同，符合共變數迴歸係數同質性假定。在線上評分機制\*電腦自我效能的組內迴歸係數同質性檢驗部分，其  $F$  值=1.690， $p=0.189$ ，亦未達 0.05 顯著水準，表示此二組迴歸線的斜率相同，亦符合共變數迴歸係數同質性假定。經由分析結果顯示，不同的評分機制對於測驗成績的影響適宜進行共變數分析。

不同的線上評分機制對於線上測驗成績影響分析的  $F$  值為 6.285， $p$  值為 0.014(見表 2)，因此在排除先備領域知識及電腦自我效能的干擾後，線上評分機制的種類會造成學習者測驗成績的顯著差異，因此支持假設 H1。從表 3 中發現，學生使用智慧型評分機制所獲得的平均測驗成績為 69.14 分，高於使用一般型評分機制的 61.70 分。如果將評分機制的評分成績與教師紙筆測驗評分的成績比較發現，智慧型評分機制的評分成績與紙筆測驗的教師評分成績相近，而一般型評分機制的評分成績與紙筆測驗的教師評分成績有很大的差距。因此線上測驗系統評分結果的正確性應被提出質疑，並有必要針對不同評分機制與教師紙筆測驗評分的結果差異進行後續分析比較。

為了瞭解各種線上評分機制的評分效力，本研究針對智慧型線上評分機制及一般型評分機制，分別與教師紙筆測驗評分結果進行成對樣本  $t$  檢定，結果如表 4。一般型評分機制與教師紙筆測驗評分的結果差異高達 5.357 分， $t$  值為 7.426，表示這兩種評分機制所產生的評分成績有顯著差異。而智慧型評分機制與紙筆測驗教師評分的結果差異則只有 0.25 分， $t$  值為 1.443，並未達到 0.05 顯著水準，表示這兩種不同的評分方式所產生的成績結果並無顯著差異。綜合成對檢驗結果可得知，在支援填充題型的測驗中，一般型評分機制與教師紙筆測驗的評分效力並不相等，而智慧型評分機制與教師紙筆測驗的評分效力並無顯著差異。

推論假設 H1 獲得支持的原因，是因為填充題答案可能有許多可替換的語意詞彙，因此電腦無法判斷出所有可替換的相似詞彙答案。此外，本研究初期進行的評分效力先期研究，共歸納出五項會影響支援填充題型的線上測驗系統評分效力降低之原因，其中包括電腦環境因素(大小寫及全半形)、相似詞彙語意交替使用、多格填充問題的集合順序、受測者中文程度低落(錯別字及同音異字)及隨機性漏字疏忽等。由於智慧型評分機制已使用智慧模糊評分技術來消弭及改進這些造成評分效力降低的問題，並模擬教師紙筆評分時的特性與規則，因此智慧型評分機制的評分結果與教師紙筆評分結果間並無顯著差異，顯示智慧型評分機制與教師紙筆評分有相同的評分效力。

此外，表 2 也顯示出線上評分機制種類對於學習者的測驗成績受到學習者不同先備領域知識程度影響分析的結果，其  $F$  值為 19.102， $p$  值為 0.000，顯示出線上評分機制種類對於學習者的測驗成績會因為學習者不同的先備領域知識程度而有顯著差異，因此支持假設 H3。進一步將先備領域知識程度與線上測驗成績進行 Pearson 相關分析，二變數間的相關係數為 0.422， $p$  值為 0.000，達到 0.05 顯著水準，顯示出先備領域知識程度與線上測驗成績結果間亦存在顯著的正相關。因此先備領域知識程度會影響線上評分機制對於線上測驗成績的表現，其中先備領域知識程度越好的學生會使其

線上測驗成績的表現也越好，而先備領域知識程度較差的學生在線上測驗成績的表現也會相對較不理想，此分析結果與過去大多數的研究結果相符合。

線上評分機制種類對於學習者測驗成績受到學習者電腦自我效能程度影響之分析結果如表 2，其 F 值為 11.433，p 值為 0.001，顯示出線上評分機制種類對於學習者的測驗成績會因不同的電腦自我效能程度而有差異，因此支持假設 H4。進一步將電腦自我效能程度與線上測驗成績進行 Pearson 相關分析，二變數間的相關係數為 0.376，p 值為 0.000，達到 0.05 顯著水準，顯示出電腦效能程度與線上測驗成績結果間存在顯著的正相關。因此電腦自我效能程度會影響線上評分機制對於線上測驗成績的表現，其中電腦自我效能較高的學生在線上測驗成績的表現也越好，而電腦自我效能較低的學生在線上測驗成績的表現也會相對較不理想。

表2 評分機制對於測驗成績影響之共變數分析結果

Source	Sum of Squares	df	Mean Square	F值	p值 ( $\alpha=0.05$ )
Corrected Model	8812.742	3	2937.581	15.693	0.000
Intercept	1458.280	1	1458.280	7.790	0.006
先備領域知識(共變數)	3575.731	1	3575.731	19.102	0.000***
電腦自我效能(共變數)	2140.146	1	2140.146	11.433	0.001***
評分機制	1176.464	1	1176.464	6.285	0.014*
Error	20216.535	108	187.190		
Total	508359.000	112			
Corrected Total	29029.277	111			

註：R<sup>2</sup> = .304 (Adjusted R<sup>2</sup> = .284); \*p<0.05, \*\*p<0.01, \*\*\*p<0.001

表3 紙筆測驗成績及線上測驗成績之結果

依變數 自變數		紙筆測驗成績			線上測驗成績		
		Mean	S.D.	N	Mean	S.D.	N
評分 機制	智慧型(第1組+第3組)	69.39	15.95	56	69.14	15.86	56
	一般型(第2組+第4組)	67.05	17.44	56	61.70	15.75	56

註：教師紙筆評分平均成績為 68.22 分(S.D. 為 16.68)

表4 教師紙筆測驗評分與線上評分機制評分結果之成對樣本t檢定

成對檢驗之評分機制項目	N	Paired Differences		t值	df	p值 ( $\alpha=0.05$ )
		Mean	S.D.			
教師紙筆評分-一般型評分	56	5.357	5.398	7.426	55	0.000***
教師紙筆評分-智慧型評分	56	0.250	1.297	1.443	55	0.155

註：\*p<0.05, \*\*p<0.01, \*\*\*p<0.001

#### 四、線上評分機制種類及回饋種類對於系統滿意度的假設驗證

為了檢驗不同的評分機制及回饋種類對於系統滿意度的影響，並排除電腦自我效能的干擾，本研究採用多變量共變數技術進行分析。在多變量共變數的基本假設檢驗部分，其常態性、獨立性等均符合基本假設。此外本研究使用 Box's M 方法檢驗變異數同質性，其 F 值=1.108，p 值=0.286，未達 0.05 顯著水準，表示各細格內樣本的變異數並無顯著差異。自我效能這個共變數對於系統滿意度各項指標的組內迴歸係數同質性檢驗結果如表 5，無論在回饋種類\*電腦自我效能、評分機制\*電腦自我效能、評分機制\*回饋機制\*自我效能的組內迴歸係數同質性檢驗上，均未達到 0.05 顯著水準，表示其迴歸線的斜率相同，符合共變數迴歸係數同質性假定，因此不同的評分機制及回饋種類對於系統滿意度的影響適宜進行共變數分析。

表 6 為線上評分機制對於系統使用滿意度影響的結果，其 F 值為 5.685，p 值為 0.000，表示在排除電腦自我效能的干擾後，使用不同的線上評分機制對於學習者的整體系統滿意度會造成顯著差異，因此支持假設 H2a。回饋種類對於系統使用滿意度影響分析，其 F 值為 4.767，p 值為 0.001，表示在排除電腦自我效能的干擾後，不同的回饋種類對於受測者的系統滿意度會造成顯著差異，因此支持假設 H2b。在評分機制種類及回饋種類交互作用後對於系統使用滿意度影響分析，其 F 值為 0.363，p 值為 0.873，表示線上評分機制種類及回饋種類交互作用，在排除電腦自我效能的影響後，對於學習者的整體系統滿意度並無顯著差異，因此未支持假設 H2c。此外，在評分機制種類與回饋種類對於學習者的系統滿意度受到電腦自我效能程度影響的分析結果顯示，其 F 值為 0.147，p 值為 0.981，表示評分機制種類與回饋種類對於學習者的系統滿

表 5 滿意度指標之組內迴歸係數同質性檢驗

檢驗變數	評分機制*自我效能		回饋種類*自我效能		評分機制*回饋種類*自我效能	
	F值	p值	F值	p值	F值	p值
滿意度指標						
智慧評分辨識能力	2.442	0.920	2.008	0.159	0.152	0.697
智慧評分正確及合理性	0.960	0.329	0.021	0.885	0.223	0.637
學習輔助與喜好	2.719	0.102	1.602	0.208	0.184	0.669
二元評分正確性	0.409	0.524	0.544	0.462	1.581	0.211
便利性	0.750	0.389	0.237	0.627	0.544	0.462

表 6 系統滿意度之多變量共變數檢驗結果

Effect	Wilks' Lambda	F值	p值
電腦自我效能(共變數)	0.993	F(5,103)=0.147	0.981
評分機制種類	0.784	F(5,103)=5.685	0.000***
回饋種類	0.812	F(5,103)=4.767	0.001***
評分機制種類*回饋種類	0.983	F(5,103)=0.363	0.873

備註：\*p<0.05, \*\*p<0.01, \*\*\*p<0.001

表 7 評分機制對各項滿意度指標之平均數及多變量變異數分析表

評分機制(自變數)	智慧型評分		一般型評分		F值	p值
	平均數	標準差	平均數	標準差		
滿意度指標(依變數)						
智慧評分辨識能力	24.393	5.783	18.857	7.247	20.492	0.000***
智慧評分正確及合理性	12.018	3.267	9.161	3.888	19.840	0.000***
學習輔助與喜好	18.714	3.172	17.393	4.335	5.773	0.018*
二元評分正確性	11.232	2.551	10.036	3.003	3.886	0.051
便利性	11.875	1.945	11.036	2.670	3.851	0.052

備註：\*p<0.05, \*\*p<0.01, \*\*\*p<0.001

意度並不會因為受測者的電腦自我效能程度而有差異，因此未支持假設 H5，其原因可能是因為在本研究中系統滿意度主要受到系統實際提供的品質、功能及特性影響，這些因素可能難以藉由測驗者本身的電腦自我效能程度而影響對於系統滿意度的認知。

後續則針對評分機制種類及回饋種類在系統滿意度出現整體差異的部分，進行個別變數的變異數分析。

#### (一) 評分機制對於系統滿意度之差異分析

從評分機制對各滿意度指標的變異數結果發現假設 H2a 幾乎獲得完全支持(見表 7)，使用不同線上評分機制的學生除了在「二元評分正確性」及「便利性」的滿意度上無顯著差異外，在「智慧評分辨識能力」、「智慧評分正確與合理性」及「學習輔

助與喜好」等三項滿意度上均有明顯差異。這可能是因為智慧型評分機制能消弭填充題型的評分效力問題，並提供填充題智慧評分及模擬紙筆教師的評分規則，因此評分結果與教師親自批改之結果相似，也比較為合理。另外，透過智慧評分的回饋說明解釋，也能提供學習者更多的學習輔助，因此不同的評分機制在「智慧評分辨識能力」、「智慧評分正確與合理性」及「學習輔助與喜好」等三項滿意度指標上會有顯著的差異。另外，「二元評分正確性」及「便利性」二個滿意度指標上其差異顯著性也接近 0.05。不過，無論在哪一種線上評分機制上，所提供的測驗地點與時間彈性也大致相同，而所包含之是非題、單選題及複選題皆是採用二元評分方法，因此在「二元評分正確性」及「便利性」二個滿意度指標以 0.05 顯著水準來看，並無顯著差異。

表 8 回饋種類對各項滿意度指標之平均數及多變量變異數分析表

評分機制(自變數) 滿意度指標(依變數)	KR回饋		綜合回饋		F值	p值
	平均數	標準差	平均數	標準差		
智慧評分辨識能力	20.161	6.962	23.089	6.981	5.753	0.018*
智慧評分正確及合理性	9.321	3.293	11.857	3.979	15.620	0.000***
學習輔助與喜好	16.661	3.455	19.446	3.722	17.044	0.000***
二元評分正確性	9.946	2.932	11.321	2.587	7.317	0.008**
便利性	10.982	2.386	11.929	2.263	4.756	0.031*

備註：\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

## (二)回饋種類對於系統滿意度之差異分析

在回饋種類對各滿意度指標的變異數結果發現(見表 8)，使用具備不同回饋種類線上測驗系統的學生在「智慧評分辨識能力」、「智慧評分正確與合理性」、「學習輔助與喜好」、「二元評分正確性」及「便利性」等五項滿意度指標上均有顯著差異。由於 KR 回饋僅提供簡單的測驗回饋，因此在各種題型的評分結果中，KR 回饋均只給予答對或答錯的簡單回饋訊息，而未透過回饋給予課程學習相關的輔助。所以使用 KR 回饋的測驗者可能對於自己所答錯的問題，在系統給予回饋後仍然不知道為何自己答錯，更無法得知正確答案的解釋及額外的知識說明。由於系統未針對評分結果給予評分解釋的回饋，因此更容易造成測驗者對於測驗系統評分正確性、評分合理性、評分成效等項目提出質疑，甚至連帶對於測驗系統便利性的滿意程度也受到影響。而綜合回饋在任何題型上，均會提供評分解釋及額外的知識輔助，倘若綜合回饋與智慧評分機制相互搭配，則對於填充題型評分過程所出現的詞彙語意相似、漏字、同音異字、全半形、集合順序等智慧評分分析，均會提供詳細的評分結果解釋與說明。此外，綜合回饋也會利用回饋時機，提供額外的相關知識內容，以輔助測驗者更有效學習。因此，在綜合回饋環境中，測驗者對於各項滿意度指標的滿意度都能顯著提高。

## 五、四組實驗組對於系統滿意度的事後比較檢驗

雖然先前的分析顯示，線上評分機制種類及回饋種類交互作用，在排除電腦自我效能的影響後，對於學習者的整體系統滿意度並無顯著差異，因而未支持假設 H2c。這有可能是因為四組實驗組之樣本數不足(各組均只有 28 人)，評分機制種類及回饋種類的個別效果均對於系統滿意度有高度影響，彼此交互作用後對於系統滿意度的提升效果便相對有限。為了瞭解線上測驗系統若同時採用智慧評分機制及綜合回饋(第 3 組)，與其他各種評分機制及回饋種類的組合，是否會有顯著的差異，因此本研究針對四個實驗組進行系統滿意度影響的多變量變異數分析的事後檢驗比較。經過分析結果顯示其 Wilks' Lambda 值=0.639，F 值=3.381，p 值=0.000，顯示不同的實驗組對於各項系統滿意度指標有顯著的差異。進一步透過 Scheffe 來檢驗各實驗組在各個滿意度指標



上的差異，在表 9 中顯示出第 2 組在各項系統滿意度指標中均低於其他各組，第 3 組在各項系統滿意度指標中均高於其他各組。此外，在五項系統滿意度指標中，第 3 組學習者在各項系統滿意度指標上均顯著高於第 2 組的學習者。由表 9 的分析可得知，不同的評分機制種類與回饋種類組合，對於「智慧評分辨識能力」、「智慧評分正確及合理性」、「學習輔助與喜好」、「二元評分正確性」及「便利性」均會產生影響，因此線上測驗系統若要同時具備良好的評分能力及使用者對於測驗系統的滿意度，則應該同時加強評分機制的智慧評分能力以及回饋內容的豐富性。

表 9 四個實驗組進行系統滿意度影響的事後檢驗比較

系統滿意度指標 實驗組	智慧評分 辨識能力	智慧評分正 確及合理性	學習輔助 與喜好	二元評分 正確性	便利性
智慧型評分機制+ KR回饋 (第1組)	23.00	10.68	17.43	10.86	11.57
一般型評分機制+ KR回饋 (第2組)	17.32	7.96	15.89	9.04	10.39
智慧型評分機制+ 綜合回饋(第3組)	25.79	13.36	20.00	11.61	12.18
一般型評分機制+ 綜合回饋(第4組)	20.39	10.36	18.89	11.04	11.69
F值	8.814	11.899	7.037	4.737	3.024
p值	0.000***	0.000***	0.000***	0.004**	0.033*
Scheffe (組別間的比較)	3>2*** 1>2* 3>4*	3>2*** 3>1* 3>4* 1>2*	3>2*** 4>2*	3>2**	3>2*

備註：\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

表 10 本研究假設檢定之結果彙整

假設項目	假設檢驗內容	結果
H1	評分機制種類對於學習者的測驗成績有差異	支持
H2a	評分機制種類對於學習者的系統滿意度有差異	幾乎獲得完全支持
H2b	回饋種類對於學習者的系統滿意度有差異	支持
H2c	評分機制種類對於學習者的系統滿意度因不同的回饋種類而有差異	拒絕
H3	評分機制種類對於學習者的測驗成績會因不同的先備領域知識程度而有差異	支持
H4	評分機制種類對於學習者的測驗成績會因不同的電腦自我效能程度而有差異	支持
H5	評分機制種類與回饋種類對於學習者的系統滿意度會因電腦自我效能程度而有差異	拒絕

## 陸、結論與建議

### 一、研究結論

綜合資料分析結果，假設檢定之結果彙整於表 10。總結分析發現，線上測驗系統若採用不同的評分機制會對學習者的成績產生影響，因此建議教育機構或企業在建立電子化學習環境時，需要檢視線上測驗系統的評分機制是否具備良好的評分效力。一個好的線上評分機制其評分結果應該與教師親自評分的結果近乎相同(事實上不可能完全相同)，因此評分機制需要具備足夠的智慧能力以處理填充題型的詞彙語意、同音異字、集合順序、全半形、大小寫、漏字等可能造成評分效力降低的評分問題，而分析結果顯示本研究所設計的智慧型評分機制能夠模擬教師評分的特性與規則，並具備與教師紙筆評分近乎相同的評分效力(統計上無顯著差異)。

在線上測驗系統的滿意度分析中顯示，線上測驗系統若採用不同的評分機制或回饋種類都將會對學習者的系統滿意度產生顯著影響，因此教育機構或企業若要使學習者對測驗系統感到滿意，則可同時針對評分機制的評分能力及測驗回饋的方式努力。評分機制種類與回饋種類的交互對於學習者的系統滿意度的影響並不顯著，但在事後分析發現，雖然評分機制種類及回饋種類個別對於學習者的系統滿意度都有顯著影響，但在四組實驗組中，由於第3組使用的線上測驗系統同時採用智慧型評分機制及綜合回饋技術，因此學習者在五種滿意度指標中均顯著高於採用一般型評分機制或KR回饋的組別。建議未來在線上測驗的發展上，應該同時採用智慧評分機制及綜合式回饋，以提升測驗系統的評分正確性及使用者的系統滿意度。

學習者本身的先備知識程度及電腦自我效能程度都會影響評分機制對於測驗成績的結果，因此教師及未來研究者可以思考如何讓測驗系統可以依據學習者的先備知識及電腦自我效能程度給予不同的測驗內容、評分方法，並進行適性化的測驗與評分。而學習者的自我效能程度並不會影響評分機制種類與回饋種類對於學習者的系統滿意度，因此教師及教育訓練人員若要提升系統滿意度，並不需要考量學習者對於電腦自我效能的程度，而需要將重心關注在評分機制的智慧評分能力及回饋的內容方式。

## 二、本研究之教育與實務意涵

本研究在教育與實務上的意涵歸納如下：

- (一) 對於教育機構教師及企業教育訓練者而言，本研究明確指出線上測驗系統若支援填充題型會影響評分效力的五大原因。
- (二) 一般線上測驗系統提供的評分機制只能提供認知型(recognition type)的測驗題型，例如是非題、單選題、複選題等，對於未事先準備的測驗者而言，容易發生猜測答案的情形，對於判斷學生對知識的瞭解程度便容易產生偏誤。本研究證實智慧型評分機制可以改善線上測驗系統支援建構型(constructed type)題型中的填充題型所造成的評分問題，並使測驗系統在支援填充題型問題時，仍舊可以具備與教師親自評分時相同的評分效力。因此，本研究的成果可做為將填充題納入線上國家考試測驗或全國性的基本能力測驗中的可行性參考。
- (三) 本研究之結論可以協助教師、教育訓練者及線上測驗系統發展者去瞭解線上評分機制及回饋種類對於線上測驗結果及系統滿意度的影響，並同時瞭解先備知識及自我效能對於線上測驗結果及系統滿意度的干擾效果，以提供改進線上測驗系統功能的的方向。
- (四) 本研究結合模糊邏輯、資料庫、人工智慧等技術，提供高品質的系統內建相似語意詞庫相似值及評分參數，以發展出具備專家教師相同批閱效力的線上測驗系統。本研究觀察到過去一般測驗系統在進行填充題題型評分時，僅能評定出「答對」或「答錯」的二元歸屬結果，而本研究則藉由模糊觀念將評分結果擴充至更彈性的範圍，評分之結果可能因為語意相似程度、同音異字、漏字等因素，而產生介於0至1間的答案正確程度歸屬函數值。因此藉由模糊觀念的設計，本系統的確能如同教師人工評分時，視學生答題內容的正確程度或相似程度，「斟酌」給予分數，而使得測驗系統更具備人性化的評分。

## 三、研究限制與後續研究發展建議

本研究之研究限制與後續研究發展建議歸納如下：

- (一) 本研究的研究方法採實地實驗法，樣本的取得是以中部某技術學院資管系的學生作為樣本來源，因此實驗結果分析的推論，尚不能延伸到所有的對象及範圍。建

議後續研究者除了採用實地實驗法之外，亦能擴大樣本取得的對象、區域及範圍，以使得結果可推論到其他的區域及範圍。

- (二) 本研究所指稱的一般線上測驗系統，主要根據表 1 所歸納的線上測驗系統功能特性，並參考國內知名測驗網站 K-bar 考試網與陽明國中官老師課程教學網之測驗系統功能，以產生與大多數線上測驗系統功能相似且具有代表性的測驗系統，以提升本研究的外部效度。本研究的實驗方法研究結果，無法推論到其他提供多媒體試題、試題反應理論及其他未包含在本研究所探討之相關變數的測驗系統上。
- (三) 本研究目前僅針對填充題型造成的測驗效力問題進行研究，目前的研究成果若應用在 e-learning 及知識管理的環境中，僅能對於 knowledge-what 層次的知識進行有效的評量。未來的研究者可針對牽涉複雜語句文法及語意結構的問答題型進行研究，以探討更高層次的知識測驗應用，並能更有效的輔助企業從事教育訓練及知識管理等工作。
- (四) 本研究目前並未將填充題型測驗與電腦適性測驗進行整合應用，未來研究者可以針對此部分進行探討，並在 e-learning 及知識管理環境中針對個別學習者，提供各種不同層面的適性應用，諸如適性的知識內容學習、適性測驗、適性評分等，以提供學習者能更彈性及更有效率的學習知識。
- (五) 本研究目前仍須藉由申訴機制及教師的輔助，來擴充智慧評分機制所需要的語意詞彙及評分規則，未來研究可採用機器學習技術從知識管理系統中自動擴充相似語意詞庫之內容。

## 致 謝

本研究受行政院國科會專案計畫(NSC 93-2416-H-004-013)補助，特此致謝。

## 參考文獻

- 考試院，"國家考試將規劃實施電腦化測驗"，考試院新聞稿，[http://w3.moex.gov.tw/examnews/exnews\\_2.asp?pgn=1](http://w3.moex.gov.tw/examnews/exnews_2.asp?pgn=1)，2003 年，資料取得時間 2003 年 9 月 3 日。
- 何榮桂，"電腦教學系統中的測驗設計"，*中等教育*，第四十一卷第二期，1990 年，頁 29~34。
- 何榮桂，"遠距測驗與評量"，*2000 網路學習理論與實務研討會論文集*，新竹市：國立交通大學，2000 年，頁 34~43。
- 何榮桂、郭再興、蘇建誠、陳麗如，"在 Internet 上建構測驗環境之可行性及相關問題之探討"，*新世紀測驗學術發展趨勢*，心理出版社，台北，1999 年，頁 125~135。
- 何榮桂、蘇建誠和郭再興，"遠距適性測驗系統架構"，*資訊與教育雜誌*，第四十二期，1996 年，頁 29-35。
- 呂怡緯，"入口網站服務品質之研究—以搜尋網站為例"，*國立台灣科技大學管理研究所企業管理學程碩士論文*，1998 年。
- 周文正，"WWW 上電腦輔助測驗系統之研製"，*中華民國第七屆電腦輔助教學研討會*，1998 年，頁 543-550。
- 周倩、簡榮宏，"網路評量系統之發展與研究"，*遠距教育*，第四卷，1997 年，頁 12~15。

- 林明達, "全球資訊網線上測驗系統之設計與製作", 國立交通大學資訊科學研究所碩士論文, 1998年。
- 林璟豐, "全球資訊網測驗題型之研究", 國立師範大學工業科技教育研究所碩士論文, 2001年。
- 洪榮昭, 電腦輔助教學之設計原理與應用, 師大書苑, 台北市, 1992年。
- 翁榮源, "適性式學習環境在氧化還原網站之應用研究", 中國化學會化學會誌, 第六十一卷, 第一期, 2003年, 頁115~125。
- 游寶達, "ICL 心智模式取向之智慧型電腦輔助診斷學習系統之研究", 民國87年度國科會『電腦輔助學習』專題研究計畫報告, NSC87-2511-S-194-009-ICL, 行政院國科會, 台北市, 1998年。
- 黃郁雯, "電腦自我效能、電腦經驗及他人支持三者與電腦態度及電腦焦慮之關係", 國立政治大學心理研究所碩士論文, 2003年。
- 黃國禎、曾秋蓉、朱蕙君和蕭經武, "智慧型線上測驗系統題型之分析與改進", 科學教育學刊, 第十卷, 第四期, 2002年, 頁423~439。
- 鼎茂, 教育與心理測驗, 鼎茂出版社, 台北市, 2000年。
- 簡茂發, "多元化評量之理念與方法", 教師天地, 第九十九期, 1999年, 頁11~17。
- 羅凱揚, "網路報紙使用者滿意度評估模式之建立", 國立台灣科技大學企業管理研究所碩士論文, 1998年。
- Alessi, S.M., and Trollip, S.R., *Computer-based Instruction: Methods and Development*, 2<sup>nd</sup>, Prentice-Hall, Englewood Cliffs, NJ, 1991.
- Bennett, R.E., Goodman, M., Hessinger, J., Kahn, H., Liggett, J. Marshall, G., and Zack, J., "Using multimedia in large-scale computer-based testing programs," *Computers in Human Behavior* (15), 1999, pp.283-294.
- Bostorm, R.P., "The importance of learning style in end-user training," *MIS Quarterly* (14:1), 1990, pp.101-109.
- Bugbee, A.C., "The equivalence of paper-and-pencil and computer-based testing," *Journal of Research and Computing in Education* (28:3), 1996, pp.282-299.
- Compeau, D., and Higgins, C. A., "Computer self-efficacy: development of a measure and initial test," *MIS Quarterly* (19:2), 1995, pp.189-211.
- Cronbach, L.J., and Snow. R.E., *Aptitudes and Instructional Methods*, Irvington, NY, 1977.
- Cyert, R.M., and March, J.G., *A Behavioral Theory of the Firm*, Prentice-Hall, Englewood Cliffs, NJ, 1963.
- DeLone, W.H., and McLean, E.R., "Information system success: the quest for the dependent variable," *Information System Research* (3:1), 1992, pp.60-95.
- Dempsey, J., Driscoll, M.P., and Swindell, L.K., "Text-Based Feedback", in *Interactive Instruction and Feedback*, Dempsey, J.V., and Sales, G.C. (Eds.), Educational Technology Publications, Englewood, NJ, 1993, pp.21-54.
- Dempsey, J.V., and Wager, S.U., "A taxonomy for the timing of feedback in computer-based instruction," *Educational Technology* (28:10), 1988, pp.20-25.
- Devedzic, V.B., "Key issues in next-generation web-based education," *IEEE Transactions On Systems, Man, and Cybernetics-PART C: Applications and Reviews* (33:3), 2003, pp.339-349.
- Dochy, F., Segers, M., and Buehl, M.M., "The relation between assessment practices and outcomes of studies: the case of research on prior knowledge," *Review of Educational Research* (69:2), 1999, pp.145-186.
- Gilman, D.A., "Comparison of several feedback methods for correcting errors by computer-assisted instruction," *Journal of Educational Psychology* (60), 1969, pp.503-508.

- Gronlund, N.E., *Assessment of student achievement*, Allyn and Bacon, Needham Heights, MA, 1998.
- Han, T., Kolen, M., and Poglmann, J., "A comparison among IRT true-and observed-score equatings and traditional equipercentile equating," *Applied Measurement in Education* (10:2), 1997, pp.105-121.
- Hwang, G.J., "A conceptual map model for developing intelligent tutoring systems," *Computers and Education* (40), 2003, pp.217-235.
- Igbaria, M., and Iilvari, J., "The effects of self-efficacy in computer usage," *Omega* (23:6), 1995, pp.587-605
- Jonassen, D.H., and Grabowski, B.L., *Handbook of Individual Differences Learning, and Instruction*, Lawrence Erlbaum Associates, Hillsdale, 1993.
- Karsten, K., and Roth, R.M. "The relationship of computer experience and computer self-efficacy to performance in introductory computer literacy Courses," *Journal of Research on Computing in Education* (31:1), 1998, pp.14-24.
- Lord, F.M., *Applications of Item Response Theory to Practical Problems*, Erlbaum, Hillsdale, NJ, 1980.
- Mark, D.R., "The next generation of computerized tests: implications for testing of advances in multimedia, intelligent tutoring systems, and language processing," *AEDS Journal* (19:2-3), 1997, pp.81-108.
- Malhotra, N. K., *Marketing research: an Applied Orientation*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- Marshall, S.P., "The Assessment of Schema Knowledge for Arithmetic Story Problems: a Cognitive Science Perspective," in *Assessing Higher Order Thinking in Mathematics*, G. Kulm (Ed.), American Association for the Advancement of Science, Washington, 1993, pp.155-168.
- McCormack, D. and Jones, D., *Building a Web-Based Education System*, Wiley, NY, 1997.
- McCormick, C.B. and Pressley, M., *Educational Psychology: Learning, Instruction, Assessment*, Longman, NY, 1995.
- Moundridou, M., and Virvou, M., "Analysis and design of a web-based authoring tool generating intelligent tutoring systems," *Computer and Education* (40), 2003, pp.157-181.
- Pitt, L.F., Watson, R.T. and Kavan, C.B., "Service quality: a measure of information systems effectiveness," *MIS Quarterly* (19:2), 1995, pp.173-187.
- Sun, K. T., "An Effective Item Selection Method by Using AI Approaches," *The Meeting of the Advanced in Intelligent Computing and Multimedia System*, Baden-Baden, Germany, 1999.
- Swafford, M., and Brown, D., "MallardTM: Asynchronous Learning on the World-Wide-Web," *Proceedings of the ASEE 96 Conference*, Washington, DC, 1996, pp.23-26.
- Tam, S. F., "Self-efficacy as predictor of computer skills learning outcomes of individuals with physical disabilities," *The Journal of Psychology* (130:1), 1996, pp.51-58.
- Tsai, T. H., Hanson, B.A., Kolen, M.J., and Forsyth, R.A., "A comparison of bootstrap standard errors of IRT equating methods for the common-item nonequivalent groups design," *Applied Measurement in Education* (14:1), 2001, pp.17-30.
- Van der Linden W. J., and Glas, C.A.W., *Computerized Adaptive Testing: Theory and Practice*, Kluwer Academic Publishers, Boston, 2000.
- Van Gorp, M.J., and Boysen, P., "ClassNet: Managing the virtual classroom," *International Journal of Educational Telecommunications* (3:2), 1997, pp.279-292.
- Wainer, H., and Dorans, N.J., *Computerized adaptive testing: a primer*, 2<sup>nd</sup>., Lawrence Erlbaum Association, Mahwah, NJ, 2000.

- Wang, A.Y., and Newlin, M.H., "Predictors of web-student performance: the role of self-efficacy and reason for taking and on-line class," *Computer in Human Behaviors* (18:2), 2002, pp.151-163.
- Weiss, D.J., *Proceedings of the 1979 Computerized Adaptive Testing Conference*, University of Minnesota Computerized, Minneapolis, 1980.
- Zadeh, L. A., "Fuzzy sets," *Inform.Control* (8), 1965, pp.338-353.

## 附錄: 本研究中使用部分問卷之量表項目

電腦自我效能部分之問題項目
1.雖然沒有人指導我如何使用這個電腦軟體或網路系統，我一樣能正確的使用它
2.即使我不曾使用過類似的電腦軟體或網路系統，我一樣能使用它
3.如果我有這個電腦軟體或網路系統的操作手冊，我就能使用它
4.如果有師長或同學先操作一次給我看，我就能使用它
5.如果遇到操作的問題有人可以請教，我就能使用它
6.如果初學時有人教導我，我就能使用它
7.如果完成指定工作或作業的時間很充裕，我就能使用它
8.若電腦軟體或網路系統有線上求助的功能，我就能使用它
9.只要在初學時有人簡略的操作一次給我看，我就能使用它
10.如果我之前有使用類似電腦軟體或網路系統的經驗，我就能使用它
線上測驗系統成效及滿意度部分之問題項目
1.使用線上測驗系統進行測驗比紙筆測驗更有效率
2.線上測驗系統可以讓同學更彈性的選擇適當的測驗地點
3.線上測驗系統可以讓同學更彈性的選擇適當的測驗時間
4.我認為線上測驗系統在是非題類型的評分結果是正確的
5.我認為線上測驗系統在單選題類型的評分結果是正確的
6.我認為線上測驗系統在複選題類型的評分結果是正確的
7.我認為線上測驗系統在填充題類型的評分結果是正確的
8.整體而言，我相信線上測驗系統評分結果的正確性
9.我認為線上測驗系統的評分結果與老師親自評分的結果一致
10.我認為線上測驗系統的評分機制及扣分是合理的
11.我認為線上測驗系統提供的測驗結果回饋內容，對於資訊提供及知識學習都會有顯著的幫助
12.我認為線上測驗系統在填充題的評分機制中，既使我輸入了跟標準答案不同，但卻與標準答案相同或相同的語意字彙，測驗系統也能正確辨識及斟酌狀況給分
13.我認為線上測驗系統在填充題的評分機制中，既使我不小心遺漏輸入其中部分答案內容，測驗系統依舊能正確的辨識及斟酌給分
14.我認為線上測驗系統在填充題的評分機制中，既使我不小心切換成「全形」輸入的答案內容，測驗系統依舊能正確的辨識及給分
15.我認為線上測驗系統在填充題的評分機制中，既使我不小心因為注音輸入法選字，將答案中的某個字選到另一個同音但不正確的字，測驗系統也能正確的辨識及並斟酌狀況給分
16.我認為線上測驗系統在填充題的評分機制中，當填充格有二個以上且答案順序可以相互顛倒時，無論我填的答案順序為何，系統都能正確辨識及給分
17.我認為線上測驗系統在填充題的評分機制中，既使答案是英文字母大寫的字，而我不小心切換成「字母小寫」並將答案輸入時，測驗系統依舊能正確的辨識及給分
18.當我對於線上測驗的評分結果有所懷疑時，我可以透過線上申訴機制機制而爭取到應得的分數
19.我喜歡利用線上測驗系統進行測驗
20.如果可以選擇紙筆測驗及線上測驗任何一種測驗方式，我會選擇線上測驗的方式
21.我認為線上測驗系統能提供歷史的個人測驗資料及電子化試卷，這可以讓我更瞭解自己學習的歷程及狀況
22.我認為線上測驗系統能夠幫助同學在電子學習環境中的學習

