

第六章 結論與建議

第一節 結論

在本研究中有二個主要的核心研究問題，第一個研究問題是「如何在多位教師共用測驗系統的環境中，讓電腦具備教師個人化的評分風格與特質，並在包含填充題型的線上測驗中，電腦自動評分的結果能具有與紙筆測驗相同的評分效力？」第二個研究問題則是「如何讓電腦輔助教師產製具有 Bloom 分類概念且符合試題編製原則的試題？」

針對第一個研究問題，本研究利用模糊函數、相似語意詞庫、集合、樣式比對、規則推論等觀念及技術來發展一個具備智慧評分的線上測驗與智慧評分子系統，以使測驗系統能具備與教師相同的批閱能力。在探討此問題之前，本研究針對線上測驗系統常見的是非題、單選題、複選題及填充題等四種試題類型，先進行了「探討影響線上測驗評分效力因素來源」的先導研究，研究結果顯示傳統紙筆測驗與線上測驗兩種測量工具在是非題、單選題、複選題之學生得分上並無顯著差異，而填充題部分之得分在紙筆測驗及線上測驗上則有顯著的分數差異，經過每一份試卷的分析比對，本研究發現填充題型在線上測驗無法擁有紙筆測驗同樣的評分效力，其主要形成的原因包括電腦環境介面、可替換之相似詞彙、填充格的集合順序、學生中文程度低落及隨機性漏字疏忽等五個來源。除了第五項影響評分效力的來源是無法藉由資訊科技來控制之外，本研究針對前四項因素來源，進行智慧評分機制的系統設計，以期線上測驗系統能具有與教師人工評分時相同的評分能力與風格。

經過本研究實作雛形系統後，隨即透過實驗設計進行智慧評分機制的評分效力實證檢驗，研究結果顯示在包含填充題型的測驗中，不同的評分機制在測驗成績的評分結果上會有顯著差異，智慧型評分與紙筆測驗評分間的成績差距，會明顯比一般型評分機制與紙筆測驗評分間的成績差距還小，因此支持 H1 假設。但「未擴充語意前智慧評分機制」的評分效力尚無法與紙筆測驗等化，若經過學生實際測驗後的語意擴充，將能使得智慧評分機制具有與紙筆測驗相同的評分效力，並支持 H2 之假設。

對於用線上測驗系統做為評分工具的教師及從事線上測驗系統發展工作者而言，可以持續透過智慧評分機制、語意詞彙關係的發展及教師評分規則參數的建立，使得線上測驗系統提供填充題題型時，仍然可具備與紙筆測驗相同的評分效力，並能保存教師個人「斟酌」給分的風格與特質。而線上測驗系統若只處理是非題、單選題、複選題等具有固定答案的測驗題型時，並不會產生與紙筆測驗結果不同的測驗結果。但在包含填充題型的線上測驗中，則需要注意其評分機制的評分效力問題。

除此之外，在多位教師共享相似語詞庫的線上測驗平台環境中，本研究能讓電腦模擬教師個人的評分風格與特質，並藉由模糊推論技術來讓系統自動整合系統內所有教師專家的評分認知及共識，提供高品質的相似語意詞庫及評分環境參數建議，降低教師建立語意詞彙及環境參數的負擔。

由於在教育領域中，Bloom 的認知領域教育目標分類理論廣泛被應用在教學課程設計及測驗試題產製的領域中。因此針對第二個研究問題，本研究將透過人工智慧、Bloom 教育目標分類理論、本體論、詞彙網路、規則推論等概念及技術，發展輔助教師產製測驗題庫子系統，以使電腦能輔助教師產製出線上學習系統所需之題庫。

在發展輔助教師產製測驗題庫子系統之前，本研究先針對教師人工出題時試題所能涵蓋 Bloom 知識向度及認知向度的範圍及所可能遭遇的困能進行先導研究，研究結果顯示當教師只能產製是非題、單選題、複選題及填充題的題型限制下，參與實驗教師實際能產製的試題大致上能涵蓋所有的 Bloom 知識類型，但在認知層次方面，屬於評鑑或創造認知層次的試題則較不容易發展，特別在創造層次的試題上並無任何參與實驗教師能利用這種題型來產製。此外，本研究亦依照試題編製原則及參與實驗教師實際產製試題的內容，結合 Bloom 分類理論發展教材知識本體及試題結構，並實作雛形系統來驗證系統成效。而在電腦產製試題的可行性方面，本研究透過實作系統及教師親自操作系統，再讓參與實驗教師進行系統成效的評估。研究成果顯示，15 位教師平均每人花費 3.9 小時，共產製 386 個不同的試題，而本研究根據相同的教材內容，由電腦花費約 5 分鐘產製了 18621 題試題，這些試題所能評量的知識類型包含了事實知識 12507 題，概念知識 3002 題，程序知識 3112 題，另外所能評量的認知層次包括記憶層次 3058 題，了解層次 32 題，應用層次 50 題，分析層次 14968 題，評鑑層次 513 題，評鑑層次試題則無法透過是非題、單選題、複選題及填充題題型來發展。

而在試題品質方面，電腦輔助教師產製試題之試題採用意願分析結果顯示，不同的教師對於特定題型及題目難易度會有特殊的認知偏好及效用認知差異，因此會造成某些教師對於特定題型給予試題較差的主觀評價，若排除教師對於特定題型的排斥因素後，參與實驗的教師對於電腦輔助教師產製題庫的合適性與認同度方面，均給予正面的評價，顯示本研究的試題能被一般教師所接受及使用。但是，如果要對於試題品質進行客觀的評估，則需要另外針對大量的試題進行額外的實證研究，以取得每個試題的難易度與鑑別度結果。

此外，電腦輔助教師產製試題子系統的系統成效評估分析結果顯示，利用電腦能夠輔助教師產製出用以評量「事實知識」、「概念知識」及「程序知識」等各種知識類型的試題，以及輔助教師產製出用以評量認知歷程的「記憶」、「了解」、「應用」、「分析」及「評鑑」等類型的試題，此結果與教師人工出題範圍一致，顯示本研究確實能輔助教師進行具備 Bloom 分類概念資訊的試題發展。而透過電腦輔助教師產製用以評量較低層次認知的試題表現較佳，若要產製評量更高層次的認知類型試題，電腦便則較不容易透過固定的結構規則來達成。

實證結果也顯示，透過電腦產製的試題能達到試題產製時間及試題變化的效益，所產製的試題也能符合試題編製原則要求、避免同一試卷中出現評量相同知識概念的重複試題、提供包含 Bloom 分類資訊試題、透過知識本體的知識連結可產生新概念試題等效益。此外，老師們也認為電腦產製試題能協助教師發展一般常見的基本層次測驗試題，因此老師們有相當高的意願利用此系統來協助其產製大量題庫及選取試卷題目，而教師所節省的時間，則可進行更高層次的試題發展。

第二節 本研究對於教育領域的影響與意涵

本研究透過知識本體、語意分析及 Bloom 理論的整合，讓電腦具備輔助教師產製符合 Bloom 理論概念的試題。此外本研究的智慧評分機制亦利用人工智慧、詞彙語意、模糊理論的整合，消除線上測驗中填充題型所造成的評分等化影響。因此本研究的成果對

於教育領域的可能影響與意涵，將說明如下。

- 一、測驗等化研究領域的突破：過去三十年來教育領域針對測驗等化的議題，以施測階段的評量等化研究為主，這些研究強調不同評量方法的等化效力問題，而並未針對測驗階段的結果進行等化研究，因此本研究在測驗評量的等化研究上做了一些突破。
- 二、線上測驗環境需確認已排除電腦環境所造成的評分誤差：此研究證實在包含填充題的測驗中，一般電腦的評分結果，與傳統紙筆測驗的評分結果並不相同，而造成評分效力差異的來源主要包括電腦環境介面、可替換之相似詞彙、填充格的集合順序、學生中文程度低落及隨機性漏字疏忽等五個來源。本研究嘗試利用人工智慧、詞彙語意關係、模糊理論等技術發展一個智慧評分機制，研究結果也證實透過本研究的智慧評分機制，能消彌電腦環境所造成的評分誤差。因此數位學習的業者及學術界的教師在建構電子化學習環境時，應重視線上測驗是否與傳統紙筆有相同的評分效力，並確認已排除電腦環境所造成的評分誤差。
- 三、試題編製過程將由教師為主電腦為輔，轉移成電腦為主教師為輔的角色：傳統教師編製測驗試題主要包括測驗設計、擬題、預試、選題、編排、建立常模等過程，電腦科技在過去通常只做為完成此過程的介面輔助工具，實際的試題擬題開發工作，仍是由教師親自負責。而本研究嘗試發展能夠協助教師執行測驗設計、擬題、預試、選題、編排試題等過程工作的系統，而研究成果也顯示，本研究透過 Bloom 理論、知識本體及語意分析技術，能夠產生大量的題庫試題，每個試題均能依照原使知識教材內容，賦予評量特定知識類型及認知能力層次的測驗試題，過程中教師除了仍需要針對電腦產製的試題進行基本的檢視與篩選試題工作外，即使連預試的工作，也可以透過線上自學者針對課程學習後的自我評量來取得每個試題的難易度及鑑別度計算結果，再由電腦即時進行試題品質自動管控，一旦某些試題已經出現無法產生鑑別力的情形時，則可由電腦進行試題的自動列管，以保持系統內的試題品質。一旦資訊科技能協助測驗編製過程的程度越來越成熟時，未來教師將有可能只需要進行人工的試題品質初步審查工作，其餘所有試題編製工作及後續試題分析及試題難易度及鑑別力的計算，皆可藉由電腦線上即時處理的特性來隨時更新及管理。
- 四、適性測驗在實務教學應用上難以實踐的主因，可透過本研究的成果來加以解決：適性測驗是根據學生的個別差異，設計最適當的測量方式及題目類型，以便獲得學生真正的學習成果。雖然近年來許多研究都指出，適性測驗有許多相對於古典測驗的優點，但是由於適性測驗需求龐大的題庫做後盾，並且必須讓每個試題包含試題反應理論中所需要的參數資料，諸如每個試題的難易度、鑑別力及誘答力等，其中光是要教師在線上課程學習環境中，針對每一個課程單元及章節內容，提供大量的題庫來進行適性測驗，基本上在實務上都是難以實現的。而本研究則能針對線上教材中的每個課程單元及章節內容，輔助教師產製大量的題庫供系統使用，因此使得適性測驗也能應用在平時的課程學習中。本研究將使得適性測驗不僅只適合應用在國家級或專業機構舉辦的能力鑑定測驗中，也能被應用在教師個人教學的課程學習測驗輔助中。
- 五、以教材與試題合一為基礎，產生覺察導向(perceived orientation)效果來輔助學習者完成知識學習：從建構理論的學習觀而言，教師並無法單純地將知識灌輸給學生，學生必須靠自己建構知識。而網路學習環境，符合以學習者為中心的學習行為模式。

此外，在 e-learning 的自學環境下，學習者往往是以自行研讀方式學習，由於沒有教師在旁指引，因此能夠依照自己的需求及學習方式來選取學習路徑，但是也容易產生學習者在超空間迷失(lost in hyperspace)的問題，Hill & Hannfin 等人(1997)也指出當學習者出現超空間迷失現象時將有礙學習。此外，學習者是否能完成所有線上課程教材內容的學習，是否有部分重要的課程教材遺漏學習，則可透過測驗活動及具有覺察導向的機制來協助學習者完成學習目標。本研究中電腦輔助教師產製的試題，均保留出題來源教材的連結資訊，因此當學習者在超空間迷失後，可能導致部分知識被學習者忽略。透過本研究發展的試題，將使得學習者在測驗回饋中，由系統提供試題與出題來源教材的超連結及回饋資訊，來進行補救學習。此外，若要使電腦能提供更有效的察覺導向能力，則線上課程知識學習系統可以記錄下學習者的學習歷程，並檢查學習者在每一個課程知識項目中是否已完成點選學習行為，並由系統給予學習引導的輔助。

- 六、協助非教育系背景教師釐清試題的測驗目的並發展符合 Bloom 理論的試題：就教育測驗的功能而言，測驗編製的過程，可協助教師釐清教學目標。傳統的測驗編製，需要教師投入大量心思來使得教師能將測驗試題與教育目標進行連結，以期能使每個試題都有明確的評量目的。然而現有國內教育體制，除了高中、國中、國小教師通常需具備教育領域相關之教育學分外，許多大專院校專業系所之教師，並未具備任何教育領域的課程編製及測驗理論等知識，因此所發展的試題，往往缺乏明確的測驗目的，也未與教學目標相結合，此外試題也未具備 Bloom 的概念資訊。而本研究使電腦能夠輔助教師進行試題的編製，而且電腦編製的過程則以 Bloom 理論為基礎，並根據教材內容產製具有評量特定向度目標的試題，因而透過電腦編製試題的過程，電腦可直接協助非教育系之專業教師，產製具有明確評量目的 Bloom 概念試題。
- 七、以評量相同知識概念的多試題方法來達成行為主義中所稱之練習律：形成性測驗是用來測知學生進步情形，僅包括教學的有限部分，如一個單元或一章，並用以測量此部分的重要學習結果。其目的是在了解學生對於所學功課的經熟程度，並提供回饋讓學生瞭解自己學習尚未達精熟之處。在自學環境中，行為主義心理學中的立即回饋的提供，及增強理論的應用等項目，可透過電腦科技來強化其在網路學習環境中的應用。而嘗試錯誤的學習方式在行為主義中也是一個非常重要的學習理論概念，但無論是在傳統教學或是電子化學習環境，課程設計者若要針對每一個知識概念，設計許多不同題幹與題型試題是不易達成的，因而也難以針對相同知識概念提供不同的試題讓學習者反覆練習，以加強對知識概念的熟悉。本研究針對每一個知識概念能產生數個不同的題幹內容及題型試題，因此當學習者對於特定知識概念不熟悉時，本研究的試題可配合行為主義心理學中的立即回饋概念，直接指引學習者進入與此試題概念相關的知識教材內容中，並進行知識概念的增強及補救學習。而補救教學的內容則依據試題中所記錄的原始出題來源教材來產生，在補救教學的內容中亦可使用本研究對此知識概念所產生的其他多組試題來重新評量學習者是否已真的瞭解，並能藉由測驗回饋訊息、補救教學、特定知識的多組測驗試題反覆練習，以達成行為主義中所稱之刺激與反應連結。
- 八、電腦提供非分數化之回饋資訊可提供自我概念瞭解的學習反思機會，並能增進學習

動機與成果：華人社會往往會以學生成績分數高低，來評價學生的努力與學習成果，因此測驗引導教學似乎也逐漸成為許多學校或補習班教師的普遍現象，而學生也容易開始將取得測驗高分視為學習的主要目標，而模糊了知識學習的過程與意義。本研究透過電腦科技的輔助，賦予每個試題及智慧評分時，系統提供的測驗回饋不只是測驗分數及簡單的答對或答錯訊息，而能提供出題來源教材的知識連結、學習者在各認知向度及知識向度的學習表現、答案被系統評為錯誤的原因解釋等回饋資訊，使得學生能明瞭自己在哪些知識內容不熟悉，對於哪些認知能力向度的學習表現較差，並進而針對本身學習過程及學習方法進行反思，以改進自己的學習方式，並激勵學生能將學習不足之處進行補救學習。

- 九、以「人本」及「適性」為中心的因材施教與因材施教理想，將會隨著資訊科技的輔助，而逐漸成為可行的主流：由於在傳統教學環境中，一位教師需要面對許多不同的學生，雖然每一位學生的程度與學習情況不同，但是教師仍是無法針對每一位學生，提供不同的教學方法與授課內容，當然也無法針對每一位學生適合其程度的試卷，以避免題目超出學習者程度，導致學習者過於挫折，而影響其學習動機。因此在教材設計上可透過網路學習方式讓學習者進行自我選擇學習路徑的非線性學習，甚至提供不同難易度的知識教材。而在試題評分設計上，電腦也不該再以教師的觀點來設限標準答案的可能範圍，而需要能接納及擴充學生可能想出的其他亦可接受的正確答案內容。而在測驗試題的設計上，則可透過大量不同難易度及鑑別度的題庫試題支援，進行適性測驗或是提供不同測驗程度的古典測驗試卷，避免試卷內容過份打擊學習者信心，或讓學習者過份高估自己能力。
- 十、傳統教師的角色及人數將因為資訊科技在學習輔助能力的精進，而有所衝擊：假設在傳統學校中，同一個年級有 10 個班級需要上「歷史」課程，全校可能就有 5 個老師需要參與授課，然而 10 個班級的學生由於授課教師能力不同，所學得的知識成效也會有明顯不同。今日若是將此「歷史」課程委由專家來製作線上課程教材及測驗評量試題，並提供所有學生利用網路直接學習，則將會造成許多教師其教學工作中的許多事務被電腦直接取代，特別是教師「傳授知識」、「編製試題」及「評量學習成效」的角色比重也將逐漸降低，因此該學校聘請的歷史老師人數似乎也將會因為教材 e 化、試題 e 化及評量 e 化後，而大幅減少。而此現象將不只在歷史課程發生，在任何課程都會面臨此一情形，而未來的許多課程，可能都是由領域專家製作好線上課程內容，再由電腦自動產製試題並自動評分。而教師在教學過程中，也將從主要的教學者角色，而轉換成輔助學習者學習的角色。而試題的編製也不再需要大量教師共同參與製作大量的題庫，電腦本身便能依照教材內容產製大量具有結構規則的試題，教師僅需要輔助線上學習系統，再編製其他較為靈活及需要深層的測驗試題。

第三節 適用於線上測驗環境的試題品質及評分等化能力評估概念模式

一般而言試題的命題編寫與學生作答結果分析(試題反應)的效度化是測驗發展的一體兩面(Haladyna, , 1999)，此二個構面也是評量試題品質的主要依據。因此本研究綜合成果及結論，本研究提出一個適用於線上測驗環境的試題品質及評分等化能力評估模

式，此模式涵蓋試題編寫及試題反應所延伸之試題品質，如圖 69 所示。在電子化學習環境的線上測驗系統中，由於學習者會於課程自學後進行學習成效的自我評量，因此電腦需具備即時評分及立即回饋的能力，所以本研究以單選題、複選題、是非題及填充題等常見的線上測驗題型進行研究。在受限於電腦可自動評分的單選題、複選題、是非題及填充題等題型限制下，無論是教師人工出題或是電腦輔助教師產製試題，均無法進行「創造」認知能力層次的試題，因此在評量線上測驗試題的內容品質時，需要確保每個試題皆具有明確的評量目標，然而測驗試題的評量目標可以利用 Bloom 分類理論來結合教學目標，使教學與評量能具有一致性，然而因為受限於線上測驗系統能夠自動評分的題型限制，因此圖 69 中，試題評量目標所能涵蓋的知識向度僅包含事實、概念與程序知識，而試題評量目標所能涵蓋的認知向度則僅包含記憶、了解、應用、分析與評鑑層次，「創造」層次則未包含在線上測驗試題內容品質的評量範圍，這與教師對於包含問答題型測驗的傳統紙筆人工評分的試題品質評估範圍不同。此外，圖 69 也顯示教師個人的教學年資經驗、課程領域授課經驗、教育背景(教育學分)、個人喜好與認知等因素，可能會對線上測驗試題內容的品質、產製試題一般能力、線上測驗試題編寫品質產生影響；其中影響到的產製試題一般能力可能包括了產製不同題型試題、產製大量變化試題、試題涵蓋重要知識及試題涵蓋高廣度知識等能力；而影響到的線上測驗試題編寫品質，主要包括了試題發展原則中的題幹品質及選項品質。而線上測驗試題內容品質、線上測驗試題編寫品質及產製試題一般能力等三項構念因素，也將直接影響到線上測驗試題的品質。

過去的文獻指出與試題或試卷本身相關的品質包括(師大書苑，民 83)：

1. 適切性(relevance)：所選用的問題類型，要能適切地測量所其問的學習成果
2. 平衡性(balance)：每一教材領域所包含的試題比率，要符合細目表所預定的比率。
3. 有效性(accuracy)：試題的編製、評分及學生作答的時間要經濟。
4. 客觀性(objectivity)：試題要明晰，答案要明確，只要是該領域專家，應可得到滿分或接近滿分。
5. 特殊性(specificity)：正確解答試題一定要具備該領域的特殊知識，若未具備該領域知識，則僅能憑猜測來作答。
6. 適當難度(difficulty)：試題難度適中，不太難也不過份容易。
7. 良好的鑑別度(discrimination)：即試題要具有區別高低學習成就者的能力。
8. 信度(reliability)：測驗結果要有一致性及可靠性，不因時間及內容之不同，而有太大改變。
9. 效度(validity)：編製選用的工具要能真正測得所欲評量之特質或學習結果。

而線上測驗試題的品質依照文獻及本研究的實驗結果，共歸納成試題平衡性、同概念試題充足性、有效性、客觀性、適當性、特殊性、試題信度、試題效度、適當難度及高鑑別度。本研究中電腦輔助教師產製的試題品質，主要藉由系統控制的方式來控管線上測驗試題內容品質、線上測驗試題編寫品質及產製試題一般能力對於試題品質的影響，此外本研究也利用教師直接針對試題進行試題有效性、適當性、客觀性、特殊性及同概念試題充足性的評估，而本研究也另外針對電腦隨機抽取的 50 個試題進行試題的信度及效度檢驗說明，並能具有足夠的信度與效度。而在試題平衡性部分，本研究則提供一個「電腦輔助教師自動選題」功能，電腦能依據教師所設定試題需求，由電腦隨機抽取符合條件的試題，以達到試題平衡性及試題評量概念不重複的目的。而試題難度及鑑

別度因為需要先透過大量樣本進行試題的施測評量後方能計算取得，因此需於後續研究中持續進行。而另一方面，試題品質不僅受到試題本身編寫的內容品質影響外，也會受到受測樣本對於每個試題反應的影響，當受測學生程度普遍較高或較低時，對於試題品質的鑑別度與難度影響將會很大。此外，學生的疲累程度、情緒，以及試卷長度、試題重複出現率、試題存放時間長度都可能影響受測者對於試題的反應結果，並間接影響了試題難易度及鑑別度結果，進而也會連帶影響試題的信度。而填充題因為在線上測驗系統中需要由受測學生透過電腦鍵盤輸入答案，因此電腦是否能正確的對於填充題答案內容進行正確的評分，是否電腦的評分機制具備與傳統測驗相同的評分效力，這都會影響到試題難度與鑑別度的計算，並間接影響了試題品質。

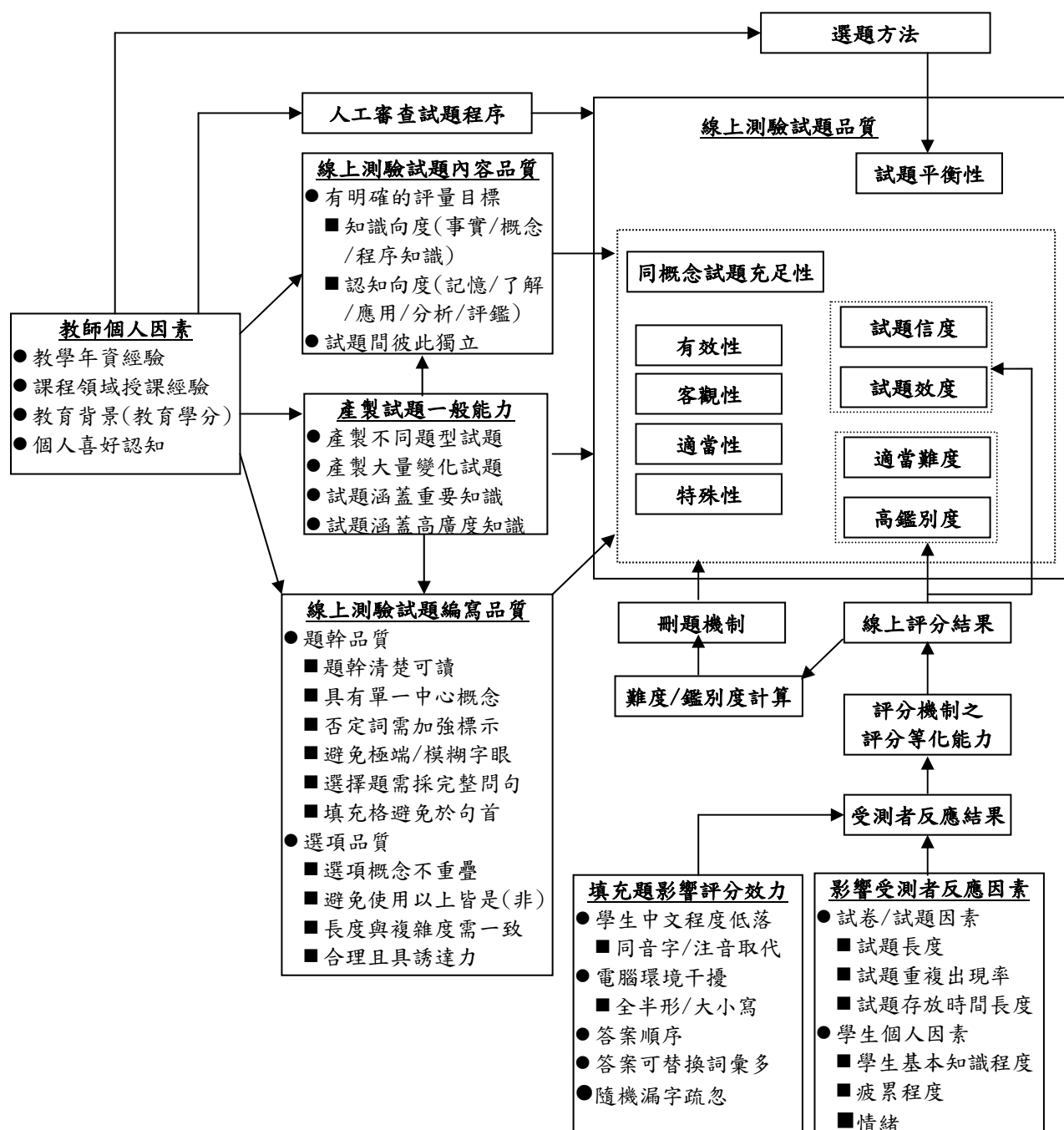


圖 69: 適用於線上測驗環境的試題品質及評分等化能力評估概念模式

因此，本研究透過輔助教師產製測驗題庫子系統，結合了 Bloom 理論、各類題型試題編製原則、語意分析、本體論及人工智慧等技術，輔助教師產製具有線上測驗試題內容品質、線上測驗試題編寫品質及產製試題一般能力的試題，此系統對於線上測驗環境試題品質及評分等化能力的成效輔助如圖 70 所示，本研究的輔助教師產製測驗題庫子系統，能協助教師提升編製試題的品質。而針對受測者反應部分的試題品質，本研究利用智慧型評分機制，來消彌線上測驗中填充題型所可能產生的評分效力影響，並使得試題的難度及鑑別度不會因為評分效力的影響，而造成難度、鑑別度的計算錯誤，以及產生錯誤的試題品質評估結果。

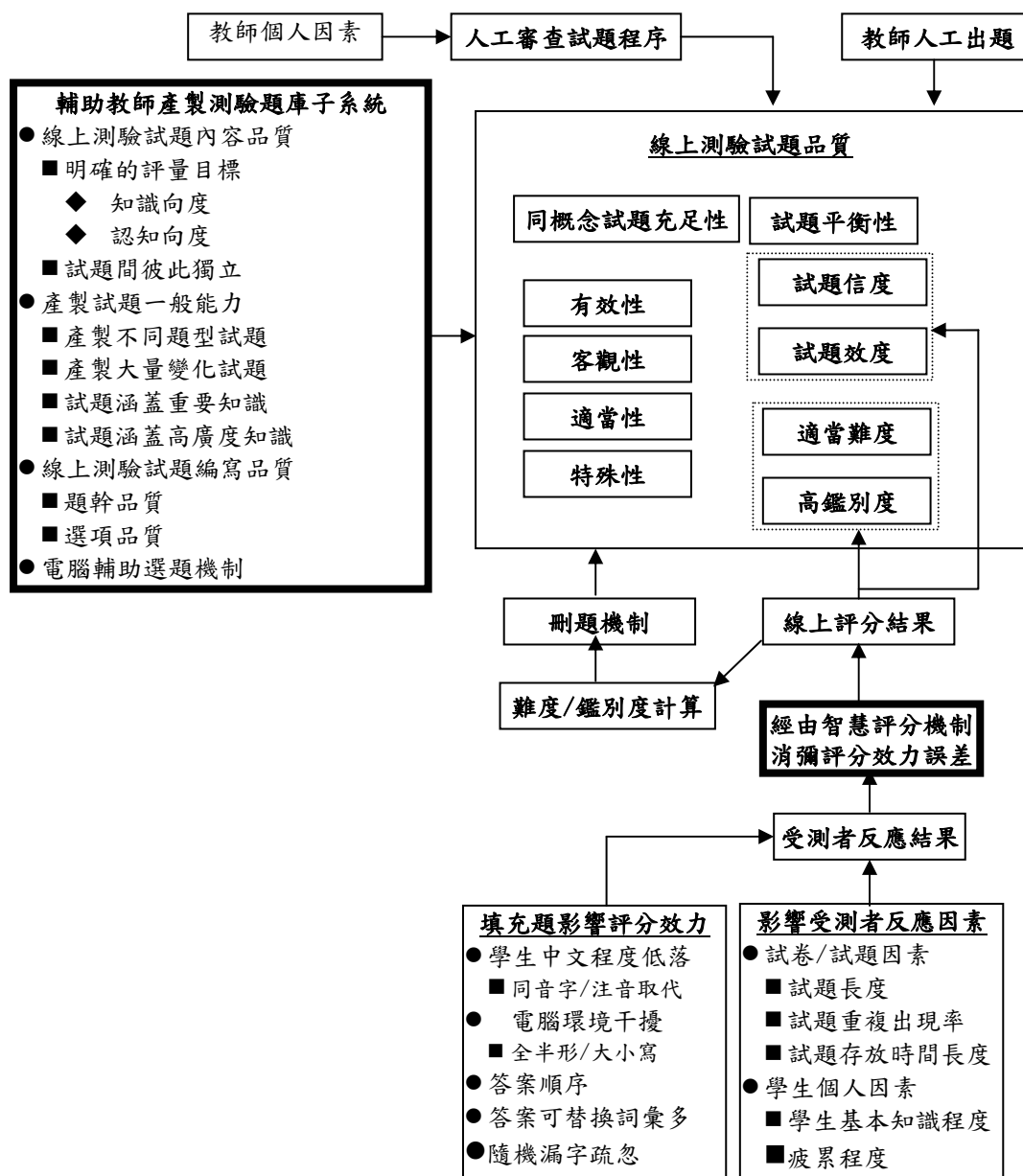


圖 70:本研究之系統對於線上測驗環境試題品質及評分等化能力的成效輔助

第四節 未來研究方向與建議

本研究的未來研究方向與建議如下：

- 一、由於研究目的與時間的限制，本研究初始儲存的教材知識，均先以人工的方式，並依照教材知識本體的結構進行知識儲存，未來研究將使系統能自動進行教材內容的斷詞、知識萃取及分類儲存的工作。
- 二、由於一般的個案實例類型知識，通常包含過多的情境背景及隱藏的意涵，因此若要透過語意分析技術萃取出知識背後所隱藏的深層意涵實屬不易，目前本研究僅能對於一般教材知識內容的文法結構及知識關係進行辨識分析。另外，某些透過多媒體影音及圖表附件來呈現知識內容的教材知識，因為其知識儲存的結構方式並非以一般的文法結構來儲存，因此也不適用於語意分析技術，所以本研究未針對這些類型知識進行試題產製，未來研究可針對情境分析及多媒體知識的方向進行試題研究。
- 三、本研究輔助教師產製的試題結果，雖然可以涵蓋 Bloom 認知目標的第一層分類概念，但尚無法完全產製出涵蓋 Bloom 知識向度及認知向度第二層子類別的試題。未來研究可針對 Bloom 分類第二層的內容，透過知識本體結構的修改，以及試題產製規則的設計來努力。
- 四、本研究因著重於智慧評分機制及輔助教師產製試題的方法設計，關於測驗系統對於學習成效影響的大樣本實證研究工作，將列入未來的研究方向。
- 五、雖然參與實驗教師認為本研究所產製的試題，能涵蓋一般教師平時所產製評量基本知識概念的試題，而且對於試題的合適性及可用性也給予高度評價，但是電腦產製試題的鑑別力及難易度，因為需要經過大量學生來對於這些試題進行實際測驗後，方能計算出每個試題的難易度與鑑別度，其另需花費冗長的研究時間，因此將於後續研究工作中再進行實證。
- 六、本研究的電腦輔助教師編製試卷功能雖能依照教師對於「題型」、「知識類型」、「認知層次」三種概念自動選出符合教師需求的試題，但為了能考量教師的實用性考量，本研究將在後續研究中透過試題實際施測後，計算每個試題的鑑別度與難度資訊，再整合「配分」、「主題範圍」、「題型」、「知識類型」、「認知層次」、「題數」、「難度」等概念，輔助教師更彈性且更實用的進行電腦自動選題的作業，並針對每份試卷給予總難度的資訊標示。
- 七、由於中文語法結構變化多元，目前系統所設計的試題產製規則並無法包含所有中文環境中的規則結構，因此本研究仍須隨著系統的測試與實驗，進行試題產製規則的發展與修正，以使得所產製的試題品質能更佳提升。
- 八、本研究對於教材知識本體及試題產製規則的設計，是以華文的課程知識學習環境為基礎。因此僅能以中文的語法結構進行知識內容的分析及試題產製的工作，尚無法對於英文文章內容進行內容分析及英文試題產製。未來研究可考慮結合英文文法規則及結構，來達成英文文章的分析及試題產製工作。
- 九、在建構型試題的研究上，本研究目前僅完成填充題的詞彙語意分析研究，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，並能更有效的輔助學校或企業從事教育訓練及知識管理等工作。
- 十、本研究因著重於電子化學習環境的學習過程測驗，並藉由測驗活動來輔助學習，與

近年試題反應理論及電腦適性測驗著重單一時間點的能力鑑定不同，因此本研究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。

- 十一、由於目前市面上的中文輸入法眾多，本研究中的模糊評分機制僅使用「注音輸入法」來探究輸入法對於評分效力的影響及改善方法，因此其他輸入法可能造成的評分效力影響，在未來的研究系統中可擴充加入不同的輸入法，並將各種輸入法之編碼資料建置於系統中，以持續提升線上測驗系統的評分效力。
- 十二、本研究目前僅針對系統內建之相似語意詞庫相似值，採用模糊理論技術推論產生，推論過程結合系統內所有教師所設定的個人化相似語意詞庫相似值，因此能聚合出高共識且高品質的系統內建相似語意詞庫相似值。而評分環境參數由於較不易產生語意上及認知上的模糊，因此本研究系統內建的評分環境參數，僅透過系統內所有教師的個人化評分參數進行加總平均而產生，未來在系統內建評分環境參數值上，亦可嘗試採用模糊理論技術來產生。
- 十三、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。

第五節 研究貢獻

本研究的研究貢獻歸納如下：

- 一、過去 30 年來教育領域針對測驗等化的議題，以施測階段的等化方法研究為主，而並未針對評量階段的結果進行等化研究，因此本研究在測驗評量過程的等化研究上做了一些突破。此研究也證實一般電腦的測驗結果，與傳統紙筆測驗的結果可能不同，因此數位學習的業者及學術界的教師在建構電子化學習環境時，應重視線上測驗是否與傳統紙筆有相同的評分效力，並確認已排除電腦環境所造成的評分誤差。
- 二、本研究在電腦輔助教師產製試題方面，以詞義關係、本體知識、中文文法結構來進行研究。本研究的雛形系統在初期研究成果中，已可獲得許多輔助人工出題的的效益，除了降低教師人工出題的負擔外，也能輔助教師產製涵蓋評量事實、概念及程序等三種知識向度及記憶、了解、應用、分析及評鑑等五種認知向度的試題。從教育領域而言，本研究的貢獻主要在於結合了 Bloom 教育目標分類概念及試題出題原則，使電腦能提供具有實質教育意涵的測驗試題。
- 三、本研究的研究結果顯示，透過電腦科技可以更有效率的產製出大量且變化的基本試題，實際減輕教師產製試題的負擔，而教師可將節省下來產製試題時間，用來設計及發展更高層次的試題。
- 四、本研究提供「電腦自動選題編製試卷」功能，教師可快速依據所需要的試題知識向度、認知向度或測驗題型，分別編製成不同測驗目的之試卷，以增加教師進行各種評量的彈性。
- 五、經由教材知識本體的設計，原本非連續的教材知識間可以依據其所屬之知識類型關

係結構來相互連結，使得教材知識不再只能以原有的知識概念呈現，透過知識結構的連結，便有可能挖掘出新的知識概念。而透過試題產製規則的設計，系統也能依據教材知識本體結構及知識內容，產生新知識概念的試題。

- 六、本研究的雛形系統架構設計，使得測驗試題能直接從線上教材中萃取，並與知識教材緊密結合，因此系統輔助教師產製的試題，便能與原始系統內的原始教材知識內容進行連結，當受測者在答錯試題時，系統便能輕易的透過這樣的測驗回饋連結，直接讓受測學生連結至該題目相關的原始教材內容中進行知識補強的學習，如此將能帶給學習者更大的學習輔助。
- 七、經由本研究教材知識本體的設計及測驗產製的成果，可顯示出將教材以有意義的方式儲存，並能儲存經過斷詞後的原始教材知識內容及詞性標記，將能使得電子學習環境的發展及應用更加多元。
- 八、本研究在線上測驗評分效力及智慧評分機制的研究方面，結合模糊邏輯、資料庫、人工智慧等技術，並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，提供不同教師在共同的測驗平台環境中共享資源。過去使用模糊機制之測驗評量研究，主要在探究原始成績、題目難度、學習程度轉換為不同等級歸屬函數，並藉以發展相關之智慧系統。然而，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，本系統的確能具備教師人工評分時的評分風格，並視學生答題內容的正確程度或相似程度，「斟酌」給予分數。
- 九、透過智慧評分機制設計，解決填充題型造成評分效力不佳的問題，並提供政府及教育單位在考慮將填充題型納入線上測驗中的可行性參考。
- 十、本研究藉由電腦智慧評分機制及輔助產製試題的雛形系統設計，解決電子化學習環境中線上測驗系統的許多缺點及問題，並讓電腦能具備一定程度的智慧能力，協助測驗的模糊評分及輔助教師產製試題，因使能提升線上測驗系統的能力。
- 十一、針對電子化學習環境，提出適用於線上測驗系統的試題品質及評分等化能力評估概念模式，以強調 Bloom 分類理論在線上測驗系統可自動評分前提下的使用範圍與限制。此外，此模式也透過測驗等化觀念，來強調線上測驗環境評分效力對於試題品質的評估影響，因此本研究也針對教育領域測驗理論在評估試題品質上提出新的觀點。