

建置網路學習小組議題報告電腦化評量機制之研究

尹玫君* 蘇彥寧**

近年來，隨著網路學習的蓬勃發展，學習者參與教學活動所產生的學習歷程檔案，除可使學習者檢視自身學習歷程外，也提供教師於評量學生學習成效時一個重要的參考依據。在網路學習的環境中，經由資訊技術的導入，已可實現系統自動記錄學習者學習歷程的目標，包含學習時間、登入次數、文章點閱數及小組議題報告等資料皆可完整保留。故能累積為數眾多的歷程檔案，可供教師瞭解學生的學習狀況，並進一步作為評量學習成效之用途。

本研究旨在發展一網路學習小組議題報告的電腦化評量機制，以解決傳統教師人工評量可能耗時費力或評量標準未能一致的問題。研究者參考網路學習小組議題報告的評量規準，並透過中文潛在語意分析與決策樹演算法建置電腦化評量系統。待系統發展完成，即經由比較電腦化評量與人工評量在評量時間耗費及評量得分結果的差異，以及邀請領域專家進行系統評估，作為驗證電腦化評量系統之方法。研究結果顯示：

- 一、電腦化評量的評量時間顯著低於人工評量；
- 二、電腦化評量與人工評量的評量結果無顯著差異；

三、電腦化評量系統的「系統介面親和度」、「系統功能有用性」、「系統整體滿意度」皆獲領域專家一致的肯定。

關鍵字：電腦化評量、小組議題報告、決策樹、潛在語意分析、網路學習

* 作者現職：國立臺南大學教育學系教授

** 作者現職：國立成功大學工程科學所資訊組博士生

通訊作者：蘇彥寧，e-mail: yenning@mail.tn.edu.tw

壹、前言

近年來隨著網際網路技術的發展與普及，網路學習的應用已日趨廣泛。一般而言，網路學習可分為同步學習和非同步學習，相較於傳統教學以同步式為主的教學活動，網路學習無論在時間及地點上，皆提供更多的彈性及自由度，並具備非同步、多方向、個別化及自動記錄學習歷程等優勢 (洪明洲, 1999)，以迎合不同學習者之需求。從教師的角色而言，網路學習的運用，可將傳統以教師為主灌輸的方式，轉化為以學生學習的要求為主，方便同儕間的互動及成長的學習方式 (陳年興、楊錦潭, 2006)。是故，對於教學者而言，如何運用網路科技及數位學習的優點，以提升學習成效 (趙美聲, 2003)，並應用其妥善規劃數位化的教學和評量策略，是一項重要的議題。

透過資訊技術的導入，網路學習環境可運用許多互動媒介，如文字、圖像、語音、影像等 (顧超光、郭世明, 2005)，並能以上述媒介為基礎，衍生相關的網路互動工具，如網路討論區、隨選視訊 (VOD)、線上同步教學影音等。因此，對於教學者及學習者而言，網路學習提供了多元的互動環境，並可打破時間及空間限制，使教與學具備更高的自由度。而學習者可透過網路學習平台所支援的功能，與同儕間進行同步、非同步的互動、交談、溝通及討論。另一方面，基於網路學習環境的同儕互動，除可減少來自同儕的壓力，亦能提供學習者間「同時不同地」及「不同時不同地」的互動優勢，加上非同步式的議題討論，學習者可以有較充裕的思考時間解決傳統課堂上，學生表達意見時片斷、簡短的問題 (陳年興、楊錦潭, 2006)。

在網路學習的領域中，即時影音互動及議題討論區，提供一個同步及非同步學習的方式，供學習者討論在學習過程中所遭遇的難題，或針對特定一主題進行相關的探討活動。蔡昂叡 (2007) 認為，學生參與議題討論的行為，是老師瞭解學生學習歷程的一個重要依據。網路學習的議題討論，可提供學生較為充裕的思考時間，可促進學生的思考、整理及深化過程。在討論過程中的互動，能讓學生有相當程度的參與感，並進一步與同儕互動與交流想法。另一方面，教師對於學生參與討論的內容及表現，應可考慮對其進行評分 (陳年興、楊錦潭, 2006)。然而過去的研究，多傾向從網站的紀錄中(如：總閱覽文章數、發表文章數、提出問題數、登入登出次數等)，分析學習者的學習行為及成效。曾憲雄、蔡秀滿、蘇東興、曾秋蓉 (2006)認為此類研究用來判斷學習者所歸屬的資訊過於短少，且應用的學習歷程分析方法過於簡單，因此在分析學習者特質與能力上，成效相當有限。相對於參與議題討論行為，網路學習小組議題報告為學習者針對小組討論的內容記錄加以分析、整理、修正並提出看法，為討論內容統整、歸納之綜合成果。故針對網路學習小組議題報告的內容進行分析，應能作為教學者於評量學生學習成效時的一項重要參考依據。

然而，針對網路學習小組議題報告的評量方式，以往教師皆以人工方式進行評分，並藉此瞭解學生學習的成效。倘若學生分組的數量眾多，或報告篇幅較長，難免造成教師在評量上之困難。針對此項問題，王信智 (2001) 認為可能遭遇兩種類型的困難：1. 評量費時；2. 計分困難。對於教師而言，充斥於網路上的大量資料，迫使其必須花上較多的時間審視資料的好壞及符不符合需求 (劉佳宗, 2007)，難免造成人力及時間上的耗費。黃彥博 (2008) 亦認為，教師在評分過程中若倚賴人為的判斷，則評分標準有時難以維持一致，故評量的信度常受到質疑。電腦自動化計分的優勢在於可解決人工計分的耗時費力及人為判斷的誤差 (黃彥博, 2008)。故若能以公平及快速的方式進行評分，將可協助教師評量網路學習小組議題報告。為有效解決上述問題，研究者運用中文潛在語意分析、決策樹演算法等技術發展一「網路學習小組議題報告電腦化評量系統」，以期節省教師評量網路學習小組議題報告所需耗費的時間，並達成以一致的評分標準進行網路學習小組議題報告評量之目的。

貳、文獻回顧

一、小組議題討論學習活動

孫春在和林珊如 (2007) 認為，小組成員間的互動過程是決定學生學習成果的一個關鍵。高台茜 (2003) 指出，教師將分組討論活動運用於教學中，可藉由討論活動使教學更為生動有趣。在其過程中，同儕互動對於學習成效的影響主要取決於同儕互動品質、學習動機、態度與人際關係的技巧。陳志憲 (2004) 指出，在同儕相互辯詰的過程中，歷經了共同思考、問題解決與決定的互動歷程，可使學生對於討論議題有更為深刻的瞭解，且可培養其反省與批判思考的能力。高台茜 (2003) 進一步指出，若要促成同儕間良好的互動，則應透過回應、提供解釋、提供學習資源、評論對方意見、鼓勵小組成員等互動行為，才能協助同儕建構知識以解決問題。因此，教師若能促成學生間良好的討論及互動，則對於提昇小組的學習成果將可有所幫助。

尹玫君 (2009) 指出，網路非同步的討論環境，較傳統教室的同步討論，可以提供學習者較多的思考時間與機會。陳年興和楊錦潭 (2006) 認為網路學習的議題討論，可提供學生較為充裕的思考時間，可促進學生的思考、整理及深化過程。在討論過程中的互動，能讓學生有相當程度的參與感，並能進一步與同儕互動與交流想法。尹玫君 (2009) 指出，學習者參與討論活動是提升教學和學習成效的要素之一。葉寶霞 (2004) 提及，網路學習環境的同步與非同步討論可促進同儕間互動，並在發表看法時增加成就感。而討論內容的整理，可以精緻化呈現別人的經驗，促進自我觀念的統整，有助於學習。Hara, Bonk & Angeli (2000) 曾透過其研究發現，大專以上學生所討論的內容多使用深層的認

主題文章

知歷程，且能援引同儕的發表內容進行討論。另外，在互動的過程中，學生小組多能使用高層次的認知策略(如：推論、判斷)，及後設認知策略(如：自我反省及自我察覺)進行互動(引自孫春在、林珊如，2007)。因此，經由上述文獻回顧可得知，小組議題討論可有助於小組成員間的互動。在互動的過程中，學習者透過網路學習環境的非同步與同步溝通工具，共同進行知識的建構、分享學習資源、交流想法意見、彼此評論答辯，進而提升小組成員的學習成效。

小組討論活動結束後，小組成員即可依據互動過程所討論的內容，進行統整與歸納，並藉此彙整為小組議題報告。黃政傑和林佩璇(1996)指出，小組報告的功能主要在呈現小組合作的成果，可易於教師檢視學習者的學習成效。而小組議題報告是為同儕間互動與討論所歸納之綜合成果，因此對於呈現小組成員的學習成效，即具備完整的代表性。

二、小組議題報告評量

評量的目的在於判斷教學活動是否達成目標、瞭解學生學習實況與作為改進教學促進學習的參考(方欽鴻，2007)。因此，評量在教學過程中佔有相當重要的地位。透過教學評量的實施，可瞭解教師教學目標達成程度、教師的教學效率，以及診斷學生的學習狀況與困難(楊銀興，2000)。尹玫君(2009)指出，評量是教學者和學習者改變或革新教育最有效的驅動力，與有效的教學和學習有直接的關係。Hoerr(2000)指出，對於學生而言，評量提供他們表現的回饋，使學生可以增強其內省的智慧(陳佩正譯，2001)。

以往，在評量學生網路學習的學習成效時，有研究針對學生的網路外在行為進行評量與探討，如劉惠如(1999)即曾以學生之網站登錄次數、線上測驗次數，以及在課程討論區文章篇數進行評量。過去的研究，亦多傾向從網站的紀錄中(如：總閱覽文章數、登入登出次數等)，分析學習者的學習行為及成效。尹玫君和劉亭言(2009)指出，以往教師僅以學習者的上站次數、討論篇數、討論字數與上站停留時間等，來評量學生的學習成效，如此是以「量」的形式進行評分，恐無法達到評量的客觀性與公平性。曾憲雄等人(2006)認為此類評量方法用來判斷學習者所歸屬的資訊過於短少，且應用的學習歷程分析方法過於簡單，因此在分析學習者特質與能力上，成效總是相當有限。陳年興和楊錦潭(2006)針對上述問題，認為可考慮對學生參與討論的內容及表現進行評分，以作為瞭解學生學習成效之基礎。尹玫君和劉亭言(2009)指出，如以學習者的外在參與行為作為認定學習者學習成效之依據，容易忽略其學習的內在價值。而網路學習小組議題報告，為學生小組互動、交流、彙整、歸納之學習成果，如能針對其內容進行評量，則應可改善僅參考學習者外在行為之不足。

在評量小組議題報告的部份，尹玫君 (2009) 曾參考有關線上討論內容分析的研究，發展出一份網路學習的小組議題報告評量表，作為評量小組學習成效的工具，如表 1 所示。

表 1 網路學習小組議題報告評量表

班級：		受評組別：						評分組別：					
評量向度	評鑑項目(高分)	評分						評鑑項目(低分)					
		高分	←	→	低分								
內容品質	1. 報告的內容能呈現深思熟慮後的觀點或整合的結果。	10	8	6	4	2	0	1. 報告的內容不能呈現深思熟慮後的觀點或整合的結果。					
	2. 報告的內容有與主題相呼應，也有聚焦焦點。	10	8	6	4	2	0	2. 報告的內容沒有與主題相呼應，且沒有聚焦焦點。					
	3. 報告的內容有相關的理論依據。	10	8	6	4	2	0	3. 報告的內容沒有相關的理論依據。					
	4. 報告的內容有理性批判與建議。	10	8	6	4	2	0	4. 報告的內容能理性批判與建議。					
	5. 報告的內容有自我檢視及反省改進的功能。	10	8	6	4	2	0	5. 報告的內容沒有自我檢視及反省改進的功能。					
語意表達	1. 報告中使用的詞句客觀恰當。	10	8	6	4	2	0	1. 報告中使用的詞句不客觀不恰當。					
	2. 報告內容段落能表達順暢。	10	8	6	4	2	0	2. 報告內容段落不能表達順暢。					
	3. 報告內容用字正確(含標點符號)，沒有造成誤會。	10	8	6	4	2	0	3. 報告內容用字不正確(含標點符號)，有造成誤會。					

資料來源：尹玫君 (2009)。應用 e-learning 合作學習的教學和評量策略之研究 (國科會專題研究計畫)。台南：作者。

在此份量表中，共分為內容品質與語意表達兩個評量向度，題項為八題。評量者可運用評量工具，就學習者所彙整的報告內容進行評量。在本評量表中，每一題項均分為零到十分。以「語意表達」向度的「報告中使用的詞句客觀恰當」為例，如評量者審閱小組議題報告內容，經評估後發現其內容詞句大多客觀恰當，即圈選較高的得分。反之，如所評估的內容詞句較不客觀恰當，即圈選較低的得分。如此，可補足以學習者外在行為(如：總閱覽文章數、登入登出次數、討論文章字數等)進行評量，而忽略學習者內在學習狀況之缺失。

三、電腦化評量

隨著資訊科技的發展與普及，電腦化評量的技術已逐漸成熟且受到廣泛的運用。若將電腦化評量運用於評估學生學習的成果，不僅可節省人工施測及評閱的時間耗費，亦能打破傳統人工施測的時空限制，使受測者隨時隨地皆可透過網路上線參加測驗，因此較傳統人工施測更具便利性 (Fuchs, Fuchs & Hamlett, 1992)。

針對現行電腦化評量系統之發展，過去已有許多研究以學習者的作業、作品、學習歷程為評量內容。經由專家系統、資料探勘、潛在語意分析等技術之使用，建置其領域所屬之電腦化評量系統，再透過相關實驗以實證系統的成效(許溫雅，2009；陳薇如，2007；黃彥博，2008；張炳雄，2003；廖聖傑，2002)。

主題文章

在相關研究所發展之電腦化評量系統架構中，張炳雄 (2003) 透過專家系統之建置，發展文書處理術科測驗的評量機制，於其系統架構中，係由使用者介面、專家診斷規則庫、計分模組等部份所組成，受評者可透過使用者介面參與文書處理術科測驗。其操作結果將會傳遞至專家診斷規則庫進行比對，再經由計分模組計算受評者所得之分數，使評量者能瞭解受評者之學習成果。許溫雅 (2009) 在其研究中，以模糊理論為基礎，發展創造力評量系統，在其系統中，使用者介面係以數位遊戲的方式呈現，蒐集參與者的操作過程，歷程中所蒐集的資料，透過匯入由模糊理論為基礎的評分機制中，藉此評量參與者的創造力指標分數。黃彥博 (2008) 以發展科學文章摘要自動化的工具為目的，在其研究中，研究者透過概念比對、潛在語意分析、關鍵詞比對等三種計分方式，發展自動化評量工具。其系統架構可分為使用者介面、評量模組等部份，學習者可經由使用者介面閱讀文章，並進行文章摘要的編輯工作。最終透過評量模組之計分，匯出學生摘要的評分結果。

回顧前述相關文獻，電腦化評量系統之架構多由使用者介面、評量模組與計分模組所組成。因此，本研究即參考此類型架構，作為實現電腦化評量系統的方法。

參、研究方法

一、研究對象

以下分別就小組議題報告及評量結果的蒐集與系統評估之對象進行敘述：

(一) 小組議題報告研究對象

本研究以某國立大學 62 位修習「資訊科技融入教學理論與實務」課程之學生為研究對象，其背景為教育學系教學科技組四年級之在學學生，皆具備基本電腦操作能力及上網能力。在課程實施時，即透過分組討論的方式，使學生自行排定組別並進行議題討論活動。而各小組所討論之主題係由該課程授課教授進行規畫與指派，使各組學生能依照教師指示進行議題內容的討論。

(二) 人工評量之研究對象

針對蒐集訓練與測試決策樹模型時所需的人工評量資料。研究者分別邀請 53 位具備教學科技背景之博士生、碩士生、國民中小學現職教師，以及大學部四年級學生，以尹玫君 (2009) 所發展之網路學習小組議題報告評量表為標準，安排評量者就研究者蒐集所得之 128 篇網路學習小組議題報告進行人工評量。人工評量實施完畢後，研究者即以蒐集所得的人工評量資料，作為後續訓練與測試決策樹模型之原始資料。

(三) 系統評估之研究對象

針對專家系統評估之部份，本研究於系統上線運作後，邀請 10 位具備網路學習等相關專長之實務與（或）研究經驗的專家學者，以及國民小學現職教師，擔任電腦化評量系統之評估委員。據以評估系統介面親和度、功能可用性以及整體滿意度，作為系統後續修正及發展之參考。

二、研究工具

為達成本研究之目的，於研究過程中所使用之研究工具有以下三項：(一) 網路學習小組議題報告評量表；(二) 軟體工具；(三) 系統易用性及滿意度評估表，茲將各種工具及編製過程陳述如下：

(一) 網路學習小組議題報告評量表

本研究所採用之「網路學習小組議題報告評量表」係由尹玫君 (2009) 所編製而成。此一量表主要分為兩個評鑑向度：內容(content)及語意表達(expression)，共計 8 個評鑑項目。研究者透過瞭解中央大學學習與教學研究所閱讀與學習研究室所發展之「Chinese Latent Semantic Analysis System」，可運用潛在語意分析技術針對文章內容的連貫性及題意適切性 (柯華葳、陳明蕾、王學誠，無日期) 進行計算。經由對照「網路學習小組議題報告評量表」的量表內容，可得知四個評鑑項目：「報告的內容有相關的理論依據」、「報告的內容有理性批判與建議」、「報告的內容有自我檢視及反省改進的功能」、「報告中使用的詞句客觀恰當」，皆屬於人為描述中帶有知識陳述、個人理解、個人反思，或敘述筆法的評鑑向度，故與潛在語意分析技術可判別的文章內容連貫性及題意適切性，存在許多差異。反之，「報告的內容能呈現深思熟慮後的觀點或整合的結果」、「報告的內容有與主題相呼應，也有聚集焦點」、「報告內容段落能表達順暢」、「報告內容用字正確(含標點符號)，沒有造成誤會」，則分別與潛在語意分析技術可判別的文章內容「連貫性」及「題意適切性」有所對應。因此，基於技術上可能造成的結果偏誤，本研究提出電腦化評量及人工評量內容適用對照表如表 2 所示：

表 2 電腦化及人工評量內容適用對照

評鑑向度	評鑑項目	電腦化評量
內容品質	報告的內容能呈現深思熟慮後的觀點或整合的結果	題意適切性
內容品質	報告的內容有與主題相呼應，也有聚集焦點	題意適切性
語意表達	報告內容段落能表達順暢	連貫性
語意表達	報告內容用字正確(含標點符號)，沒有造成誤會	題意適切性

(二) 軟體工具

本研究所使用之軟體工具，包含中央大學中文潛在語意分析系統及決策樹程式，分別敘述如下：

1. 中文潛在語意分析系統

本研究所採用之中文潛在語意分析系統，係為中央大學學習與教學研究所閱讀與學習研究室所發展之「Chinese Latent Semantic Analysis System」，該系統係透過中央研究院 2006 年所發展之現代漢語平衡語料庫，作為字詞矩陣的建置工具。而該系統所使用之語料庫具備五百萬詞，是為目前繁體中文中最為豐富的一個語料庫 (Chen, Wang, & Ko, 2008)。在本研究中，研究者透過此系統之使用，作為計算小組議題報告內容與議題之題意適切性，以及瞭解報告內容段落連貫性的工具。

2. 決策樹程式

在本研究中，研究者採用 Roach (2006) 所開發及公佈的 Decision Trees 程式，以進行決策樹的訓練工作。該程式以 Python 為發展語言，而其程式碼係遵循決策樹演算法進行撰寫，故為一正規的決策樹程式。因此，本研究即透過此套正規的決策樹程式輸入訓練資料，以發展與建立決策樹模型。待模型建置完成，研究者將可依照模型的內容，進行判斷程式的撰寫，作為發展電腦化評量系統的基礎。

(三) 系統易用性及滿意度評估表

待系統開發完成，於上線運作後，必須經由相關領域專家學者進行系統整體的評估工作，以瞭解「網路學習小組議題報告電腦化評量系統」之易用性及整體滿意度，作為檢視系統是否達成研究目的及後續程式修正、發展之參考。本研究所使用之「系統易用性及滿意度評估表」共分為三個面向，包含系統介面親和度、系統功能可用性以及系統整體滿意度，共 15 題。問卷採李克特式 (Likert) 五點量表，分別為非常同意、同意、普通、不同意以及非常不同意五個選項。各面向的得分依序為：5 分、4 分、3 分、2 分、1 分，得分越高，代表專家委員對於該系統的滿意度越高。於問卷內容效度的部分，本問卷採用專家鑑別法以評估量表之內容效度。研究者於問卷初稿編製完成後，即邀請電腦化評量、網路學習、系統開發相關領域專家，就本評量表之題意、架構、設計及文句用字提供修正意見，作為修訂評量表的依據，並依此完成本研究所需之「系統易用性及滿意度評估表」。

三、建置決策樹模型

在建置決策樹模型的過程中，研究者將蒐集所得之 128 篇網路學習小組議題報告作為樣本，經由中央大學中文潛在語意分析系統計算個別議題報告之「題意適切性」、「連貫性」數值。如此，即已取得訓練及測試決策樹模型之部份原始資料。

訓練及測試決策樹模型之另一部份原始資料為人工評量之得分。研究者分別邀請 53 位具備教學科技背景之博士生、碩士生、國民中小學現職教師，以及大學部四年級學生，以尹玫君(2009)所發展之網路學習小組議題報告評量表為標準，安排評量者就研究者蒐集所得之 128 篇網路學習小組議題報告進行人工評量。人工評量實施完畢後，研究者共回收 1147 份人工評量的評分資料，即以此作為分析評分者一致性之原始資料。如單一評分者評量遺漏值過多，則予以篩除，因此各評分項目篩除後所得人數與人工評量資料分別為：1. 「報告的內容能呈現深思熟慮後的觀點或整合的結果」51 人，共 1023 筆人工評量資料；2. 「報告的內容有與主題相呼應，也有聚集焦點」51 人，共 1073 筆人工評量資料；3. 「報告內容段落能表達順暢」50 人，共 953 筆人工評量資料；4. 「報告內容用字正確(含標點符號)，沒有造成誤會」51 人，共 1023 筆人工評量資料。篩選後的資料透過 Cronbach's Alpha 分析，以瞭解人工評分者的一致性，四個評分項目所得之 Cronbach's Alpha 值如表 3 所示。

表 3 評分者一致性考驗

評分項目	評分人數	Cronbach's Alpha
報告的內容能呈現深思熟慮後的觀點或整合的結果	51	.85
報告的內容有與主題相呼應，也有聚集焦點	51	.84
報告內容段落能表達順暢	50	.80
報告內容用字正確(含標點符號)，沒有造成誤會	51	.78

經篩除遺漏值過多的評分者評量資料，以及進行評分者一致性考驗後，各評分項目的人工評量資料數量如表 4 所示。研究者將各項目所得之資料隨機抽取 70% 為訓練資料，用以進行決策樹模型的建立；另外 30% 的人工評量資料分派為測試資料，作為測試決策樹模型之用。由於測試資料不參與決策樹模型之建立，因此其對決策樹模型而言係為未知的新資料，因此可用於模型的測試及修剪，使決策樹模型不致過度發展，且能降低模型的誤判率。

主題文章

表 4 經篩除後之人工評量資料數量

評分項目	總數	訓練量	測試量
報告的內容能呈現深思熟慮後的觀點或整合的結果	1023	716	307
報告的內容有與主題相呼應，也有聚集焦點	1073	751	322
報告內容段落能表達順暢	953	667	286
報告內容用字正確(含標點符號)，沒有造成誤會	1023	716	307

待決策樹訓練及測試資料蒐集完畢後，研究者即以 Roach (2006) 所開發及公佈的 Decision Trees 程式，匯入訓練資料以建置決策樹模型。本研究各評分項目所得之決策樹規則數，如表 5 所示。

表 5 各評分項目所得之決策樹規則數

評分項目	所得規則數
報告的內容能呈現深思熟慮後的觀點或整合的結果	50
報告的內容有與主題相呼應，也有聚集焦點	50
報告內容段落能表達順暢	48
報告內容用字正確(含標點符號)，沒有造成誤會	49

四、網路學習小組議題報告電腦化評量系統

本研究所發展之電腦化評量系統係以 Web-Based 的形式呈現。評量者可透過網路連線的方式，以個別的帳號密碼登入本系統。圖 1 為系統的登入頁面。



網路學習小組議題報告
電腦化評量系統

Computerized Assessment System for Topic Report of
Web-Based Cooperative Learning

帳號: 密碼:

指導教授：國立臺南大學教授 尹玟君 博士 系統開發：國立臺南大學教育學系碩士班 蘇盛寧

建議使用 IE 6.0 以上，解析度設為 1024x768，256 色以上 >>系統管理
本研究承蒙中央大學學習與教學研究所 何華威教授、陳明雷博士 授權使用 CLSA 程式，謹此誌謝

圖 1 系統登入頁面

建置網路學習小組議題報告電腦化評量機制之研究

登入成功後，系統即會引導評量者進入電腦化評量系統的首頁(如圖 2 所示)，並提供相關功能之連結於左方選單。



圖 2 電腦化評量系統首頁

本研究所發展之電腦化評量系統，其核心可分為：「報告單筆評量」、「報告批次評量」兩項重要功能，以下即分別就兩項系統功能進行說明。

「報告單筆評量」之功能提供評量者單次評量單筆議題報告的服務。上傳議題報告的方式依據評量者需求之不同，可分為：1.以上傳檔案的方式匯入；2.以手動填入文字的方式匯入；評量者可依自身需求選擇適合的匯入方式上傳議題報告。「報告批次評量」的功能，在於提供一個單次匯入即能評量多筆議題報告的服務。在本研究所提出之電腦化評量系統中，評量者可單次評量一至十筆的議題報告。

待評量者匯入議題報告相關資料後，系統即會自動與中央大學學習與教學研究所閱讀與學習研究室所發展之中文潛在語意分析系統進行連線，進行議題報告主題與內容間的「題意適切性」及「連貫性」數值計算。待取得中文潛在語意分析系統回傳之數值後，系統即將此份數值以四捨五入法取至小數點後第一位，再經由決策樹評量模組評定此份議題報告之相關得分，並將其匯出為分析報告，以供評量者後續之使用。

肆、研究結果與討論

本節旨在探討人工評量耗費時間與電腦化評量在評量耗費時間之差異情形。在資料蒐集的部份，研究者分別就人工評量與電腦化評量之資料蒐集說明如下：

一、人工評量與電腦化評量在評量耗費時間之分析

(一) 人工評量

在人工評量的部份中，研究者分別邀請 53 位具備教學科技背景之博士生、碩士生、國民中小學現職教師，以及大學部四年級學生，以尹玫君 (2009) 所發展之網路學習小組議題報告評量表為標準，使評量者就研究者所提供之 128 篇網路學習小組議題報告進行人工評量。人工評量實施完畢後，研究者共回收 890 份人工評量的時間耗費資料，即以此作為與電腦化評量時間耗費比較之原始數據。

(二) 電腦化評量

由於人工評量實施完畢後共回收 890 份人工評量的時間耗費資料，為使資料數量能一致，因此在電腦化評量的部份，研究者以 128 篇網路學習小組議題報告作為測試樣本，將其逐一匯入本研究所發展之電腦化評量系統，以取得電腦化評量之時間耗費資料。待取得 890 份電腦化評量的時間耗費資料後，研究者即進行獨立樣本 t 考驗，以瞭解人工評量與電腦化評量在評量時間耗費之差異情形。

(三) 差異分析結果

研究者將所得之電腦化評量及人工評量時間耗費資料，透過獨立樣本 t 考驗進行分析，以瞭解兩組樣本在評量耗費時間上的差異情形。表 6 呈現電腦化評量與人工評量在評量耗費時間之平均數、標準差與 t 值。

由表 6 得知，電腦化評量的平均評量時間較人工評量為低，達 .001 的顯著水準且為高度效果量(effect size $\eta^2=1.87$)。顯示電腦化評量與人工評量的評量耗費時間有顯著的差異，且電腦化評量較人工評量節省評量所耗費的時間。

因此，從上述統計結果得知，在本研究所提出的「網路學習小組議題報告電腦化評量系統」中，電腦化評量於評量時間耗費上顯著較人工評量節省時間。

表 6 電腦化評量與人工評量時間耗費之 t 考驗摘要表

組別	平均數 (秒)	標準差	t 值
電腦化評量 (N=890)	11.31	8.96	39.35***
人工評量 (N=890)	67.74	41.84	

註：*** $p < .001$

二、人工評量與電腦化評量在評量得分之分析

研究者分別就「報告的內容能呈現深思熟慮後的觀點或整合的結果」、「報告的內容有與主題相呼應，也有聚集焦點」、「報告內容段落能表達順暢」、「報告內容用字正確(含標點符號)，沒有造成誤會」等四個評量項目，透過獨立樣本 t 考驗，瞭解人工評量與電腦評量之評量得分差異。詳細的分析流程茲說明如下：

(一) 評分資料的取得

研究者以表 4 所列各評分項目之測試資料表示人工評量的結果。研究者所匯入的測試資料數量分別是 307 份「報告的內容能呈現深思熟慮後的觀點或整合的結果」、322 份「報告的內容有與主題相呼應，也有聚集焦點」、286 份「報告內容段落能表達順暢」、307 份「報告內容用字正確(含標點符號)，沒有造成誤會」。將其逐一匯入電腦化評量系統，取得各人工評量分數所對應之電腦化評量分數，以作為分析人工評量與電腦化評量在評量得分差異情形之原始資料。

(二) 差異分析結果

研究者將所得之電腦化評量及人工評量之評量得分資料，透過獨立樣本 t 考驗進行分析，以瞭解兩組樣本在評量得分上的差異。以下分別就「報告的內容能呈現深思熟慮後的觀點或整合的結果」、「報告的內容有與主題相呼應，也有聚集焦點」、「報告內容段落能表達順暢」、「報告內容用字正確(含標點符號)，沒有造成誤會」等四個評量項目，逐一進行說明。

1. 報告的內容能呈現深思熟慮後的觀點或整合的結果

由表 7 得知，電腦化評量的評量得分平均數較人工評量略高，但在統計上未達顯著差異($t=0.45, p>.05$)，顯示電腦化評量與人工評量在「報告的內容能呈現深思熟慮後的觀點或整合的結果」評分項目的評量得分並無顯著的差異，電腦化評量結果與人工評量結果並無差異。

主題文章

表 7 電腦化評量與人工評量得分之 t 考驗摘要表 (一)

組別	平均數	標準差	t 值
電腦化評量 (N=307)	5.21	2.82	.45
人工評量 (N=307)	5.10	2.90	

2. 報告的內容有與主題相呼應，也有聚集焦點

由表 8 得知，電腦化評量的評量得分平均數較人工評量略低，但未達顯著差異($t=-1.46, p>.05$)，顯示電腦化評量與人工評量在「報告的內容有與主題相呼應，也有聚集焦點」評分項目的評量得分並無顯著的差異，電腦化評量結果與人工評量結果並無差異。

表 8 電腦化評量與人工評量得分之 t 考驗摘要表 (二)

組別	平均數	標準差	t 值
電腦化評量 (N=322)	5.26	2.88	-1.46
人工評量 (N=322)	5.61	3.16	

3. 報告內容段落能表達順暢

由表 9 得知，電腦化評量的評量得分平均數較人工評量略高，但未達顯著差異($t=0.64, p>.05$)，顯示證電腦化評量與人工評量在「報告內容段落能表達順暢」評分項目的評量得分並無顯著的差異，電腦化評量結果與人工評量結果並無差異。

表 9 電腦化評量與人工評量得分之 t 考驗摘要表 (三)

組別	平均數	標準差	t 值
電腦化評量 (N=281)	5.23	2.31	.64
人工評量 (N=281)	5.10	2.72	

4. 報告內容用字正確(含標點符號)，沒有造成誤會

由表 10 得知，電腦化評量與人工評量的評量得分平均數，在統計上未達顯著差異($t=0.03, p>.05$)，顯示證電腦化評量與人工評量在「報告內容用字正確(含標點符號)，沒有造成誤會」評分項目的評量得分並無顯著差異。，電腦化評量結果與人工評量結果並無差異。

表 10 電腦化評量與人工評量得分之 t 考驗摘要表 (四)

組別	平均數	標準差	t 值
電腦化評量 (N=303)	5.89	2.31	.030
人工評量 (N=303)	5.88	2.80	

三、系統易用性及滿意度評估結果

本研究經由相關研究文獻及專書等資訊，建置網路學習小組議題報告電腦化系統，為探究此一電腦化評量系統是否具有預期功能，進行專家評估。研究者分別邀請 10 位擁有具備網路學習等相關專長之實務與（或）研究經驗的專家學者，以及國民小學現職教師，擔任電腦化評量系統評估委員，據以評估系統介面親和度、功能可用性以及整體滿意度。問卷中針對系統介面親和度、系統功能有用性、系統整體滿意度三大面向的評估結果顯示如下：

(一) 系統介面親和度

在系統介面親和度的部份，六個題項的滿意度合計皆達 60%至 100%之間（包含同意及非常同意），此向度所屬之六個題項皆未呈現「不同意」與「非常不同意」的數值，故結果顯示系統的介面親和度之「本電腦化評量系統容易操作」、「本電腦化評量系統的網頁內容顯示清晰」、「本電腦化評量系統的網頁轉換流暢穩定」、「本電腦化評量系統的色彩配置感覺舒適」、「本電腦化評量系統的文字字體大小適當」、「本電腦化評量系統的文字行距適當」方面受到專家相當程度的肯定。但在「本電腦化評量系統的文字字體大小適當」此題項中，有 40%的受訪者勾選普通，其比例明顯較其他題項為高，顯示本研究系統在字體大小設定的部份，仍有改進的空間，故研究者後續將修改加大部份頁面的字體大小，以提供更為舒適的操作環境。

(二) 系統功能有用性

在系統功能有用性的部份，七個題項的滿意度合計皆達 80%至 100%之間（包含同意及非常同意），此向度所屬之七個題項皆未呈現「不同意」與「非常不同意」的數值，顯示系統的功能有用性之「本電腦化評量系統的小組議題報告匯入功能很方便」、「本電腦化評量系統提供的小組議題報告之評量非常客觀及公平」、「本電腦化評量系統的議題報告評量功能可滿足評分需求」、「本電腦化評量系統能協助使用者迅速瞭解各小組的學習成果」、「本電腦化評量系統的批次評量功能（執行一次即可評量多筆報告），可滿足評分的便利性」、「透過本電腦化評量系統能有效提昇評量小組議題報告的效率」、「本電腦化評量系統可以讓使用者很方便的匯出小組議題報告評量結果」等方面受到專家高度的肯定。

(三) 系統整體滿意度

在系統整體滿意度的部份，三個題項的滿意度合計皆達 100%(包含同意及非常同意)，此向度所屬之三個題項皆未呈現「普通」、「不同意」與「非常不同意」的數值，顯示在「我對此套電腦化評量系統整體感到滿意」、「我很樂意使用此套電腦化評量系統來輔助評量」、「我很樂意將此套電腦化評量系統應用於議題報告的評量」方面受到專家高度的肯定。

伍、結論與建議

一、結論

本研究依據實證分析結果，歸納與說明研究結論如下：

(一) 小組議題報告的電腦化評量的評量時間顯著低於人工評量

在本研究所發展之網路學習小組議題報告電腦化評量系統中，電腦化評量進行評量網路學習小組議題報告所耗費的時間，顯著的較人工評量低，顯示電腦化評量在評量耗費的時間上顯著較人工評量節省。

(二) 小組議題報告的電腦化評量與人工評量的評量分數沒有顯著差異

在本研究所發展之網路學習小組議題報告電腦化評量系統中，電腦化評量與人工評量在「報告的內容能呈現深思熟慮後的觀點或整合的結果」、「報告的內容有與主題相呼應，也有聚焦點」、「報告內容段落能表達順暢」、「報告內容用字正確(含標點符號)，沒有造成誤會」上之評量得分，未達顯著水準。結果顯示電腦化評量結果與人工評量結果具有一致性。

(三) 網路學習小組議題報告電腦化評量系統受到專家的肯定

本研究發展之網路學習小組議題報告電腦化評量系統，不論在系統介面親和度的向度、系統功能有用性的向度，和系統整體滿意度的向度，均獲得專家的肯定。

二、建議

(一) 電腦化評量可視為一輔助教師評量的工具

電腦化評量之目的之一即在希望能達成「省時省力」、「評分標準能達一致」。因此，評量工具能否準確反映受評者的學習狀況，是為評量者所關注的焦點。傳統上，教師多就其領域專業的背景，採人工評量的模式進行學生作業、

作品的批閱工作。此種方式雖未必「省時省力」，但在「反映受評者的學習狀況」的部分仍具備一定程度的準確度。

因此，未來如欲於實際進行電腦化評量，即可能面臨「評量結果能否真實反映受評者的學習狀況」的挑戰。故研究者建議，評量者在進行電腦化評量時，除應就電腦化評量工具之評量正確度進行調校，同時亦可納入電腦輔助評量的觀點，將電腦化評量工具視為一輔助教師評量的工具，經由電腦化評量與人工評量之共同實施，以取得可「真實反映受評者的學習狀況」的評量結果。

(二) 電腦化評量可應用於不同階段的學習者

本研究所提之「網路學習小組議題報告電腦化評量系統」，係應用於大學階段的學生。建議未來可納入不同學齡階段之學習者，經由電腦化評量工具的應用，使其能自動評估學習者的學習狀況。如此，對於輔助教師瞭解學習者的學習狀況，應可有所助益，並能適度節省老師批閱報告的時間。

(三) 評量內容的限制

目前本研究所提之「網路學習小組議題報告電腦化評量系統」，已具備文本內容的評量能力。然學生議題報告的內容，可能透過圖片、語音，以及影片的形式進行呈現。因此，此類型的內容格式，本研究所提之電腦化評量系統無法正確反映學習者的學習狀況。故研究者建議，未來研究可朝圖片、語音、影片之內容分析進行探討，並進一步嘗試發展內容評估工具。如此，對於突破電腦化評量僅限於文本格式的限制，應可具相當程度的助益。

參考文獻

- 尹玫君(2009)。應用 e-learning 合作學習的教學和評量策略之研究 (國科會專題研究計畫)。臺南：作者。
- 尹玫君、劉亭言(2009)。網路合作學習中的線上討論評量表。論文發表於 2009 海峽兩岸卓越論文學術研討會，2009/07/04，中國上海：華東師範大學。
- 方欽鴻(2007)。從數學競試中分析高中生數學能力之分析研究。**雄工學報**，8，1-36。
- 王信智(2001)。應用資訊檢索技術於科學寫作作品評量之探究。未出版之碩士論文，國立台南師範學院資訊教育研究所，臺南市。
- 洪明洲(1999)。網路教學，臺北市：華彩。
- 柯華蕙、陳明蕾、王學誠(2009)。Writing Evaluation。取自：

主題文章

http://www.lsa.url.tw/modules/lisa/lisa_writing_evaluation.php

- 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。臺北：心理出版社。
- 高台茜(2003)。網頁小組討論之小老師鷹架輔助對小組互動品質的促進效果研究。論文發表於第十一屆國際電腦輔助教學研討會。2003/04/24 ~ 2003/04/26。臺北：國立臺灣師範大學。
- 張炳雄(2003)。專家診斷系統應用於文書處理術科評量機制之研究。未出版之碩士論文，國立臺灣師範大學工業教育學系，臺北市。
- 許溫雅(2009)。應用模糊理論於創造力評量系統。未出版之碩士論文，國立臺南大學數位學習科技學系，臺南市。
- 陳年興、楊錦潭(2006)。數位學習理論與實務。臺北：博碩文化。
- 陳志憲(2004)。討論區架構對網路合作學習人際互動之影響。未出版之碩士論文，國立臺南師範學院資訊教育研究所，臺南市。
- 陳佩正譯(2001)。Thomas R. Hoerr (2000)原著。多元智慧融入教學與領導。臺北：遠流。
- 陳薇如(2007)。運用學習歷程探勘之適性化數位學習評量管理系統。未出版之碩士論文，華梵大學資運管理學系，宜蘭市。
- 曾憲雄、蔡秀滿、蘇東興、曾秋蓉(2006)。網路資料探勘 (教育部顧問室通訊科技人才培育先導型計畫)。臺北：教育部。
- 黃彥博(2008)。科學文章摘要自動化計分方式的比較研究。未出版之碩士論文，國立臺南大學測驗統計研究所，臺南市。
- 黃政傑、林佩璇(1996)。合作學習。臺北：五南。
- 楊銀興(2000)。傳統評量與新式評量之比較及國小教師對實施新式評量相關問題覺知情形之研究。未出版之博士論文，國立臺灣師範大學教育學系，臺北市。
- 葉寶霞(2004)。以網際網路資源強化學習成效之研究—國一數學互動網。未出版之碩士論文，國立臺灣科技大學資訊管理研究所，臺北市。
- 廖聖傑(2002)。從學習歷程檔案建構決策樹以支援網路教學。未出版之碩士論文，國立中山大學資訊管理研究所，高雄市。
- 趙美聲(2003)。數位學習之教學策略分析。收於中山大學主編。WISCS 2003 網

路教學系統平台與內容標準化學術研討會論文集(頁 149-152)。高雄：中山大學。

劉佳宗(2007)。利用機器學習摘要概念為基礎之文件摘要自動建立方法。未出版之碩士論文，國立成功大學資訊管理研究所，臺南市。

劉惠如(1999)。整合式網路教學之教學設計與評量。未出版之碩士論文，國立中山大學資訊管理研究所，高雄市。

蔡昂叡(2007)。以本體論為基礎之學習討論區主題分析器。未出版之碩士論文，國立交通大學資訊科學與工程研究所，新竹市。

顧超光、郭世明(2005)。應用多媒體教材於網路輔助教學之成效評估。收於國立台臺技術學院主編。2005 數位設計研討會論文集(頁 58-67)。臺中：國立臺中技術學院。

Chen, M.L., Wang, H. C., & Ko, H.W. (2008). The construction and validation of Chinese Semantic Space by using Latent Semantic Analysis. Retrieved from <http://140.115.78.41/CJP08A17R1.pdf>

Fuchs,L.s., Fuchs, D., & Hamlett, C.L.(1992). Computer applications to facilitate curriculum-based measurement. *Teaching Exceptional Children, Summer*, pp.58-60.

Roach (2006), Building Decision Trees in Python, Retrieved Nov 20,2009,from http://onlamp.com/pub/a/python/2006/02/09/ai_decision_trees.html?page=1

The Study on Developing a Computerized Assessment System for Report of Web-based Learning

Yin, Mei-Chun * **Su, Yen-Ning ****

The purpose of this study is to develop a Computerized Assessment System for a Report of Web-based Learning. The researchers reviewed the evaluation criterion of a Web-based Learning report, used the Latent Semantic Analysis (LSA) and Decision Trees to build the Computerized Assessment System. As the system was developed, the researchers compared the rating results in the time spent and the grades between the Computerized Assessment System and human assessment. Then they proceeded with the questionnaire assessment to get relevant data. The research results are summarized as follows : 1.The Computerized Assessment saves more time than the human assessment, 2. There is no significant difference in rating results between the Computerized Assessment and the human assessment, 3. “User friendliness”, “System Usability” and ”Use Satisfaction” for the Computerized Assessment System are highly approved by the experts.

Keywords: Computerized Assessment, Report, Decision Trees, Latent Semantic Analysis, Web-based Learning

* Yin, Mei-Chun, Professor, Department of Education, National University of Tainan

** Su, Yen-Ning, Ph.D. Student, Department of Engineering Science, National Cheng Kung University (Corresponding Author)