

智慧型線上論證自動評量系統的實作與成效評估

黃振榮、王宇武、廖家健、陳俊華、翁傳翔、朱玉珍、沈宏彥¹

《摘要》

隨著網際網路科技與多媒體應用的蓬勃發展，傳統教學方式已經無法滿足現代人的需求，許多新穎的科技被引進校園，用以輔助並提升教學成效，例如數位學習(e-learning)就是一個不錯的例子。而電腦中介傳播(Computer-mediated communication, CMC)在其中亦扮演了重要的角色，它是建構於網際網路之上的電腦系統，因其具有包括互動性、共時性、文字化的訊息形式、匿名性等特性，得以讓數位學習系統順利運行，並使學習者透過電腦經由網路在不同地方能夠進行資訊的交換、增進學習者之間的互動，以及利用線上群組進行討論。

近年來論證(Argumentation)能力的訓練被發現為當前教育的一大重點，透過批判性思考以及使用論據證明論點，能夠有效地獲取知識，因此受到教育領域的重視。隨著數位學習的發展，論證的研究議題也跟著進入數位化。文獻指出，利用網際網路作為傳輸媒介，開發的線上論證系統有其特殊性及必要性，例如學習成就較低或被動型的學生即能在線上論證系統中較能克服主動參與討論的畏懼感。當教師設定好論證題目，學生透過線上論證系統進行論證，為了瞭解學生的論證品質，老師必須檢驗其論證內容是否達到論證模組的等級標準。我們發現過去雖然在文獻上已有許多研究使用 Toulmin 論證模型來檢驗學生的論證能力，但都只能由教師根據論證元素檢驗學生的論證，檢驗完之後仍須經由老師評估其論證能力等級，因此並未達到系統自動化的目標，造成老師的教學負擔並未相對減輕，因而對於使用線上論證系統輔助教學感到興趣缺缺。此外，經由專家判定的論證能力等級有可能因專家知識背景的不同，而產生某種程度上的人為因素誤差。還有學生使用完系統之後，亦需等待專家的評量才能得知其論證能力等級，無法得到及時的回應。

¹ 本文作者黃振榮為國立東華大學理工學院副院長；王宇武、廖家健、陳俊華、沈宏彥為國立東華大學學習科技研究所學生；翁傳翔、朱玉珍為國立東華大學資訊科學系學生。

為了解決以上所述的一些問題，本研究採用 Toulmin 的論證模型，開發一線上論證自動評量系統。我們利用資料探勘的技術找出能夠辨識辯證的結構化特徵，並根據學生論證過程的情況，將學生的論證能力分為五個等級。

本研究提出之線上論證自動評量系統採用中央研究院研發的中文斷詞系統 (Chinese Knowledge and Information Processing, CKIP) 先萃取出學習者討論內容的關鍵詞，接著使用資料探勘技術將萃取出來的討論內容關鍵詞與事先由老師在專家系統資料庫中設定好的五個等級關鍵詞資料進行比對，並由系統自動計算該論證品質的等級。由於事先放在專家系統資料庫中的資料可能無法囊括所有可能的答案，所以當系統計算出的分數較判定為五個等級所需的門檻值為低時，表示該答案是專家系統資料庫無法判別等級的特例，此時該系統會以 E-mail 通知老師登錄系統評分，除能避免論證內容被系統誤判外，亦能將新的答案視需要放進專家系統資料庫當作未來的參考答案，進而使系統具有自我調適的能力。此外，系統會根據學生不同的論證等級給予相對應的回饋訊息，例如達到較低論證等級的學習者給予較多較詳細的回饋訊息，以引導學習者能夠進一步提出論據、反證等，加強其論證能力，進而達到輔助教學的成效。而在分組討論的過程中，有部分學生可能在聊天或是談論與主題無關的話題，本系統會偵測出來並給予回饋訊息加以制止。為了驗證本研究建構之線上論證自動評量系統對於學生論證過程是否能提供實質幫助，並能正確將學生的論證分級，我們選定國小五年級二個班級學生為實驗組及控制組，在控制組我們安排在老師指導下完成前測的部分，並以這班學生的學習成果做為系統的訓練樣本，訓練樣本不足或不適用的部分，由老師加以補充及修正。實驗結果顯示，我們建立的線上論證自動評量系統除了可以檢查偵測學生在討論區內的交談內容是否與教學活動所設定之議題相關，並且可以判定學生論證的辯證等級。當評量系統偵測到學生在活動的過程中方向有所偏離，回饋規則建構機制可即時發出回饋訊息向學生提出建議或提示。此外，藉由評量系統提供的輔助學習機制，國小學生於自然科學的辯證學習成效亦獲得提升。

關鍵字：線上論證；數位學習；學習輔助；資料探勘；專家系統

壹、研究動機與研究問題

近年來有研究指出在教室中經由科學質詢(National Research Council[NRC], 1996; Kuhn & Resier, 2005)學生有能力來建構以證據為本的解釋，對於科學教育(Lemke, 1990; Rosebery et al, 1992; Schauble et al, 1995)成功的關鍵。論證過程也可以用來測驗學生對於科學概念的了解程度(McAlister, 2001)。因此論證能力訓練近年來對於科學教育極為重要(Driver et al, 1994; Duschl & Osborne, 2002; Kuhn, 1993; Sandoval & Millwood, 2005; Zohar & Nemet, 2002)。完整的辯證分析必須依賴以下幾個知識的觀點：語言上的限制、領域的相關性、概念的相關性及論述的結構。

論證架構的研究大多採用 Toulmin's 模型(Toulmin, 1958)來決定及辨識辯證的結構化特徵，例如主張、資料、理由、支持、資格及反證。透過分析學生論證的結構可以了解學生對其論述的吸收程度(Driver et al., 2000)及處理資訊的能力，並了解學生以他們的生活經驗來提出辯證的推理能力(Simon et al, 2004; Osborne et al, 2007; Osborne et al, 2006; Aufschnaiter et al, 2005; Simon et al, 2002)。

利用網際網路作為傳輸媒介，開發的線上論證系統有其特殊性及必要性，例如學習成就較低或被動型的學生即能在線上論證系統中較能克服主動參與討論的畏懼感。當教師設定好論證題目，學生透過線上論證系統進行論證，為了瞭解學生的論證品質，老師必須檢驗其論證內容是否達到論證模組的等級標準。我們發現過去雖然在文獻上已有許多研究使用 Toulmin 論證模型來檢驗學生的論證能力，但都只能由教師根據論證元素檢驗學生的論證，檢驗完之後仍須經由老師評估其論證能力等級，因此並未達到系統自動化的目標，造成老師的教學負擔並未相對減輕。此外，經由專家判定的論證能力等級有可能因專家知識背景的不同，而產生某種程度上的人為因素誤差。還有學生使用完系統之後，亦需等待專家的評量才能得知其論證能力等級，無法得到及時的回應。

資訊技術的發展日新月異，雖然電腦仍無法取代教師的地位，但卻可以適度地分擔老師的工作。所以我們希望能針對學生語意表達能力的診斷，建構一個能夠有效支援教師對學生辯證能力評量的工具，並且透過系統提供學生適當的學習建議，以減輕教師在診斷評量和教學的負擔，而能有更多的時間及精力去設計對學生的學習更有幫助教學策略。

貳、文獻回顧與探討

在文獻中有許多結構化的辯論支援環境，並應用在科學辯證中，例如 Collaboratory Notebook (Edelson, Pea, & Gomez, 1996)、CaMILE (Guzdial, Turns, Rappin, & Carlson, 1995)及 Knowledge Forum/CSILE (Scardamalia, Bereiter & Lamon, 1994)。這些可被視為有關知識的累積與發展的學習環境。SpeakEasy (Hoadley, Berman & S. Hsi, 1995)、Sensemaker (Bell, 1997)及 BGUILE (Tabak, Smith, Sandoval & Reiser, 1996)歸類為較大的質詢環境。(Marije et al, 2007)提出建構論證圖，可以有效加強學生論證的組織能力，以上所述的這些線上環境強調資訊的交換或立論的建構。

幾個不同的方法被用來辨識辯證的基本特徵及在小組會談中學生辯論的結構 (Forman *et al* 1998; Kelly *et al* 1998; Bell *et al* 2000)。除了這些特殊的環境，基本的線上非同步的討論區的使用，與支援教室辯證環境可以達到相同的效果 (Kelly *et al* 2002; Resnick *et al* 1993)。由 (Kuhn *et al* 2005)針對非正式推理及辯論所開發的評量，及 Osborne 等人 (Walker *et al* 2006; Sadler *et al* 2005)提出有關社會科學辯證的評量研究。近來的一些研究像 (Gordon *et al* 1994; Prakken *et al* 2001)著重在勸導性的對話，(Kraus *et al* 1998; Parsons *et al* 1998)聚焦在辯論過程中的磋商，及 Hulstijn *et al* 2000 正式的對談模式，Parsons 等人則對勸導、資訊的蒐尋及質詢對談定義其簡單的結構並探究其屬性。

一、電腦協作學習(Computer Supported Collaborative Learning, CSCL)

是學習範疇裡日漸興盛的一個分支，其關切的是人們在電腦的幫助下如何一起學習。協作學習(Collaborative Learning)和合作學習(Cooperative Learning)是有所區隔的。合作指學習由個人完成，之後每個人貢獻他們個別的結果，然後展示出個別結果的組合，成為小組的成果。協作學習是在人與人之間共同建構知識，個人是以團體成員的方式投入，而非個人學習活動，成員之間以協商與分享共同完成。

二、Moodle(Modular Object-Oriented Dynamic Learning Environment)

澳洲人 Martin Dougiamas 所最先發展，其已在符合 GNU 規範下釋出。以模組化物件導向動態學習情境為其特色，是自由軟體中著名的課程管理系統 (Moodle; Margetson, 1994)。

三、學習風格

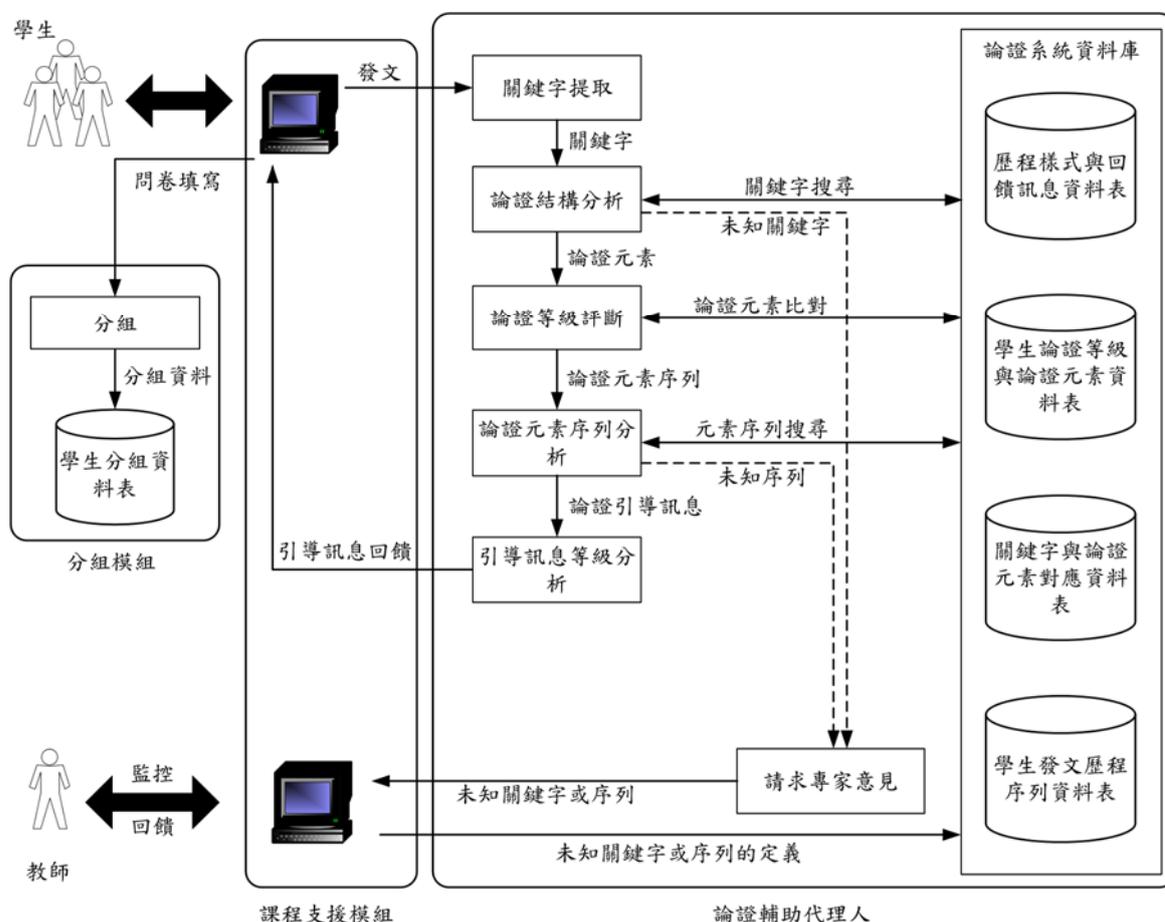
根據學生不同的學習風格來給予不同的學習型態，能有效提昇學生的學習成效。有關學習風格的評量工具很多，在本研究中我們選用了 Grasha-Riechmann 學生學習風格量尺(The Grasha-Riechmann Student Learning Style Scales, GRSLSS)，由

GRSLSS 可分辨出以下六種社會學習風格，獨立型、依賴型、競爭型、合作型、逃避型和參與型。

參、研究方法及步驟

本計畫所使用的學習平台之系統架構如圖一，由分組模組、課程支援模組及辯證輔助代理人三大部分所組成。首先學生需填寫學習風格問卷，並依照學生的學習風格加以分組。接下來系統可以開始接受學生發文，並針對每次發文進行處理，其處理過程如下：

1. 學生發文後，系統首先會將其關鍵字分離出來，並將關鍵字使用同義字取代。
2. 搜尋關鍵字對應的論證元素，若無法得到對應元素，便會要求老師做出定義。
3. 將得到的論證元素與此學生已經完成的論證元素合併後，判斷出其目前的論證等級，同時產生最新的論證元素序列。
4. 依照學生目前的論證等級與論證元素序列，搜尋對應的引導訊息，若無相對應的訊息，也會要求老師做出定義。
5. 依照引導訊息等級，判斷是否需要回饋給學生。



圖一 智慧型線上論證自動評量與學習輔助系統

構成系統的分組模組、課程支援模組及辯證處理代理人三大部分說明如下。

一、分組模組

將學生分組的模組。它利用 Kumar 所設計的學習風格分類問卷可評定出每位學生的學習風格，再將每一位學生分類至五個學習風格群集之一，這五類學習風格群集如

表一所示。理論上每一組應包含五種群集各一位學生。這種分組的方式可預期對於學習是有幫助的，例如「參與型」及「合作型」的學生會推動「逃避型」的學生前進，而「獨立型」和「合作型」的學生可以成為「依賴型」學生指引對象進而促進「依賴型」學生的進步。

表一 五種學習風格群集

群集	學習風格
群集 1	依賴型、參與型、競爭型
群集 2	參與型、依賴型、合作型
群集 3	合作型、參與型、獨立型
群集 4	獨立型、合作型、參與型
群集 5	逃避型

二、課程支援模組

圖二中顯示的是課程支援模組中使用者的介面。它可以協助教師及學生收集補充學習資源，並且可以讓教師在討論區中設定討論的議題。



圖二 課程支援模組中使用者的介面

當學生送出要發表的意見後，系統會回饋引導訊息，若此引導訊息需要立即通知，系統會自動跳出一個警告訊息(如圖三)，以提醒學生應注意的細節。



圖三 系統回饋引導訊息

若系統無法在資料庫中搜尋到學生發文的關鍵字，或是無法找到與學生論證歷程相同的元素序列，系統便會要求教師提供關鍵字或是元素序列的定義及回饋所需的訊息。

三、論證輔助代理人

(一) 關鍵字提取

關鍵字提取是由同義詞取代單元與中文知識與資訊處理程序(CKIP)所組成。

i. 同義詞置換

相同的意思，不同的學習者其用詞常會不同，所以要先建立同義詞資料庫，在進行斷詞之前，將所有同義詞先置換成統一的用詞，以利於後續的分析工作。

ii. CKIP

中研院建構的斷詞系統，可將贅詞排除而將動詞及名詞提取出來，做為後續處理之用，為國內外中文自然語言處理及其相關研究提供基本的研究資料與知識架構。

(二) 論證結構分析

關鍵字與論證元素對應表(表二)中包含主張、理由、資料、支持與反證等元素，是由 Toulmin 模型(Toulmin, 1958)，以及根據(Lazonder, 2003)所提到應用引導文字(Sentence Openers)當作半結構化的對話工具，並考量時下同學間慣用語辭做為進一步的參考。

表二 關鍵字與論證元素對應表

論證元素	關鍵字	
Claim (主張)	我認為... 我覺得... 我的觀點是...	我的看法是... 我的想法是...
Data (資料)	我有份資料... 這個資料顯示... 從這些資料...	證據是...
Warrant (理由)	因為... 而且... 我同意，因為...	原因是... 我有一個例子... 老師曾說過...
Backing (支持)	文獻上指出... 根據...理論 我有一個理由... 根據報導 指出...	支持的理由是... 我解釋 一下... 課堂上有
Rebuttal (反證)	為什麼...你有什麼根據? 但是有個問題... 另一個問題是...	有一個反例... 這樣想法的例外情形是... 我有另一個想法...

使用關鍵字對應的方式，可以讓本系統達到一般化，不需要為每一個討論議題，設計特定的資料庫。一旦建立好完整的對應表，便可以廣泛的應用在任何不同的討論議題上，而且本系統也提供老師自行新增關鍵字，因此也適用於特殊領域的議題

(三) 論證等級評斷

將學生目前發文的論證元素與歷次發文所累積的論證元素合併後，依照論證元素與論證等級對應表(表三)的規則，評斷學生目前的論證等級，例如：學生回答了他的主張-C，跟論據-W，所以論證元素是 CW，而 CW 在我們所設定的論證等級是為等級二。

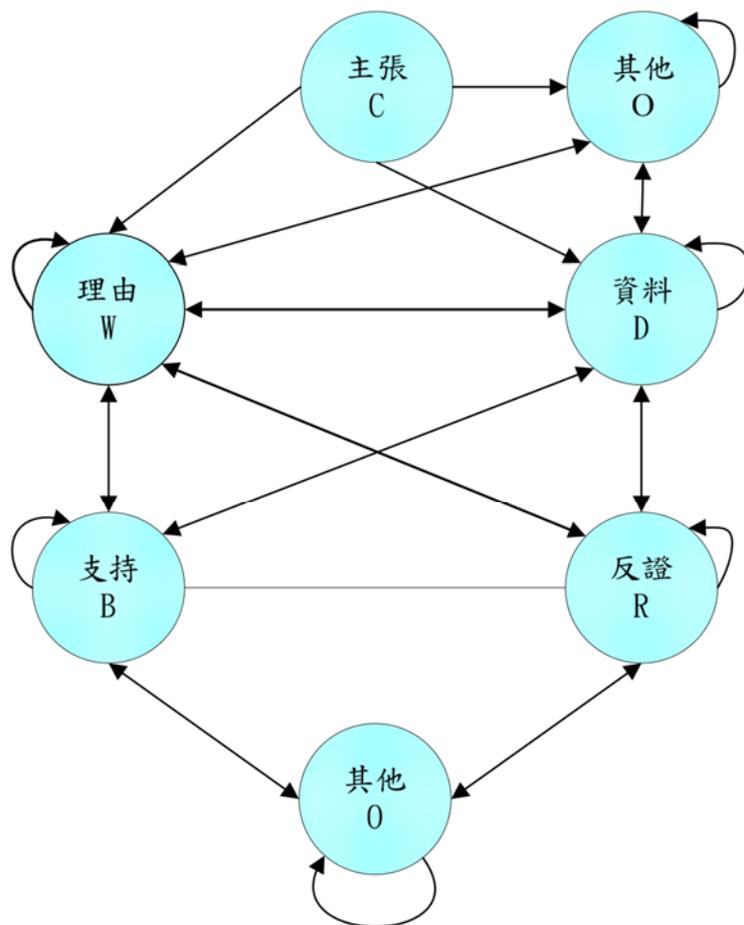
表三 論證元素與論證等級對應表

論證等級	論證元素要求	例子
Level 1	C	[我認為](C)外太空有生物存在
Level 2	CW/CD	[我認為](C)外太空有生物存在，[因為](W)我有看過這類似的新聞。
Level 3	CDW/CDR/ CDB/CWB	[我認為](C)外太空有生物存在，[因為](W)曾有太空人有拍到外星物質的照片，[根據 xx 研究](B)指出，有發現類似生物生活的痕跡。
Level 4	CDWR/CDBR/ CDWB/CWBR	[我認為](C)外太空有生物存在，[我有份資料](D)有說到外星球是有生物存在，[因為](W)曾有太空人有拍到外星物質的照片，[根據 xx 研究](B)指出，有發現類似生物生活的痕跡。
Level 5	CDWRB	[我認為](C)外太空有生物存在，[我有份資料](D)有說到外星球是有生物存在，[因為](W)曾有太空人有拍到外星物質的照片。 [根據 xx 研究](B)指出，有發現類似生物生活的痕跡。 [我有另一個想法](R)為什麼有的星球可以有生物生存，有的不行。

(四) 論證元素序列分析

由於論證能力的提升，需要較長的學習時間，因此每個學生在過程中勢必會有多次的發文，而每次發文的內容都可能包含一到多個論證元素，甚至也有可能不包含任何論證元素，因此為了可以準確地掌握學生的論證歷程，本計畫設計了論證元素序列，此序列是以每次發文所產生的論證元素所組成，其路徑如圖四所示。

一開始討論時系統會要求學生必須對問題做出表態，因此第一次發文皆會由『主張』出發，之後依照學生自己的論證能力自由發文。由於系統無法預測學生發文的內容，因此除了『主張』這個元素外，每個元素都可以連接到其相鄰的元素包括自己在內，若學生重複提出主張則後續的主張皆會被認定為『其他』。



圖四 論證元素序列路徑圖

為了可以讓學生充分發表自己的意見，本系統並不會限制學生的發文次數，導致元素序列可能會無限制的成長，為此我們必須將序列做適當的切割，況且系統需要對學生目前的學習狀態做出即時的引導，因此也不可以使用學生整個發文歷程來判斷所需要的引導，於是我們提出適當的演算法來進行切割，只保留最新的部分做為判斷的依據，其演算法如圖五

<i>Post_Sequence</i> [] = 此次發文的元素序列	
<i>Current_Level</i> = 學生目前的論證等級	
<i>Current_Element</i> [] = 學生目前提出過的論證元素	
<i>Current_Sequence</i> [] = 切割後保留的元素序列	
<i>PATTERN_TABLE</i> = 儲存元素序列樣式的資料表	掃描所有發文元素
for (<i>i</i> =0 ; <i>i</i> <Count(<i>Post_Sequence</i> []) ; <i>i</i> ++) { if (<i>Post_Sequence</i> [<i>i</i>] NOT IN <i>Current_Element</i> []) {	
<i>Current_Level</i> += 1	若論證等級提升則清除元素序列
<i>Current_Element</i> [] += <i>Post_Sequence</i> [<i>i</i>]	
<i>Current_Sequence</i> [] = EMPTY	否則將此元素加入序列
} else {	
<i>Current_Sequence</i> [] += <i>Post_Sequence</i> [<i>i</i>]	
}	若搜尋到與此序列相同的樣式且最後一個元素不是『其他』的話，則在顯示提示後清除元素序列
}	
If (<i>Current_Sequence</i> IS IN <i>PATTERN_TABLE</i>)	
{ ShowSomeHint()	
if <i>Current_Sequence</i> [] != `O`	
<i>Current_Sequence</i> [] = EMPTY	
endif	
}	

圖五 論證元素序列切割演算法

經由上述演算法切割處理過後，得到的元素序列就是學生最新的發文狀態。

(五) 引導訊息等級分析

得到最新的論證元素序列後，系統便會參考學生目前的論證等級，從歷程樣式資料表中搜尋與論證元素序列相同的樣式，便可以得到相對應的引導訊息。但並非每種樣式都需要給學生引導，若學生是正常討論且其論證能力也有所提生，那麼系統會認為此學生方向是正確的，因此不會有任何引導。但若產生序列顯示學生偏離主題或是無法進一步提升論證能力，那麼系統會依照學生目前的論證等級來回饋不同程度的引導訊息。假若系統在資料表中無法搜尋到想同的樣式，系統也會要求老師對於新的樣式做出定義並且提供引導訊息。

舉例來說假設要求學生討論的題目是：「在真空中，是否能聽見物體碰撞的聲音？」，且學生完成如表四的討論串：

表四 模擬學生討論串

學生 A(A1)：我覺得不可以聽到聲音。
 學生 B(B1)：我覺得可以聽到聲音。
 學生 A(A2)：因為之前上課中老師有說。
 學生 A(A3)：所以我想真空中是不行聽到聲音的。
 學生 B(B2)：老師有說過嗎？我不確定耶。學生
 A(A4)：我看你還是改一下答案好了。
 學生 B(B3)：也許吧，我不知道啦！學生 A(A5)：好像是因為真空中沒有介質所以才聽不到。

則系統處理的過程會如表五所示：表五 系統處理過程的模擬

發文	論證元素	論證等級	論證歷程	目前序列	是否提示
A1	C	Level 1	C	C	否(初次提出主張，不需提示)
B1	C	Level 1	C	C	否(初次提出主張，不需提示)
A2	W	Level 2	CW	CW	否(論證等級提升，因此不需提示)
A3	C	Level 2	CWC	C	否(雖然重覆提出主張，由於等級為 2，因此不提示)
B2	O	Level 1	CO	O	是(由於等級為 1，且偏離主題，因此需要提示)
A4	O	Level 2	CWCO	CO	是(重覆提出主張，且又偏離主題，因此需要提示)
B3	O	Level 1	COO	OO	是(連續的偏離主題，因此需要更明確的提示)
A5	D	Level 3	CWCOD	D	否(論證等級提升，因此不需提示)

肆、實驗過程與內容分析

一、前測獨立樣本 T 檢定

以 SPSS 獨立樣本 T 檢定檢驗實驗組與控制組各 31 位學生的前測成績平均值是否呈現顯著差異，結果如表六、表七。

表六 實驗組與控制組前測成績組別統計量

Class	個數	平均數	標準差	平均數的標準誤
實驗組	31	70.48	10.905	1.959
控制組	31	69.35	12.162	2.184

表七 實驗組與控制組前測成績獨立樣本檢定

	變異數相等的 Levene 檢定		平均數相等的 t 檢定						
	F 檢定	顯著性	t	自由度	顯著性 (雙尾)	平均差異	標準誤差異	差異的 95% 信賴區間	
								下界	上界
假設變異數相等	.404	.527	.385	60	.702	1.12	2.93	-4.73	6.99
不假設變異數相等			.385	59.30	.702	1.12	2.93	-4.74	6.99

由表六得知，實驗組平均成績為 70.48，控制組平均成績為 69.35；由表七得知，變異數相等的 Levene 檢定未達顯著 ($p=.527>0.05$)，表示兩組的變異數具同質性，因此採用假設變異數相等這一系列的值，結果顯示平均數相等的 t 檢定未達顯著 ($p=.702>0.05$)，顯示兩組學生的前測成績無顯著差異，在平台學習之前對學習單元的概念認知幾近相同。

二、前測單一樣本 T 檢定

將實驗組與對照組的前測成績分別以檢定值為 70 作單一樣本 T 檢定，如表八、表九所示，實驗組單一樣本 T 檢定未達顯著 ($p=.807>0.05$)，控制組單一樣本 T 檢定未達顯著 ($p=.770>0.05$)，顯示兩組學生的前測平均成績無顯著差異，在平台學習之前，兩組的基本能力相當。

表八 實驗組前測成績單一樣本檢定

實驗組	檢定值 = 70					
	t	自由度	顯著性 (雙尾)	平均差異	差異的 95% 信賴區間	
					下界	上界
Score	.247	30	.807	.484	-3.52	4.48

表九 控制組前測成績單一樣本檢定

控制組	檢定值 = 70					
	t	自由度	顯著性 (雙尾)	平均差異	差異的 95% 信賴區間	
					下界	上界
Score	-.295	30	.770	-.645	-5.11	3.82

三、後測獨立樣本 T 檢定

以 SPSS 獨立樣本 T 檢定檢驗實驗組與控制組各 31 位學生的前測成績平均值是否呈現顯著差異，結果如表十、表十一。

表十 實驗組與控制組後測成績組別統計量

Class	個數	平均數	標準差	平均數的標準誤
實驗組	31	81.45	13.366	2.401
控制組	31	72.42	12.966	2.329

表十一 實驗組與控制組後測成績獨立樣本檢定

	變異數相等的 Levene 檢定		平均數相等的 t 檢定						
	F 檢定	顯著性	t	自由度	顯著性 (雙尾)	平均差異	標準誤差異	差異的 95% 信賴區間	
								下界	上界
假設變異數相等	.002	.965	2.70	60	.009	9.03	3.34	2.34	15.72
不假設變異數相等			2.70	59.94	.009	9.03	3.34	2.34	15.72

由表十得知，實驗組平均成績為 81.45，控制組平均成績為 72.42，兩組的後測成績均比前測進步。由

表十一得知，變異數相等的 Levene 檢定未達顯著 ($p=.965>0.05$)，表示兩組的變異數具同質性，因此採用假設變異數相等這一系列的值，結果顯示平均數相等的 t 檢定達顯著 ($p=.009<0.05$)，顯示兩組學生的後測成績有顯著差異，平台的學習輔助機制有助於學生的學習成效。

四、控制組前後測獨立樣本 T 檢定

以 SPSS 獨立樣本 T 檢定檢驗控制組 31 位學生的前、後測成績平均值是否呈現顯著差異，結果如表十二、表十三。

表十二 控制組前後測成績組別統計量

	個數	平均數	標準差	平均數的標準誤
前測	31	69.35	12.162	2.184
後測	31	72.42	12.966	2.329

表十三 控制組前後測成績獨立樣本檢定

	變異數相等的 Levene 檢定		平均數相等的 t 檢定						
	F 檢定	顯著性	t	自由度	顯著性 (雙尾)	平均差異	標準誤差異	差異的 95% 信賴區間	
								下界	上界
假設變異數相等	.014	.907	-.96	60	.34	-3.06	3.19	-9.45	3.32
不假設變異數相等			-.96	59.75	.34	-3.06	3.19	-9.45	3.32

由表十二得知，控制組前測成績為 69.35，後測成績為 72.42，後測成績比前測成績進步少許；由表十三得知，變異數相等的 Levene 檢定未達顯著

($p=.907>0.05$)，因此採用假設變異數相等這一系列的值，結果顯示平均數相等的 t 檢定未達顯著

($p=.0341>0.05$)，顯示控制組學生的前後測成績無顯著差異，平台在沒有學習輔助機制下，對學生的學習成效幫助有限。

五、實驗組前後測獨立樣本 T 檢定

以 SPSS 獨立樣本 T 檢定檢驗實驗組 31 位學生的前、後測成績平均值是否呈現顯著差異，結果如表十四、表十五。

表十四 實驗組前後測成績組別統計量

Class	個數	平均數	標準差	平均數的標準誤
前測	31	70.48	10.905	1.959

後測	31	81.45	13.366	2.401
----	----	-------	--------	-------

表十五 實驗組前後測成績獨立樣本檢定

	變異數相等的 Levene 檢定		平均數相等的 t 檢定						
	F 檢定	顯著性	t	自由度	顯著性 (雙尾)	平均差 異	標準誤 差異	差異的 95% 信賴區間	
								下界	上界
假設變異 數相等	.536	.467	-3.54	60	.001	-10.96	3.09	-17.16	-4.77
不假設變 異數相等			-3.54	57.67	.001	-10.96	3.09	-17.17	-4.76

由表十四得知，實驗組前測成績為 70.48，後測成績為 81.45，後測成績比前測成績有大幅度進步；由表十五得知，變異數相等的 Levene 檢定未達顯著 ($p=.467>0.05$)，因此採用假設變異數相等這一系列的值，結果顯示平均數相等的 t 檢定未達顯著 ($p=.0001<0.05$)，顯示實驗組學生的前後測成績有顯著差異，平台在有學習輔助機制下，能有效提升學生的學習成效。

伍、結論

在本研究中，我們提出了以論證元素序列分析技術所建立的智慧型線上論證自動評量與學習輔助系統。本系統能自動檢查偵測學生在討論區內的交談內容是否與活動所設定之議題相關，並判定學生的論證等級。當系統偵測到學生在活動過程中的方向有所偏離時，回饋規則建構機制會即時發出回饋訊息向學生提出建議或提示。由實驗結果顯示，本研究中所提出的智慧型線上論證自動評量與學習輔助系統在協助國小學生於自然科學的辯證學習活動上確實有所助益。由於本研究中有針對學生的學習風格進行分組，未來系統將考慮在論證過程中，結合各學生的學習風格給予適性化的回饋。

陸、參考文獻

Aufschnaiter, C. Von, Erduran, S., Kraus, M. E., Osborne, O., Rogge1, C. & Simon, S., (2005) Argumentation and the learning of science, in European Science Education Research Association Conference, Barcelona, Spain, August.

- Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher*, 23, 5-12.
- Driver, R., Newton, P., & Osborne, J., (2000) Establishing the norms of scientific argumentation in classrooms, *Science Education*, vol. 84(3), pp. 287-213.
- Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, 38, 39-72.
- Edelson, D. C., Pea, R. D., & Gomez, L., (1996) Constructivism in the Collaboratory, in *Constructivist learning environments: Case studies in instructional design*, B. C. Wilson, Ed. Englewood Cliffs, NJ: Educational Technology Publications, pp. 151-164.
- Kuhn, L., and Resier, B., (2005) Students constructing and defending evidence based scientific explanations, *National Association for Research in Science Teaching*, Dallas, TX, April.
- Lazonder, A.W., Wilhelm, P., Ootes, S. A.W. (2003). Using sentence openers to foster student interaction in computer-mediated learning environments. *Computers & Education*, Vol. 41, No. 3, 291-308
- Lemke, J. L., (1990) *Talking science: Language, learning and values*. Norwood, Ablex Publishing Company.
- Marije, V. A., Jerry, A., & Gellof K., (2008) How students structure and relate argumentative knowledge when learning together with diagrams. *Computers in Human Behavior* Vol. 24(3), pp. 1293-1313.
- National Research Council, (1996) *National science education standards*, Washington, DC: National Research Council.
- Osborne, J., Erduran, S., & Simon, S., (2007) *Talking to learn and learning to talk*, in *Association for Science Education*, Birmingham, January.
- Osborne, J., Erduran, S., & Simon, S., (2006) *Ideas, evidence and argument in science education: supporting teacher practice*, in *Conference of the American Educational Research Association*, San Francisco, USA. April.
- Rosebery, A., Warren B., & Conant, F.. (1992) Appropriating scientific discourse: Findings from language minority classrooms, *The Journal of the Learning Sciences* vol. 2, pp. 61-94.
- Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' used of evidence in written scientific explanations. *Cognition and Instruction*, 23, 23-55.
- Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J., (1995) "Students' understanding of the objectives and procedures of experimentation in the science classroom," *The Journal of Learning Sciences*, vol. 4(2), pp. 131-166.
- Simon, S., Osborne, J., & Erduran, S., *Ideas*, (2004) *Evidence and Argument in Science Project*, in *Association of Tutors in Science Education, Annual Day Conference*, Bedford, UK. March.

Simon, S., Erduran, S., & Osborne, J., (2002) Enhancing the Quality of Argument in School Science, in Conference of the National Association for Research in Science Teaching, New Orleans, USA. April.

Toulmin, S., (1958) The uses of argument. Cambridge: Cambridge University Press.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39, 35-62.

Implementation and Performance Evaluation of an Intelligent Online Argumentation Assessment System

Chenn-Jung Huang、Yu-Wu Wang、Jia-Jian Liao、Chun-Hua Chen、
Chuan-Hsiang Weng、Yu-Jen Chu、Hung-Yen Shen

《Abstract》

Recent researches indicated that students' ability to construct evidence based explanations in classrooms through scientific inquiry is critical to successful science education. The amount and quality of explanations from these reasoned arguments that students construct can show their understanding of science concepts. The practice of argumentation is then an essential teaching activity within science education because it can provide the opportunity to develop young students' ability to construct argument.

Most of the researches on student discourse depend heavily on Toulmin's model to determine and identify the structural features of arguments, such as claims, data, warrants, backings, and qualifiers, and process argumentation in the literature. Analyzing the structure of students' arguments let us understand how students assimilate the desired practices of argumentation and process the information about the structure and field of reasoning that students use when they provide arguments based on their experiences from living environment.

Structured argumentation support environments have been built and used in scientific discourse in the literature. They can be considered as learning environments that concerns about knowledge accumulation and development. These mentioned environments emphasized either on exchanging information or on constructing arguments for presentation. To the best our knowledge, no research work in the literature addressed the issue of automatically assessing the student's argumentation quality. The teaching load of the teacher that uses the online argumentation support environments is not alleviated and accordingly the willingness of the teacher to apply e-learning to argumentation training is seriously affected.

In this work, an intelligent argumentation assessment system based on machine learning techniques is proposed. Learners' arguments on discussion board were examined to detect whether the learners address the expected discussion issues and to determine the argumentation level achieved by the learner's argument. A feedback rule construction mechanism is used to issue feedback messages to the learners in case the argumentation assessment system detects that the learners go in the biased direction or stay idle for a while.

To verify the effectiveness of the argumentation assessment system proposed in this work, two fifth grade classes at an elementary school participated in the experiments. After the instructor made a thorough description of how to build a toy plane in traditional classroom teaching, the learners were placed into separate groups and then worked together to make toy planes by using Styrofoam plates. The students were expected to focus on the discussion of the design issues such as airfoil size, center of gravity, throwing angle, throwing strength, and prow shape. The argumentation processing agent recorded whether the students addressed the five key design issues in learner portfolio database. Each student was asked to take a pre-test and a post-test right before and after the argumentation activity, respectively. The instructor can accordingly discover the groups with low argumentation level in their discussions, and reassign the students with poor performance to other heterogeneous groups to effectively facilitate their knowledge constructing ability. The experimental results exhibit that the proposed work is effective in classifying each student's argumentation level and assisting the students in learning the core concepts taught at a natural science course on the elementary school level.

Keywords: online argumentation, e-learning, learning assistance, data mining, expert systems.

