

# 以答題信心度為基礎之線上診斷評量系統

黃國禎 朱蕙君

國立臺南大學數位學習科技系

王榕榆

暨南國際大學資訊管理學系

## 摘要

一般線上評量系統的測驗方式，是由學生回答單選或複選題；由於忽略了學生答題過程的不確定因素，因此在診斷學生的學習困難時，不易作出精確的研判。針對這個問題，已有學者嘗試以答題信心度為基礎，來開發線上評量系統，並獲得不錯的成果。然而，由於目前已提出的評量模式，是由學生主觀認定其信心度，因此在診斷的準確性方面仍有很大的改善空間。在本研究中，我們嘗試加入答題時間參數，來加強其診斷效果，以期更準確地分析學生遭遇的困難。為了驗證這個新評量機制的分析成效，我們以國小學童中文字辨識課程進行實驗；實驗結果證明，本研究提出之「考量時間因素之信心度診斷評量機制」較原有的機制能提供更精確的診斷結果。

關鍵字：國語文教學、補救教學、診斷測驗、電腦輔助學習、網路學習

## 壹、緒論

本章共分兩節來介紹以答題信心度為基礎之線上測驗與評量系統；第一節首先敘述本研究的背景與動機，第二節說明研究的目的。

### 一、研究背景與動機

行政院教育改革審議委員會（1996）於總諮詢報告書提出「發展適性適才的教育」的建議，強調政府應於國民教育階段規劃並建立完整的補救教學系統，以彌平學生的個別差異，使個體的學習困難獲得改善（王錦如，2000）。目前國民基本教育由於班級學生人數眾多及教師授課時間限制等因素，使得教師不易深入地瞭解學生的個別差異。目前國內多數學校的班級，一名教師通常須指導二、三十名學生，並在同一個班級以一致的課程進度教授統一的教材，往往無法考慮到在學習方面有困難或是理解速度較慢的學生。因此，如何提出一套有效

的教學策略與評量方式，來協助教師瞭解學生的個別差異，進而提高學生的學習興趣及學習成就，已成為目前亟需解決的課題（余民寧，2001）。

要落實補救教學的理念，必須有精確而有效的學習困難診斷機制，才可針對學生的學習困難提供適切的補救方向及教材，提昇學習的品質。然而，學習困難點的發生，除了概念本身或先備知識的不足，也有可能是對類似概念產生混淆所造成；若無適當的診斷策略來提供改善的方向，同樣的錯誤可能重複地發生，進而影響後續的學習。在評量過程能否蒐集更多的答題資訊來協助資料分析，是影響學習診斷結果精確與否的重要因素之一。由於一般的線上測驗與評量系統受限於單選題或多選題的出題形式，使得能夠蒐集到的測驗資訊非常有限。為了改善這個問題，Lo, Wang 與 Yeh（2004）提出一個以答題信心度為基礎的線上評量機制，以改善過去在評量過程中無法蒐集完整測驗資訊的缺點，並在英文介系詞的診斷方面獲得不錯的成果。然而，由於該評量模式是由學生主觀認定其答題信心度，因此在診斷的準確性方面仍有很大的改善空間。為解決這個問題，本研究透過導入答題信心度的診斷評量機制，並融入答題反應時間參數，讓系統可針對測驗歷程與答題資訊進行分析，以更精確地找出個別學生所遭遇的困難，並提供補救方向。

## 二、研究目的

本研究嘗試提出結合答題時間及答題信心度的評量機制，並建立線上診斷測驗系統，以解決傳統測驗模式蒐集測驗資訊不易的問題；同時，希望透過信心度評量機制的導入與改良，幫助個別學生找出迷失概念，並提供適性化的學習教材，以改善其學習成效。

另外，在教師方面，我們也希望藉由線上系統對於班級學生資料的分析，協助教師掌握班上學生的學習狀況，以作為改善教學的依據；而線上試題管理與試題參數的調整機制，更提供了教師掌握測驗方向的良好工具，使診斷測驗更具成效。

## 三、研究範圍

本研究主要針對易產生混淆的概念進行診斷，以提供精確的補救學習教材。以英文介系詞中的「at」、「on」、「in」為例，三者均可用來表達「時間」及「位置」的觀念。這些介系詞會隨著不同片語的搭配，或在句中扮演角色，而有不同的用法；因此，往往會導致學習者的誤用。同樣的，在中文字的結構方面，也有偏旁相似的字，例如「倍」、「陪」及「部」，或是一字多音及同音字，容易造成學習者的混淆。另外，一個中文字的定義，除了形音義的定義之外，也會隨著詞句的內容，以及在句中的角色，而產生詞性的變化，例如轉化為介詞、助詞或名詞。針對這些可能造成混淆的文字，本研究嘗試運用改良的答題信心度評量機制，來取得更多的答題資訊，以改善傳統測驗模式無法診斷易混淆文字或概念的缺點。

## 貳、文獻探討

在本章將回顧測驗、評量與補救教學的發展，並針對本研究的應用領域－「國語文教學的文字辨識」作詳細的說明。

### 一、測驗、評量與診斷

陳玫秀（2000）指出，評量的意義，是在可能的範圍內，由各方面蒐集有關學習者的資料，以作為提供適當且有意義的教育方案的依據。評量的方式可包括正式的測驗及觀察，其著眼點在衡量學生的能力。當教育人員瞭解學生需要接受補救教學時，就需要更進一步的診斷，而「診斷」的目的，在依學生的個別需求及心智發展狀況，調整教學內容及方式，以引導學生完成學習目標。

評量不僅可以讓教師瞭解學生的起點行為，也可以評估教師的教學效果及學生的學習困難；同時，更可增進學生對教學目標的瞭解，以激發其學習動機（郭生玉，1997）。評量的方式有許多種，包括安置性評量、形成性評量、診斷性評量及總結性評量等（陳英豪、吳裕益，1995，歐用生，1989）。安置診斷評量是在教師進行學生編班、分組與實施教學之前實施，旨在全盤瞭解學生學習的內外因素，以進行有效的安置；內在因素包括身體、才智、情緒和教育方面等因素，外在因素則包括對校內及校外環境的瞭解。形成性評量可在教學過程中隨時進行，例如教學活動告一段落時，或每節下課前幾分鐘，其目的在評估學生是否已達成學習目標，以作為教師修正教材及教法的回饋資料。診斷性評量則是在教學過程中實施，用來診斷造成學生學習困難的原因，以實施補助教學。總結性評量是在課程或單元結束後實施，其用途在評定學習成就，以作為加深、加廣教材或升級進修的依據。

這四種評量方式不僅目的不同，所以使用的時機也不相同。在教學前，可利用安置性評量來測得學生的程度；在教學中，可利用形成性評量提供回饋，以瞭解每一個階段學生和教師的學習與教學狀況，而診斷性評量則可診斷學生的學習困難所在，在教學後，可利用總結性評量評定學生的學習成果（余民寧，2001）。

當學生在學習上持續遭遇困難，而無法以形成性評量的補救措施解決時，就需要進行診斷性評量，以找出產生學習困難的原因（郭生玉，1997）。診斷測驗是屬於診斷性評量的一種，其目的在分析個人在某一學習領域表現的優、缺點，指出造成學習困難的可能原因，並根據學習者的條件與需要設計補救教學方案（吳裕益、王佳文，1997）。目前，診斷測驗已朝著結合認知科學和心理計量學的方向發展，而成為獨立的研究領域，稱為「認知診斷評量」（Cognitively Diagnostic Assessment, 簡稱 CDA）（Nichols, 1994）。CDA 是以當代的測驗理論的基礎，結合認知心理學對學習理論的研究與發現，用更精緻的模式，來探索三大課題：

1 受試者在測驗領域中所使用的認知程序（cognitive process）和知識結構（knowledge

structure) 為何？

2 這些認知程序和知識結構在測驗過程中如何應用？

3 高能力者與低能力者的表現有何不同？

Snow 和 Lohmann (1989) 指出，認知心理學能夠提供學生在學習問題上的回饋，而這些回饋可作為教師在教學上的參考與改進依據。有別於傳統測驗只提供測驗結果，認知診斷可解決學生在學習上的困惑，甚至是能針對個人問題提供更完善的解釋，以達到個人化的評量效果。因此，CDA 的方法有別於傳統以邏輯分類、內容說明和統計推論為基礎的評量方法。另外，由於 CDA 可以提供學生在解決問題時所需的策略、知覺概念間的關係及用以瞭解某學科的原則等訊息，在未來的教學與課程改革中，將能提供有用的資訊，以協助教育決策者制訂適合學生個別需求的教學活動及課程內容。

目前已有多個以 CDA 施測的線上系統被開發，例如 Marshall (1995) 提出的「故事問題解決者」(Story Problem Solve, 簡稱 SPS) 及 Langley, Wogulis 與 Ohlsson 等學者 (1990) 提出的「自動認知模型系統」(Automated Cognitive Modeler, 簡稱 ACM)，這些都是依據認知心理學理論，並以人工智慧技術來模擬人類解決數學問題的認知程序與知識結構。

除此之外，國內外學者更針對電腦化診斷測驗進行多方面的研究，包括「學生模組」(student models) 的建立，這類的研究主要考慮學習環境、課程、認知模組及學習表現等因素，並透過系統計量處理建立「學生模組」，以作為診斷學生學習困難的參考依據 (Anderson, 2000, Corbett, Anderson & O'Brien, 1995)。Mislevy (1995) 則以機率的結構來呈現學生學習問題所發生的頻率，由於學生在其知識、技巧和策略上皆有不同的型態的產生，因此針對不同的問題型態就有不同的發生機率。在問題解決技巧的診斷評量方面，Gitomer, Steinberg 及 Mislevy 等學者 (1995) 提到，透過智慧型教學系統 (Intelligent Tutoring System, 簡稱 ITS) 可以推論學生處理問題的能力狀態，以及學生本身對於所遭遇的問題是否有足夠瞭解，因此他們認為，在智慧型教學系統中，可以運用認知診斷、基礎機率的推論、教育評量及心理計量學等策略，來發展學生模組。

同時，Britton 及 Tidwell (1995) 則嘗試以認知結構測驗 (Cognitive Structure Testing, 簡稱 CS) 來分辨專家與新手的差異。認知結構測驗是一套具有效率且便於使用的系統，可以針對學習者對於概念的誤解或欠缺，以及學習過程中所產生的概念混淆，提供完善的資訊，以協助學習者瞭解並改善個人的知識結構，以期達到專家程度的境界。Hwang (2003) 的研究則認為，迷失概念的診斷可藉由分析概念之間的關係來達成，因此提出智慧型教學評量系統的構想，該系統可透過概念關係圖協助學生找出需要補救的概念，提供適當的補救學習路徑。

另外，將診斷測驗應用於語文教學領域的研究也有一些不錯的成果，例如胡依斐 (1988) 將國語診斷測驗分為語音、詞彙、詞法語意及語用等四個連續測驗，每一測驗均包含理解力和表達力兩種測驗，以準確地評估受試者在語言各方面的發展，鄭美芝 (2002) 針對國民小

學低年級注音符號能力診斷測驗與補救教學效益進行探討，發現一般學生對篇章的認讀表現優於單音的認讀，而學習困難學生則在單音認讀的表現優於篇章的認讀。陳姿君（2002）則在進行國中英語閱讀能力診斷測驗編製的研究中發現，英語閱讀能力診斷測驗可有效區分國一普通班與資源班學生，並可作為篩選資源班學生之工具；其研究也指出，國一資源班及低成就學生在認字語音、詞彙理解及句子理解方面均呈現出學習上的困難。

綜觀以上國內外診斷測驗發展與應用的實例可得知，診斷的精確與否，與評量的範圍、目的及評量技術有很大的關係，若能蒐集更多有用的評量資料，將更能精確地分析學生的困難，並進行教學方式的調整（曾清文，2004）。

## 二、補救教學

由過去多個研究結果中顯示，若能深入地診斷學生在學習時遭遇的問題，才能針對學習者需求提供正確的補救教學處方，並在下一階段的學習中獲得進步（吳裕益、王佳文，1997，陳玟秀，2000）。陳姿君（2002）等學者在英語閱讀能力診斷測驗中也發現，良好的測驗與評量機制，有助於瞭解學生的學習困難，因此對於改善教學與學習成效有很大的幫助。

補救教學（Remedial teaching）是指對學生學習失敗的原因加以確認之後，作有效的教學措施以謀補救（高廣孚，1997），其主要目的是協助未達成教學目標者或學習有困難者再學習，以實踐「帶好每一位學生」的理想。補救教學的實施是一種「評量-教學-再評量」的歷程，期望受輔導學生的學習進度可以跟上原班級的教學進度（陳長春，1992）。由於非所有的學生都必須進行補救教學，為了讓所有學生可以同時進入下一階段的學習，對於不需要進行補救教學的學生，可能必須放慢學習腳步（林秀如，2004），因此，為了避免在補救教學過程中影響不同程度學生的學習，應該針對個別學生的差異提供個人化的教材及進度，才可維持整體良好的教學品質（Hwang, 2007）。

## 三、國語文字教學

語文教學在小學課程中一直佔有非常重要的比例，同時也是眾所認定的一種學習活動的基礎能力，而由「熟習課文中生字的形、音、義並能活用」，逐步拓展到「對文字結構、簡易六書造字原理的認識」，乃至「文字的書寫正確，不寫錯字、別字」的學習目標規劃可看出，「文字教學」是語文領域中非常重要的一部分。因此，能運用正確的文字來表達想法，是每一位小學生應該具備的語文能力（丘慶鈴，2003）。

### （一）識字

符顯仁（1983）研究指出：「一般所謂識字，就是要認識每個單字的字形、字音和字義，而在看讀說寫的時候，沒有發生錯誤的意思。」李白芬（1995）的研究指出「由於國字是單

音節的緣故，沒有拼音文字的形式變化，因此語音的應用範圍容易受到限制，所以同音字特別多。」如查字典「ㄅ」一個音，就有「盃背杯悲卑摺俾碑唄裨裨鵠」等字，學習時容易產生別字誤用的情形。

而筆畫相像也是造成識字困難的原因，一般分為形聲相似字（如瞻、瞻）、形體相似字（如崇、崇）、形義相似字（如履、屨）三種。另外還有依部分筆畫些微差別而音義迥異的字，例如土土、未未、胃胃、抄抄、余余等，往往使學習者感到困擾，造成辨認的困難。

## （二）別字

「別字」即指字書上本來有的字，因錯用而寫出以甲代乙的字。有關別字部分的研究，雖然多數學者提出「音同形異」、「音近形異」的錯誤類型分類（林昌炫，1979，符顯仁，1983；莊明廣，1994；陳金明，1995）；然不論「音同形異」或「音近形異」因素所產生的別字，本質上皆單純因為「字音」的關係而致誤，可合併視為「音別字」所產生的別字錯誤類型。

丘慶鈴（2003）分析各專家學者的研究分類，並將別字之錯誤類型依導致錯誤的因素與漢字的形音義特性綜合整理歸納為單因素、雙因素、三因素和其他四大類，如表 1 所示。音別字、形別字和歧義別字皆是因為和字的音、形或義其中一個因素而造成別字的發生，是單因素所造成的別字類型，音義相近字、形音相近和形義相近別字，則是合併漢字特點「形音義緊密連結」的其中兩個特質而形成的別字，應視為「雙因素」所造成的別字；兼具形音義三種漢字特點造成混淆的別字類型，應視為「三因素」所形成的別字類型。另有一些特殊情況產生的別字，無法就漢字「形、音、義」三方面做配對探討的，則歸於其他一類（丘慶鈴，2003）。

表 1 別字之錯誤類型整理

單因素	形別字	形似音異致誤	如把「失敗」寫成「失販」。
	音別字	1 音近形異致誤	如把「應該」寫成「因該」。
		2 音同形異致誤	如把「而已」寫成「而以」。
歧義字	字義混淆致誤	如把「一味」寫成「一味」。	
雙因素	音義相/近別字	音義相近致誤	如把「做」寫成「作」。
	形音相/近別字	形似音近致誤	如把「既然」寫成「即然」。
	形義相/近別字	形似音同致誤	如把「收穫」寫成「收獲」。
三因素	形音義/相近別字	形音義皆相近致誤	如把「練習」寫成「鍊習」。
其他	1 無法看出原因的別字	如把「些」寫成「也」。	
	2 詞序顛倒致誤	如把「蝴蝶」寫成「蝶蝴」。	
	3 一字多音別字	如把「曾經」寫成「層經」。	
	4 方言致誤	如把「縫衣」寫成「宏衣」。	

資料來源 丘慶鈴，2003。

由以上中國文字特性與別字錯誤類型可得知，國小學生在學習識字過程中容易產生概念混淆和迷思的情形；因此，本研究將以國小學生識字的學習診斷為應用目標，進行測驗內容及實驗設計，以驗證所提出的評量模式之成效。

## 參、以信心度為基礎之線上診斷評量機制

在本研究中，我們嘗試提出答題時間信心度的評量模式，並以中文字辨識進行實際的測試。以下將詳細介紹以學生信心度自評為基礎的診斷評量機制，以及本研究提出的融合答題時間信心度的診斷評量機制及學習狀態診斷演算法。

### 一、以學生信心度自評為基礎的診斷評量機制

為改善傳統線上評量系統的分析能力，Lo, Wanga 與 Yeh (2004) 嘗試由教師事先設定好學生在作答時應該具備的信心度，稱為信心水準 (Confidence Levels, CL)，再由學生在進行答題時設定回答結果的信心度 (Confidence Rating, CR)。CL 與 CR 兩個項目的值域範圍介於  $1 \sim -1$ ，每道試題的答案選項中，若 CL 的值設定為正值 (介於  $0 \sim 1$ )，表示其在答案選項中為正確的答案選項；若為負值 (介於  $-1 \sim 0$ )，則為錯誤的答案選項。

學生對每道試題的答題信心度分數 (Confidence Score, CS) 可利用公式  $CS_q = CL \times CR$  來求得 (Shortliffe & Buchanan, 1975)，若信心度分數為正值，表示學生認定某選項為對，而此選項在試題中確實是對的選項 ( $CL > 0$  和  $CR > 0$ )；或是學生認定某答案不是正確的，而此選項確實為錯誤的答案選項 ( $CL < 0$  和  $CR < 0$ )。換言之，若信心度分數為負值，表示學生在作答過程中做了錯誤的選擇，亦即學生的回答，與教師設定的信心水準不一致，例如將對的答案選項設定為錯的選項 ( $CL > 0$  和  $CR < 0$ )，或是錯的答案選項設為對的選項 ( $CL < 0$  和  $CR > 0$ )。

### 二、融合答題時間信心度的診斷評量機制

在測驗的過程中，除了讓學生自行設定答題信心度，系統也記錄了答題時間做為調整信心度的參考，而調整的幅度是依據統計經驗法則中的常態面積。根據統計學的大數法則及中央極限定理，當樣本值個數足夠時 (大於 30 個)，樣本值的平均數即接近母體平均數 (陳順宇, 1996)；且當資料呈常態分配時，約有 68% 的觀測值落於  $(\bar{X} - s, \bar{X} + s)$  的區間內，95% 的觀測值落於  $(\bar{X} - 2s, \bar{X} + 2s)$  的區間內，而 99.7% 的觀測值落於  $(\bar{X} - 3s, \bar{X} + 3s)$  的區間內。因此，本研究將以此分布作為調整答題信心度的依據。修正的自評信心度公式為  $CR' = CR \times (1 - \text{依時間被扣除的答題信心度面積})$ ；若學生答題時間小於或等於全班平均數，則答題時間權重值維持為 1；若答題時間超過全班平均數但未超過 1 個標準差，則答題時間權重值會從 1 下降為 0.5 (即  $1 - 0.5$ )，其中扣除的 0.5 代表學生失去 0.5 的答題信心度面積；若學生答題時

間超過平均數且落在 1 個標準差至 2 個標準差之間時，則為  $CR' = CR \times (1 - 0.5 - 0.34)$ ，其中 0.5 及 0.34 為學生失去的答題信心度面積。答題信心度的常態面積分佈如圖 1 所示。

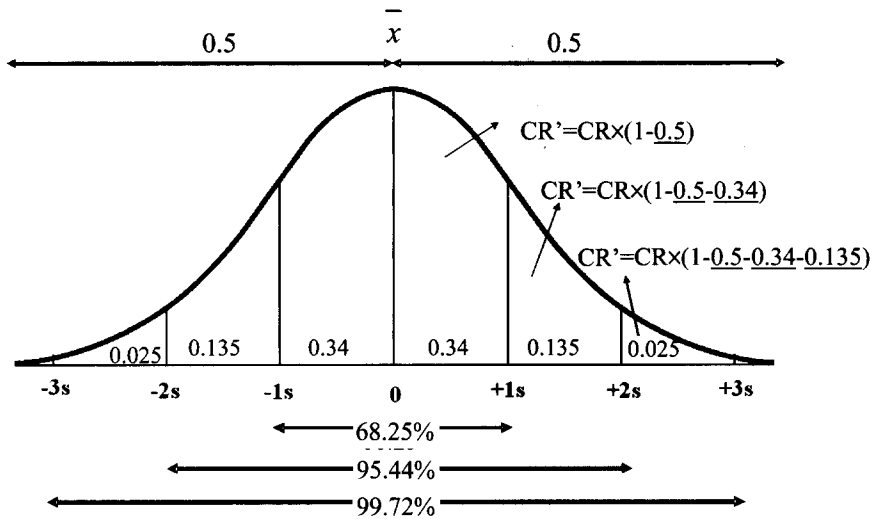


圖 1 答題信心度常態面積分佈圖

因此，本研究融合時間之答題信心度的修正規則如下：

$$\text{if } x \leq \bar{x} \text{ then } CR' = CR \times 1$$

$$\text{if } \bar{x} < x < \bar{x} + s \text{ then } CR' = CR \times (1 - 0.5)$$

$$\text{if } \bar{x} + s < x < \bar{x} + 2s \text{ then } CR' = CR \times (1 - 0.5 - 0.34)$$

$$\text{if } \bar{x} + 2s < x \text{ then } CR' = CR * (1 - 0.5 - 0.34 - 0.135)$$

其中  $x$  代表學生作答時間； $\bar{x}$  代表全班作答時間平均數； $s$  為全班作答時間標準差； $CR'$  為依據答題時間調整過後的學生答題信心度； $CR$  為未調整前的學生答題信心度。

另外，為了協助教師設定信心水準，本研究依 Hwang, Chen, Hwang 與 Chu (2006) 提出的方法，對於明確的答案及錯誤給予 +1 及 -1 值；傾向非常適用、適用及勉強適用的答案分別給予 +0.75、+0.5 及 +0.25 的值；傾向嚴重錯誤、錯誤及些許錯誤的答案分別給予 -0.75、-0.5 及 -0.25 的值。

### 三、學習狀態診斷演算法

在求得每道題目中的信心度分數 ( $CS_q$ ) 後，再與先前測驗相同概念的試題信心度累計分數 ( $CS_{old}$ ) 合併計算，以獲得新的信心度分數 ( $CS_{new}$ )，如公式(1)-(3)所示。本方法源自數學的邏輯推論，用來累積多個證據的決策效果 (Shortliffe & Buchanan, 1975)。



$$CS_{new} = CS_{old} + CS_q(1 - CS_{old}), \text{ when } (CS_{old}, CS_q > 0) \quad (1)$$

$$CS_{new} = CS_{old} + CS_q(1 + CS_{old}), \text{ when } (CS_{old}, CS_q < 0) \quad (2)$$

$$CS_{new} = (CS_{old} + CS_q) / (1 - \min(|CS_{old}|, |CS_q|)), \text{ otherwise} \quad (3)$$

在分析學生的學習狀態時，以  $n$  個答案選項的信心度分數陣列及  $n \times n$  維的信心度分數矩陣表來表示學生可能遭遇的問題（如圖 2 所示）。從信心度分數陣列的產出值中，可觀察出學生對概念的學習狀態，而從信心度分數矩陣表，可檢視學生對概念誤解的原因。學習診斷流程的分析法則如圖 3 所示，其內容說明如下

		學生作答			
		己	以	己	
測驗 概念	己				此CS值可診斷測驗的概念與學生作答的概念是否產生混淆
	以				
	己				此CS值可診斷出學生不瞭解的概念

圖 2 信心度分數矩陣表

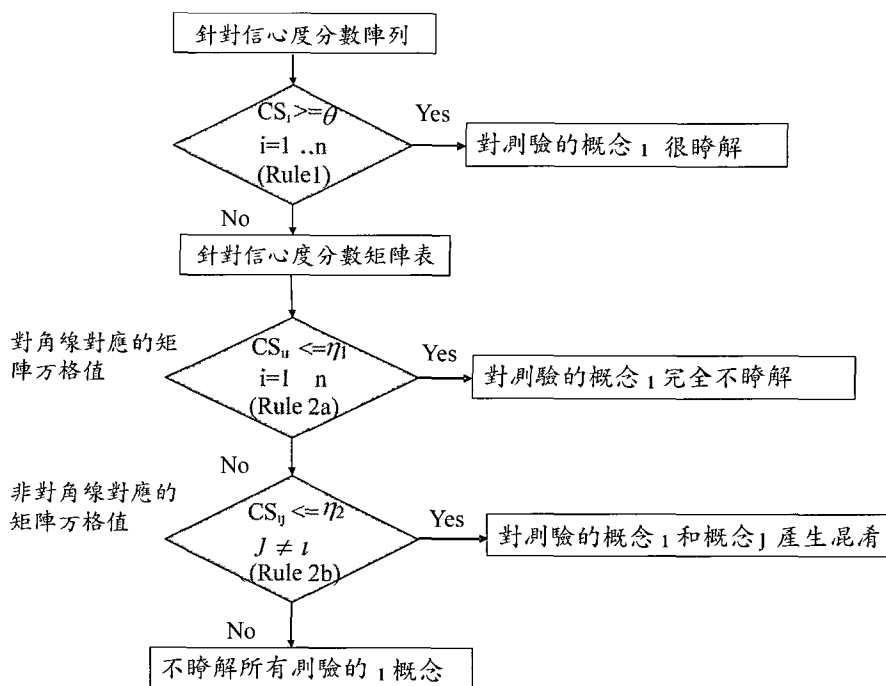


圖 3 學習診斷流程圖

法則 1：在信心度分數陣列的概念選項中，如果信心度分數值大於或等於概念門檻值  $\theta$ ，即表示學生對此概念學習狀況良好，不需要補救教材。

法則 2：針對在法則 1 中沒有通過門檻值的概念，分析學生沒有通過門檻值的原因及種類為何。

法則 2a：若信心度分數矩陣表中，信心度分數的值小於定義的誤解門檻值  $\eta_1$ ，即判斷學生對概念不瞭解，並提供相關的補救學習教材，否則，檢視法則 2b。

法則 2b 信心度分數矩陣表中，若非對角線對應方格的信心度分數值小於定義的誤解門檻值  $\eta_2$ ，則判定學生在作答概念和測驗概念間發生混淆 (Lo, Wang, & Yeh, 2004)，並提供相關補救教材。

#### 四、實例說明

以下將以實例說明從測驗、診斷及回饋學生補救教材的流程：

- (一) 假設題庫中有 10 道試題，用來測驗學生對概念「已」、「以」、「己」的瞭解程度。針對每道試題包含的概念，會由教師先對答案選項設定好學生應具備的信心水準 (confidence levels, CL)，在表 2 為某位學生對每道試題答案選項的信心度 (confidence ratings, CR)，其中底線部分為依照其作答時間所進行的信心度權重值調整。
- (二) 表 3 為依照信心度分數運算公式 (1) ~ (3) 求得該學生對 10 個試題的信心度分數陣列。假設對概念瞭解的門檻值為  $\theta = 0.8$ ，3 個概念中只有「己」陣列值為  $0.9938 > \theta$ ，因此必須針對「己」之外的概念探討陣列值未達到門檻值的原因。表 4 為依作答時間調整信心度的陣列表，受到時間調整的試題分別為  $Q_3$ 、 $Q_6$ 、 $Q_8$ 、 $Q_9$ ，其調整的幅度可參考表 2 中的時間權重值，在修正後的陣列表中仍只有「己」陣列值為  $0.9808 (> \theta)$  是通過門檻值的。

表 2 學生測驗結果範例資料

試題題號	測驗概念	作答時間 (秒)	CL/CR'		
			己	以	已
Q1	己	20/s	0.8/0.6	-0.7/-0.8	-0.5/0.1
Q2	以	24/s	-0.7/0.8	0.7/-0.9	-0.8/0.7
Q3	己	42/s	-0.8/-0.9* <u>0.5</u>	-0.6/0.7* <u>0.5</u>	0.8/0.2* <u>0.5</u>
Q4	己	25/s	0.7/0.8	-0.7/-0.6	-0.6/-0.8
Q5	以	22/s	-0.9/-0.2	0.8/-0.7	-0.7/0.6
Q6	己	52/s	-0.5/-0.8* <u>0.16</u>	-0.6/-0.7* <u>0.16</u>	0.7/0.6* <u>0.16</u>
Q7	己	25/s	0.5/1	-0.6/-0.8	-0.6/-0.9
Q8	以	42/s	-0.8/-0.8* <u>0.5</u>	0.8/0.9* <u>0.5</u>	-0.6/-0.2* <u>0.5</u>
Q9	己	42/s	-0.6/0.7* <u>0.5</u>	-0.9/0.5* <u>0.5</u>	0.7/-0.6* <u>0.5</u>
Q10	己	23/s	0.8/0.9	-0.9/-0.4	-0.7/-0.8

表 3 所有試題的信心度分數累加陣列值

試題題號	測驗概念	CSq=CL × CR		
		已	以	已
Q1	已	0.48	0.56	-0.05
Q2	以	-0.56	-0.63	-0.56
Q3	已	0.72	-0.42	0.16
Q4	已	0.56	0.42	0.48
Q5	以	0.18	-0.56	-0.42
Q6	已	0.4	0.42	0.42
Q7	已	0.5	0.48	0.54
Q8	以	0.64	0.72	0.12
Q9	已	-0.42	-0.45	-0.42
Q10	已	0.72	0.36	0.56
信心度分數陣列值		0.9938	0.7343	0.6791

表 4 具作答時間調整的信心度分數累加陣列值

試題題號	測驗概念	CSq=CL × CR'		
		已	以	已
Q1	已	0.48	0.56	-0.05
Q2	以	-0.56	-0.63	-0.56
Q3	已	0.36	-0.21	0.08
Q4	已	0.56	0.42	0.48
Q5	以	0.18	-0.56	-0.42
Q6	已	0.064	0.0672	0.0672
Q7	已	0.5	0.48	0.54
Q8	以	0.32	0.36	0.06
Q9	已	0.21	-0.225	-0.21
Q10	已	0.72	0.36	0.56
信心度分數陣列值		0.9808	0.4914	0.5633

(三) 圖 4 是依學生自評獲得的信心度分數矩陣表，依據圖 3 的學習診斷流程，在誤解門檻值設定  $\eta_1 = \eta_2 = 0$  的情況下，圖 4 的對角線方格「以」概念的信心度分數值 ( $CS_{22}$ ) 為 -0.4186，小於門檻值 ( $\eta_1 = 0$ )，故判斷學生對「以」概念不瞭解，並不再對「以」概念的其他非對角線方格進行診斷，而直接提供「以」概念的補救教材給學生 (如圖 6 所示)。而在對角線的「已」信心度分數矩陣值為 0.16 大於門檻值 ( $\eta_1 = 0$ )，故判定學生對「已」概念瞭解，仍會依照學習診斷法則 2b 檢視「已」和其他概念是否有混淆的情形發生。在檢視非對角線的信心度分數值中，可看出在測驗學生「已」概念時，學生對選項「以」未做正確的選擇，故信心度分數值為 -0.45，沒有通過門檻值 ( $\eta_2 = 0$ )，因此認定學生在「已」和「以」概念之間發生混淆，並提供補救教材讓學生進行學習 (如圖 7 所示)。

- (四) 在圖 5 的信心度分數矩陣表中，其和圖 4 的不同點在於其信心度分數是經過作答時間調整後的結果，故可能會產生和圖 4 不一樣的學習診斷結果。由於圖 5 對角線方格的「以」和「已」概念方格的信心度分數值分別為-0.8069 和-0.0795，二者皆未通過門檻值 ( $\eta_1 = 0$ )，故依據學習診斷法則 2a 可判定學生對「以」和「已」二個概念都不瞭解。
- (五) 由以上(三)、(四)項的分析結果可得知，加入作答時間來調整信心度值，會產生不同的診斷結果，至於何者才可確實反應出學生的學習狀態，則為後面章節所要進行驗證的部分。

		學生作答		
		己	以	已
測驗 概念	己	0.968	0.9151	0.8892
	以	0.3291	-0.4186	-0.71
	已	0.7103	-0.45	0.16

「以」的概念信心度 =  
-0.4186 <  $\eta_1$  對  
「以」概念有不瞭解!

-0.45 <  $\eta_2$  , 「己」概  
念被「以」概念混淆!

圖 4 以學生自評為基礎之信心度分數矩陣表

		學生作答		
		己	以	已
測驗 概念	己	0.9680	0.9151	0.8892
	以	-0.2109	-0.8069	-0.7911
	已	0.5268	-0.3437	-0.0795

「以」的概念 =  
-0.8069 <  $\eta_1 = 0$  , 對  
「以」概念有不瞭解!

「己」的概念 =  
-0.0795 <  $\eta_1 = 0$  , 對  
「己」概念有不瞭解!

圖 5 以作答時間進行修正的信心度分數矩陣表

## 迷失的概念－補救教學

- 「以」，音：一 √
  - 用、拿。如：「以毒攻毒」、「以物易物」、「以逸待勞」、「以卵擊石」、「以其人之道還治其人之身」。
  - 以為：當做。如：「原來是你，我還以為是姜先生呢！」；
  - 認為。如：「這辦法你以為如何？」
  - 因為。如：「以人廢言」、「不以貧窮而氣餒。」
  - 連詞。相當於「而」。如：「治世之音安以樂。」
  - 介詞。相當於「之」。它的後面常接時間詞或方位詞。如：「以前」、「以後」、「以往」、「以東」、「以上」、「以下」。

圖 6 學生對概念不瞭解時的補救教材

## 混淆概念之教材對照

- 「以」
- 「以」，音：一 √
- 1 用、拿。如：「以毒攻毒」、「以物易物」、「以逸待勞」、「以卵擊石」、「以其人之道還治其人之身」。
  - 2 以為 當做。如：「原來是你，我還以為是姜先生呢！」，認為。如：「這辦法你以為如何？」
  - 3 因為。如：「以人廢言」、「不以貧窮而氣餒。」
  - 4 連詞。相當於「而」。如：「治世之音安以樂。」
  - 5 介詞。相當於「之」。它的後面常接時間詞或方位詞。如：「以前」、「以後」、「以往」、「以東」、「以上」。
- 「已」
- 「已」，音：一 √
- 表示過去之詞。如：「已經」、「已然」。
- 停止。如：「雞鳴不已」、「讚美不已」、「後悔不已」。
- 1 動詞 完畢、完成。
  - 2 副詞 已經。如：「時間早已過了。」
  - 3 代名詞 此、如此。

圖 7 發生混淆概念時的補救教材

## 肆、以信心度為基礎之線上診斷測驗系統

依據本研究提出之信心度診斷評量機制，我們開發一套線上診斷評量系統。在設計遠距測驗系統時，為考量系統運作的穩定性與操作的便利性，本系統採用三層式主從架構（three-tier client/server system architecture）來規劃測驗的環境。此架構優點在於使用者端可分擔伺服器在測驗過程中的部分計算工作，因此可適用於大量學生施測的情境。系統開發環境主要是以 Windows2000 Server 為作業系統平台，網頁服務使用了微軟的 Internet Information Server 5.0（IIS）作為系統的網頁伺服器，後端則利用微軟的 MS SQL Server 來當作系統的資料庫伺服器。另外，動態網頁開發語言使用 ASP 3.0、FrontPage 2003、VB Script、ActiveX 來達到測驗過程中與使用者互動的使用者介面。

學生在身份確認完畢後可開始進行測驗，系統會出現試題讓學生進行作答。學生在決定試題選項對與錯的同時，也必須視圖中圓圈大小來決定自己對答案的信心，由小圓圈到大圓圈代表信心度由低到高之意，如圖 8 所示。



圖 8 學生進行作答畫面

若發現學生在某些概念的學習上有迷失的情形，系統將針對測驗的結果進行分析，以瞭解導致概念學習迷失的原因，並提供個人化的補救教材，如圖 9 所示。

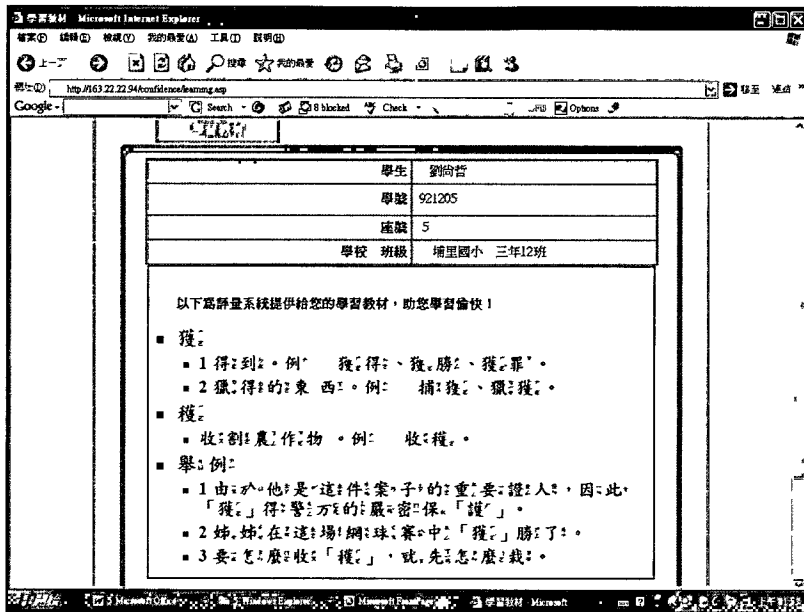


圖 9 提供補救教材的畫面

教師可線上查詢、分析班級的學習診斷結果，而學習診斷分析結果的呈現如圖 10 所示。另外，透過教師介面，教師可設定每道試題及學生在作答選項中應具備的信心水準，信心水準的範圍由-1 至+1 正值代表試題的正確選項，負值代表錯誤選項，如圖 11 所示。

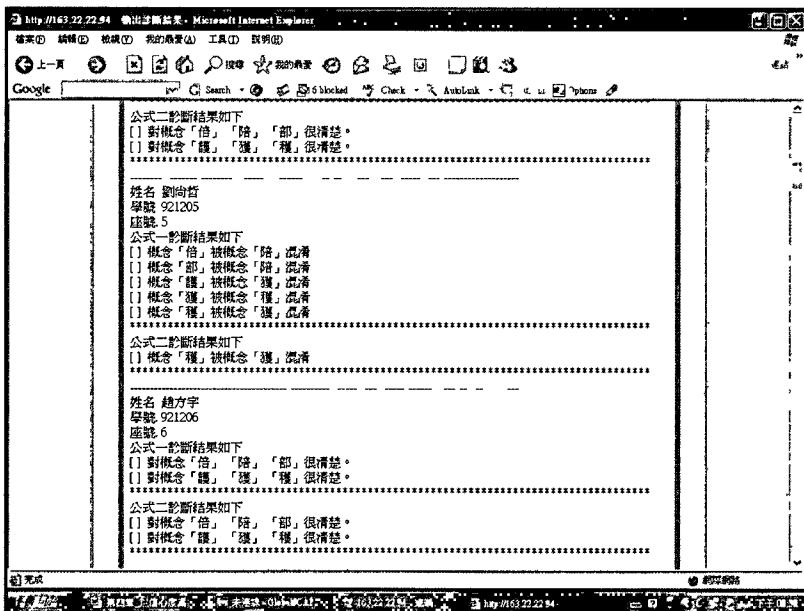


圖 10 教師查詢學生的學習診斷結果介面

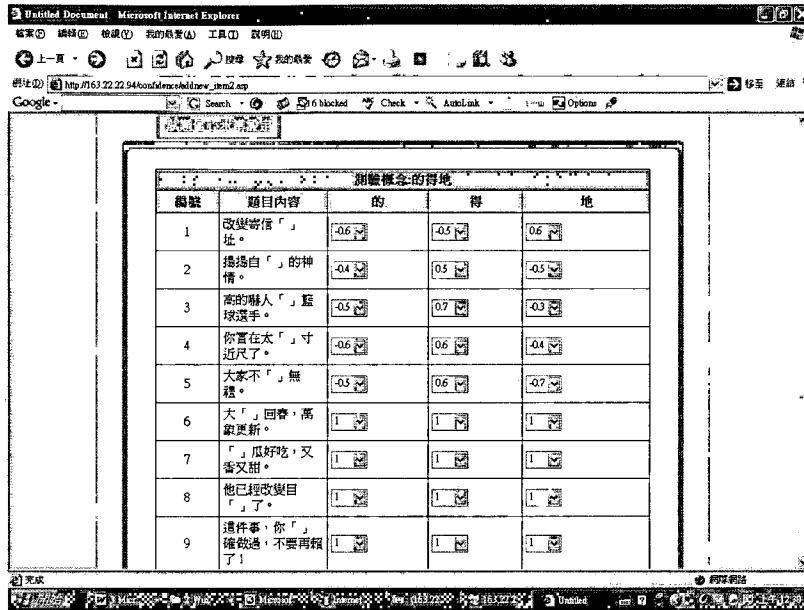


圖 11 教師設定試題選項之信心水準介面

## 伍、實驗與資料分析

為了驗證本研究提出的「融合時間因素之信心度診斷評量機制」的效果，我們嘗試將新的方法及先前學者（Lo et al., 2004）提出的方法分別與教師的評量結果進行比較。以下將介紹整個實驗的設計及實驗結果的數據分析。

### 一、實驗對象與流程

本研究的實驗對象為南投縣埔里鎮的埔里國小二個班級以及大成國小一個班級的學生。為瞭解評量方法在辨識不同文字概念組合的適用性，由每個班級測驗二組不同的概念組合。受試班級基本資料如表 5 所示。

表 5 施測班級基本資料

班級	班級人數	文字辨識概念組合
埔里國小 3 年 12 班	37 人	「護獲穫」、「倍陪部」
埔里國小 4 年 5 班	35 人	「地得的」、「倍陪部」
大成國小 3 年甲班	30 人	「地得的」、「密祕蜜」



為了掌握試題的信度及效度，本測驗由領域專家提供「護獲穫」、「倍陪部」、「地得的」及「密祕蜜」等四組試題各 12 題，並事先由南投縣中年級的國小學童 98 名進行預試；經 SPSS 進行相關分析後，刪除與測驗主題相關程度未達 0.5 ( $\gamma < 0.5$ ) 的題目，最後保留「護獲穫」、「倍陪部」、「地得的」及「密祕蜜」等組各為 10 題、9 題、10 題、8 題。在信度方面，「護獲穫」、「倍陪部」、「地得的」及「密祕蜜」等各組試題的 Cronbach  $\alpha$  係數分別 0.88、0.86、0.87 及 0.84，均達到 0.01 的顯著水準；另外以學生國語科的學期平均成績與各組題目的預試成績進行關聯效度分析，獲得  $\gamma$  值分別為 0.43、0.36、0.38 及 0.35，均達到 0.01 的顯著水準。

本實驗的操作環境為本研究開發的線上評量系統。在全班學生施測完畢之後，由系統依二種不同的評量方式進行分析；同時，由教師對學生進行實際訪談及分析，以確認系統依二種不同模式提出的學習診斷報告與學生實際學習狀況是否相符。表 6 為教師經訪談學生的學習狀態後，與二種不同評量模式進行比較的結果，其中[x]代表該結論教師的看法與系統不符，[v]則代表相符。

表 6 老師訪談學生學習狀態結果示例

姓名：王小明
學號：921205
座號：5
評量機制一的診斷結果如下：
[x]：概念「倍」被概念「陪」混淆
[v]：概念「部」被概念「陪」混淆
[v]：概念「護」被概念「獲」混淆
[v]：概念「獲」被概念「穫」混淆
[v]：概念「穫」被概念「獲」混淆
評量機制二的診斷結果如下：
[v]：概念「穫」被概念「獲」混淆

## 二、實驗數據分析

教師對三個施測班級的學生完成訪談後，為瞭解二種評量方式較與教師診斷結果的符合程度，我們提出相似度比較公式來計算二種評量機制的解答與教師解答的相似度，如公式(4)所示。

$$Score(S_{ij}) = \max \left\{ \frac{C_i - f_i - m_i}{t}, 0 \right\} \dots\dots\dots(4)$$

其中  $S_{ij}$  代表第  $i$  個方法和第  $j$  位學生的解答， $i \in \{1,2\}$ ，其中  $i=1$  時表示「融合時間觀念

信心度評量機制」在某位學生的正確性得分， $i=2$  時為「不具時間觀念信心度評量機制」的得分； $C_i$  代表解答與正確答案符合的個數； $f_i$  代表誤判的個數； $m_i$  代表遺漏的解答個數； $t$  代表正確解答的總個數。例如：教師的判斷結果為 { 不瞭解「的」的用法，不瞭解「得」的用法，將「地」與「的」混淆 }，而方法一判斷結果為 { 不瞭解「的」的用法 }，則  $Score(S_{11}) = \max\left\{\frac{1-0-2}{3}, 0\right\} = \max\left\{-\frac{1}{3}, 0\right\} = 0$ ，故計算結果顯示方法一在 1 號學生的正確性得分為 0。

### 三、實驗結果

將三個受試班級的教師訪談結果與二個診斷評量機制的結果進行比對，並利用正確性評分公式來進行正確性估算，並取全班正確性分數之平均數來做比較，如表 7 所示。

表 7 診斷評量機制正確性平均比較表

施測班級	評量機制一 (考慮時間因素)	評量機制二 (不考慮時間因素)	差異值
大成國小 3 年甲班	0.88	0.41	0.47
埔里國小 3 年 12 班	0.69	0.48	0.21
埔里國小 4 年 5 班	0.80	0.60	0.20

由表 7 之比較數據中，評量機制一的正確性水準大約為 69~88%，評量機制二維持在 41~60%；因此評量機制一在正確率的表現較佳，大約高出 20~47%。由此可知，評量機制一加入時間觀念的評量結果，更能反映出學生作答時間過長與信心度降低的關係，而這部分正是評量機制二所忽略的。

以大成國小 3 年甲班為例，由表 8 可知評量機制二「誤判個數」、「遺漏解答個數」的總數較高以及「方法與正確解答相符個數」總數較低；同時，評量機制一在「遺漏解答個數」總數較少與「方法與正確解答相符個數」較高。因此，評量機制一的正確性總分高於評量機制二 8.34 分，也是表 7 中二種機制平均差異達 47% 的原因所在。由於評量機制一能以時間因素推導學生信心度不足的狀態，使得學習狀態診斷結果更為符合學生實際情況，因此使「遺漏解答個數」、「方法與正確解答相符個數」和評量機制二呈現出明顯的差距。因此，可知評量機制一在評量診斷方面確實較評量機制二精確。

表 8 大成國小 3 年甲班正確性估算結果

學生	正確解答個數	A 與正確解答相符個數	B 與正確解答相符個數	A 誤判個數	B 誤判個數	A 遺漏解答個數	B 遺漏解答個數	A 正確性 ( $S_{1j}$ )	B 正確性 ( $S_{2j}$ )
S07	3	3	2	0	0	0	1	1 00	0 50
S09	2	2	1	0	1	0	0	1 00	0 00
S11	5	4	4	0	0	1	1	0 60	0 60
S13	3	3	2	0	1	0	0	1 00	0 33
S33	3	3	3	0	0	0	0	1 00	1 00
S41	5	5	2	0	0	0	3	1 00	0 00
S43	2	2	0	0	0	0	2	1 00	0 00
S46	4	3	4	0	0	1	0	0 50	1 00
S56	1	1	0	0	1	0	0	1 00	0 00
S63	6	4	2	1	1	2	4	0 17	0 00
S79	3	3	3	0	0	0	0	1 00	1 00
S86	1	1	1	0	0	0	0	1 00	1 00
S88	2	2	1	1	0	0	1	0 50	0 00
S91	3	3	3	0	0	0	0	1 00	1 00
S97	1	1	0	0	0	0	1	1 00	0 00
S102	2	2	0	0	0	0	2	1 00	0 00
S105	2	2	2	0	0	0	0	1 00	1 00
S108	1	1	0	0	0	0	1	1 00	0 00
<b>總個數</b>	<b>49</b>	<b>45</b>	<b>30</b>	<b>2</b>	<b>4</b>	<b>4</b>	<b>16</b>	<b>15.77</b>	<b>7.43</b>

註 此處 A 代表評量機制一「融合時間因素之信心診斷度評量機制」，B 代表評量機制二「不考慮時間因素之信心度診斷評量機制」

## 陸、結論

本研究嘗試提出一個改良式診斷評量的方法；我們以答題信心度為基礎，並融入答題時間參數的考量，來改善原來以學生自評為唯一考量的診斷評量機制，以期更準確地分析學生在學習上可能遭遇的困難，尤其是混淆的觀念。為了驗證這個新的評量機制的成效，我們以中文字辨識為應用，針對國小學童進行實驗；結果發現本研究提出之「考慮時間因素之信心度診斷評量機制」較原有以學生自評為唯一考量的機制能提供更精確的診斷結果。

本研究提出的改良方法，在開發為資訊系統後，將有助於協助瞭解個別學生發現混淆的觀念，例如不同英文介係詞的使用方式及時機、中文同音異型或意義相同的字詞、甚至其他學科可能的混淆的概念。依本方法開發的系統，可支援教師在課堂授課時瞭解學生在學習方面的問題，以協助個別學生進行補救學習，也可由全班學生的問題分佈狀況，來瞭解教授課時需要加強的地方。另外，本方法也可建置到教學系統中，以分析學生的學習狀況，並配合

教學系統直接提供個人化補救學習的管道。

由於線上補救教學提供一個有效解決學習差異的管道，同時也可減輕教師的教學負擔，因此目前已受到廣泛的重視。未來我們規劃將本研究提出的評量機制應用於不同學科，包括數學、英語、自然科學及工程類的課程等，以分析這樣的評量方式在不同性質的課程中可發揮的功效，並作為後續改善評量機制的依據。

## 參考文獻

- 王錦如(2000)。國小學童學習困難的診斷與輔導—運用動態評量模式。台中師範學院國民教育研究所碩士論文，未出版，台中市。
- 丘慶鈴(2003)。避免小學生寫錯別字之教學策略。新竹師範學院臺灣語言與語文教育研究所碩士論文，未出版，新竹市。
- 行政院教育改革審議委員會(1996)。教育改革總諮議報告書。台北市 行政院教改會。
- 余民寧(2001)。教育測驗與評量(6版)。台北市 心理出版社。
- 吳裕益、王佳文(1997)。國小數學診斷測驗之編製。載於教育測驗新近發展趨勢學術研討會論文集(pp 193-215)，台南市，台南師範學院。
- 李白芬(1995)。國小學童認知風格及其錯字錯誤類型之關係研究。國立嘉義師範學院初等教育研究所碩士論文，未出版，嘉義市。
- 林秀如(2004)。以資料探勘發掘學習錯誤概念之關聯的預防教學—以英語介係詞為例。中華大學資訊管理學系碩士論文，未出版，新竹市。
- 林昌炫(1979)。國中學生作文錯別字分析。中國語文，267，39-50頁。
- 胡依斐(1988)。編訂兒童各階段國語發展診斷測驗擬議。輔仁大學語言學研究所碩士論文，未出版，台北縣。
- 高廣孚(1997)。教學原理。台北市 五南圖書出版有限公司。
- 符顯仁(1983)。中國文字面面觀(2版)。台北市 莊嚴出版社。
- 莊明廣(1994)。國民小學五年級學生國語科習寫的錯別字研究。載於進修(推廣)部學士學位進修班獨立研究專輯第三輯(191-215頁)。
- 郭生玉(1997)。心理與教育測驗(9版)。台北市 精華書局。
- 陳玫秀(2000)。國民中學資源班學生數學科診斷測驗編製之研究。國立高雄師範大學特殊教育學系碩士論文，未出版，高雄市。
- 陳金明(1995)。識字教學與兒童認知發展。載於第一屆小學語文課程教材教法國際學術研討會論文集(pp 363-373)。臺東市，國立臺東師範學院。
- 陳長春(1992)。加強補救教學的意義。中縣文教季刊，13，17-18。
- 陳姿君(2002)。國中英語閱讀能力診斷測驗編製之研究。彰化師範大學特殊教育學系在職進修專班碩士論文，未出版，彰化市。
- 陳英豪、吳裕益(1995)。測驗與評量。高雄市 復文圖書。
- 陳順宇(1996)。統計學。台北市 華泰書局。
- 曾清文(2004)。國民小學四年級學生數學領域乘法能力診斷測驗之探討。國立台中師範學院教育測驗統計研究所碩士論文，未出版，臺中市。
- 歐用生(1989)。質的研究。台北市 師大書苑。
- 鄭美芝(2002)。國民小學低年級注音符號能力診斷測驗與補救教學效益之探討。臺南師範學院國民教育研究所碩士論文，未出版，臺南市。

- Anderson, J R (2000) *Cognitive Psychology and Its Implications* (5th ed ) New York Worth Publishing
- Britton, B K & Tidwell, P (1995) Cognitive Structure Testing A Computer System for Diagnosis of Expert – Novice Differences In P D Nichols, S F Chipman, & R L Brennan (Eds ) *Cognitively Diagnostic Assessment* (pp 251-278) Hillsdale, NJ Erlbaum
- Corbett, A T, Anderson, J R , & O'Brien, A T (1995) Student Modeling in the ACT Programming Tutor In Paul D Nichols, Susan F Chipman, & Robert L Brennan (Eds ) *Cognitively Diagnostic Assessment* (pp 19-41) Hillsdale, NJ Erlbaum
- Gitomer, D H , Steinberg, L S , & Mislevy, R J (1995) Diagnostic Assessment of Troubleshooting Skill in an Intelligent Tutoring System In Paul D Nichols, Susan F Chipman, & Robert L Brennan (Eds ) *Cognitively Diagnostic Assessment* (pp 73-101) Hillsdale, NJ Erlbaum
- Hwang, G H , Chen, J M , Hwang, G J , & Chu, H C (2006) A time scale-oriented approach for building medical expert systems *Expert Systems with Applications*, 31(2), 299-308
- Hwang, G J (2003) A concept map model for developing intelligent tutoring systems *Computers & Education*, 40(3), 217-235
- Hwang, G J (2007) A gray forecast approach for developing testing and diagnostic systems *IEEE Transactions on System, man and Cybernetic Part C* , 37(1), 98-108
- Langley, P, Wogulis, J , & Ohlsson, S (1990) Rules and principles in cognitive diagnosis In N Frederiksen, R Glaser, A Lesgold, & M G Shafto (Eds ) *Diagnostic Monitoring of Skill and Knowledge Acquisition* (pp 217-250) Hillsdale, NJ Erlbaum
- Lo, J J , Wanga, H M , & Yeh, S W (2004) Effects of confidence scores and remedial instruction on prepositions learning in adaptive hypermedia *Computers & Education*, 42(1), 45-63
- Marshall, S P (1995) *Schemas in Problem Solving* Cambridge, MA Cambridge University Press
- Mislevy, R J (1995) Probability-Based Inference in Cognitive Diagnosis In P D Nichols, S F Chipman, & R L Brennan (Eds ) *Cognitively Diagnostic Assessment* (pp 43-71) Hillsdale, NJ Erlbaum
- Nichols, P D (1994) A framework for developing cognitively diagnostic assessment *Review of Educational Research*, 64(4), 575-603
- Shortliffe, E H , & Buchanan, B G (1975) A model of inexact reasoning in medicine *Mathematical Biosciences*, 23, 351–379
- Snow, R E , & Lohman, D F (1989) Implications of cognitive psychology for educational measurement In R L Linn (Ed ) *Educational Measurement* (3rd Ed ) (pp 263-331) New York Macmillan

## 致謝

本研究部分由中華民國行政院國家科學委員會補助，計畫編號 NSC 95-2524-S-024-002 及 NSC 95-2520-S-024 -003 。

## 作者簡介

黃國禎，國立臺南大學數位學習科技系，教授/理工學院院長

Gwo-Jen Hwang is a professor of Department of Information and Learning Technology and Dean of the College of Science and Engineering, National University of Tainan, Taiwan

E-mail gjhwang@mail.nutn.edu.tw

朱蕙君，國立臺南大學數位學習科技系，博士班研究生

Hui-Chun Chu is a doctoral student of Department of Information and Learning Technology, National University of Tainan, Taiwan

E-mail carolchu@mail.nutn.edu.tw

王榕榆 暨南國際大學資訊管理學系，碩士班研究生

Rung-Yu Wang is a graduate student of the Department of Information Management, National Chi Nan University, Taiwan

E-mail dggbdok@yahoo.com.tw

收稿日期 95 09 05

修正日期 96 06 08

接受日期 96 11 15

# Development of an On-Line Learning-Problem Diagnostic System Using a Confidence-Scoring Mechanism

Gwo-Jen Hwang Hui-Chun Chu

Department of Information and Learning Technology  
National University of Tainan

Rung-Yu Wang

Department of Information Management  
National Chi Nan University

## Abstract

In conventional testing systems, students usually take a test by selecting one or more answers for each test item. Nevertheless, the confidence degree of individual student in answering each test item is usually ignored, which might affect the accuracy of diagnosing student learning problems. Researchers have attempted to cope with this problem by developing test systems that allow students to show their degree of confidence regarding the answer to each test item. Such a confidence scoring mechanism has been proved to be effective. However, the existing confidence scoring mechanism is mainly based on the subjective judgments of the students, which may reduce the degree of diagnostic accuracy. In this study, we attempt to develop a new confidence-scoring model by taking the "time spent answering each test item on the Internet" factor into consideration while determining the confidence degree for individual students to answer each test item. To evaluate this innovative approach, an experiment has been conducted using a Chinese word-identification course. Experimental results showed that our approach can provide more accurate learning suggestions to students, and hence is more helpful in improving their learning performance than previously proposed approaches.

**Keywords:** Chinese language instruction, remedial instruction, diagnosis test, computer-assisted learning, web-based learning