

採用開放式申論題之教學輔助系統的回饋時機

Feedback Timing for Intelligent Tutoring System Based on Open-Ended Questions

廖君興 李蔡彥
國立政治大學資訊科學系
{g9430, li}@cs.nccu.edu.tw

張俊彥
國立台灣師範大學科學教育研究所暨地球科學系
changcy@ntnu.edu.tw

【摘要】本研究在實作一個以開放式問答題的作答方式為基礎的線上輔助學習系統，藉以提升評量及學習的效果。我們以地球科學中極光的發生為例，將問題拆解細分成許多小概念，建立階層化使用者知識模型，以評估學生於各個概念的分別表現。此類系統的目標希望能在學生作答時，即時分析學生的答案，嘗試協助學生思考原本未能答好的概念，在適當的時機提供回饋。在過去的文獻中的相關的研究，多缺乏針對此類問題，分析學生之答題風格及最佳之回饋時機。本論文除了實作了一個能即時分析開放式答案之系統外，並以實驗及個案分析的方式探討此類系統在設計上可採用的回饋時機。

【關鍵詞】 自動評分, 開放性問題, 使用者模型, 智慧型教學系統

***Abstract:** In this work, we attempt to design an intelligent tutoring system based on open-ended questions to improve the learning and evaluation effects. We adopt the occurrence of polar light in Earth Science as an example of open-ended question. The answer to this question is decomposed into several concepts requiring knowledge from different fields. The proposed system aims to provide real-time feedback to students based on the open-ended text that a student has entered. In the literature, not much research has studied the answering styles and best feedback timing for this kind of open-ended questions. In addition to implement an intelligent tutoring system that can analyze open-ended answers in real time, we also conducted experiments to study the appropriate feedback timing for users.*

Keywords: Automatic Grading, Open-ended Question, User Modeling, Intelligent Tutoring System

1. 前言

數位學習的發展已行之有年，而且有許多學校都已利用數位教學平台協助學生學習，但美中不足的地方是，大部分的測驗都還是以課堂上的紙筆測驗為主，線上測驗為輔助；而線上測驗所包含的題型則多半還是以是非題、單一選擇題或多重選擇題為主，利用開放式申論題為主的測驗形式仍然較為少見。形成這種現象有一個主要的原因在於批改所需花費的成本。一般而言，以人力來對開放式申論題的答案進行評分，是件十分

辛苦費時的工作。如採電腦閱卷，則仍是相當具挑戰性的工作，主要的原因在於這個自動評分的工作需要用到自然語言處理以及學習者建模(User Modeling)等技術，而這些技術仍未臻成熟，信效度有待檢驗。儘管如此，我們仍認為有必要進一步研究如何將開放式申論題的測驗形式加入線上學習系統，使開放式申論題成為線上學習或測驗的一部分。設計一個可信賴的開放式問題自動評分系統，將是能否將開放式申論題加入線上學習的關鍵之一。我們認為這種具有人工智慧的自動化評分系統，不管是對教學者或是學生來說，都應有相當的助益。

在以往開放式申論題的測驗中，一道題目通常會隱含許多不同的概念。有的時候學生會因為緊張，或是對題目的理解度不足，導致在答案中缺乏了對某個概念的說明，或是對概念的描述模糊不清。負責評分的老師或是研究人員往往無從得知學生只是因為表達能力不佳而描述不清，還是因為對這個概念不夠了解，所以模糊帶過。在以輔導學生學習為主的評量測驗中，如能了解學生的問題所在，進而加以輔導，將進一步提升測驗的效益。然而，在現今的測驗模式之下，除非個別指導，老師或是研究人員只能根據答案中其他概念的描述，來推斷學生是否具備某特定概念。換言之，在傳統的測驗或學習模式之中，老師通常只有學生最後回答出的答案，而無法取得學生的思考歷程及推理邏輯，更無法即時在學生思考如何應答的過程中給予適當的引導。

為了解決這樣的問題，我們在之前發展的線上開放式申論題測驗系統(Huang et al., 2006)中，加入了即時分析與回饋的機制設計。這個系統提供了額外的兩項協助：(1) 當學生在線上作答的同時，系統每隔一段時間便會即時回傳學生目前的答案至伺服器端，紀錄學生的答題歷程，以供後續分析與回饋；(2) 系統會對當前學生的答案作各個子概念的預先評分，系統便會根據目前學生所獲得的分數判斷目前學生所缺乏的概念。我們將在本論文中描述系統與實驗設計及初步實驗的結果。

2. 相關研究

開放性問題之測驗方式的目的之一，在提升學生創造性問題的解決能力，尤其是要評估學生在回答問題背後所具備的能力。Treffinger 等人(2008)指出，要研究 CPS 和問題解決風格(problem solving style)，我們必須先定義清楚問題解決的本質及它和創意間的關聯。「問題」是一個位於我們所在的位置以及我們想要去的地方之間的障礙(Johnson, 1972)，而「問題解決」是指我們尋求要如何去到我們想去的地方所產生的思考和行為。一個問題領域或任務的定義可以是明確界定出它的結構，也可以是模糊不清的。模糊不清的問題需透過創造利解決問題或發現問題。

要能適時的協助學生解決問題、瞭解學習概念，我們必須建立使用者的模型。而使用者模型的建立，是許多智慧型人機介面或智慧型教學系統必備的模組。早期的使用者模型，依賴著由人手動建立的知識庫來儲存猜測的資訊和觀察得來的行為，並手動的建立出它們之間的對應關係。近年來，複雜人類行為的模塑方法逐漸走向統計模型的研究方向，預料人類行為的某些方面，如目標、行動及偏好等。UPSAM(User Problem-Solving Ability Modeler)(Wang et al., 2005)也是其中的一項研究。UPSAM 的目的，是在自然科學的領域中建立並維護一個專家的知識模型。這個知識模型裡包含了專家本身具備的知識，以及解決問題可能會用到的字彙集合。UPSAM 透過這個知識模型，來進行對學生在開放式申論題的答案進行自動評分。

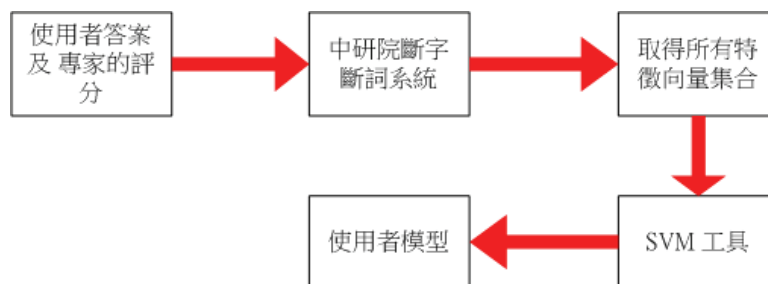


圖 1 建立使用者知識模型的流程

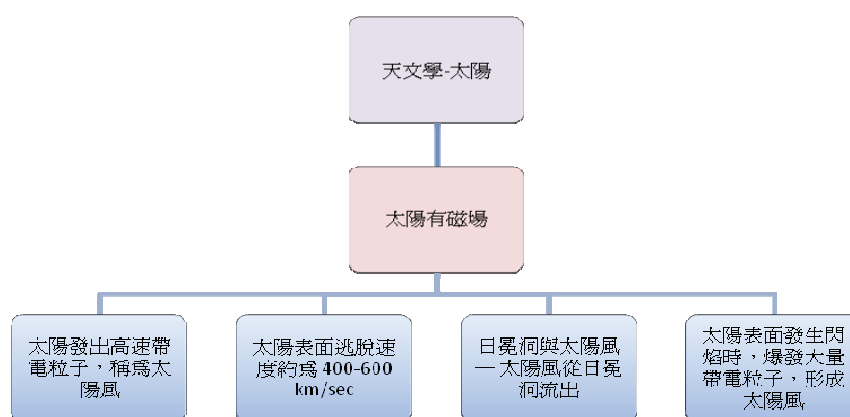


圖 2 天文學－太陽的概念樹模型圖

Idea Storming Cube(Huang et al., 2008)是在提供一個基於網際網路的協同式創意學習工具，讓學生進行有趣的腦力激盪。透過類似遊戲進行的模式，它提供了學生創意性思考的支援，讓學生們有多面向思考的習慣。而 WOTAS(Web-based Online Testing and Analysis System)(Huang et al., 2006)，是一個以 UPSAM 的評分功能為基礎，基於 web 的線上測驗和分析的系統，替學生在開放式問答測驗中的表現及回應進行評分。除了評分之外，WOTAS 同時也讓教師和研究人員可以對測驗的結果進行分析。

VIBRANT(Wang et al., 2006)則是一個在科學教育方面，和學生進行腦力激盪的教學系統。這個研究主要在設計一個讓學生可以腦力激盪之教學系統，以支援解決質性問題。認知導向與社會導向的支援則是藉助於過去的一些經驗法則所設計產生的一些回饋。VIBRANT 將 ITS 研究的領域更進一步推向開放式的科學問題的解決，也就是在教育上越來越受重視的創造性解題模式。VIBRANT 同時也著重於支援產生概念的系統行為，為腦力激盪回饋提供了一個介面，讓學生們得到相關學習資源的通盤性觀點。

3. 系統設計

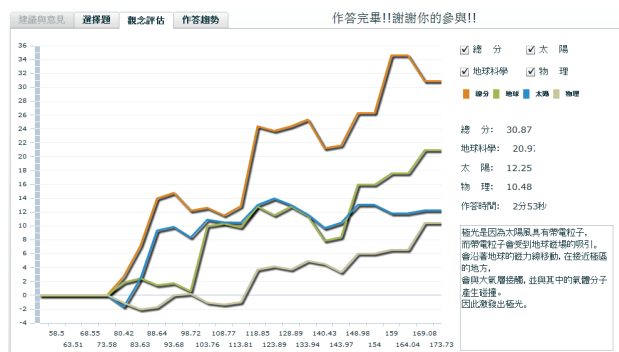
本論文所使用的系統，主要是由前端的使用者介面及後端的幾個模型所組成。前端的使用者介面裡包含了定時器(Timer)，定期將使用者目前的答案傳送給後端的使用者知識模型進行評分，並將分數傳遞給互動模型中的答題風格模型，決定是否需要回饋。如果需要回饋，則再交由階層式概念模型來決定要回傳何種回饋訊息，使用者介面再將其顯示出來。

1. 請說明地球極光發生的機制，為什麼只能在極區看見極光？

作答區域:

極光是因為太陽風具有帶電粒子，而帶電粒子會受到地球磁場的吸引。會沿著地球的磁力線移動，在接近極區的地方，會與大氣層接觸，並與其中的氣體分子產生碰撞。因此激發出極光。

送出



(a)

(b)

圖 3 學生在作答時的(a)輸入介面及(b)作答完所看到的評分結果

伺服器接收到前端使用者介面傳過來的使用者答案之後，會交給使用者知識模型進行評分的動作。圖 1 中簡單的說明了建立使用者知識模型的流程。本研究中採用的開放式簡答題，是第一屆國際地球科學奧林匹亞台灣區競賽時曾經使用的題目：「為什麼只有在極區才能看見極光？」。我們透過過去測驗的數據，取得學生對此問題的答案以及專家為這些答案所給的分數。我們將這些歷史答案透過中研院的斷字斷詞系統進行字詞分析後，得到許多分割開的單字或詞組集合。我們會給予這個集合中的每個元素一個代碼，以供日後預測使用者分數時使用。每一使用者答案均被視為一個特徵向量，我們將此特徵向量及其得到的分數交給以 SVM (Support Vector Machine) 方法建立的機器學習工具，成為我們所需要的使用者知識模型。

在先前的研究中，我們已經為這個問題建立了一個使用者的知識模型(Huang et al, 2006)。本研究除了接續使用此使用者知識模型外，並另外將回答此問題所需的科學概念，分為地球科學、天文學—太陽、及物理知識三個方面的子概念，分別建立使用者知識模型。在建立使用者知識模型的時候，我們依據概念學習的先後順序，來建立階層化的使用者知識模型。圖 2 為專家所提供階層化概念的範例，而我們根據這些概念建立了階層化的概念樹。其中，對各個概念樹的根節點，我們利用機器學習的方法，進一步建立使用者的知識模型。建立方法和整體知識模型的建立方法相同，唯一不同的部分在於專家替使用者答案進行評分時，是分別以地球科學、天文學太陽、物理知識模型來進行評分。也就是說同一個使用者的答案會有四個不同的分數。在本系統中，使用者整體知識模型評估出來的分數，可用來決定何時提供回饋；而要回饋的內容，則是由三個概念模型的分數來決定。

4. 實驗設計

本系統是以 Web 應用程式的方式開發，因此使用者可以從任何可以連上網際網路的地方使用。使用者端的程式是以 Flex 語言開發，機器學習部分則採用 weka 系統。針對使用者知識模型的建立，我們根據專家對正式測驗的評分結果進行學習。我們對於系統中使用的四個使用者知識模型進行了交叉驗證。將所有資料切成 10 等份，每次 9/10 用來訓練模型、1/10 的資料用來做測試。做 10 次之後取其平均，各個知識模型所得到的相關係數在.70 到.83 的範圍間，最低的是物理知識模型為.70，地球科學與太陽為.72 及.74，最高的是整體模型為.83。

表 1 受測學生依照思考時間分群組的數據表

	人數	平均分數	分數標準差	平均思考秒數
30 秒以下	14	2.36	1.15	11.46
30~60 秒	6	2.68	1.25	40.02
60 秒以上	5	2.74	1.99	94.14
總體	25	2.51	1.32	34.85

表 2 思考時間較長的學生，作答時最先輸入的句子及其獲得的分數表

學號	答案	思考時間	單句分數	最後分數	作答時間
9716	那是因為我們居住的星球之磁力線引導之進入	70	0.36	0.43	82.5
9722	因為	60	0.06	1.54	191.13
9723	太陽會發出一些帶電粒子。	190	0.94	3.62	411.29
9727	當太陽風內的待電力子穿過范艾倫帶時，	70	1.60	5.60	532.74
9736	因為太陽風被磁層引至極區。	80	1.85	2.52	112.66

為了瞭解學生作答過程中是否會特定的風格或是情境，我們設計了實驗來收集學生答題風格的資料，並且對實驗所獲得的資料進行分析，以找出我們想得到的資訊，包含可能需要提供回饋的時機點，以及提供回饋的類型。

本次實驗邀請台北市立某高中 25 位高一的學生作為實驗對象。在實驗中，我們想要找出學生在回答問題時，是否會有些固定的行為模式，而這些行為模式是否可以做為我們提供回饋的依據。實驗時間為一節課（50 分鐘）。施測內容包含極光先備知識（10 分鐘）、第一次簡答題（至少 5 分鐘）、及提供測驗結果。

圖 3(a)為實驗時學生作答的介面，學生可以在作答區域內鍵入他們的答案，當學生按下送出按鈕時，表示已經作答完畢，答案將被送到後端伺服器。之後，學生可以看到他們在這次測驗中各個概念以及總和的電腦評分。在學生作答時，每隔 5 秒系統會抓取作答區域內的答案，傳送到後端伺服器。伺服器端收到學生的答案之後，會利用之前建立的階層式概念模型為學生的答案評分，並且將答案及分數儲存起來做為日後分析之用。在作答完畢之後，學生可以看到自己所得到的分數曲線，以及在不同概念裡所得到的評分(如圖 3(b)所示)。

5. 實驗結果

為了找出回饋的時機點，我們嘗試找尋學生在作答過程中有可能遇到困難的情況。為了找到這些時機點，我們針對之前 25 位學生的資料進行個案分析。首先，我們注意到有些學生在作答的時候，會思考比較久才開始作答，而有些學生則是很快就開始作答。我們想知道思考比較久才開始作答的學生，他們在作答過程中的行為是否會與較快開始作答的學生有所不同，而最後的所得到的成績是不是也會有差異。

表 1 的分類是依據學生在作答一開始的思考時間，並且計算學生們的平均分數。思考了一分鐘以上才開始作答的學生有五位，一分鐘內就開始作答時共有 20 位。學生面

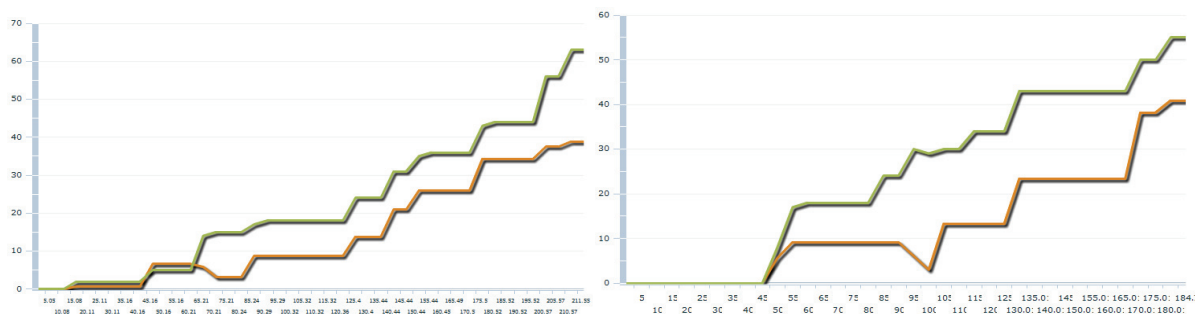


圖 4 為學生 9739(左)及 9735(右)的作答過程曲線，綠線為字數變化、橘線為分數變化，橫軸為時間

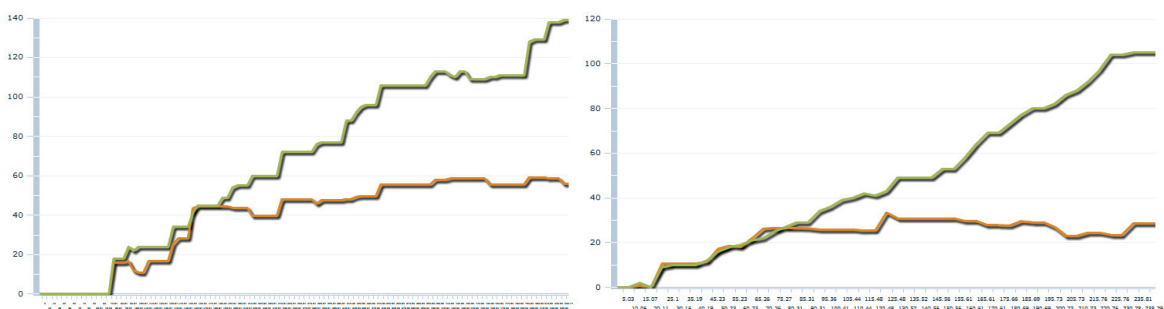


圖 5 為學生 9727(左)及 9731(右)的作答過程曲線，綠線為字數變化、橘線為分數變化，橫軸為時間

對題目卻不開始作答，我們假設有兩種不同的情形，一種是學生正在思考該怎麼作答可以寫的比較流暢，或者是該怎麼說明自己的想法，而另一種則是不知道答案，當然也就不知道該怎麼回答。

我們想知道在思考比較久的學生裡，表現較差和表現不錯的學生在作答的過程中有沒有甚麼差異。表 2 列出了這五位學生在作答時第一個傳回系統的答案，也就是學生在經過思考後開始作答的第一句話，以及系統對於第一句話的評分還有學生最後的表現和總共作答時間。從這些資料中可以看出，最後的表現中等以上的三位學生(9723, 9727, 9736)，在作答時第一句就會有不錯的分數；相反的，表現比較差的兩位學生則差強人意。其中，9716 的答案則很明顯的可以看出並沒有回答到重點，而扣除了他的思考時間之後，其實他真正花在作答上的時間，只有 12.49 秒。從這些觀察中，我們發現回饋設計是很重要的，而且回饋的時間必須要在學生還在作答的時候才有意義，因為要是學生很快的放棄作答，那麼系統能給予的幫助就非常有限了。

一般而言，學生在作答的時候，分數應該隨著學生輸入的字數而慢慢的增加，如圖 4 中學生 9739 及學生 9735 的作答過程的曲線圖，橫軸代表的是時間，綠色的線是字數的變化，橘色的線代表了分數的變化。當學生輸入的字數增加，但分數卻未相應的增加時，這個時候學生可能正是遇到了瓶頸。如圖 5 中，學生 9727 及學生 9731 正是這類型學生作答過程的曲線圖。這兩位學生作答到了中後期之後，雖然作答字數持續有再增加，但是分數卻未相對應的增加。

6. 回饋設計建議

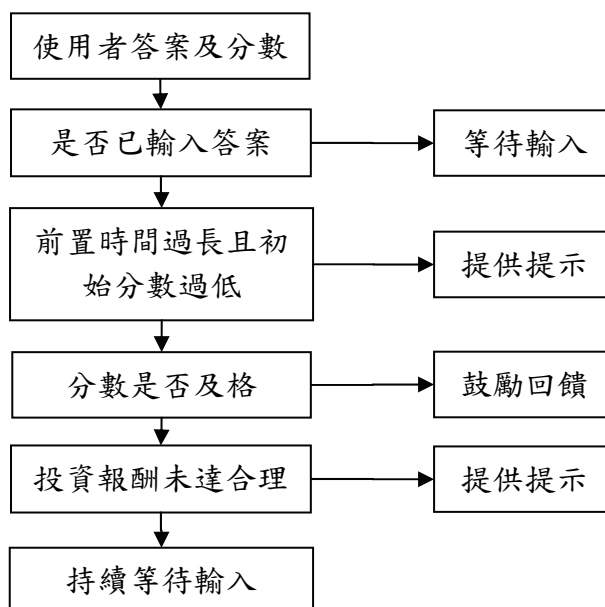


圖 6 使用者風格模型判斷之流程圖

從上面的例子來看，系統提供回饋與否，與學生分數的變化有相當的關係。因此，我們假設一段時間內學生的分數如果沒有增加，可能就是需要協助。但是一段時間應該是要間隔多久，而分數增加的比率又要是多少會比較合理呢？為了掌握這兩個關鍵點，我們對 25 位同學的作答過程進行分析，主要是看在多少的時間區間與分數變化下給予回饋是比較適當的時機點。在給予回饋的時候，會偏向於不要太快就給予學生回饋，而又會希望不會漏掉那些真正需要回饋的學生為主。我們將時間間隔的長度及應該增加的分數多少當成兩個參數，時間間隔的範圍是從 20 秒到 120 秒；應該增加的分數從 0.1 到 2.0 分。用不同的參數去測量在這樣的條件下，有多少的學生能得到系統的回饋，而平均給予第一個回饋的時間又是在何時。另一方面，我們希望能給予學生多一點的作答時間，避免太早給予學生回饋，中斷學生思考。根據對實驗個案的觀察與分析，我們認為在此實驗案例中，將回饋時間設在 65 秒，而在分數差異設在 1.2 分為較佳。也就是說，每隔 65 秒學生的分數至少應增加 1.2 分。如否，則系統應可以進行回饋。

根據我們對個案數據的觀察及上述的討論，我們對使用者風格模型的建立與系統回饋時機可以有進一步的設計。在使用者答題風格的模型中，我們可以觀察到下列兩件事：第一是使用者是否於一開始作答時就遭遇了瓶頸，第二是學生在作答時的投資報酬率是否有達到我們預期的水準。我們所設計的回饋時機判斷流程圖如圖 6 所示。我們假設學生一開始就可能遇到瓶頸的情況，也就是圖 6 中「前置時間過長且初始分數過低」的階段。若是學生遇到了瓶頸，則系統可以馬上提供回饋。否則，系統將進一步根據分數是否達到及格標準，如是，則系統將會提供鼓勵的訊息，讓學生知道他已經答得不錯。若尚未及格，則將進一步觀察答題的投資報酬率是否達到標準。我們所定義的時間報酬率為「分數變化/單位時間」，也就是說固定時間內的分數變化比率。如果投資報酬率高於標準，那麼系統就可以知道使用者是朝正確的方向再作答，所以可以繼續的讓使用者作答下去，而不需要提供回饋。相反的，如果投資報酬率低於標準，那麼使用者目前很有可能是遇到困難，在這個時候系統便可給予使用者提示。

7. 結論

我們設計了一個以開放式申論題為基礎的線上學習輔助系統，用以提升學習的效果。藉由將開放式問題所涉及的概念細分成許多概念組合，並建立階層化使用者知識模型以進行評估，讓系統得以了解學生於各個概念的表現。同時我們也對於各個概念設計相對應的提示，在學生陷入瓶頸的時候能夠給予適當的協助。我們也設計了實驗來收集學生在作答時的答題風格資料，並且嘗試找出我們認為理想的回饋時機點。根據這些收集到的數據，我們提供了系統進一步設計回饋機制的建議。

在未來的研究中，我們將針對上述歸納建議的回饋時間，搭配系統對子概念自動評分的結果，回饋使用者較為不足的內容，進而設計成一個能自動回饋的智慧型教學系統。我們將對此系統設計進一步的實驗，瞭解回饋時機是否有個別化的差別，而回饋的方式是否有不同的效果。

參考文獻

- Huang, C.-C., Chang, C.-Y., Li, T.-Y. & H.-C. Wang (2008) *A Collaborative Support Tool for Divergent Thinking: Idea Storming Cube*. Presented at 2008 Annual International Conference of National Association for Research in Science Teaching.
- Huang, C.C., Wang, H.C., Li T.-Y. & Chang, C.Y. (2006) An Online Testing and Analysis System for Students' Creative Problem-Solving Ability in Sciences. Presented at the Tenth Global Chinese Conference of Computers in Education, China.
- Johnson, D. M. (1972). *Systemic introduction to the psychology of thinking*, New York: Harper & Row.
- Liao, C.-H., Li, T.-Y., Chang, C.-Y. (2009) *Inferring Appropriate Feedback Timing from Answering Styles for Open-Ended Questions*. Presented at the 17th International Conference on Computers in Education, Hong Kong.
- Treffinger, D. J., Selby, E. C. & Isaksen, S. G. (2008) Understanding individual problem-solving style: A key to learning and applying creative problem solving. *Learning and Individual Differences, Elsevier*.
- Wang, H.C., Li, T.Y. Huang, C.C. & Chang, C.Y. (2006) VIBRANT: A Brainstorming Agent for Computer Supported Creative Problem Solving. PM. Ikeda, K. Ashley, and T.-W. Chan (Eds.), *LNCS 4053*, pp. 787 – 789.
- Wang, H.-C. Li, T.-Y. & Chang, C.-Y. (2005) *A user modeling framework for exploring creative problem-solving ability*. Presented at 12th International Conference on Artificial Intelligence in Education.