

ISSN 2306-5966

教育人力與專業發展

educators and professional development

雙月刊

2014年6月15日
第31卷 第3期

數位學習與評量

教室內數位閱讀學習與評量

謝進昌

國家教育研究院測驗及評量研究中心副研究員

根據 Internet World Stats (2012) 的統計，全球有近三分之一人口，超過 24 億的網路使用人口，就台灣而言，有高達 75% (約 1753 萬) 網路使用人口。上網者大多是以搜尋資訊、瀏覽網頁訊息、收發電子郵件、網路社群參與等占多數，而這些行為多得依靠閱讀來進行訊息擷取、理解。雖然，新一代族群經常接觸這些網路、數位科技，但並不代表他們具有良好的網頁導航或訊息搜尋、來源評估技能 (Henry, 2006)，網路或數位環境的閱讀就像是一把雙面刃 (林珊如, 2010)，是需要透過良好教學、評量策略來培養學生數位閱讀能力。

網際網路是 21 世紀全球化經濟發展的核心要素，因應全球知識經濟時代的演變，學校有責任確保每位學生有能力進行數位環境的閱讀，而數位閱讀不單僅從靜態紙本獲取資料轉為動態連結，它更強調心智模式、新素養建立，關心學生適應網路新科技變化的學習能力，著重如何透過數位閱讀來學習 (digitally read to learn)、甚至是透過數位閱讀來尋找答案或解決問題 (digitally read to problem solving) 的能力。整體而言，本文目的在導引讀者了解數位閱讀理解獨特性及例舉範例說明教師如何在教室內執行數位閱讀學習與評量，以為實務教學參考。

一、紙本與數位閱讀差異

數位閱讀 (digital reading) 概念是雷同於線上閱讀 (online reading)、電子

閱讀 (electronic reading) 或超文本閱讀 (hypertext reading) 等，意指閱讀者透過網際網路、多媒體、搜尋引擎等科技進行閱讀行為，以達學習或者問題解決的歷程；另一方面，紙本 (或書面) 閱讀 (print reading) 或稱離線閱讀 (offline reading) 則指閱讀者僅利用紙本文本或非數位環境進行閱讀。學者主張數位閱讀理解是比紙本閱讀理解需要更多能力 (Coiro & Dobler, 2007)，而此論點也已獲得國際閱讀協會 (International Reading Association, 2009) 支持，認同數位閱讀理解的重要性及其獨特本質，其中，除了閱讀媒介差異外，數位文本 (digital text) 與紙本文本 (print text) 也具有幾項顯著結構性差異或轉換，例如：已從靜態文本走向動態視窗與界面、內容從線性的安排走向超連結的非線性形態、從文字描述走向多媒體與類現實體驗、從作者主導文本內容走向線上討論或社交場域等等 (OECD, 2011)。

隨著閱讀結構的改變，也影響閱讀者的閱讀形態，在紙本閱讀中，讀者利用目錄來擷取欲閱讀的篇章或內容，但在數位環境下，閱讀者會使用搜尋引擎或超連結功能來找到欲閱讀訊息；在紙本閱讀中，強調的多半是就單一文本內容，進行概念連結與整合，而在數位閱讀下，更強調要如何從多個數位訊息管道進行訊息整合；在紙本閱讀中，閱讀的文章多半是具有結構良好 (well-structured) 的內容，而在數位閱讀中，閱讀者較常

接觸到結構不完整 (ill-structured) 的內容；在紙本閱讀中，閱讀者多屬被動接受作者對於內容的詮釋，而在數位閱讀中，閱讀者同時能接受來自其他人對於文本內容的評論或意見 (Brown, 2002)。

二、數位閱讀理解能力

數位閱讀者所面臨的環境，不僅接收的訊息、閱讀理解歷程愈趨複雜、多元，需要額外的能力或策略，才足以因應未來環境的變化，根據 Leu 及其團隊的研究結果，提出數位閱讀理解是至少包含五個元素，也大致符合或已逐漸融入美國近年來所推廣的美國各州共同核心標準 (CCSS) (Leu et al., 2013)。它們分別是：

- (一) 擬定或掌握重要問題或議題 (identify problems or questions)：在網路閱讀環境中，問題解決是一重要核心，而閱讀者能否真正掌握或瞭解問題內涵則扮演著重要關鍵。
- (二) 找到訊息 (locate)：閱讀者要有能力去閱讀並找到符合己身或解題所需的訊息，其中，此能力包含如利用適當的關鍵字進行搜尋、從搜尋引擎 (如 Google) 檢索結果找到適當連結並做出推論、能有效從網頁內找到相關訊息等。
- (三) 批判性評估訊息 (evaluate information critically)：閱讀者要有能力去閱讀與評估網路的訊息，雖然，在紙本閱讀環境中，評鑑或評估能力本身就占有一席之地，但是網路訊息更為複雜、多元或更易存在的偏誤，因此，

網路評估的能力重要性是更甚以往，而其類型可以是評鑑網頁或數位社交工具內容是否具合理性、準確性、可信度或存在偏誤等面向。

- (四) 整合訊息 (synthesize)：就如同紙本閱讀一般，閱讀者需要整合文本內訊息或概念，但在數位環境中，閱讀者更常面臨的是多個線上訊息的整合，在閱讀期間，要不斷決定下一個閱讀網頁、下一個參訪的網站、下一個回應的訊息等，因此，良好的整合是需要閱讀者同時掌握自身閱讀歷程覺知與文本內容理解。
- (五) 溝通信息 (communication)：在數位環境中，閱讀與回應已變成不可分離的元素，閱讀者不僅要具有接收訊息能力，同時要利用如部落格、電子郵件、臉書等數位社交工具進行分享訊息的能力。

三、教室中數位閱讀學習與評量

現行臺灣對於學生數位閱讀評量的投入，是以 2015、2016 年正式參與 PISA 及 PIRLS 線上閱讀評量，但這類國際評比模式多半不適合教師於教室中執行，其因是這類評量試題都是由一群具有資訊、閱讀、認知心理、或其它學科領域等專長的團隊所執行開發 (範例題可參考網站：http://pisa.nutn.edu.tw/sample_era_tw.htm)，此舉對於學校教師而言，無異是艱難任務，但細究大型評量與班級評量模式內涵，本身就具有不同特性，前者關注群體表現，需大量施測，所以，

易使用與計分的選擇題型是偏好的選項，後者則具備學生數少的特性，因此，以關注學生個體表現、實作評量、或者課程本位延伸之班級評量及教學活動則成為可行的項目。

表 1 數位閱讀教學範例是以發展學生找到訊息的能力，其下包含利用適當關鍵字進行搜尋、從搜尋引擎檢索結果找到適當連結等，從簡短文本引起學生動機及問題，進而銜接至利用網路搜尋引擎找尋答案的情境，教學發展中，教師除要導引突顯不同關鍵字可能找到不同搜尋

結果外，更要教導學生選擇出適當的連結，而其中包含一些如從標題判斷、關鍵字會以特別顏色（形式）出現在文本中、從網址來輔助判斷（com 商業、edu 教育、org 政府組織等）等技巧，都可適當引導。最後，教師可利用分組極速搶答的遊戲來檢核學生是否習得本次能力指標，其中，搶答時間的限制是需要的，因為，網路閱讀情境著重的是培養學生在面臨大量訊息下，能快速及適當的選擇出進一步閱讀的連結。

表 1 數位閱讀教學與評量範例

| 學習領域 | 國語文、自然 | |
|------------------|--|---------|
| 活動主題 | 北極熊的抉擇 | |
| 能力指標 | 找到訊息：利用適當關鍵字進行搜尋、從搜尋引擎檢索結果找到適當連結 | |
| | 能力指標 | 說明 |
| 教 學 流 程 | <p>* 準備活動</p> <p>北極是地球最冷地方之一，是一片終年冰雪覆蓋的銀白色世界。那裡住著陸地上最大的肉食動物 - 北極熊，為了適應當地寒冷天氣，北極熊除了有很厚脂肪外，還有特別的毛皮大衣。</p> <p>北極熊最常捕捉海豹，冬天時，北極熊會找尋浮冰上海豹呼吸用的呼吸孔，連續幾小時一動也不動，靜待海豹從呼吸孔探頭呼吸時，一口捉住。對於躺在浮冰上的海豹，北極熊也有特別的狩獵技巧，它發揮自己的游泳專長，無聲無息地從水中秘密接近海豹，有時還會推動一塊浮冰作為掩護，等接近目標時，突然躍出水面，緊緊捉住毫無防備的海豹。</p> <p>近年來由於全球暖化，北極氣溫上升、冰層融化，對於北極熊的生存帶來很大危機。北極熊需要靠海上的浮冰，才能獵捕牠們最愛吃的海豹。但全球暖化讓北極溫度在 50 年內上升了攝氏 4 度，海面上的浮冰逐年減少，狩獵時間變短，食物也跟著變少了，為了生存下去，北極熊應該如何抉擇呢？</p> <p>一、老師可能導引的問題：</p> <ol style="list-style-type: none"> 1. 北極熊為何是陸地上最大肉食動物，它有多大呢？ 2. 它的毛皮大衣有何特別之處？ 3. 北極熊愛吃的食物有哪些？ 4. 為什麼北極氣溫會上升，什麼是全球暖化？ | 引起動機及問題 |

| | | |
|-------------|--|---|
| <p>教學流程</p> | <p>*發展活動 一、如果要找到這些問題的答案，您習慣使用哪一種搜尋引擎？ <input type="checkbox"/> Google <input type="checkbox"/> 雅虎 <input type="checkbox"/> 蕃薯藤 <input type="checkbox"/> 其它：_____</p> <p>二、搜尋時，您所使用的關鍵字是什麼？ (以第1個問題為例) <input type="checkbox"/> 北極熊 <input type="checkbox"/> 北極熊的身材 <input type="checkbox"/> 最大肉食動物 <input type="checkbox"/> 北極熊的特徵 <input type="checkbox"/> 其它：_____</p> <p>三、為什麼您會想使用這個關鍵字呢？ _____</p> <p>四、老師示範使用「北極熊的身材」關鍵字，找出下列結果(2014.05.26)： 1. 是不是每筆資料都能找到答案呢？哪些訊息是有用的，讓我們試著來刪除不必要訊息。有找到要的連結嗎？</p> <p>鳳凰旅遊-遇見野生北極熊+幸福極光-鳳凰旅行社 www1.phoenix.com.tw/images/polarBear/ ◀ 白絨絨胖胖的身材，北極熊可愛的模樣讓人直呼太可愛，可以在真正野生世界裡，近距離的靠近牠們，體驗真實的感動驚奇，一生肯定要去一次...</p> <p><input type="checkbox"/> 保留 <input type="checkbox"/> 刪除 因為：_____</p> <p>誰是北極冰原世界的王者？- 歡迎光臨康軒教師網知識+ www.945enet.com.tw/knowledge/poll_result.asp?SA=True&ID=3 ◀ 解析：原來是身材壯碩的北極熊！獵食季節到了，北極熊媽媽會帶領北極熊寶寶四處尋找食物，海豹是北極熊最爱的食物，為了獵食海豹，北極熊會使出最厲害的破冰...</p> <p><input type="checkbox"/> 保留 <input type="checkbox"/> 刪除 因為：_____</p> <p>德超人氣動物北極熊努特滿週歲囉-香港大紀元 www.epochtimes.com.hk/7/12/7/56169.htm ◀ 德國的超人氣動物明星北極熊努特，已經滿1週歲了！目前牠的身材壯碩，體重高達110公斤，與去年3月首次亮相的嬌小可愛模樣相去甚遠，不過仍不減德國人對牠的...</p> <p><input type="checkbox"/> 保留 <input type="checkbox"/> 刪除 因為：_____</p> | <p>從不同搜尋引擎找尋答案</p> <p>利用不同關鍵字找尋答案</p> <p>從檢索結果找出適當連結 (參考技巧： 1. 關鍵字會以特別顏色標註在文本中 2. 從標題判斷 3. 從網址來輔助判斷 (com 商業、edu 教育、org 政府組織等)</p> <p>比較不同關鍵字的搜尋結果</p> |
|-------------|--|---|

| | | |
|-------------|--|-----------------------------|
| <p>教學流程</p> | <p>五、老師使用另一關鍵字「北極熊的特徵」，找出下列結果：</p> <p>北極熊的外型特徵 - Yahoo!奇摩知識+ https://tw.knowledge.yahoo.com/question/question?qid=1008041904800 ◀ 2008年4月19日 - 說到北極熊的外型特色那當然是毛色、體型、腳掌啦! 1.毛色, 白色的毛有在雪地中偽裝的功能。2.體型, 北極熊作為最大型的陸上肉食動物, 是因為再寒冷的 ...</p> <p><input type="checkbox"/>保留 <input type="checkbox"/>刪除 因為：_____</p> <p>北極熊-中文百科在線 www.zwbk.org/zh-tw/Lemma_Show/98000.aspx ◀ 2010年12月20日 - 外形特征, 體型. 單從物種角度說, 北極熊是目前世界上最大的熊科動物, 也是最大的陸地食肉動物 (如果要計入亞種, 科迪亞克棕熊體型則更為龐大) ...</p> <p><input type="checkbox"/>保留 <input type="checkbox"/>刪除 因為：_____</p> <p>[PPT] 關於北極熊 bbti.she.nkfust.edu.tw/bioclases/biology/Homework/.../polar%20bear.pp... ◀ 北極熊無法儲備足夠的脂肪度過夏季的無冰期, 因此到了夏季末時, 北極熊會因過度消瘦而影響牠們繁殖能力。 北極熊特徵. ◦身體長度可以達到2.5米以上。 ◦體重可 ...</p> <p><input type="checkbox"/>保留 <input type="checkbox"/>刪除 因為：_____</p> <p>* 綜合活動 一、學生分組就其它問題 (如第 3 個問題) 利用關鍵字搜尋資料。並發表。 1. 曾使用哪些關鍵字? 哪個關鍵字是最適合的? 2. 找到哪些資料? 哪幾個連結是適合的?</p> <p>* 學習評估活動 - 急速搶答 (保留 V.S. 刪除) 老師想知道北極溫度上升情況, 輸入關鍵字「北極氣溫上升」, 請每一組用最快速度幫老師決定是不是要保留下來繼續閱讀這個連結? (註: 若要個別計分, 也可限時顯示 PPT, 學生個別作答) 以 PPT 逐一呈現搜尋結果, 學生分組舉牌「保留 或 刪除」, 例如:</p> <p>北極-全球暖化放大器/許冕雄研究員中央研究院環境變遷研究 ... www.tcap.ndc.gov.tw/climate-change-aspects/.../507-north-pole.html ◀ 1980年代以來, 北極溫度上升趨勢高居全球第一, 全球平均氣溫大約上升0.4C, 北極平均溫度則上升了約1.1C, 將近3倍之多。相同的現象也發生於1920-1940年的暖 ...</p> <p style="text-align: center;"> </p> <p>為何臭氧洞出現在南極上空 - IPCC www.gcc.ntu.edu.tw/chinese/Education/30subjects/ozone%20hole.htm ◀ 加上南極高原本來就海拔高氣溫低, 因而形成極低的低溫環境。臭氧層所在的20公里高度上氣溫常常在-80℃以下 (比北極要低得多)。南極大氣渦旋中的空氣上升過程 ...</p> <p style="text-align: center;"> </p> | <p>學生分組練習</p> <p>學生分組練習</p> |
|-------------|--|-----------------------------|

四、結語

數位閱讀理解能力的培養是逐漸受到國內重視，除了國際大型評比計畫的參與，教育部也積極推廣一系列研習課程，除以課文本位為主體的閱讀理解教學（課文本位的閱讀理解教學網頁，2014）外，PIRLS（或稱 ePIRLS）團隊也於未來規劃「數位閱讀素養課程研習手冊」（課程研發小組，2014年4月），企圖透過暑

期研習方式，提昇教師的數位閱讀教學知能。本文希冀作為一個觸發，除簡要說明數位閱讀內涵外，更期望引起教師對於數位閱讀教學重視及不斷反思各種相關議題，何謂數位閱讀能力？與紙本閱讀能力的差異為何？等，只有教師真正掌握其內涵後，才得以編擬出適合自身班級學生的數位閱讀教學活動。

參考文獻

- 林珊如（2010）。數位時代的閱讀：青少年網路閱讀的爭議與未來。《圖書資訊學科》，8（2），29-53。
- 課文本位的閱讀理解教學網頁（2014）。課文本位的閱讀理解教學網頁。2014.05.15 檢索自：
http://140.127.56.86/pair_System/Search_index.aspx?PN=PlanInfo
- 課程研發小組（2014年4月）。數位閱讀素養課程研習手冊。PIRLS 會議資料。
- Brown, G. J. (2002). Beyond print: Reading digitally. *Library Hi Tech*, 19(4), 390-399.
- Coiro, J., & Dobler, E. (2007). Exploring the online comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, 42(2), 214-257.
- Henry, L.A. (2006). SEARCHing for an answer: The critical role of new literacies while reading on the Internet. *The Reading Teacher*, 59(7), 614-627.
- International Reading Association. (2009). *New literacies and 21st-century technologies: A position statement of the International Reading Association*. Retrieved June 1, 2013, from http://www.reading.org/Libraries/position-statements-and-resolutions/ps1067_NewLiteracies21stCentury.pdf
- Internet World Stats (2012). *World internet users and population stats*. Retrieved May 26, 2014 from the world wide web: <http://www.internetworldstats.com/stats.htm>
- Leu, D. J., Forzani, E., Burlingame, C., Kulikowich, J., Sedransk, N., Coiro, J., & Kennedy, C. (2013). The new literacies of online research and comprehension: Assessing and preparing students for the 21st century with Common Core State Standards. In S. B. Neuman, & L. B. Gambrell (Eds.), *Quality reading instruction in the age of Common Core Standards*(pp.219-236). Newark, DE: International Reading Association.
- OECD(2011). *PISA 2009 Results: Students on Line: Digital Technologies and Performance* (Volume VI). Paris, France: Author.

創新電腦化測驗題型與測量模式的結合

陳柏熹

國立臺灣師範大學教育心理與輔導學系副教授

摘要

隨著電腦科技的進步，測驗的題型、作答方式與試題媒材也有明顯的改變。從傳統的紙本劃記或書寫型式的測驗題型，擴展到應用資訊設備來點選答案、圖形圈選、拖曳、搖桿操作、電腦繪圖，甚至於表現動作再以攝影機記錄的方式來作答。然而，應用電腦科技於測驗中的主要目的是為了提升測驗的信度與效度，因此，是否達到提升測驗信度與效度的目的才是評價創新電腦化題型的關鍵。本文的觀點是將電腦化測驗題型與測量模式結合，讓創新的電腦化測驗題型有驗證效度與評估信度的理論基礎，並以兩項測驗為例來闡述如何將創新電腦化測驗題型與測量模式作結合。

關鍵字：電腦化測驗、測量模式、試題反應理論、信度、效度

一、創新電腦化測驗題型

電腦科技在近二十年來有相當快速的發展，相關的資訊技術也廣泛地應用到當代的許多心理與教育測驗中。舉凡大型教育測驗中的托福（Test of English as a Foreign Language, TOEFL）與 GRE（Graduate Record Examinations）考試、國際教育評比 PISA（the Programme for International Student Assessment）、語言能力認證中的漢語水平考試（Hanyu Shuiping Kaoshi, HSK）與華語文能

力檢定（Test of Chinese as a Foreign Language, TOCFL），以及許多職業證照考試，例如，美國醫師證照考試（United States Medical Licensing Examination, USMLE）、護士證照考試（National Council Licensure Examination for Registered Nurse, NCLEX）、會計師、建築師證照考試…等，不勝枚舉，都是採用電腦化測驗題型。除了上述的成就測驗（achievement test）之外，性向測驗、性格測驗與職業興趣量表也都開始有電腦化版本，例如，著名的通用性向測驗（General Aptitude Test Battery, CAT-GATB）與大五人格量表（Revised NEO Personality Inventory, NEO PI-R）都有電腦化適性測驗或電腦化施測與解釋版本。其中，美國陸軍職業性向電腦化適性測驗（Armed Services Vocational Aptitude Battery, CAT-ASVAB）可以算是最早發展出電腦化適性測驗（computerized adaptive testing, CAT）且研究較完整的電腦化適性測驗。

然而，上述測驗的題型大部分是選擇反應題型（select response item），例如，選擇題（multiple choice item）或李克特量表（Likert type scale），其電腦化版本只是將此題型從紙本形式變成電腦作答而已，並沒有利用電腦多媒體技術來提高測驗的效度或信度。拜現代電腦科技之賜，測驗除了可以使用這兩種題型，還

有許多不同類型的作答方式與題目呈現形式。Parshall、Davey 與 Pashley (2000) 以及 Parshall 與 Harmes (2007) 曾提出當代電腦測驗題型的多向度分類架構圖 (taxonomy of innovative item)，將電腦化測驗題型從評量結構 (assessment structure)、複雜度 (complexity)、似真性 (Fidelity)、互動程度 (interactivity)、媒體類型 (media inclusion)、作答反應方式 (response action)、計分方式 (scoring method) 等七個向度來進行分類，以下分別簡述之：

(一) 評量結構

評量結構是指題目之間的結構關係，通常可以依此將題目分成獨立題 (discrete item)、情境任務題組 (situated task)、虛擬實境活動 (simulated environments)。獨立題是傳統單一題目的問題，又可以分為選擇反應題

(selected response item) 與建構式反應題 (constructed response item) 兩大類。選擇反應題型如是非題、選擇題或配合題等，其具有作答快速與計分客觀的優點，但比較容易讓受測者產生猜答案的行為而影響測量的信度。建構反應題型如填空題、簡答題、計算題、申論題等。由於受測者難以在這種題型中猜測答案，因此可以改善選擇反應題型容易被猜對的誤差，但作答較費時，且計分也比較容易因為評分者的主觀因素影響評分者一致性信度。圖 1 是多元智慧電腦化適性測驗的題目 (陳柏熹, 2008)，屬於選擇反應題型，受測者需分別針對每個題目點選答案。圖 2 為節能減碳素養電腦化測驗系統 (Chen, Huang, Liu & Chen, 2014) 的建構反應試題，受測者需打字回答問題，雖然是採用題組式 (testlet) 題型，但各題目間仍是獨立試題



圖 1 多元智慧電腦化適性測驗之選擇反應題

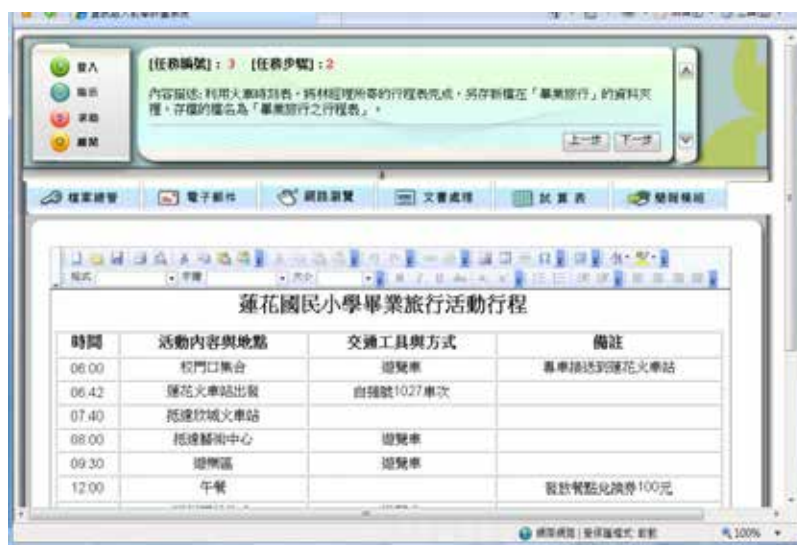
請根據發光亮度相同的燈泡規格表，舉出社區居民選用LED燈具的二個原因。

| 種類 | 傳統燈泡 | 省電燈泡 | LED燈泡 |
|-------|---|---|---|
| 外觀 |  |  |  |
| 耗電功率 | 60 瓦 | 18 瓦 | 9 瓦 |
| 可使用壽命 | 1,000 小時 | 6,000 小時 | 25,000 小時 |
| 購買價格 | 25 元 | 150 元 | 300 元 |

圖 2 節能減碳素養電腦化測驗之建構反應題

情境式任務是一系列具有關連的情節式 (scenario) 測驗題組，受測者需要多個步驟才能完成題目所要求的任務，如圖 3 所示。在這類題目中，各子題之間可以是具有結構式 (structured) 的路徑關係，其前項任務是下一個任務的基礎，也可以是非結構式 (structured) 的開放型任務，由受測者自行決定如何進行。圖 3 中小學資訊能力評量系統 (張國恩、劉

遠禎、王曉璿、蕭顯勝、宋曜廷、陳柏熹，2008) 的情境式任務題組就是屬於結構式的情境任務，受測者需要依據題目的要求逐一進行各項任務。圖 4 的電腦化創造力測驗 (陳柏熹，2011) 則是屬於非結構的開放式任務，受測者可以根據題目的要求，自行決定要如何進行操作，設計出一項具有創新性與實用性的家具，以達成評量目標。



【任務編號】：3 【任務步驟】：2
內容描述：利用大綱時刻表，將林經理所寄的行程表完成，另存新檔在「畢業旅行」的資料夾裡，存檔的檔名為「畢業旅行之行程表」。

| 時間 | 活動內容與地點 | 交通工具與方式 | 備註 |
|-------|---------|-----------|-------------|
| 06:00 | 校門口集合 | 遊覽車 | 專車接送至蓮花火車站 |
| 06:42 | 蓮花火車站出發 | 自強號1027車次 | |
| 07:40 | 抵達竹城火車站 | | |
| 08:00 | 抵達藝術中心 | 遊覽車 | |
| 09:30 | 遊樂區 | 遊覽車 | |
| 12:00 | 午餐 | | 發放餐點兌換券100元 |

圖 3 中小學資訊能力評量系統中的結構式情境任務題型 (圖片來源：中小學資訊能力評量機制發展與推廣計畫成果報告)

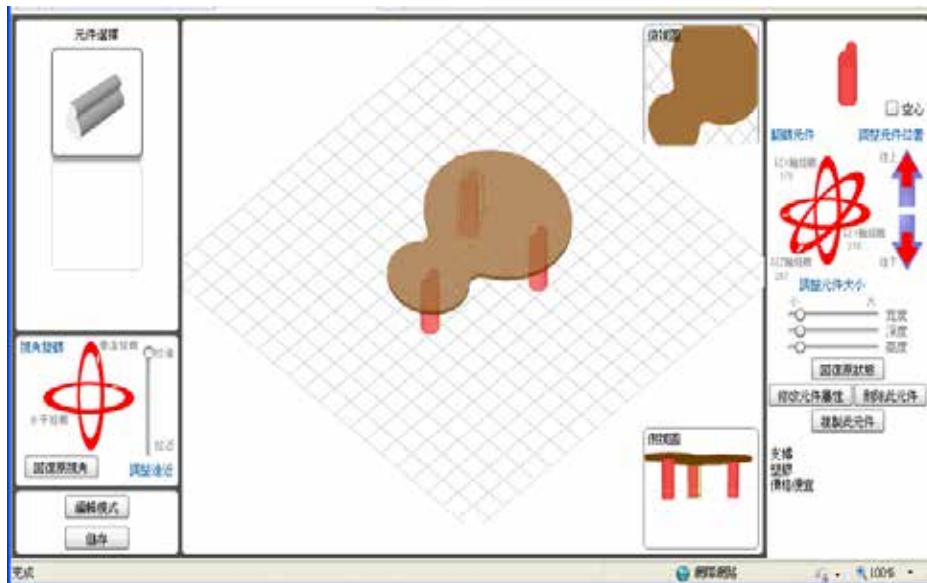


圖 4 電腦化創造力測驗中的非結構式情境任務題型

虛擬實境活動是讓受測者在模擬真實狀況的電腦虛擬實境測驗系統中進行實作活動，例如：飛機駕駛、汽車、儀

器操作，如圖 5 所示。受測者必須進行實際的操作，就好像在真實的生活環境中進行實務操作一般。



圖 5 汽車駕駛能力（左）與手眼協調性（右）的虛擬實境評量題型
（左圖取自林正堅（2011）；右圖取自蕭顯勝、林建佑、邱敬尊（2011））

（二）複雜度

題目複雜度是指受測者在回答題目時所需考慮的訊息種類與數量多寡。例如：在題目中的不同位置呈現多項文字與圖形，讓受測者進行判斷。還有些題目需使用圖形組合、按鈕操作、表單輸入或數值計算、影片播放等複雜的電腦

操作活動等。題目的複雜度愈高，認知負荷量就愈大。傳統僅包含一段文字或一張圖形的選擇題是屬於低複雜度的題目，而情境任務題組通常是較高複雜度的題目。有些國家的醫學證照考試採用電腦化的模擬個案臨床測驗，就是屬於

高複雜度的測驗，受測者需要判斷許多不同的訊息內容，包含病人的主訴與外觀症狀、醫學檢驗報告、用藥紀錄等，針對這些訊息進行診斷或後續檢查，並開藥物或安排住院治療。例如：美國醫師證照考試電腦化模擬個案測驗（United States Medical Licensing Examination step 3 computer-based cases simulation, USMLE step 3 CCS）（USMLE, 2014）。而圖4的電腦化創造力測驗也是屬於高複雜度的題型，該測驗中有許多的元件操作方式與調整按鈕，受測者需要了解這些元件如何使用，才能有效地完成題目的任務。

（三）似真性

似真性是指題目內容與真實生活環境的相似程度，題目呈現出愈多與真實生活情境相似的題材內容，或是要求受測者作出與生活情境相似的實作活動，題目的似真性就愈高。傳統以文字為基礎的選擇題其似真性較低，而虛擬實境實作活動是屬於高似真性的題目。題目的似真性愈高，表示愈接近生活情境中的能力展現，測量的效度就愈高。以圖5中的汽車駕駛測驗為例，若是全部採用文字型選擇題，只能測量出受測者的汽車駕駛常識或交通規則，即使受測者在該測驗中得到高分也不能保證他已經會開車了。若使用圖5中的高似真性汽車駕駛虛擬實境測驗，則較能測量出受測者在面對道路駕駛情境時操作方向盤與控制車輛的能力及應變技巧。若該題目情境與使用的設備材料似真性愈高時，所測量到的能力就愈接近真實的汽車駕駛能力。

（四）互動程度

互動程度是指根據作答者的反應，即時呈現出與此反應有關的資料、情境或後續題目，讓受測者進行下一階段的答題。傳統測驗的互動程度較低，通常只呈現題目要求受測者回答答案就結束了。而互動性較高的測驗會要求受測者回答一連串有邏輯關連的問題，不宜將這些題目拆開使用。在美國的醫師證照考試的模擬個案測驗中，題目會先提供有關病人的症狀與檢查數據，接著會讓受測者進行診斷、用藥、進行後續追蹤檢查等一系列問題，且受測者回答完每個問題後，會得到回答後的結果，前項處理的結果就是進行後續處理的依據（USMLE step 3 CCS; USMLE, 2014）。在Liu, Lin與Wang（2012）的統計概念模擬輔助學習系統中，受測者可以藉由操作資料數據立即看到統計圖表上的變化，並藉此來進行後續的概念學習與判斷作業，也是屬於高互動性的學習評量系統。

互動式測驗的試題資訊會隨著受測者的反應而改變，因此計分方式相當複雜，目前此類測驗尚未找到適合的心理計量模式。較常採用的作法是將受測者的作答反應與專家群所建立的最佳反應（optimal response）進行比對，並分別從多個向度來計分。以美國的醫師證照電腦化模擬個案測驗為例，就是根據專家群所建立的最佳反應，將受測者的作答反應依資訊考量的完整性、時效性、治療的有效性、避免風險等多個向度來進行評分。

(五) 使用的媒體類型

腦化測驗中常見的題目媒體類型主要有文字、圖形或照片、影片、聲音、動畫等。媒體的使用通常與題目所欲測量的能力有關。例如：在溝通能力測驗中，受測者在進行溝通時可能需要面對訊息傳遞者或訊息接受者，因此可以在題目中預先錄製一段影片（或動畫），並要求受測者在看完這段影片後去判斷影片中的主角所要傳遞的訊息、或是同理影片中主角的感受、或是回答出如何

說服他人或安慰他人等表達語句。在圖6的電腦化大學生基本素養測驗（陳柏熹，2014）中，其溝通表達素養就是要求受測者在看完一段動畫後，進行上述的判斷。此外，在該測驗的美感素養試題中，由於主要是評量受測者對美感訊息的理解與表達能力，因此大部分題目都是以生活中的照片、圖案或影片為媒體來呈現，並要求受測進行判斷。



圖 6 電腦化大學生基本素養測驗

(六) 作答方式

由於當代電腦科技的快速發展，目前受測者在電腦上作答的方式相當多，包括用滑鼠點選、螢幕觸控、鍵盤輸入數值或文字、以搖桿控制、在圖形中圈選位置（hot spot）、圖案拖曳或組合排列、用麥克風錄音等。早期電腦化測驗的作答方式通常是以鍵盤輸入或滑鼠點選，這類測驗大多是以認知能力測驗為主。近代則可以用攝影機來擷取受測者的肢體動作反應來進行作答，或甚至於用眼動儀追

蹤眼球運動，以眼睛控制的方式來作答。這種較新的作答技術通常被應用在實作技能的評量上。圖7是以動作影像擷取作答方式來進行肢體動作智慧的電腦化測驗（陳柏熹、歐詠芝，2011b）。在此測驗中，題目刺激與受測者的身體會呈現在電腦螢幕上，受測者必須根據題目的要求來移動身體，使螢幕中的身體影像能碰觸到特定圖案或避免碰觸到某個外框，以評量出受測者的反應速率與平衡感。



圖 7 肢體動作電腦化測驗的動作影像擷取作答方式

(七) 計分方法

計分方式是指如何根據受測者的作答反應計算出每一題的原始分數。獨立題的選擇反應題型計分方式較單純，通常是採用二元計分（dichotomous scoring）。而獨立題的建構反應題型、情境任務題組或虛擬實境活動等，由於評量結構與作答方式都較複雜，通常是多元計分（polytomous scoring）。

二、創新電腦化測驗題型對測驗效能與測量模式的影響

表 1 是電腦化測驗題型的分類向度對測驗效能與測量模式的主要影響層面。雖然每一種分類方式中的不同題型都可能同時影響到測驗的信度、效度與分析時所採用的測量模式，但有些分類向度對測量模式的影響較大，有些則是對測驗效度的影響較大。分述如下：

題目的評量結構類型主要與所使用的測量模式有關。例如：在獨立題中，每個題目彼此是互相獨立的，因此適合

使用試題反應理論（item response theory, IRT），因 IRT 中的基本假設是局部獨立性，也就是對相同能力水準的受測者而言，各題目的答對機率是互相獨立的。然而，在情境任務題型中，前面題目可能會影響後面題目的答題表現，已經違反局部獨立性假設，此時就須要採用題組反應模式（testlet response model）

（Wainer, Bradlow, & Du, 2000; Wang & Wilson, 2005），或是將各個步驟的作答結果與成績合併成一題，改採部份給分模式（partial credit model）（Master, 1982）來進行分析。若是使用非結構式情境任務題或虛擬實境題型，由於需要有評分者來評量受測者的實作技能表現，可能會受到評分者主觀因素的影響，因此可以採用多面向模式（many-facet model）（Linacre & Wright, 1993），將評分者嚴苛度與不同實作技能面向都納入測量模式中進行分析。

題目的複雜度也與測量模式有密切關聯。在低複雜度的情況下，影響作答反應的能力向度較單純，因此適合以單向度

測量模式（單向度 IRT 或單一因素的因素分析）進行分析。當題目的複雜度增加時，所包含的訊息種類與數量也增加了，影響作答反應的能力向度或因素也隨之增加，因此較適合採用多向度試題反應模式（Mckinley & Reckase, 1983; Hattie, 1981; Adams, Wilson & Wang, 1997）或多因素的因素分析來進行分析，如此才能瞭解不同因素對作答結果的影響。除此之外，若將題目所包含的各類訊息視為影響題目難易度的預測變項，也能採用線性對數模式（linear logistic test model, LLTM）（Fischer, 1973）來進行分析，以了解不同訊息對題目難易度的影響力。

似真性主要是影響測驗的效度。當題目情境與使用的設備材料似真性愈高時，所測量到的能力就愈接近真實生活中所展現出來的能力。因此，題目的似真性是決定測驗預測效度（predict validity）的重要因素。

題目的互動程度則是對測驗的信度與效度都會產生影響。由於目前尚未找到適合的心理計量模式可對高互動程度的題目進行分析，且仰賴專家群所建立的最佳反應來與受測者反應進行比對的做法，其客觀性與有效性也缺乏實徵研究證據。因此，目前這種因素對測驗效能的影響力還需進一步研究來提供證據。

媒體類型主要會影響測量的效度。題目中的媒體類型與真實生活愈接近，題目就愈能反映出受測者在日常生活中的情境，其所引發的作答反應就會愈接近受測者的真實能力。不過，閱讀或觀看媒體通常會增加作答時間，因此當測驗的作答時間限制較嚴格時，複雜媒體的題型可能會影響到測量的信度。

電腦化測驗的作答方式必須配合測驗目標，才能有效測量出受測者的能力，提升測驗的效度。例如：若要測量受測者對機械手臂的操作能力，應先了解實際生活狀況中的機械手臂通常是以何種方式來控制？若是以搖桿來控制，則搖桿將會是該項電腦化測驗中較佳的作答方式。圖 5 與圖 6 的幾種作答形式都是實際生活中展現出該測驗目標能力時最常使用的方式。

計分方式與測量模式有密切的關係。若題目為二元計分，則較適合採用二元計分的試題反應模式來進行分析，例如：Rasch 模式（Rasch, 1986）、2PL 模式或 3PL 模式（Birnbaum, 1968; Lord, 1952）等。當題目為多元計分時，則較適合採用評定量尺模式（rating scale model, RSM）（Andrich, 1978）或部份給分模式（partial credit model, PCM）（Masters, 1982）。

表 1 電腦化試題分類向度對測驗效能與測量模式的影響

| | 主要影響測驗的層面 | | |
|-----------|-----------|------|------|
| | 測量模式 | 測驗信度 | 測驗效度 |
| 電腦化試題分類向度 | | | |
| 評量結構 | ✓ | | |
| 複雜度 | ✓ | | ✓ |
| 似真性 | | | ✓ |
| 互動程度 | | ✓ | ✓ |
| 媒體類型 | | | ✓ |
| 作答反應方式 | | | ✓ |
| 計分方式 | ✓ | | |

三、電腦化測驗題型與測量模式的結合

為了讓創新電腦化測驗題型能發揮其測量功能，電腦化測驗編製者需思考如何選擇適當的題型，以及如何選擇適當的測量模式來進行分析。以下以兩項電腦化測驗的研發為例來進一步闡述。

(一) 電腦化大學生基本素養測驗

1. 測驗目標

該測驗的目的是評量大學生在創新領導、問題解決、終身學習、溝通合作、公民素養、美感素養、科學思辨、資訊素養、生涯發展等九項通識教育基本素養上的學習成果，屬於生活基本能力的素養（Chen, Hsu, Huang, Li, Chen, Yang & Yeh, 2013）。測驗中每一種素養皆包含認知能力與情意態度兩個面向，本文僅分析認知能力層面之試題。

2. 選擇電腦化測驗題型

有關電腦化大學生基本素養測驗的題型選擇整理如表 2 所示。在評量結構方面，由於該測驗的目標特質與日常生活基本素養有關，因此需要有一「日常生活的情境」做為試題題材，引導學生展現出與生活情境有關的基本素養，因此適合採用情境式任務題組。但由於評量目標為認知能力，而非實作技能，因此將任務導向的實作題型改成比較適合評量認知能力的非任務型情境測驗題組。又為了能在學生作答完後即時回饋測驗結果，因此採用可以進行電腦自動計分的選擇反應題型，包括選擇題與叢集式是非題（cluster true-

false item）兩種題型，如圖 6 所示。

在複雜度方面，該測驗中各素養欲評量的能力皆為單向度能力，為了避免受到其他能力的干擾，因此選擇低複雜度的題型。而在試題似真性方面，為了使測量結果較能反映出學生在日常生活中的實際能力，因此選擇中等似真性的題型，將生活中常看到的影像、照片或圖表都納入作為題目的材料。但礙於經費有限，因此並沒有發展成高似真性的虛擬實境題型。在互動程度方面，為了讓能力估計單純化，提高計分的客觀性，因此採用低互動性的試題，每個題目在作答完後並不會根據測者的答案而變化後續題目所提供的資訊。

在媒體類型方面，為了提高測驗的效度，該測驗配合各素養的需求而使用各種不同媒材的試題，例如：溝通合作素養多採用人際互動影片或動畫作為題材，美感素養多採用生活中常見的圖像或影片作為題材，問題解決素養多採用圖表呈現問題讓受測者進行問題解決。在作答方式上，由於該測驗是由大學生自行上網作答，且所有試題皆為選擇反應題型，因此該測驗使用大學生最熟悉的滑鼠點選方式，以避免受測者對作答方式不熟悉而影響測量結果。

關於計分方式，由於該測驗中有些素養的題目並無唯一的正確答案，而僅有較佳選項、次佳選項與不佳選項之區別，因此不宜全部採用二元計分模式，需使用多元計分，例如較佳選項記為 2 分，次佳選項記為 1 分，不佳選項記為 0 分。

表 2 電腦化大學生基本素養測驗在電腦化試題分類向度上的題型選擇

| 電腦化試題分類向度 | | 情境測驗題組 |
|-----------|--|--------|
| 評量結構 | | |
| 複雜度 | | 低 |
| 似真性 | | 中 |
| 互動程度 | | 低 |
| 媒體類型 | | 多樣 |
| 作答反應方式 | | 滑鼠點選 |
| 計分方式 | | 多元計分 |

3. 選擇測量模式

為了能符合上述的情境測驗題組結構與多元計分方式，該測驗採用多元計分的題組反應模式進行分析（Wang & Wilson, 2005）。資料來源為 2012 年 11 月至 2013 年 6 月該測驗進行預試時所蒐集之 1049 位大學生作答反應，該預試採用不等組變動共同題設計（Non-equivalent group with variable anchor test, NEVAT）（Chen, Kuo & Sung, 2011）。表 3 為大學生基本素養測驗之多元計分的題組反應模式分析結果，表中數值為主要欲評量的能力質變異數與題組效果變異數，其

中，題組效果變異數愈大，表示題目間的相關係愈高，對真實能力估計的影響就愈大。在估計能力時若未採用題組模式將題組效果分離出來，將會錯估受測者能力值並高估能力估計的精準度（Wang & Wilson, 2005）。從表中可以看出，公民素養、美感素養與資訊素養中皆有部分題組有較高的題組效果。究其原因為該題組中皆有叢集是非題。若將叢集是非題重新計分為一題多元計分題，則題組效果明顯下降。表 3 最後一列是將資訊素養的叢集是非題重新計分後的題組效果分析結果。

表 3 大學生基本素養測驗之多元計分的題組反應模式分析結果

| 基本素養名稱 | 能力 σ^2 | 題組 1 | 題組 2 | 題組 3 | 題組 4 |
|---------------|---------------|----------------------------|----------------------------|----------------------------|----------------------------|
| | | $\sigma^2_{\gamma_{1Y:Y}}$ | $\sigma^2_{\gamma_{2Y:Y}}$ | $\sigma^2_{\gamma_{3Y:Y}}$ | $\sigma^2_{\gamma_{4Y:Y}}$ |
| 溝通合作 | 0.21 | 0.35 | 0.91 | 0.86 | - |
| 科學思辨 | 0.27 | 0.66 | 0.56 | 0.42 | - |
| 公民素養 | 0.27 | 0.41 | 0.45 | 1.14 | - |
| 終身學習 | 0.47 | 0.59 | 0.46 | 0.84 | 0.92 |
| 美感素養 | 0.40 | 0.87 | 0.31 | 1.17 | - |
| 創新領導 | 0.36 | 0.36 | 0.92 | 0.92 | - |
| 問題解決 | 0.19 | 0.29 | 0.64 | 0.39 | - |
| 問題解決 | 0.19 | 0.29 | 0.64 | 0.39 | - |
| 生涯發展 | 0.37 | 0.36 | 0.75 | 0.64 | - |
| 資訊素養 | 0.15 | 0.08 | 0.05 | 1.12 | 3.42 |
| 資訊素養 (recode) | 0.17 | 0.18 | 0.10 | 0.33 | 0.98 |

(二) 電腦化創造力測驗

1. 測驗目標

該測驗主要是評量學生在產品設計的創造歷程與創造成品中所展現出來的創新性與實用性（陳柏熹，2011a）。創造歷程之評分有七項指標，包含功能聯結（含元件選取、材質選取）、表徵轉換（材質用途創新、元件變化）、心向旋轉（元件翻轉）、組合（組合獨創、組合多樣）四大面向。創造成品之評分有八項指標，分別為外觀創新、功能創新、不對稱性、精緻化、多元功能、平衡感、使用便利性、堅固耐用性等。

2. 選擇電腦化測驗題型

電腦化創造力測驗的題型選擇整理如表四所示。在評量結構方面，該測驗的目標特質是設計產品的創造力，為了能適用於不同年齡受測者，所以使用一般人在日常生活中都會接觸到的「家具」設計做為試題題材，因此也是適合採用情境式任務題組。而其任務是使用三項幾何元件材料，實際設計出一項具有創新性及實用性的家具，因此是屬於實作技能類的任務。如表 4 所示。

在複雜度方面，由於創造力為複雜的多向度能力，在產品設計領域被定義為設計出具有創新性與實用性的產品，並被群體所認可（葉玉珠，2004; Amabile,

1996），且可以展現在創造的歷程中，也能展現在創造出來的成品中，因此選擇高複雜度的題型，給予許多控制元件讓受測者可以從不同角度、調整物件的外觀形狀材質，並組合各項物件。

在試題似真性方面，為了使測量結果較能反映出學生在產品設計的實際能力，因此選擇高似真性的題型，將施測介面規劃成與市售產品設計軟體類似的介面，以模擬產品設計人員的工作。在互動程度方面，為了能紀錄受測者的產品設計歷程的各項動作，因此採用高互動性的試題，使受測者在進行每一項操作時都能立即得到回饋，以便調整後續的操作。

在媒體類型方面，由於本測驗所使用的設計元件都是幾何元件，較為單純，因此媒體類型都是使用圖形。在作答方式上，由於該測驗主要目的是要求受測者根據所提供的幾何圖形元件進行移動、變化、組合等操作，因此採用滑鼠點選、圖形拖曳、圖形組合等作答方式。此外，為了使評分者了解作品特性與創作理念，在設計結束時也讓受測者加入文字敘述。

關於計分方式，由於創造力並無唯一的正確答案，而是要從七項創造歷程指標與八項創造成品指標來評估受測者的創造力，因此使用多元計分，每項指標皆為 0-2 分，評分規準參見陳柏熹（2011a）。

表 4 電腦化大學生基本素養測驗在電腦化試題分類向度上的題型選擇

| 電腦化試題分類向度 | |
|-----------|---------------------|
| 評量結構 | 開放式情境測驗題 |
| 複雜度 | 高 |
| 似真性 | 高 |
| 互動程度 | 高 |
| 媒體類型 | 圖形 |
| 作答反應方式 | 滑鼠點選、圖形拖曳、圖形組合、文字輸入 |
| 計分方式 | 多元計分 |

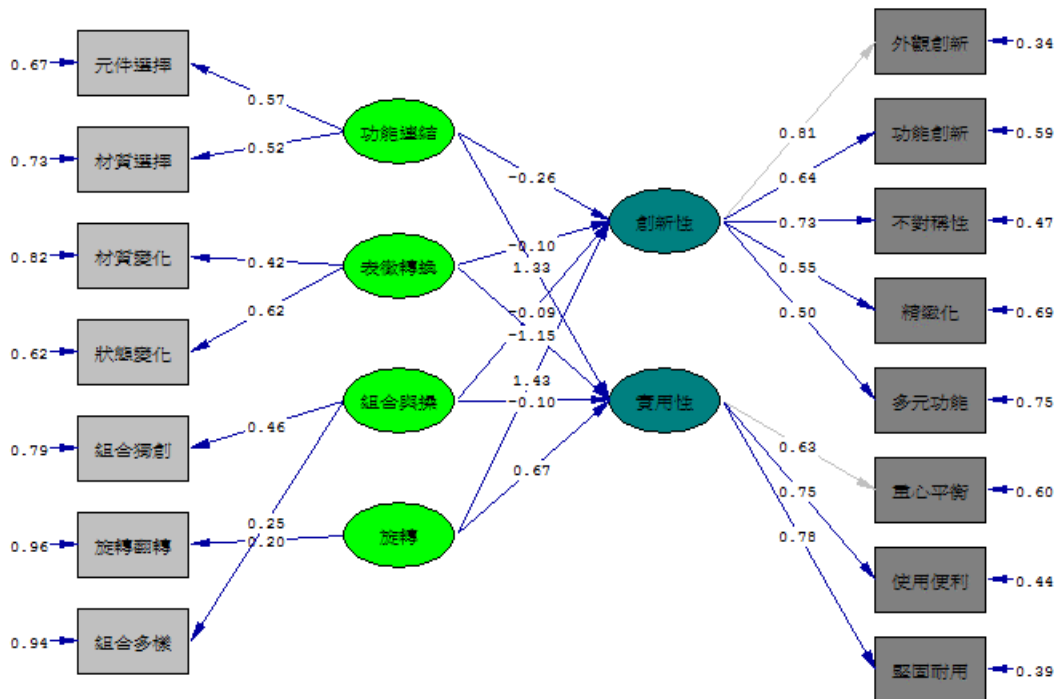
3. 選擇測量模式

為了能符合上述的高複雜度、多種作答反應及多元計分方式，該測驗採用多向度能力之多元計分模式來進行分析，並使用結構方程模式（structure equation model）（Jöreskog & Sörbom, 1993）來探討模式與資料的吻合程度，受測者為217位大學生。初始模式採用測量目標中的評分指標架構，其參數分析結果如圖7所示。其中*表示該項參數之t檢定值顯著不等於零（ $p < 0.05$ ）。從創造歷程評分的測量模式中可以看出，「旋轉翻轉」項目信度頗低（ $0.03 = 1 - 0.97$ ），且受潛在變項之影響也未達顯著水準（ $\lambda = 0.20$ ， $p > 0.05$ ），表示該指標測量效果不佳，無法達到應有之測量功能，因此該指標在後續分析中予以刪除。而「組合多樣性」雖然受組合操弄潛在變項之影響也不大（ $\lambda = 0.25$ ， $p > 0.05$ ），但根據修改指標顯示其在「表徵轉換」潛在變項中之測量效果頗佳。經研究團隊討論其評分規準發現該項評分內涵與「表徵轉換」有關，因此保留該指標併入表徵轉換的潛在變項，改稱為「轉換變化」。創造成品評分測量模式分析結果顯示，各項評分指標之 λ 值皆大於0.5，且t檢定皆顯著不等於零（ $p < 0.05$ ），顯示成品評分向度之測量模式尚可。但根據修改指標顯示，「外

觀創新」與「不對稱性」、「功能創新」與「多元功能」兩組指標皆有顯著之相關，顯示此兩組指標在評分時可能會互相干擾，經研究團隊討論發現此兩組指標內的兩項指標之內涵頗為接近，因此決定兩組指標中都只保留一項指標，刪除「不對稱性」與「多元功能」。

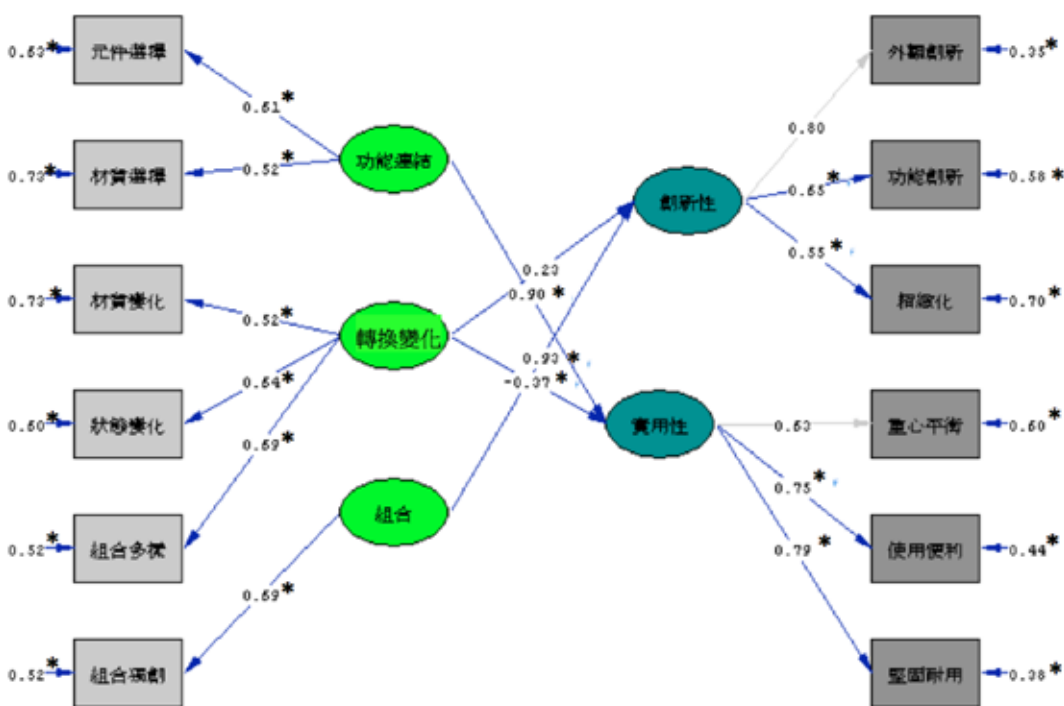
從結構關係模式來看，潛在自變項與潛在依變項之參數值 γ 皆不高，且皆未達顯著。但由於刪除其中某些 γ 參數時，會影響到其他 γ 參數之估計。因此，本研究之後續分析僅將參數絕對值 < 0.2 之 γ 先行刪除，其他 γ 仍予以保留。

初始模式之模式符合度資料如表五所示。從表五中可以看出，各項模式符合度指標皆不理想，例如：因素負荷量沒有介於0.5~0.95之間，卡方值與自由度比例大於3，GFI、AGFI、NFI、IFI等整體符合度指標皆小於0.90，以及各變項的信度皆不理想。顯示初始模式有許多不太恰當之模式假設，應予以修改。本研究根據上述分析，將創造歷程之變項修改後形成模式二，再將創造成品之變項修改後形成模式三。最後再根據修改指標之建議與研究團隊討論之結果修改結構模式（潛在自變項與潛在依變項間的關係），形成最終模式。



Chi-Square=273.35, df=76, P-value=0.00000, RMSEA=0.110

圖 7 創造歷程與創造成品的初始模式分析結果



Chi-Square=89.41, df=47, P-value=0.00019, RMSEA=0.065

圖 8 創造歷程與創造成品的最終模式分析結果

該測驗之最終模式參數分析結果如圖 8 所示。從歷程評分與成品評分的測量模式中可以看出，所有測量模式的因素負荷量皆達顯著，各項評分指標之 λ 值皆大於 0.5，且 t 檢定皆顯著不等於零 ($p < 0.05$)。但仍有將近一半的測量指標其測量誤差接近 0.5。由於本研究乃以人為評分來建立指標分數，因此其測量信度不如客觀答題反應的量表，未來需要加強評分者的訓練以提高測量信度。

從結構關係模式來看，創造歷程之「功能連結」對創造成品「實用性」的迴歸係數達 0.90，顯示此歷程可以提升創造成品之「實用性」；而創造歷程之「組合方式」對創造成品「創新性」的迴歸係數達 0.93，顯示此歷程可以提升創造成品之「創新性」。創造歷程之「轉換變化」對創造成品「創新性」有正向影響 (0.23)、但對創造成品「實用性」則

有負向影響 (-0.37)，顯示「轉換變化」會提升創造成品之「創新性」，但會降低創造成品之「實用性」。不過此歷程對創造成品「創新性」的影響係數未達顯著，未來還需要更多創造歷程的評量指標來了解其對創造成品創新性的關係。

最終模式之模式符合度資料如表五所示。從表 5 中可以看出，各項模式符合度指標大致能符合要求，例如：測量模式之因素負荷量都介於 0.5~0.95 之間，卡方值與自由度比例小於 3，GFI、AGFI、NFI、IFI 等整體符合度指標皆大於 0.90 等。顯示最終模式已經與實際資料大致相符。雖然「改變轉換」對創造成品「創新性」的影響未達顯著，但根據過去相關研究 (李珮瑜, 2011)，創新性高低不同者，其在「變化轉換」的相關指標上皆有顯著差異，因此該項迴歸線予以保留。

表 5 初始模式與最終模式之模式符合度指標

| | 模式符合度指標 | 初始模式 | 最終模式 |
|-----------|-------------------|-------------------|-------------------|
| 基本適配度指標 | 沒有負的誤差變異 | 是 | 是 |
| | 參數間的相關絕對值不會太接近 1 | 不符合 | 是 |
| | 因素負荷量 0.5-0.95 | 不符合 | 是 |
| 整體模式適配度指標 | χ^2 未達顯著 | 251.92 (P=0.0000) | 96.26 (P=0.0002) |
| | χ^2 比率 <3 | 251.92/76=3.31 | 是 (96.26/47=2.04) |
| | GFI 指數 >0.9 | 0.86 | 是 (0.94) |
| | AGFI >0.9 | 0.77 | 0.89 |
| | RMSEA <0.05 | 0.11 | 0.065 |
| | RMR/SRMR <0.05 | 0.092 | 0.071 |
| | Q-plot 殘差分佈斜度 >45 | 否 | 是 |
| | NFI >0.9 | 0.83 | 是 (0.90) |
| | IFI >0.9 | 0.87 | 是 (0.94) |
| | NHFI >0.9 | 0.82 | 是 (0.92) |
| 模式內在品質 | 個別項目的信度 >0.5 | 否 | 半數符合 |
| | 潛在變項平均變異量 >0.5 | 否 | 否 |
| | 估計的參數都達顯著 | 否 | 大致符合 (1 項不符) |
| | 修正指標是否小於 3.84 | 否 | 是 |

四、結論

本文主要介紹當代電腦化測驗之題型，並參考 Parshall 與 Harmes (2007) 的分類向度舉例說明近代幾項創新的電腦化測驗題型。以目前資訊科技之進度速度來看，電腦化測驗題型種類繁多，也頗接近日常生活的情境，讓受測者可以使用許多便利的作答方式來展現其潛在特質，提高了測量工具的效度。

然而，電腦化測驗題型僅是蒐集受測者外顯行為之工具，測驗的主要目的仍然是評估受測者的潛在特質，而非只是蒐集外顯行為。因此，測驗編製者仍須根據測驗的理論架構與所使用的電腦化

測驗題型的特性，尋找適當的測量模式，並驗證該測量模式是否能有效地解釋受測者的作答反應，此即建構效度的驗證過程。若測量模式所預期的結果與受測者實際作答反應的吻合度不佳，則需考慮調整題型或調整測量模式。當測量模式能有效解釋受測者的作答反應時，創新的電腦化測驗題型才能真正發揮測量的功能。

致謝

本文感謝教育部「邁向頂尖大學計畫」與科技部「跨國頂尖研究中心計畫」(NSC103-2911-I-003-301) 支持。

參考文獻

- 林正堅 (2011)。汽車駕訓之動感模擬系統設計。勤益新聞網。2013年5月31日檢索。
網址：[http://enews.ncut.edu.tw/web/news/news.php?NewsCategoryID=13 & NewsID=127&Year=2011](http://enews.ncut.edu.tw/web/news/news.php?NewsCategoryID=13&NewsID=127&Year=2011)
- 李佩瑜 (2011)。潛在類別分析與二階段群集分析分群效果之比較研究。國立臺灣師範大學教育心理與輔導學系碩士論文。
- 葉玉珠 (2004)。科技創造力測驗的發展與常模的建立。測驗學刊，51 (2)，127-162。
- 張國恩、劉遠禎、王曉璿、蕭顯勝、宋曜廷、陳柏熹 (2008)。中小學資訊能力評量機制發展與推廣計畫結案報告。教育部。
- 陳柏熹 (2008)。多元智慧電腦化適性測驗的發展：不同 MIRT 模式之適配度比較。第 47 屆台灣心理學會，台北。十月。
- 陳柏熹 (2011)。創造力電腦化測驗系統之發展暨學生創造歷程之研究：IRT 取向 (第二年)。國科會期中報告 (NSC 98-2511-S-003-011-MY3)。
- 陳柏熹、歐詠芝 (2011)。肢體動作電腦化測驗的研發與效度研究。2011 心理與教育測驗學術研討會。台北。十月。
- 陳柏熹 (2014)。電腦化大學生基本素養測驗簡介。評鑑雙月刊，第 47 期。
- 蕭顯勝、林建佑、邱敬尊 (2009)。應用擴充實境技術發展電腦化動作技能測驗系統。第五屆台灣數位學習發展研討會論文 (TWELF 2009)。台灣：台南大學，9 月。
- Amabile, T. M. (1996). *Creativity in context*. CO: Westview.

- Andrich, D. (1978). Rating formulation for ordered response categories. *Psychometrika*, 43, 561–573.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Chen, K.-L., Huang, S.-H., Liu, S.-Y., Chen, P.-H. (2014). Energy Literacy of Secondary Students in Taiwan: A Computer-based Assessment. The Third International Conference on E-Learning and E-Technologies in Education (ICEEE 2014), Malaysia: Kuala Lumpur.
- Chen, P.-H., Hsu, C.-Y., Huang H.-Y., Li, P.-W., Chen Y.-S., Yang C.-Y., Yeh, T.-T. (2013). Development of the general literacy test for university students. Paper presented at the 78th International Meeting of Psychometric Society 2013 (IMPS 2013). Amsterdam, Netherland.
- Chen, P. H., Kuo, J. W., & Sung, Y. T. (2011). Influence of Pre-Test Design on the Precision of the Parameters Estimation in the Multidimensional Items Bank. Paper presented at the 76th Annual Meeting of the Psychometric Society (IMPS 2011). Hong Kong, China.
- Fischer, G. H. (1973). The Linear logistic model as an instrument to educational research. *Acta Psychologica*, 37, 359-374.
- Hattie, J. (1981). *Decision criteria for determining unidimensional and multidimensional normal ogive models of latent trait theory*. Armidale, Australia: The University of New England, Center for Behavioral Studies.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Mooresville, IN: Scientific Software, INC.
- Linacre, J. M. and Wright, B. D. (1993). *A user's guide to FACETS: Rasch-measurement computer program*. Chicago, IL: MESA Press .
- Liu, T. C., Kinshuk, Lin, Y. C., & Wang, S. C.(2012). Can verbalisers learn as well as visualisers in simulation-based CAL with predominantly visual representations? Preliminary evidence from a pilot study. *British Journal of Educational Technology*, 43(6), 965-980.
- Lord, F. M. (1952). A theory of test scores. Psychometric Monograph, No. 7.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Mckinley, R. L., & Reckase, M. D. (1983). MAXLOG: A computer program for the estimation of the parameters of a multidimensional logistic model. *Behavior Research Methods & Instrumentation*, 15, 389-390.
- Parshall. C. G. & Harmes, J. C. (2007). Designing templates based on a taxonomy of innovative items. (2007). In D. J. Weiss (Ed.). *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*. Retrieved 2014/5/25 from www.psych.umn.edu/psylabs/CATCentral/
- Parshall. C. G., Davey, T. & Pashley. P. J.(2000). Innovative Item Types for Computerized Testing. In Wim J. van der Linden and Gees A.W. Glas (Eds.). *Computerized Adaptive Testing: Theory and Practice*, pp 129-148.
- USMLE (United States Medical Licensing Examination).(2014). Tutorial and Practice Items for Multiple Choice Questions and Primum Computer-based Case Simulations (CCS). Retrieved 2014 from <http://www.usmle.org/practice-materials>.
- Wainer, H., Bradlow, E. T., & Du, Z. (2000). Testlet response theory: An analog for 3PL model using in testlet-based adaptive testing. In W. van der Linden & C. A. W. Glas (Eds.), *Computerized Adaptive Testing: Theory and Practice*. (pp. 245-269). London: Kluwer.
- Wang, W.-C. & Wilson, M. R. (2005). The Rasch testlet model. *Applied Psychological Measurement*.
- Wang, W.-C., Wilson, M. R., & Adams, R. J. (1997). Rasch models for multidimensionality between items and within items. In M. Wilson, G. Engelhard & K. Draney (Eds.), *Objective measurement: Theory into practice*. (Volume 4, pp. 139-155). Norwood, NJ: Ablex.

教師科展專業知識分享社群平台系統開發與評估

吳穎涓

國立中央大學網路學習科技研究所副教授

施彥宏

國立中央大學網路學習科技研究所

一、前言

「探究」(inquiry) 是現代科學教育的共同核心，美國國家研究委員會 (National Research Council) 在2000年頒佈的「美國國家科學教育標準」(National Science Education Standards; NSES) 中強調「探究」是科學學習的核心 (NRC, 2000)，也特別出版了「探究與國家科學教育標準：教與學的指引」(Inquiry and National Science Education Standards: A Guide for Teaching and Learning) 一書提供科學教師作為實施探究教學之參考；而我國教育部在民國九十二年所發表的「科學教育白皮書」中提及「科學教育是經由科學性的探究活動，使學生獲得相關的知識與技能，養成科學思考的習慣，依照科學方法從事探討與論證，運用科學知識與技能解決問題，進而形成對科學本質的認識，並建立科學精神（科學態度），此外，在我國的「97年國民中小學九年一貫課程綱要」針對自然與生活科技學習領域的基本理念亦提及「自然與生活科技之學習應以探究和實作的方式來進行，強調手腦並用、活動導向、設計與製作兼顧及知能與態度並重。」。雖然「科學教育白皮書」與「國民中小學九年一貫課程綱要」均將「探究」視為是我國科學教育的核心，但是國內科學教育學者段曉林教授在與 Abd-El-Khalick 等

學者 (2004) 共同發表了一篇探討美國、英國、澳洲、黎巴嫩、以色列、委內瑞拉及台灣這幾個國家中「探究」在科學教育所扮演的角色的國際期刊論文中提及「因為在我國的中小學中科學評量還是以傳統紙筆測驗為主，因此強調的是學生的語文表達與邏輯思考能力，而非過程技能表現，即使是在實驗室中也是如此，只有少數的教師重視學生的探究技能。」

(Abd-El-Khalick et al., 2004)，因此，對國內的科教學者及研究者而言，如何提升國內學習者的探究能力，特別是進行「開放性探究」的能力，是一個重要且值得國內科教學者及研究者關注的議題。

而許多科教學者指出「科學展覽」(science fair) 是在科學教室中最常見的「開放性探究」學習活動 (例如：黃鴻博, 1999; Bell et al., 2005; Abd-El-Khalick et al., 2004)，「科學展覽」在包含我國在內的許多國家都行之有年 (Bencze & Bowen, 2009)，我國每年都有大量的中小學師生參與其中，而國小「科學展覽」作品品質低落的現象日益嚴重，這樣的現象應該值得國內的科學學者及研究者特別重視。而無庸置疑，科學教師對於探究教學 (包含科學展覽專題的指導) 的專業知能是影響學習者進行探究學習學習成果的重要因子之一 (NRC, 2000)，許多

國內外相關文獻指出許多教師在指導學習者進行「開放性探究」活動，特別是指導學生進行「科學展覽專題」時會面臨許多的挑戰，其中包括自己本身相關專業知能的不足，以及時間、資源與協助不足等困境（黃鴻博，1999; Anderson, 2002），特別是相關專業知能不足的部分，相關文獻指出只有少數的中小學科學教師知道如何有效地引導學習者進行探究（Justi & Gilbert, 2002）。此外，許多中小學科學教師自己本身都沒有透過探究來學習科學的經驗（Windschitl, 2004），而這或許是導致學生「科學展覽專題」成果品質低落的重要因素。而目前尚無研究特別聚焦於促進教師「科展探究教學」的專業成長，因此本研究針對這個議題進行探討。

二、文獻探討

（一）網路學習社群

「學習社群」，意指來自各不同領域的人們，具有共同的學習興趣或學習目標，在一共同的空間裡，彼此分享經驗、交流資訊或分享知識（Jonassen, Howland, Moore & Marra, 2002）。而一群專業工作者所組成的學習與成長團體，成員們基於對專業的共同信念、願景或目標，透過協同探究的方式，致力於精進本身的專業素養，已持續達成專業服務品質的提升與卓越，此種社群型態便稱為「專業學習社群」。而「教師專業成長社群」指的是社群成員為一群志同道合的教師所組成，彼此有著共同的信念及願景，且關注於學生學習成效的提升，以合作的方式共同進行探究及問題解決（教育部，2009）。

近年來，隨著資訊科技的蓬勃發展，利用資訊科技增進「教師專業成長」成為許多學者關心的重要議題，而透過網際網路（network）或社群（communities）是促進教師專業成長便是一種有效的方式之一（Pennel & Firestone, 1998; Sumsion & Patterson, 2004）。於網路上，因為擁有共同目標或感興趣的討論議題，久而久之，便逐步形成特定的族群，然而，透過網路資訊平台進行資訊的交流，藉此過程逐步建構出知識，「網路學習社群」便因而產生。「網路學習社群」指的是一群有著共同的理想、相同的興趣或一致的目標，基於上述之理由而群聚再一起的人們，透過網路的協助，社群成員在此找尋共同的目標，並可進行娛樂、知識分享及電子商務…等活動，經過一段時間後，進而形成線上的學習社群（Rheingold, 2000; De Souza & Preece, 2004），「網路學習社群」可以做為人與人之間溝通及經驗分享的媒介，所有成員皆可能為知識的接收者或授予者，擁有不同背景及知識的成員，朝著共同的目標，彼此分享經驗、專業的知識和技能，透過共享的方式，達到更大的效益（Collins & Bielaczyc, 1997），且線上的學習社群也提供了社群成員們可以在任何的時間及任何的地點進行學習活動（Shrivastava, 1999）。利用「網路學習社群」來增進專業知能的過程中，社群成員們藉由網路學習社群提供的空間彼此溝通、互動，成員之間也可以避免面對面所產生的壓力，如此，對於意見的表達，更能表達出原意，也較不易讓自己想要發表的意見受到侷限，增加成員彼此表達的機會，且網路學習社群的運作可以是非同步的方式進行，所有

社群成員發表的內容皆能完整保存於此，成員們能夠更加彈性的選擇自己適合的時間來進行瀏覽（黃旭盛，2010）。

而「網路學習社群」也逐漸被用來幫助教師進行專業成長（Duncan-Howell, 2010），而在「網路學習社群」中，參與教師會進行「知識分享」（knowledge sharing），「知識分享」指的是個人或團體間自願性地將知識傳遞給其他人的行為（Lee, 2001），所以，一個「網路學習社群」通常也被視為是一個「網路知識分享社群」（online knowledge sharing community）。

（二）知識管理與知識分享

Davenport 與 Prusak（1998）提出對「知識」的定義，並對解釋了「資料」（data）、「資訊」（information）、「知識」（knowledge）的不同。「資料」指的是一群不連續的數據，透過觀察與收集後，對事物的客觀描述；「資訊」是經過處理後具有意義的資料；「知識」則是融合了結構化的經驗、價值、系統資訊、專家觀點，且提供了架構來評估及整合新的經驗及資訊；而知識是由資訊經過轉化所產生，資訊是資料經過處理所衍生出來的。

而知識可依屬性分為「內隱知識」（Tacit）及「外顯知識」（Explicit），其中「內隱知識」無法用言語進行表達，是屬於主觀且無法具體化的知識，包含了認知與技能此二種元素；而「外顯知識」則可用文字或數字，表達客觀的知識（Nonaka & Takeuchi, 1995）。Nonaka & Takeuchi（1995）指出知識的轉移與創造包含下列四種形式，稱為「知識螺旋」（Spiral of knowledge）：（一）「共同

化」，是組織成員間內隱知識的移轉；（二）「外部化」，指的是組織成員的內隱知識轉換為外顯知識；（三）「結合」，指的是組織成員結合不同來源的外顯知識並形成新的外顯知識；（四）「內部化」，則是組織成員將外顯知識理解、吸收後成為內隱知識。

1990年開始，因資訊科技的蓬勃發展，「知識管理」（knowledge management）逐漸獲得重視，許多的大型企業及學術界紛紛著手進行知識管理相關活動，後續也有不少學者對知識管理做出了定義，讓其更加地明確。O' Dell & Grayson（1998）指出，「知識管理」是一種策略，適時的提供知識給需要的成員，以幫助成員增加績效的持續性過程，也可以提升組織內部的能力及對組織外部的競爭優勢，此過程包含了知識的獲得、創造、分享、整合、存取、改進到淘汰..等步驟（O' Dell & Grayson, 1998; 林東清, 2007）。隨著時代的演進及發展，知識被視為一種新的資產、競爭力與優勢。在這數位化的時代，知識得以迅速的累積，如何做好有效的管理，以利於知識分享，便是值得關注的焦點之一。一個網路學習社群的形成，在長期的運作下，透過網路成員彼此間資訊的交流與知識的分享，其後伴隨而來的便是龐大的資料及知識，如何有效地整合、管理學習社群內擁有的資源並加以妥善運用，有賴於一套完善的知識管理機制加以輔助，提升知識的品質，讓學習者能更有效的運用於學習活動中。

而「知識分享」指的是自願性地將資訊傳遞給其他人的一種行為，是一種個人、團體或組織將知識轉移與另一人、

團體或組織的活動 (Lee, 2001; Russell, 1996)。個體間彼此擁有各自的知識，若要將知識進一步的發展及加值是十分有限的，因此，必須透過分享的行為，讓彼此的知識互通有無，相互的吸收、擷取，以此擴充知識的內涵，所以，知識分享扮演著舉足輕重的角色。

為幫助教師形成「網路知識分享社群」，有些研究者也開發了「知識管理系統」(knowledge management systems; KMSs) (例如：Carroll et al., 2003)，而為了促進「網路知識分享社群」進行知識管理與分享，「網路知識分享社群」中所使用之知識管理工具有其重要性，Spector (2002) 從社群知識管理的角度提出了四種知識管理工具：(1) 溝通 (Communication)：用以使個人及社群的訊息能有效的傳遞的工具，例如：email、討論區，或是群體訊息。(2) 協調 (Coordination)：用以快速地排定社群行程或是小組工作的工具，例如：共享日曆。(3) 合作 (Collaboration)：用以提供成員分享和交換工作文件，以促進小組合作的工具，例如：工具分享及共同的工作空間。(4) 控管 (Control)：用以自動稽核文件以及控制系統版本的工具，例如：版本及組態控管工具。此外，也有一些研究者則是從個人知識管理 (personal knowledge management) 的觀點提出在「網路知識分享社群」中使用者進行個人知識管理的知識管理工具，例如：Tsui (2002) 提出的個人知識管理工具有：(1) 搜尋工具 (index/search tools)：透過網路或索引功能，來輔助資訊搜尋。(2) 整合式搜尋工具 (meta-search tools)：可以同時利用多

個搜尋引擎進行資訊搜尋。(3) 相關連結工具 (associative links)：透過超連結取得網路資源。(4) 資訊取得/資訊分享工具 (information capturing and sharing tools)：可將資訊重新組織或進行文件管理。(5) 概念圖或心智圖工具 (concept/mind mapping)：組織及連結不同資訊，並可協助個人在社群中集思廣益，共享心智模型。(6) 電子郵件之管理、分析及整合工具 (e-mail management, analysis and unified messaging)：可用以改善或整合不同的通訊系統。(7) 語音辨識工具 (voice recognition tools)：透過語音工具的協助，幫助使用者傳達指令給系統。(8) 合作及同步工具 (collaboration and synchronization tools)：透過此工具可用以促進知識分享。(9) 學習工具 (learning tools)：用以輔助使用者學習，如學習歷程追蹤。

(三) 科展探究教學的學科教學知能

「學科教學知能」(pedagogical content knowledge; PCK) 的概念首先是由 Shulman 在 1986 年所提出，Shulman (1986) 指出「學科教學知能」是一種融合學科內容與教學方法的知識，來自於教師個人專業上的理解，是一種以學科知識為基礎之知識的轉化 (transformation of knowledge)，有別於學科知識與一般的教學知識。後續許多研究者也根據 Shulman 的研究針對 PCK 的概念提出看法 (例如：Grossman, 1990; Magnusson, Krajcik, & Borko, 1999; Cochran, DeRuiter, & King, 1993)，雖然研究者們的看法稍微有些差異，然而研究者們都同意 PCK 是一種數個教學知識融合轉型而成的知

識（例如：Grossman, 1990; Magnusson, Krajcik, & Borko, 1999），只是對於是哪幾種知識整合而成的看法不一，例如：Grossman（1990）認為 PCK 是由「學科教材知識」（subject matter knowledge; SMK）、教學知識（Pedagogy knowledge; PK）、和教育情境知識（Context knowledge; CK）所共同融合轉型而成的。而 Cochran, DeRuiter, & King（1993）則提出 PCKg（pedagogical content knowing），認為 PCKg 是教師對於教學知識（pedagogy）、學科知識（subject matter knowledge）、學生特質（student characteristics）與學習環境脈絡（environmental context of learning）的統整理解。

而 Magnusson 等人（1999）則特別針對科學教師的 PCK 進行探討，他們以 Grossman（1990）的研究為基礎進而指出科學教師的 PCK 的組成（components）包括：

1. 科學教學取向（Orientation to teaching science）：
指教師對於特定科學內容教學目標的理解，這些理解會引導教師所做的教學決定，而教師的科學教學取向可能從以傳輸學科內容為導向的教學、注重科學過程的教學，到探究活動導向的教學。
2. 對於科學課程的知識（Knowledge of science curricula）：
教師的科學課程的知識有兩種，包括對於教學目標的理解（knowledge of goals and objectives）以及對特定科學課程的理解（knowledge of specific curricular programs）。

3. 對於學生科學理解的知識（Knowledge of students' understanding science）：
指教師對於幫助學習者發展特定知識所需要的知識，包括對於學生學習時必須具備何種先備條件的知識（knowledge of requirements for learning）以及關於學生學習困難處的知識（knowledge of areas of student difficult）。
4. 對於科學評量的知識（Knowledge of assessment in science）：
指老師能夠使用適當的評量方法，了解學生學習的過程和結果，以便作為教學改進的參考，包括對於科學學習評量向度的知識（knowledge of dimensions of science learning to assess）與對於評量方法的知識（knowledge of methods of assessment）。
5. 對於教學策略的知識（knowledge of instructional strategies）：
可分成特定學科教學策略（knowledge of subject-specific strategies）和特定主題的教學策略（knowledge of topic-specific strategies）兩種，而特定主題的教學策略又包括特定主題的表徵（topic-specific representation）和特定主題的活動（topic-specific activities），特定主題的表徵是指為了促進學生學習，教師表徵特定內容或原則之方法的知識，也包括一位教師創造表徵的能力去幫助學生發展特定概念或關係之理解；而特定主題的活動是關於可以協助學生理解特定概念與關聯性的活動的知識。

綜言之，過去的相關研究證實了

「網路知識分享社群」對於教師專業成長的幫助，因此，透過「網路知識分享社群」可能可以有效促進教師「科展探究教學」的專業成長，然而目前尚未有以促進教師「科展探究教學」專業成長為目標的「網路知識分享社群」，因此，本研究開發一個「教師科展教學網路社群平台」（Teacher Science Fair Instruction Knowledge Management System; TSFI-KMS），而教師的 PCK 一向被視為是教師專業成長的重要指標（NSTA, 1998），因此，教師對於「科展探究教學」的「學科教學知能」即為本研究所開發的「教師科展專業知識分享社群平台」中所要進行知識管理與分享的內容知識，雖然由網路社群平台所創立的專業知識管理對於網路社群還有成員都十分重要，但是大多數的教師社群平台系統都沒有從知識管理（knowledge management, KM）的角度來設計，因此，本研究所開發的系統平台在設計時整合了社群知識管理系統工具（knowledge management tools; KM tools）（Spector, 2002）與個人知識管理系統工具（personal knowledge management tools; PKM tools）（Tsui, 2002），此外，本研究也針對使用者的不同偏好設計了社群為主（community-

based）與個人為主（personal-based）兩種操作模式，而本研究也進一步針對國小科學教師對於 TSFI-KMS 所提供的知識管理工具之有用性與其對於兩種操作模式的使用偏好進行評估。

三、系統開發

（一）系統概念

本研究中建置的「教師科展教學網路社群平台」，其目的為讓對科展有興趣的國小教師能夠在此分享科展相關知識，使有經驗的教師能夠給予生手教師指導科展上的建議，希望藉此提升教師們的專業知能，本平台特色如下：

1. 為了能讓教師們符合自身的網站使用習慣，平台分別提供了「以社群為主」及「以個人為中心」兩種操作模式（如圖 1），希望能讓教師在使用網站上更加便利。
2. 平台也提供了教師們可以共同進行專案的空間，讓教師們可以更密切地進行溝通及合作。
3. 「以社群為主的操作模式」中討論區的是與國小教學現場教師討論，並結合科學教師「探究教學 PCK」進行設計。

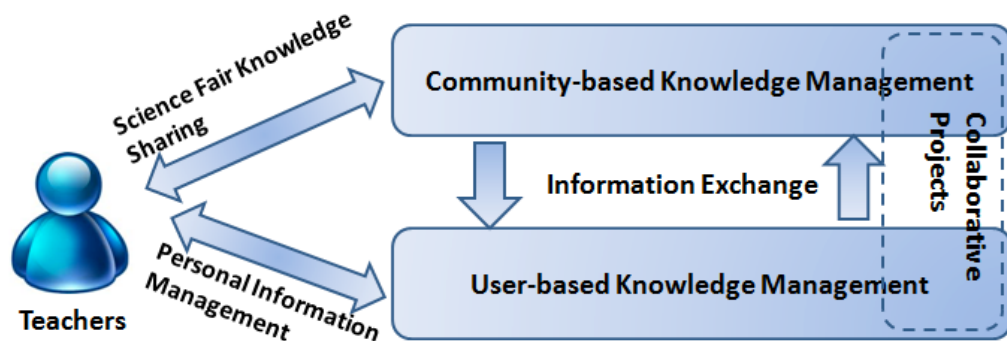


圖 1 系統概念圖

(二) 系統架構

本研究之「教師科展專業知識分享平台」，一共分成七大模組，分別為「網站成員管理模組」、「管理員功能模組」、「專案執行與互動模組」、「即時資訊通知模組」、「科展檔案管理模組」、「科展資訊交流模組」、「文章搜尋模組」，圖 2 為本研究「教師科展專業知識分享

平台」之系統架構圖。由圖中可知，使用「教師科展專業知識分享社群平台」的角色有二，分別為「網站會員」及「網站管理員」。網站會員可以透過系統中的模組，與其他成員進行科展相關資訊的交流，並且可以透過專案執行模組讓教師們密切的合作；網站管理員可以管理系統公告以及網站內的科展相關文章。

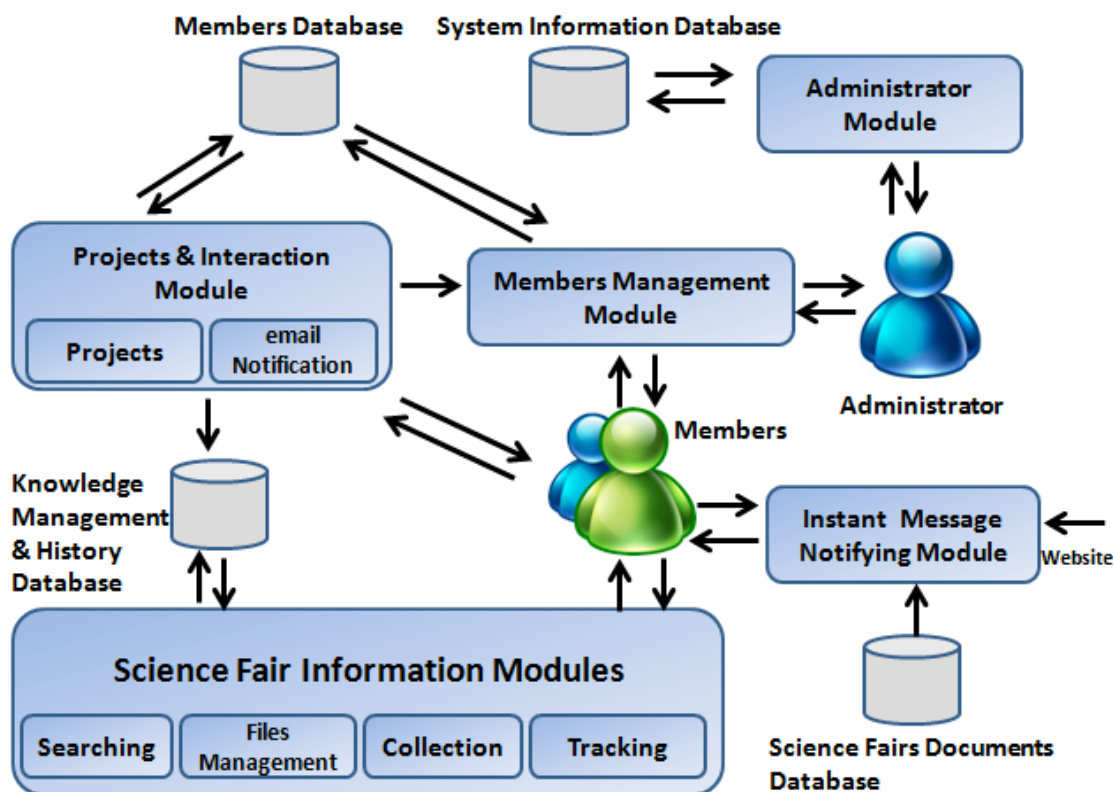


圖 2 系統架構圖

圖 3 與圖 4 為 TSFI-KMS 系統的畫面截圖，其中圖 3 是 TSFI-KMS 社群操

作模式的介面，而圖 4 則是 TSFI-KMS 個人操作模式的介面。



圖 3 TSFI-KMS 社群操作模式介面



圖 4 TSFI-KMS 個人操作模式介面

四、系統評估

本研究開發的「教師科展專業知識分享平台」，使用對象為國小教師且不限定教師的教學科目，本研究採用線上問卷填答的方式進行，參與實驗且回收的總問卷數為 284 份，刪除其中 102 份填答不完整的問卷，完整進行問卷填答並回收之有效問卷為 182 份。系統評估的時間為期兩週。本研究主要分成三個部分進行，第一部分為「教師網站使用及科展相關經驗調查」，於教師進入「教師科展專業知識分享社群平台」進行操作前，先請教師填寫其基本背景資料以及網站使用與科展相關經驗之背景變項；第二部分「操作教師科展專業知識分享平台」，讓國小教師實際進行操作，體驗平台的運作，為了能讓此次實驗回收的問卷資料能更真實的反映教師使用平台的感受，於進行第三部分的「系統評估」前，在此設計了驗證機制，教師必須瀏覽網站超過 10 分鐘且點擊網站的次數達 30 次以上，目的是希望教師們能確實地進行網站瀏覽後再進行問卷填答。第三部分

「系統評估」，本研究依系統評估需求，進一步透過問卷調查得知國小教師們對於此平台的知覺有用性及知覺易用性以及教師偏好的操作介面模式。

為了瞭解國小教師對 TSFI-KMS 的知覺有用性、易用性與使用意願，本研究共使用兩份問卷，其中之一參考 Phang et al. (2009) 所開發的問卷來評估系統整體有用性、易用性及使用意願，另一份問卷則分別對網站中的「以社群為主」與「以個人為中心」的操作模式進行開發，用以評估教師對此二種操作模式各別的有效用性、易用性及使用意願。系統整體評估的部分則參考 Phang et al. (2009) 所開發之問卷，並將問卷依向度的題項進一步依題項內容的敘述歸類成評估系統有用性、易用性及使用意願：知識追蹤 (knowledge tracking fulfillment) 與社會互動 (social interactivity) 歸類至有用性、系統易用性 (ease of use) 與系統可靠度 (system reliability) 歸類至易用性、知識搜尋與貢獻 (knowledge seeking/contribution) 歸類至使用意願 (如表 1)。

表 1 系統整體評估量表信度檢驗

| Scale | Items | α | Example |
|---------|-------|----------|---|
| 知識搜尋、貢獻 | 3 | 0.92 | 我會想要使用「教師科展專業知識分享社群平台」找尋有用的資訊。 |
| 易用性 | 4 | 0.9 | 「教師科展專業知識分享社群平台」的操作模式不會令我感到困擾。 |
| 系統可靠度 | 3 | 0.82 | 「教師科展專業知識分享社群平台」是穩定的系統。 |
| 知識追蹤 | 3 | 0.93 | 「教師科展專業知識分享社群平台」能讓我追蹤文章動態。 |
| 社會互動 | 3 | 0.92 | 「教師科展專業知識分享社群平台」提供的功能，可以幫助我與其他網站成員進行互動。 |
| 整體問卷 | 16 | 0.95 | |

此外，本系統中「以社群為主」的操作模式，主要參考 Spector (2000) 提出的四項知識管理工具，依本研究之需求，對其中的「溝通」、「協調」、「合作」工具進行問卷開發，此三個向度的題項數量皆為 4 題（下表 2），且經檢驗後信度（Cronbach's alpha, α ）分別為 0.92、0.88 與 0.87；「以個人為中心」的操作模式則參考 Tsui (2002) 所提出之個人知識管理工具，依本研究之需求，對其

中的「搜尋工具」、「整合式搜尋工具」、「相關連結」、「資訊取得/資訊分享工具」、「電子郵件之管理、分析及整合」、「協作及同步工具」、「學習工具」進行問卷開發，此七個向度的題項數量各為 2 題、3 題、2 題、4 題、2 題、3 題、2 題，且信度（Cronbach's alpha, α ）分別為 0.85、0.88、0.88、0.91、0.93、0.87、0.79，而整個問卷的信度為 0.96，可見此問卷有良好的信度。

表 2 知識管理工具評估量表信度檢驗

| | | Items | α | Example | |
|-------|---------------|---------|----------|----------------------------|--------------------------|
| 以社群為主 | 溝通 | 4 | 0.92 | 可以幫助我傳遞訊息給其他社群成員。 | |
| | 協調 | 4 | 0.88 | 可以幫助安排我與社群其他成員共同線上討論的時間。 | |
| | 合作 | 4 | 0.87 | 可以幫助社群成員在小組合作時互相交流檔案或工作文件。 | |
| | | 搜尋工具 | 2 | 0.85 | 可以幫助我利用關鍵字進行搜尋找到我要的資料。 |
| | | 整合式搜尋工具 | 3 | 0.88 | 可以讓我搜索並比較科展社群與我的窩中的相關資料。 |
| 以個人為主 | 相關連結 | 2 | 0.88 | 可以提供我外部網站的即時資訊。 | |
| | 資訊取得/資訊分享工具 | 4 | 0.91 | 可以讓我分類儲存我的檔案或資訊。 | |
| | 電子郵件之管理、分析及整合 | 2 | 0.93 | 可以列出我所收到或發出的訊息（短消息）。 | |
| | 協作及同步工具 | 3 | 0.87 | 可以讓我跟其他成員進行專案合作。 | |
| | 學習工具 | 2 | 0.79 | 可以提供我使用平台的活動歷程紀錄。 | |

而操作介面的使用偏好評估使用，則是將 TSFI-KMS 不同操作模式的使用偏好分為四個題項：兩者都喜歡、較喜歡社群為主的操作模式、較喜歡個人為主的操作模式與兩者都不喜歡。

五、研究結果與討論

(一) 教師對於 TSFI-KMS 的整體評估

由表 3 中數據得知，國小教師對於

TSFI-KMS 整個系統的知覺有用性平均值為 5.13，其平均值高於六點量表平均值 3.5，因此得知，參與本研究的國小教師認為，TSFI-KMS 對於增加教師指導科展的專業知能是有用的。此外，國小教師對於 TSFI-KMS 的知覺易用性的平均值為 5.1，其平均值亦高於六點量表平均值 3.5，亦即參與本研究的國小教師傾向同意 TSFI-KMS 的整體操作上是容易使用

的。而國小教師對於 TSFKS 的使用意願調查結果平均值為 5.1，因此，參與本研

究的國小教師亦有高度的意願使用 TSFI-KMS。

表 3 TSFKS 各知覺表現分析

| | Mean | S.D. | Range |
|-------|------|------|-------|
| 知覺有用性 | 5.13 | 0.49 | 3-6 |
| 知覺易用性 | 5.1 | 0.52 | 1-6 |
| 使用意願 | 5.1 | 0.54 | 2-6 |

(二) 教師對於 TSFI-KMS 的知識管理工具「知覺有用性」分析

參與教師在對「以社群為主的操作模式」中提供的 KM tools 其各別向度的得分如表 3，「溝通」、「協調」、「合作」的平均得分分別為 5.32、5.06、5.32，

因此，根據這些老師的填答，表示教師們認為系統所提供的 KM tools 對於增加教師指導科展的教師專業知能是有用的，而在未來進行系統改善時，可以針對 TSFI-KMS 中「協調」的部分進行加強。

表 4 「社群知識管理工具」有用性分析

| | Mean | S.D. | Range |
|----|------|------|-------|
| 溝通 | 5.32 | 0.58 | 3-6 |
| 協調 | 5.06 | 0.66 | 2-6 |
| 合作 | 5.32 | 0.54 | 2-6 |

而分析受試者在「以個人為中心的操作模式」中提供的 PKM tools 的回答（如表 5），發現每一向度的得分均大於 5，表示教師們認為系統所提供的 PKM tools 對於增加教師指導科展的教師專業知能是有用的。綜合上述結果來看，此二類不同的知識管理工具中，「協同合作」的部

分的平均得分略低於其他知識管理工具，而此部分在 TSFI-KMS 中所對應的模組為「專案執行模組」，因此，未來改善系統時，可對「專案執行」的部分進行加強，以提供一個更符合教師需求的專案執行空間。

表 5 「個人知識管理工具」有用性分析

| | Mean | S.D. | Range |
|---------------|------|------|-------|
| 搜尋工具 | 5.26 | 0.49 | 2-6 |
| 整合式搜尋工具 | 5.09 | 0.6 | 2-6 |
| 相關連結 | 5.04 | 0.6 | 2-6 |
| 資訊取得 / 資訊分享工具 | 5.09 | 0.56 | 2-6 |
| 電子郵件之管理、分析及整合 | 5.09 | 0.56 | 2-6 |
| 協作及同步工具 | 5.08 | 0.54 | 3-6 |
| 學習工具 | 5.08 | 0.53 | 2-6 |

(三) 教師對於 TSFKS 的使用喜好程度與使用偏好分析

調查教師對於 TSFI-KMS 不同操作模式的使用偏好分為四個題項進行調查依序是「社群為主」、「個人為中心」、「兩者都喜歡」、「兩者都不喜歡」，其次數分配見表 6，由表中得知偏好使用「以社群為主的操作模式」佔總人數的

49%，接近總受測人數的一半。因此，參與本研究的國小教師對於「以社群為主的操作模式」及「以個人為中心的操作模式」都表示喜歡，但偏向使用「以社群為主」的操作模式，雖然如此，還是有許多受訪教師（26%）偏好「以個人為中心的操作模式」，這也表示兩種操作介面有同時存在的必要性。

表 6 教師使用 TSFI-KMS 之使用偏好分析

| 使用偏好 | n(%) |
|-------|---------|
| 社群為主 | 89(49%) |
| 個人為中心 | 26(14%) |
| 兩者都喜歡 | 67(37%) |

六、結論

本研究參考學者所提出的社群知識管理工具與個人知識管理工具，建置「教師科展教學之網路社群平台」，希望能促進教師在科展教學的專業成長，而考量使用者使用偏好，平台提供「以社群為主」與「以個人為主」兩種操作模式，讓教師能依自身偏好自行選擇。而透過系統初步評估也顯示受測教師對於教師科展教學之網路社群平台所提供之兩類

知識管理工具的有用性均有十分正向的回應，而「以社群為主」與「以個人為主」兩種操作模式的設計也有其功能，希望未來有更多科學教師能利用 TSFI-KMS 進行科展教學之專業成長。

致謝

本研究在中華民國行政院科技部的資助下完成，補助計畫編號為 101-2628-S-008-001-MY3，僅此致謝。

參考文獻

- 林東清（2007）。知識管理。台北：智勝文化事業有限公司。
- 黃鴻博（1999）。以學生為中心科學探究活動指導模式之研究。教育部八十八年度中小學科學教育補助專案研究計畫報告書。
- 黃旭盛（2010）。透過同儕互評建立討論區發言指數之研究。國立東華大學網路與多媒體科技研究所碩士論文，未出版，花蓮縣。
- Abd-El-Khalick, F., BouJaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., et al. (2004). *Inquiry in science education: International perspectives. Science Education*, 88(3), 397 – 419.
- Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, 13(1), 1-12.
- Bell, R., Smetana, L., & Binns, I. (2005). Simplifying inquiry instruction. *The Science Teacher*, 30-33.
- Bencze, J. L. & Bowen, G. M., (2009). A national science fair: Exhibiting support for the knowledge economy. *International Journal of Science Education*, 31, 2459-2483.
- Carroll, J. M., Choo, C. W., Dunlap, D. R., Isenhour, P. L., Kerr, S. T., MacLean, A., & Rosson, M. B. (2003). Knowledge management support for teachers. *Educational Technology Research and Development*, 51(4), 42-64.
- Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical content knowing: An integrative model for teacher preparation. *Journal of Teacher Education*, 44(4), 263-272
- Collins, A., & Bielaczyc, K. (1997). Dreams of technology-supported learning communities. *Proceedings of the Sixth International Conference on Computer-Assisted Instruction, Taiwan, 1997*.
- Davenport, T.H., & Prusak, L. (1998), *“Working Knowledge : How Organizations Manage What They Know”*, Boston: Harvard Business School Press.
- De Souza, C. S., & Preece, J. (2004) *“A Framework for Analyzing and Understanding Online Communities” Interacting with Computers*, 16,579-610.
- Grossman, P. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teachers College Press.
- Lee, J. (2001). The impact of knowledge sharing, organizational capability and partnership quality on IS outsourcing success. *Information & Management*, 323-335.
- Jonassen,D. H.,Howland, J., Moore, J., & Marra, R.M.(2002). *Learning to Solve Problems with Technology: A Constructivist Perspective. (2nd ed.)* Upper Saddle River, New Jersey.
- Justi, R. & Gilbert, J.K. (2002).Modeling, teachers’ views on the nature of modeling, and implications for the education of modelers. *International Journal of Science Education*, 24(4), 369–387.
- Magnusson, S., Krajcik, J. & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In G.-N. Lederman (Ed.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp.95-132). Dordrecht: Kluwer Academic Publishers.
- National Science Teachers Association. (1998). *Pathways to the Science Standards: Guidelines for Moving the Vision into Practice*. Arlington, va.: NSTA Press.
- National Research Council (NRC). (2000). *Inquiry in the national science education standards: A Guide for teaching and learning*. Washington, DC: National Academy Press.
- Nonaka, I. & Takeuchi, H. (1995). *The Knowledge Creating Company*. Oxford University Press.
- National Science Teachers Association (NSTA). (1998). *NSTA/NCATE standards for science teacher preparation*.

- Pennel, J., & Firestone, W. (1998). Teacher-to-teacher professional development through state-sponsored networks. *Phi Delta Kappan*, 79(5), 354.
- Rheingold, Howard. (2000). *The virtual community: Homesteading on the electronic 80 frontier*. London: MIT Press.
- Russell, R. H. (1996). Providing access: The difference between sharing and just reporting corporate information. *Information Strategy: The Executive's Journal*, 12:28-33.
- Shrivastava, P. (1999). December. Management classes as online learning communities. *Journal of Management Education*, 23, 691-702.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4-14.
- Spector, J.M., (2002). Knowledge management tools for instructional design. *Journal of Educational Technology Research and Development*, 50(4), 37-46.
- Sumsion, J., & Patterson, C. (2004). The emergence of community in a preservice teacher education program. *Teaching and Teacher Education*, 20,621-635.
- Tsui, E. (2002), “Technologies for personal and peer-to-peer (P2P) knowledge management”, CSC Leading Edge Forum (LEF) Technology Grant Report, available at www.csc.com/aboutus/lef/mds67_off/uploads/P2P_KM.pdf
- Windschitl, M. (2004). Folk theories of “inquiry:” How preservice teachers reproduce the discourse and practices of a theoretical scientific method. *Journal of Research in Science Teaching*, 41(5), 481–512.

人類行為訊號處理 (BSP)：一個新興的跨領域研究，其未來在教育領域自動評估的應用

李祈均

國立清華大學電機工程學系助理教授

一、前言

人類行為訊號處理 (Behavioral Signal Processing, BSP) (Narayanan & Georgiou, 2013) 是一個新興的跨領域研究，目的是以統計計算方式來對人類行為的訊號進行量化、分析、及建模。這個研究領域產物被稱為 – behavior informatics – 先進的跨領域整合式人類行為演算法。這些演算法的研究成果可提供不同領域的專家更客觀 (objective)、更有效率 (efficient)、及更大規模 (large scale)

的方法來分析及評斷人類行為。此概念是來自於人類的行為會表現在顯性和隱性的線索 (overt and covert cues)，這些外顯行為的線索經過演算法萃取及模型，可以提供專家來進行分析與決策。很多的研究都集中在開發相關的數學演算法以用來對於人主觀評價 (subjective evaluation) 的評量過程進行系統式的自動化。

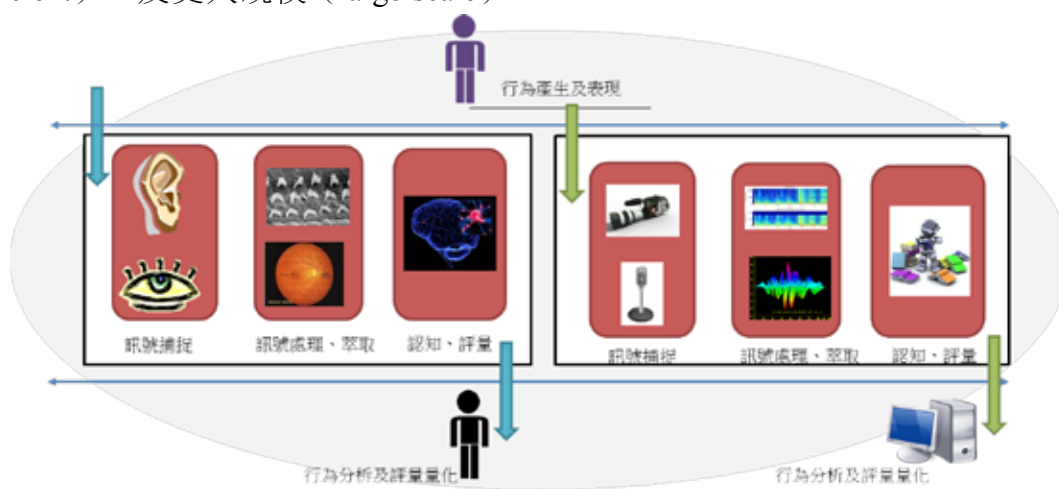


圖 1 行為分析及評量：專家處理的三個要件及電腦訊號處理的三個相對應要件

圖 1 很統括的把這種系統的可行性概念給表達出來。舉一個在精神醫療例子來說，A (病人) 去 B (醫生) 進行診斷及治療。因為 A 的精神狀態導至 A 行為表現即會透露出一些被醫生注意到的行為線索。醫生會注意到的行為線

索通常是依據該科的診斷書 (diagnostic instrument) 或是專業上訓練和經驗判斷。依此，B 可以對 A 的精神狀態透過觀察 A 的行為做出一個主觀的判斷，並接續著提供相關的治療。這種 B 的主觀判斷能力可以抽象化的分成三個主要要

件：行為訊號捕捉、行為訊號分析、及行為訊號辨識。行為訊號捕捉管道可透過人的感官（sensing），譬如視覺及聽覺。這些訊號接著會被人的感官神經進行處理及分析後（processing），接著把處理過後的特徵（features）送往大腦進行辨識（recognition）。另一個簡單的生活化例子如下：A 生氣時會說話變得很大聲，這個‘大聲’的特徵值經過 B 感官，即耳朵和聽覺神經的捕捉跟處理，接著將此特徵值送往 B 大腦。而 B 的大腦早已學習過‘大聲 = 生氣’的判別，因此可以很自然的可以對 A 的情緒進行自動辨識。

現代數位科技對於這三個不同要件已有長足的發展，因此也慢慢整合出一套電腦可以完成的端至端（end-to-end）

‘人類行為訊號處理’系統。數位科技的進步使的訊號捕捉的方式變的普及化，譬如錄音錄影設備可以方便及快速的錄製大量的聲頻及視頻；日漸便宜的儲存媒介也讓這些資料可以做長期及大量的保存。訊號處理技術的進步也使的這些影像聲音的捕捉更加清晰、低噪音，並且也同時發展出各種不同的計算方法來對原始聲頻及視頻串流進行快速的‘特徵’計算（feature computation），這些特徵的計算是仿效人類感官的處理（perceptual processing）。最後，統計模型、人工智慧、機器人學習等數學及工程領域的發展，隨著資料量的增加、特徵更精準的萃取、以及更快速的電腦運算，也產生了許許多多複雜且高階的演算法使機器能從資料中‘學習’並進行以數學統計為基礎的自動模式識別（pattern recognition）。這三個在不同科技領域的發展慢慢在匯流形成一個新興的研究方向，讓電

腦也可以以高可靠度的方式做一些一向以來皆由人類大腦在執行的的主觀判別。

這篇文章首先要介紹這個新興領域對於跨領域人類行為的研究遠景及動機，簡略的介紹既有的精神醫療方面的研究，以及最後如何將此概念移植到教育領域。透過密合的跨領域合作，即資訊科技及教育專家，在教育領域裡的各式各樣評量或也可以慢慢的採用電腦透過人類行為的客觀模型之演算法。

人類行為訊號處理（Behavioral Signal Processing, BSP）（Narayanan & Georgiou, 2013）是一個新興的跨領域研究，目的是以統計計算方式來對人類行為的訊號進行量化、分析、及建模。這個研究領域產物被稱為 – behavior informatics – 先進的跨領域整合式人類行為演算法。這些演算法的研究成果可提供不同領域的專家更客觀（objective）、更有效率（efficient）、及更大規模（large scale）的方法來分析及評斷人類行為。此概念是來自於人類的行為會表現在顯性和隱性的線索（overt and covert cues），這些外顯行為的線索經過演算法萃取及模型，可以提供專家來進行分析與決策。很多的研究都集中在開發相關的數學演算法以用來對於人主觀評價（subjective evaluation）的評量過程進行系統式的自動化。

二、研究遠景及動機

人類行為訊號處理的技術發展，主要目的之一是在於幫忙不同領域的專家解決對於人行為評量既有方法的問題。這種依靠人（專家）來對其他人的行為進行評量的過程被統稱為人工觀測方法（manual observational approach）（Margolin et

al., 1998)，幾個長期困擾的問題如下（Baucom & Iturralde, 2012）：

（一）主觀評量 (subjective evaluation)

人工觀測方法，因為是‘人工’，所有的評量皆屬於主觀（subjective）評量。所有主觀評量都會存在著因為人的因素（human factors）而產生不確性的問題（issue of reliability）。例如，疲勞，經驗不足，觀念不同，及其他相關非跟行為評量本身相關因素都會造成人工評量結果存在著不可避免也不容易控制的變數。

（二）耗時 (time-consuming)

人工觀測方法，因為是‘人工’，也相對比較耗時。再加上主觀評量有的先天問題，在各別專業領域，譬如醫療或心理，相當多的研究都著重在如何訓練人員進行有一致性（consistent）及可靠（reliable）的評量。解決主觀評量的問題包含了對個別人員進行紮實訓練，亦或是運用多位評量員來減低個別評量員可能造成的問題。這樣為了確保一致性的方式進一步的使得人工觀測方法更加費時。

（三）不容易擴展 (non-scalable)

由於前兩項的問題，利用人力來進行行為評量很難進行大規模應用。無法擴展成大規模應用的後果導致在行為研究方面僅能維持相對小量的樣本數來加以分析，或是將評量方式改以問卷或其他非行為類評量方式以便執行大規模應用。在現今數位科技的時代，行為訊號捕捉的工具已相當普遍，資料量的累積也逐漸變得快速且大量。透過行為評量可以比問卷或其他類似的測量對要測量的行為或能力進行更精準即客觀的評斷。

透過人類行為訊號處理的技術發展，目的在於提供自動演算法來解決以上述的相關問題。從一些既有的專家評量中，讓電腦演算法可以從這些既有行為資料中先學習如何模擬專家觀察的標準。在確保可信度（reliability）後，這套自動化評量系統即可進一步達成一致性（consistent），可複製性（repeatable），及大量且快速的評量（scalable）。人工智慧演算法的可信度也會隨著資料量的不斷累積而變得更準確。

同時，這樣的技術發展及運用其實更可以讓專家做出明智的決策（informed decision），而並非要讓專家被電腦給取代。因為在多數的領域，行為的評量只是介於最後專家決策與一開始人類行為產生中間的一個過程（見圖2）。譬如說，醫生依據診斷的評量書的結果來進行合適的臨床醫療行為。現今大多數人工行為評量的方式其實是在這個過程中導致了資訊瓶頸（information bottleneck）－使的多數的資訊並無法被最後的決策者給使用。透過自動化的評量，電腦可以提供各領域專家快速而且更客觀的行為評量，加上其他相關資料以便進行最後總和的決策。



圖 2 專家決策過程：評量後依結果做決策

三、醫療上的應用

人類行為訊號處理領域在過去幾年中對精神醫療方面有深入的研究。我們在這簡略的介紹其中兩個在國外子領域的發展：婚姻諮詢（couples counseling and therapy）、泛自閉症症狀（autism spectrum disorder, ASD）。

（一）婚姻諮詢

長年有婚姻生活有問題的夫妻尋求夫妻心理治療（couple therapy）在國外已是常態。臨床心理學家已建構對這些夫妻數種不同的治療方式，譬如 Traditional Behavioral Couple Therapy（Neil S Jacobson & Margolin, 1979）、Integrative Behavioral Couple Therapy（Christensen et al., 2004; N. S. Jacobson, Christensen, Prince, Cordova, & Eldridge, 2000）。在作為治療的一部分，夫妻都需要經歷解決問題的談話互動（problem-solving interaction）。在這十分鐘的談話互動中，夫妻必須各選一個他們婚姻中的問題提出來討論並尋求解決的方法。臨床心理學家會將這個對談互動進行錄音錄影。專家們為了研究治療的效果，夫妻互動對談的模式，及將其行為來預測未來產生離婚或家庭暴力的可能，臨床心理學家會請數名人員（通常3至5名），經過訓練後，以觀看錄音錄影檔的方式來對夫妻的互動進行行為評量。

行為評量的準則是根據兩份已被認可的行為評量表：Couples Interaction Rating System（CIRS）（Heavey, Gill, & Christensen, 2002）和 Social Support Interaction Rating System（SSIRS）（Jones & Christensen, 1998）。這兩份量表對夫妻

各別在解決問題的互動的行為進行評分，總共評分項目有33項。以下舉例其中一個評分項目，內容是直接引用SSIRS量表：

“**Blame:** Blames, accuses, or criticizes the partner, uses critical sarcasm; makes character assassinations such as, “you’re a real jackass,” “all you do is eat,” or “why are you such a jerk about it?” Explicit blaming statements (e.g., “you made me do it,” or “you prevent me from doing it”), in which the spouse is the causal agent for the problem or the subject’s reactions, warrant a high score.”

這樣的行為評量方法已行之有年，但臨床心理學家對於這樣的人工觀測方法感到非常耗時而且有時並無法達到必要的可靠度。第一個首要的問題是必須要徵求多名評量員，然後再根據量表內不同的行為定義對評量員加以訓練。所以臨床心理學家對於可以將這樣乏味且耗時的工作自動化覺得不但可以促進研究效率，更可以對這些有心裡有障礙的夫妻們提供更適時的幫忙。

在人類行為訊號處理的研究中，已有數篇文章提出相關技術來對這些專屬的量表進行自動化。這些系統有的是單純藉由語音（acoustics）（Black et al., 2013），有的是語音加用詞（acoustic + lexical）（Black, Georgiou, Katsamanis, Baucom, & Narayanan, 2011; Georgiou, Black, Lammert, Baucom, & Narayanan, 2011），有的是透過對夫妻彼此頭部的動作（head motion）（Xiao, Georgiou, Lee, Baucom, & Narayanan, 2013）。這些初步研究都已顯示自動化系統的可行性，系統準確率也都至少大約七成左右。

（二）泛自閉症症狀

美國 Centers for Disease Control and Prevention (CDC) 預估在美國差不多每 88 個小孩中就有一個有泛自閉症症狀。這是個嚴重的且昂貴的社會問題。在研究及臨床診斷自閉症的過程中，常被當成是黃金標準 (gold standard) 的是 Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2000) 標準化觀察行為量表。當父母帶小孩進醫院進行診斷，臨床心理學家必須要花 30 至 45 分鐘之間的時間來跟小孩互動，並在互動的過程中將小孩不同的行為根據 ADOS 量表進行記錄。

這樣的過程不只是耗時，即診斷時間方面，還有很多其他潛在不確定因素。譬如，臨床心理學家在這個狀態下不只是第三方觀察者，他們同時也是在跟小孩互動的其中一方。這樣的診斷設計必須要求被認可進行診斷的醫療人員經歷更長時間的訓練 – 這尤其重要因為這些診斷會影響小孩子接下來所要接受的醫療行為。自動化部分量表內容不儘可以節省時間，其實更可以提供家長一個‘早期發現’的機會。在研究方面，部分量表的行為其實相對的不容易透過人，即診斷的醫師，來進行評量。譬如說 ADOS – Module 3 (內含有 28 個不同行為評量項目) 的其中一項行為評量是 Speech Abnormality，在量表內的形容引用如下：

“slow and halting; inappropriately rapid; jerky and irregular in rhythm . . . odd intonation or inappropriate pitch and stress, markedly flat and toneless, . . . consistently abnormal volume”

對於語音變化的量化其實透過電腦的運算比人工判別客觀。事實上已有些初

步的人類行為訊號處理的研究指出，透過電腦量化語音及用詞 (Bone et al., 2013) 或肢體行為 (Delaherche et al., 2013) 後，可以對小孩的自閉症的嚴重程度評分 (autism severity score) 進行自動化的評量。此外因為是透過電腦的運算，可以直接對小孩和臨床心理學家在診斷的過程建立一個共同行為的量化及模型 (joint modeling) (Bone et al.)。這樣雙邊行為的同時觀測甚至可以對小孩的自閉症的嚴重程度評分的準確率更加提升。

四、教育上的應用

人類行為訊號處理技術在對於教育相關的評量在概念上其實是跟精神疾病醫療應用相當類似。因為在教育評量領域裡所遭遇的問題跟第二段所形容的人工觀測評量是相仿的。雖然在教育科技的發展目前還沒有像醫療應用的多面向且縝密的跨領域整合，但在國外教育類科技的發展中，已經有些初步技術對‘第二外語’學生的發音進行自動評量。近期也有數篇論文在研發針對老師的網路上教學的好壞評比研發自動化的評量系統。

第二外語的學生學習英文，在發音上會被母語的影響而導致部分發音的不準確。英語教學的一個重點在確保這些第二外語的學生的發音得準確度。雖然沒有像醫療用的制式量表，但英語老師，甚或一般母語為英文的人皆可以判別發音的正確與否，並對不準確的發音進行修正。一直以來這些方式都是透過人工的方法，使的外語學習普及化的速度並不快速。

對於該培訓的能力及內容都已經過長時間的淬煉及證實。在既有的課程設計及評量標準中如何幫忙教育專家注入更客觀、更迅速、及最重要的 – 更有用，的自動化資訊技術是未來在教育上應用的關鍵。

對於師資培育，可以建立一套系統來提供跟現有人工方式不一樣的行為表現評分標準，相對比較客觀也比較有可能進行大規模的應用。對於學生的學習效率，國外文獻也漸漸指出學生與老師上課中 *emotional and social engagement* 是一個重要因素 (Blanchette & Richards, 2010; Martin & Dowson, 2009)。如何準確的評量出教學或學習效率，其實學生或老師在課堂上互動行為是非常關鍵的。在人類行為訊號處理技術慢慢的成熟後，數位科技也可以漸漸進入校園及教室。藉由這些技術根據師生行為的互動來進行數位紀錄，以及用演算法自動判別不同教學及學習

的狀況 - 藉此也可以更準確的掌握教學效用。

人類行為訊號處理技術 (Behavioral Signal Processing) 是其中一個新興的‘以人為中心的計算 (Human-centric Computing)’的研究領域。其他相關領域，譬如人類社交訊號處理 (Social Signal Processing) (Pentland, 2007)、計算行為科學 (Computational Behavioral Science) (Oberleitner, Abowd, & Suri, 2013)、情感計算 (Affective Computing) (Picard, 2000) 也都是國外新興的研究領域。專門對於人的行為、情緒、社會互動透過客觀的行為捕捉、處理、統計建模來研究人彼此之前的行為互動。這些新興領域的發展對於將來把這些技術實際應用於醫療或教育的評量行為皆會有很大的幫助。

參考文獻

- Alwan, A., Bai, Y., Black, M., Casey, L., Gerosa, M., Heritage, M., . . . Lee, S. (2007). *A system for technology based assessment of language and literacy in young children: the role of multiple information sources*. Paper presented at the Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on.
- Baucom, B. R., & Iturralde, E. (2012). *A behaviorist manifesto for the 21st century*. Paper presented at the Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific.
- Black, M. P., Georgiou, P. G., Katsamanis, A., Baucom, B. R., & Narayanan, S. S. (2011). "You made me do it": *Classification of blame in married couples' interactions by fusing automatically derived speech and language information*. Paper presented at the Proceedings of Interspeech, Florence, Italy.
- Black, M. P., Katsamanis, A., Baucom, B. R., Lee, C.-C., Lammert, A. C., Christensen, A., . . . Narayanan, S. S. (2013). Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. *Speech Communication, 55*(1), 1-21.
- Black, M. P., Kazemzadeh, A., Tepperman, J., & Narayanan, S. S. (2011). *Automatically Assessing the ABCs: Verification of Children's Spoken Letter-Names and Letter-Sounds*. *ACM Transactions on Speech and Language Processing, 7*:4, 15:11 - 15:17.
- Black, M. P., Tepperman, J., & Narayanan, S. S. (2011). *Automatic Prediction of Children's Reading Ability for High-Level Literacy Assessment*. *Ieee Transactions on Audio Speech and Language Processing, 19*(4), 1015-1028. doi: Doi 10.1109/Tasl.2010.2076389

- Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion, 24*(4), 561-595. doi: Pii 914096069 Doi 10.1080/02699930903132496
- Bone, D., Lee, C.-C., Black, M. P., Williams, M. E., Lee, S., Levitt, P., & Narayanan, S. The Psychologist as an Interlocutor in Autism Spectrum Disorder Assessment: Insights from a Study of Spontaneous Prosody. *Journal of Speech, Language, and Hearing Research.*
- Bone, D., Lee, C.-C., Chaspari, T., Black, M. P., Williams, M. E., Lee, S., . . . Narayanan, S. (2013). *Acoustic-Prosodic, Turn-taking, and Language Cues in Child-Psychologist Interactions for Varying Social Demand.* Paper presented at the Interspeech, Lyon, France.
- Cheng, D. S., Salamin, H., Salvagnini, P., Cristani, M., Vinciarelli, A., & Murino, V. (2014). Predicting online lecture ratings based on gesturing and vocal behavior. *Journal on Multimodal User Interfaces, 1-10.*
- Christensen, A., Atkins, D. C., Berns, S., Wheeler, J., Baucom, D. H., & Simpson, L. E. (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology, 72*(2), 176-191. doi: Doi 10.1037/0022-006x.72.2.176
- Delaherche, E., Chetouani, M., Bigouret, F., Xavier, J., Plaza, M., & Cohen, D. (2013). Assessment of the communicative and coordination skills of children with Autism Spectrum Disorders and typically developing children using social signal processing. *Research in Autism Spectrum Disorders, 7*(6), 741-756. doi: DOI 10.1016/j.rasd.2013.02.003
- Georgiou, P. G., Black, M. P., Lammert, A. C., Baucom, B. R., & Narayanan, S. S. (2011). *That's aggravating, very aggravating": Is it possible to classify behaviors in couple interactions using automatically derived lexical features?* Paper presented at the Affective Computing and Intelligent Interaction.
- Heavey, C., Gill, D., & Christensen, A. (2002). Couples interaction rating system 2 (CIRS2): University of California, Los Angeles. Retrieved from <http://christensenresearch.psych.ucla.edu/>
- Jacobson, N. S., Christensen, A., Prince, S. E., Cordova, J., & Eldridge, K. (2000). Integrative behavioral couple therapy: an acceptance-based, promising new treatment for couple discord. *J Consult Clin Psychol, 68*(2), 351-355.
- Jacobson, N. S., & Margolin, G. (1979). *Marital therapy: Strategies based on social learning and behavior exchange principles*: Psychology Press.
- Jones, J., & Christensen, A. (1998). Couples interaction study: Social support interaction rating system. University of California, Los Angeles. Retrieved from <http://christensenresearch.psych.ucla.edu/>
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord, 30*(3), 205-223.
- Margolin, G., Oliver, P. H., Gordis, E. B., O'Hearn, H. G., Medina, A. M., Ghosh, C. M., & Morland, L. (1998). The nuts and bolts of behavioral observation of marital and family interaction. *Clin Child Fam Psychol Rev, 1*(4), 195-213.
- Martin, A. J., & Dowson, M. (2009). *Interpersonal Relationships, Motivation, Engagement, and Achievement: Yields for Theory, Current Issues, and Educational Practice. Review of Educational Research, 79*(1), 327-365. doi: Doi 10.3102/0034654308325583
- Narayanan, S., & Georgiou, P. G. (2013). Behavioral Signal Processing: Deriving Human Behavioral Informatics From Speech and Language: Computational techniques are presented to analyze and model expressed and perceived human behavior-variedly characterized as typical, atypical, distressed, and disordered-from

- speech and language cues and their applications in health, commerce, education, and beyond. *Proc IEEE Inst Electr Electron Eng*, 101(5), 1203-1233. doi: 10.1109/JPROC.2012.2236291
- Oberleitner, R., Abowd, G., & Suri, J. S. (2013). Behavior Imaging[®]'s Assessment Technology: A Mobile Infrastructure to Transform Autism Diagnosis and Treatment *Imaging the Brain in Autism* (pp. 371-380): Springer.
- Pentland, A. (2007). Social signal processing. *Ieee Signal Processing Magazine*, 24(4), 108-111. doi: Doi 10.1109/Msp.2007.4286569
- Picard, R. W. (2000). *Affective computing*: MIT press.
- Salvagnini, P., Salamin, H., Cristani, M., Vinciarelli, A., & Murino, V. (2012). *Learning how to teach from "Videolectures": automatic prediction of lecture ratings based on teacher's nonverbal behavior*. Paper presented at the Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on.
- Xiao, B., Georgiou, P. G., Lee, C.-C., Baucom, B., & Narayanan, S. S. (2013). *Head motion synchrony and its correlation to affectivity in dyadic interactions*. Paper presented at the Multimedia and Expo (ICME), 2013 IEEE International Conference on.

線上科學語言閱讀理解測驗之建置與發展

陳新豐

國立屏東教育大學教育學系助理教授

摘要

本研究旨在建置線上科學語言閱讀理解測驗（Online Reading Comprehension of Science Language Testing, ORCSLT）並探討科學語言閱讀理解測驗紙筆版本與線上版本的差異情形，研究目的有3，（1）探討 ORCSLT 的系統界面，（2）探討並建立 ORCSLT 的信效度及相關特徵，（3）探討科學語言閱讀理解測驗紙筆與線上版本的差異情形。研究對象包括 8 個班級 223 位國小學童，信度樣本 105 位，同時效度樣本 217 位。研究工具包括科學語言閱讀理解測驗（Reading Comprehension of Science Language Testing, RCSLT）、ORCSLT 及中文閱讀理解測驗等 3 種。研究結

果 ORCSLT 的信度方面，（1）中年級 ORCSLT 的 $\alpha=.79$ ，高年級 ORCSLT 的 $\alpha=.74$ ；（2）中年級 ORCSLT 的重測信度 $=.49$ ，高年級 ORCSLT 為 $.64$ 。效度部分中（1）中年級 ORCSLT 與中文閱讀理解測驗以及中年級紙筆版 RCSLT 的相關達中上相關（ $.64, .65$ ），（2）中年級版本無年級差異，高年級版本 6 年級優於 4 年級，（3）性別間的差異分析發現女生皆優於男生，（4）試題反應理論的測驗訊息中發現線上版較紙筆版困難且具年級差異。

關鍵字：科學語言、閱讀理解、線上測驗

Developing A On-Line Test for Assessing Students' Reading Comprehension of Scientific Language

Shin-Feng Chen

National Pingtung University of Education Assistant Professor

Abstract

Language helps the learners to know the world, so the scientific language plays an important role in the learning of scientific concepts. With the development of communicative technology, on-line computerized assessment is no longer metaphysical knowledge, but practical knowledge. The main purpose of this study was to developing a on-line test for assessing students' reading comprehension of scientific language. There were three goals to achieve, including to explore how to set the system of on-line scientific language reading comprehension test, to explore the reliability, validity, and the related characteristics, and to explore the differences between the paper test and on-line test of the scientific language reading comprehension. The participants were 223 elementary school students. The number of sample students for reliability was 105, and that for concurrent validity was 217. Three instruments were used in this study,

inclusive of RCSLT, ORCSLT, and Chinese Reading Comprehension Assessment. There were four main findings in this study. First, the alpha of MORCSLT was .79, and the alpha of HORCSLT was .74. The retest reliability of MORCSLT was .49, and that of HORCSLT was .64. As for the validity, the correlation between MORCSLT and the Chinese Reading Comprehension Assessment was above medium. Second, there was no grade variance as for the edition for middle graders, but the sixth graders were superior to the fourth graders as for the edition for the high graders. Third, girls were superior than boys in both paper assessments and on-line assessments. Fourth, from the viewpoint of item response theory, on-line assessments were harder than paper assessments, and there were significant difference between different graders when doing online assessments.

Key words: scientific language, reading comprehension, on-line testing

一、前言

語言是學習者認識世界的中介，學習者以它來認識世界，讓學習者在世界的知識中移動以及發展知識，一方面學習者學習概念上的語言是依賴關於世界的結構，另外一方面語言常影響學習者的觀察、描述世界的類型與方式（Schellings, Aarnoutse, & Leeuwe, 2006）。科學語言在教育研究上的應用已漸成趨勢（陳新豐、方金雅、王靜如、林曉雯，2012；Lin, Wang, Kao, & Chen, 2009；Wang, Lin, Kao, & Chen, 2009），科學語言在科學的學習中亦形重要，因此對於科學語言的研究需要投入更多的關注。

陳世文、楊文金（2006）指出，科學文本就像是電影，而文本當中的術語猶如電影中的演員，如果想使觀眾理解電影的內容，劇情脈絡的發展的重視就相對變得重要，科學文本中的術語通常是具有其特定的科學意涵或定義，如果缺乏術語的解釋將使學生無法建構出有系統的科學概念，反而使學生形成片面、零碎的科學知識，因此，我們經常可見科學文本透過技術性術語來描述或涵蓋自然現象及其背後的原理。

隨著資訊科技的發展與進步，電腦化測驗已不再只是理論上的議題，而是實務與理論兼具的知識議題，已應用在不少國家的職業證照或是選才測驗裡，例如：美國的軍職性向電腦化適性測驗（CAT version of the Armed Services Vocational Aptitude Battery, CAT-ASVAB）、美國的護士證照考試（National Council Licensure Examination, CAT-NCLEX-PN/RN）、藥師證照考試（North America Pharmacist Licensure Exam, NAPLEX）、美國商學研

究生入學檢定考試（Graduate Management Admission Test, GMAT），還有大家耳熟能詳的美國研究生入學考試/資格考試托福考試（Graduate Record Exam），也是採用電腦化測驗，而這些都是電腦化測驗成功的範例，極有學習的價值與意義。根據 Weiss（2010）的網站指出：目前有關電腦化適性測驗的研究主要有五大主題，包括內容平衡（Content Balancing）、試題曝光率（Item Exposure）、多重量表（Multiple Scales）、CAT 的終止標準（Terminating CAT）、線上 CAT（The Worldwide Weband CAT）等，其中目前電腦化測驗甚至是電腦化適性測驗在網際網路的使用更是值得學術與實務界關注的焦點，因此本研究擬建置科學語言閱讀理解的線上測驗系統，並比較不同的試題表徵方式對於國小學童閱讀理解能力的影響。

許多認知方面的研究結果發現，紙筆測驗與電腦化測驗上的成就並沒有差異，在參數的估計上，難度在紙筆與線上是沒有明顯的不同。參數估計上之差異，除了比較傳統的試題參數之外，還加入以試題反應理論為基礎的試題參數估計以及受試者能力之估計，利用變異數分析，比較在實際資料中，受試者人數的不同是否亦會造成參數估計上之不同。經過資料的收集及分析，本研究主要獲致以下的結論。當人數大致相等時，傳統紙筆測驗與線上電腦化測驗的試題參數估計是一致的。但也有一些研究的結果呈現不一致的結果（陳新豐，2005）。Mazzeo 和 Harvey（1988）年在研究中曾發現，在大

部分和電腦化測驗與傳統的紙筆測驗的測驗表現中，受試者的反應在平均得分的表現上都沒有達到顯著的差異。但是傳統的紙筆測驗的得分總是會略高於電腦化測驗上的平均得分。當施測的時間有所限制時，Greaud 和 Green（1986）曾表現，在速度測驗時，傳統的紙筆測驗與電腦化測驗的平均得分上，呈顯著的差異，在作答時間上，電腦化測驗也會比傳統的紙筆測驗還要久。

Englert, Mando 和 Zhao（2004）探討了電腦化與紙筆測驗對低成就的兒童在描述的能力上的關係。根據研究發現，我們發現低成就的學生在紙筆測驗和電腦化上的表現時，低學習成就的學生在用筆來回答電腦螢幕的問題，比回答在紙上的問題的還來的好，而且根據數據上的顯示，電腦化作答比傳統的紙筆測驗更來的有效。Tancock 和 Segedy（2004）用行動研究的方式對於二年級的學生做紙筆與電腦化上的做比較，發現用閱讀的方式將文章分數個等級和一群學生利用書面閱讀的方式，一群學生則在電腦上閱讀，發現書面閱讀的學生在對於文章內容的瞭解，優於在電腦上面閱讀的學生，在態度方面，學生則較喜歡在電腦上閱讀優於書面上的閱讀。而本研究除了探討不同試題表徵方式其測驗特徵是否有所不同之外，進一步探討科學語言閱讀理解工具不同試題表徵下對於國小學童閱讀理解能力是否有所差異進行探討而這亦是本研究主要的研究動機。

綜上所述，本文旨在利用不同表徵的科學語言閱讀理解測驗來探討國小學童的閱讀理解能力的表現，主要有3個研究問題，說明如下。

- 一、建置線上科學語言閱讀測驗系統界面為何？
- 二、線上科學語言閱讀理解測驗的信效度及相關特徵為何？
- 三、科學語言閱讀理解測驗紙筆以及線上版本中閱讀理解的表現為何？

二、文獻探討

本研究旨在建置線上科學語言閱讀理解測驗，並探討不同的試題表徵下國小學童科學語言閱讀理解的能力之表現為何？理論基礎部分依閱讀理解、科學語言、電腦化測驗、線上測驗及紙筆測驗與電腦化測驗的成就差異5個部分說明如下。

（一）閱讀理解

Hoover 和 Gough（1990）提出閱讀的簡單觀點（simple view of reading, SVR）是針對在學習閱讀之中的主要歷程，而 SVR 的觀點中明白指出閱讀理解（Reading Comprehension, RC）即為聽覺理解（Listening Comprehension, LC）以及字彙編碼（Decoding, D）的交互作用，用公式表示即為 $RC=LC*D$ 。因此閱讀理解與聽覺理解以及字彙的編碼有著密不可分的關係。另外，在佛羅里達的閱讀研究中心（FCRR）所建立的 PMRN 中認為閱讀成份主要包括5大部分，分別是閱讀理解（Comprehension）、字彙（Vocabulary）、流暢度（Fluency）、注音（Phonics）、音韻覺識（Phonological Awareness）等（FCRR, 2011）。其中的閱讀理解代表從文章獲取其意涵的能力，閱讀理解的發生主要是由於讀

者與文本之間的交互作用。閱讀之前（before reading），讀者的特徵，包括識字能力（word reading ability）與字彙（vocabulary）會直接影響到學生理解能力的水準。閱讀期間（during reading），學生依賴個人的認知和語言處理的能力，

在理解策略中的知識與技能來建構意義。閱讀之後（after reading）學生運用其它的策略和延伸的技能去建構對於文本的了解（understand），而學生閱讀理解的過程可以用下圖 1 來加以表示。

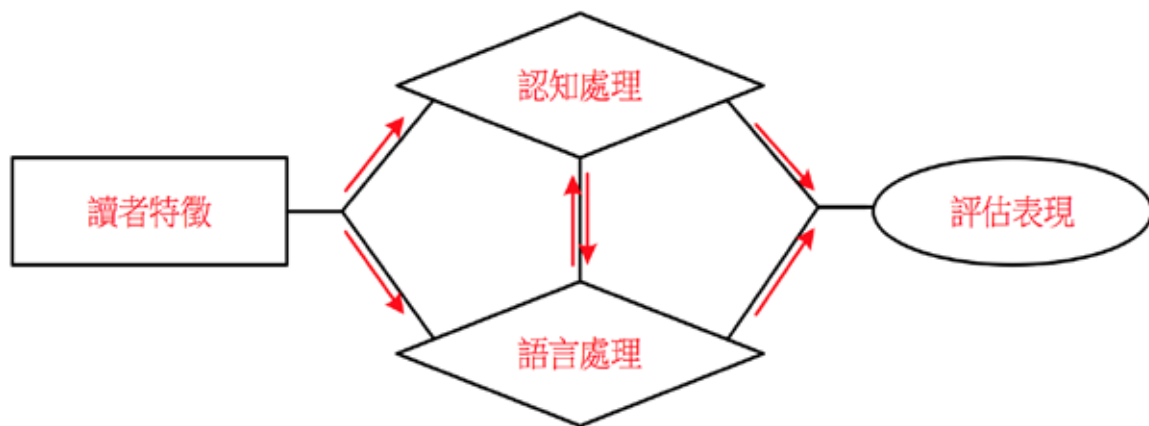


圖 1 閱讀理解的過程 (FCRR, 2011)

綜上所述，影響閱讀理解表現的分別是在閱讀前學習者的特徵，在閱讀中學習者的認知以及語言上的處理以及閱讀後的評估、監控的歷程。本研究所探討的是以科學語言下的閱讀理解能力為主，探討不同試題表徵下國小學童科學語言的閱讀理解能力。

（二）科學語言

科學語言在科學文本的閱讀理解上扮演著相當重要的角色，尤其科學語言為了呈現科學上使用的方法以及科學理論，與日常生活上的語言而有所不同，Halliday (1993) 的研究指出，科學語言的一個基本組成就是科學術語，缺少了科學術語，不可能完整呈現系統性的科學知識。因此科學語言包括了特殊的科學詞彙、語意及語法，而科學語言的運用是影響學習科學知識的重要因素 (Fang,

2006; Love, 2009; Wang et al., 2011b; Yang & Chen, 2008)。但並非科學語言即科學術語，影響學習者在科學教育中所使用相關語言的運用尚有許多重要的因素，而並非單純只有科學術語的成份 (Osborne, 2010)，因此，了解國小學童在科學教育中科學語言的理解程度即是本研究的動機。

Yore (2011) 提出科學教育三個階段的發展，從第一代以課程文本的理解為主，至第二代情境導向，直到第三代綜合科學課程素養以情境內容來了解課程內容，並且需要以學習者的後設認知、彈性推理、心智習性、語言（包括聽、說、讀以及寫等成份）、ICT 等為基礎，來達成第一代以及第二代的目標。其中科學語言即是在學習者的科學教育學習中扮演著基礎成份的角色，因此了解學習者在科學語言的閱讀理解成份是一個重要

的且需要釐清的工作，而這也正是本研究的主要目的。

（三）電腦化測驗

因應資訊時代的來臨，教學評量與現代科技的結合，已是新世紀教育的主要趨勢，電腦測驗便在這樣的時代背景下應運而生。Bunderson、Inouvy 和 Olsen（1989）曾指出電腦測驗的發展，可分為四代，依序為第一代的電腦化測驗（Computerized Testing, CT）其編製的理論基礎為古典測驗理論（Classical Test Theory, CTT）；第二代電腦化測驗則為電腦化適性測驗（Computerized Adaptive Testing, CAT），其測驗編製的理論基礎則為試題反應理論（Item Response Theory, IRT）；第三代則為系統連續性測量（Continuous Measurement, CM），其編製的理論基礎為多向度試題反應理論（Multiple Dimension IRT, MIRT）；第四代則為智慧性評量（Intelligent Measurement, IM），其中主要是應用人工智慧理論，建立專家系統、學生個人學習模式。綜合 Bunderson 等人（1989）所提出電腦化測驗的發展可以歸納出幾個要點，1. 前兩代測驗通常是靜態、單一時間點及考生成就狀態的描述；後兩代測驗則更偏重個體隨著學習經驗動態改變的紀錄與解釋，2. 將以往專業、嚴謹而針對機構所需求的測驗，轉向個體或少數人的需求設計，3. 在作業形式上，力求實用而接近生活經驗。而目前由於資訊科技的蓬勃發展，電腦化測驗的發展特色主要包括下列 4 項，1. 大部分電腦化測驗的測驗題目還是以選擇題或者是非題為主；2. 若要考慮在高風險測驗中實行電腦化測驗，所花費關於

測驗安全性的措施是非常地昂貴的；3. 在電腦化測驗的作業中，要去辨識受試者的身份是具有難度的；4. 在目前雲端服務下所進行的 e-Testing 或者是 e-Learning 的學習環境中，電腦化測驗提供非常恰當的測驗機制以及環境。

因此，本研究主要的目的即是欲建立在網路環境中的國小學童科學素養閱讀理解的電腦化測驗，期待能在目前的數位化時代中扮演著資訊提供者的角色。

（四）線上測驗

在專業分工的時代，透過「分工」發揮「專業」的功能，將研究的成果放在網站上，可以提供高品質的試題供教師輔助教學用並且發揮最大的邊際效果。藉由網路的傳遞，可以達到「系統開放、資源共享」的效果，如果將試題內容或施測結果掛在網路上，可以彼此交換心得，他人對試題的意見或使用後的經驗，更可做為修正試題的參考。以一己之力供眾人使用，可以共同分享成果；以眾人之力做一己之事，可以事半功倍，資源網路化可以使評量工具更具經濟效益（陳新豐，2002a，2005）。目前電腦化測驗甚至於電腦化適性測驗無論在理論以及實務上的發展都已經成熟了（陳柏熹，2006；陳新豐，2010）而測驗立體化即是一個電腦化測驗的發展趨勢，亦即透過 HTML 所提供階層性的文件結構，及文件間的相互超連結，將測驗的前因後果，同時完整的呈現出來，供人自由檢索，並可透過跳躍式的連結，與其他相關的網址連結，這時測驗的面貌將不同於傳統的測驗手冊所能呈現的功能，所以若能建立線上測驗，至少可以達到以下的目標，即 1. 作答線

上化：可以使學生直接在網路上，透過標準化試題評量出自己學習的缺點及結果，可得立即的回饋。2. 結果自動化：透過國際網路動態網頁的設計技術，針對受試者填答的資料進行計分，並針對計分後的結果，提供對應的建議。3. 應用多元化：透過不同的編排原則，將試題加以重組，根據不同目的而形成不同的試題組合，使得在使用測驗時能有多樣的選擇。4. 資源網路化：透過網路可以使大量的教師或學生，不管所處的地理位置在那裡，都可以使用及應用本測驗，打破時空的限制（陳新豐，2002b，2007）。

（五）紙筆測驗與電腦化測驗的成就差異

許多認知方面的研究結果發現，紙筆測驗與電腦化測驗上的成就並沒有差異，陳新豐（2005）的研究發現在參數的估計上，難度在紙筆與線上是沒有明顯差異，鑑別度上則是線上版優於紙筆版，當人數大致相等時，傳統紙筆測驗與線上電腦化測驗的試題參數估計是一致的。Englert, Mando 和 Zhao（2004）探討了電腦化與紙筆測驗對低成就的兒童在描述的能力上的關係。根據研究發現，我們發現低成就的學生在紙筆測驗和電腦化上的表現時，低學習成就的學生在用筆來回答電腦螢幕的問題，比回答在紙上的問題的還來的好，而且根據數據上的顯示，電腦化作答比傳統的紙筆測驗更來的有效。由以上的研究結果可以發現，許多的研究在認知方面發現，紙筆測驗與電腦化測驗上的成就並沒有差異，但也有一些研究的結果呈現不一致的結果，而且是否科學

閱讀理解上，利用電腦施測以及紙筆施測所得到的結果是否會有不一致的結果，這是本研究主要的研究動機之一。

三、研究方法

本研究旨在建置線上科學語言閱讀理解測驗，並探討不同的試題表徵下國小學童科學語言閱讀理解的能力之表現為何？理論基礎部分依閱讀理解、科學語言、電腦化測驗、線上測驗及紙筆測驗與電腦化測驗的成就差異 5 個部分說明如下。

（一）研究對象

本研究的研究對象主要是高雄市以及屏東縣各 1 所學校，共有 8 個班級，其中男生 118、女生 105，四年級 97、六年級 126，合計 223 位國小學童。樣本抽取的方法是採用 3 階段的分層隨機抽樣，亦即以縣市屏東縣、高雄市為第 1 層，之後隨機抽取 1 所學校，隨機抽取學校之後再依年級分 4 年級隨機抽取 2 班、6 年級隨機抽取 2 班。

信度樣本中本研究選取 105 位國小學童為信度考驗之樣本，這 105 位受試者於接受第 1 次施測之後 3 週後再接受重測。

同時效度樣本中本研究選取 217 位國小學童做為同時效度考驗之樣本，施以「中文閱讀理解測驗」（林寶貴、錡寶香，1999），求其與本測驗之效標關聯效度之情形如何？以建立線上科學語言閱讀理解測驗效度資料。

(二) 研究流程

本研究實施的流程分別為 1. 選定標定母群（高雄市、屏東縣）、2. 抽取具代表性的樣本、3. 實施第 1 次線上中年級科學語言閱讀理解測驗、4. 第 1 次線上高年級科學語言閱讀理解測驗、5. 中文閱讀理解測驗、6. 間隔三週後實施第 2 次線上中年級科學語言閱讀理解測驗、7. 第 2 次高年級科學語言閱讀理解測驗、8. 間隔三週後實施紙筆版中年級科學語言閱

讀理解測驗以及高年級紙筆版的科學語言閱讀理解測驗。其中第 3、4、6 以及 7 步驟中，線上科學語言閱讀理解測驗無論是中年級或者是高年級的版本，4 年級與 6 年級的國小學童皆要參與，亦即 4 年級的學童需要參與線上科學語言閱讀理解測驗的中年級以及高年級的版本，6 年級的國小學童亦同，以利年級差異的比較分析，詳細研究流程圖如下圖 2 所示。

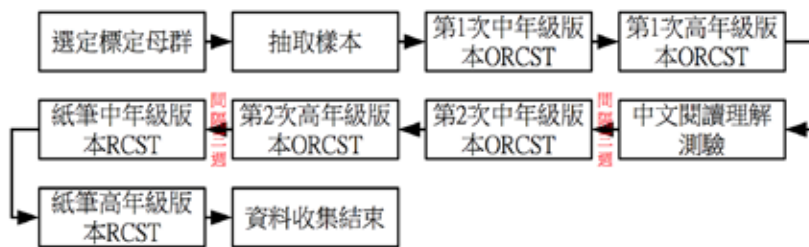


圖 2 研究流程圖

(三) 研究工具

本研究所採用的研究工具主要有 3，分別是科學素養閱讀能力評量工具、線上科學語言閱讀理解測驗以及中文閱讀理解測驗，分別說明如下。

1. 科學素養閱讀能力評量工具

科學素養閱讀能力評量工具（The Instrument of Reading Comprehension on Science Text, RCST）是由 Wang 等人（Wang et al., 2011a）所發展的科學閱讀理解問卷，做為診斷國小學童科學閱讀理解能力，其診斷項目包括能辨識科學語彙、能辨識主題、能辨識調查方法與證據、第 1 次高年級版本 ORCST 紙筆中年級版本 RCST 中文閱讀理解測驗紙筆高年級版本 RCST 能依據證據形成解釋與預測。測驗題目為四選一的選擇題型，包括辨識主題、分析訊息、訊息推論以及理解語彙等四個部分，RCST 包括中年級

以及高年級 2 種版本，題目的內容領域為海洋、科學史、地震、物理、流體力學、大氣科學等領域知識。內部一致性信度為高年級為 .80，中年級版本為 .74，重測信度為高年級版本為 .76，中年級版本為 .72。

2. 線上科學語言閱讀理解測驗

研究自行研發之線上科學語言閱讀理解測驗（Online RCSLT, ORCSLT），主要是將 Wang 等人（Wang, et al., 2011a）所研發的科學素養閱讀能力評量工具，建置於網際網路的施測環境中，並且即時將回答的資料送回至伺服器的資料庫，利用網際網路的特性，降低人工閱卷的勞力，可使閱卷的效率提高。就測驗環境而言，結合網際網路來進行的測驗，可具有超越時空、隨選隨測、高彈性的施測環境。再從網際網路的進步所帶來的遠距教學

為例，遠距教學的評量趨向於偏重學習經驗動態改變的記錄與解釋（李台玲，2001），再加上「動態評量」理念與意義，運用網路進行的電腦化測驗將逐漸重視「評量即學習」的理念，充份運用網路環境的優勢，讓被評量者在評量後可以立即深入思考學習上的缺失，獲得學習成效，焦點擴及學生在學習歷程中的學習與思考過程，而不只是測驗結果。

系統平台是採用 Apache 網頁伺服器，系統使用者資料庫是採用 MySQL，系統介面是採用 PHP、HTML 以及 Flash CS5.5 為主來加以設計，試題內容則與 RCST 一致，分為中年級以及高年級 2 個版本，中年級有 25 題，而高年級閱讀理解的題目有 26 題。

3. 中文閱讀理解測驗

林寶貴、錡寶香（1999）所編製，施測對象為國小二年級至六年級，施測的時間一般來說 45 至 65 分鐘，是採團體施測，主要的功能為評量國小二年級至六年級學的閱讀理解能力，測驗設計架構之相關能力包括：音韻處理能力（12）、語意能力（13）、語法能力（12）、了解文章基本事實（23）、比較分析（13）、抽取文章大意（12）、及推論（15），合計 100 題。重測信度 .89、庫李信度 .90、內部一致性信度 .96、折半信度 .95；效度部分，以「中華國語文能力測驗」為同時效度驗證之效標依據，同時效度全測驗相關是 .80，各閱讀理解次能力的相關介於 .69-.83 之間；各項次能力之相關介於 .79-.98 之間，而全測驗得分與各項次能力之相關介於 .85-.96 之間，其信度與效度方面可謂相當良好。

4. 資料分析

本研究分別針對研究目的進行量化

的資料分析，茲分別說明如下。針對國小學童在科學語言閱讀理解測驗得分表現，分別以描述性的平均數、變異數、偏態係數以及峰度係數等值加以說明，在信度考驗部分則以內部一致性係數、折半信度、重測信度、相關係數等，在年級以及性別的差異方面將運用 t 檢定來進行平均數差異的檢定，至於在線上科學語言閱讀理解測驗的特性將會利用試題反應理論的測驗特徵曲線加以說明。

四、研究結果與討論

本研究旨在探討國小學童在不同表徵下的科學語言閱讀理解測驗的能力表現為何？以下分為線上科學語言閱讀理解測驗系統、描述性資料內容、信度考驗、效度證據以及測驗特徵等 5 個部分說明如下。

（一）線上科學語言閱讀理解測驗系統

1. 系統架構

線上科學語言閱讀理解測驗系統是以 PHP+HTML 為架構，並建置於網際網路的環境中，測驗系統主要的界面則是以 Adobe Flash CS5.5 設計而成，受試者的填答資料庫則是以 MySQL 為主，系統架構圖如下圖 3 所示。

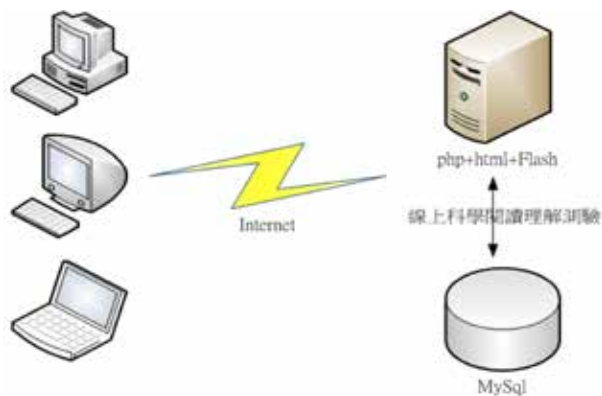


圖 3 線上科學語言閱讀理解測驗系統架構圖

2. 施測流程

受試者進入測驗系統後必須選擇基本資料以及輸入姓名之後，才能進入系統，而系統的基本資料包括縣市、學校、班級、座號、性別以及需要使用者輸入姓名。輸入基本資料的畫面如下圖 4 所示，

系統的界面為了考慮使用者易於登入且容易辨識，除了使用者姓名需要輸入之外，其餘皆以下拉式點選的方式進行，並且在縣市、學校方面，系統會自動根據所選縣市出現該縣市鄉鎮的學校供使用者選擇，減少使用者輸入錯誤的可能性。

圖 4 線上科學語言閱讀理解測驗基本資料輸入畫面

受試者輸入基本資料之後即可進入測驗的選擇畫面，並且在選擇測驗的界面會出現使用者登入的資料以供使用者確認，若有錯誤可以讓使用者回到前一頁重新輸入正確的使用者名稱，而本研究是以

閱讀理解測驗中年級以及高年級二個版本為研究的內涵，選擇測驗的畫面如下圖 5 所示。選擇所要進行的測驗內容即可進入施測畫面。

圖 5 線上科學語言閱讀理解測驗選擇畫面

3. 線上科學語言閱讀理解測驗內涵

線上科學語言閱讀理解測驗分為中年級以及高年級的版本，而測驗內容的設計是與國小自然與生活科技的教學文本不同，亦即測驗文本不會出現在中年級的自然與生活科技的閱讀文本之中，以下即以線上科學語言閱讀理解測驗系統之施測說明、題本內涵、答題設計以

及回饋設計等 4 部分說明如下。

(1) 施測說明

下圖 6 是線上科學語言閱讀理解測驗的施測說明，說明本測驗的題型、題數並出現例題說明後，讓使用者在進入正式作答之前對於測驗作答的方式有初步的認識與了解。



圖 6 線上科學語言閱讀理解測驗施測說明(中年級)

(2) 題本內涵

下圖 7 為線上科學語言閱讀理解測驗的題本內涵，版面設計主要是以一頁

出現的文本為原則，若是文本過長可以藉由右邊的捲軸加以上下拉動閱讀。



圖 7 線上科學語言閱讀理解測驗題本內涵(中年級)

(3) 答題設計

下圖 8、9 為線上科學語言閱讀理解測驗的答題設計，有 3 種方式分別是只看文章、回答問題以及只看問題，其中只看文章是答題時可以再次觀看文本後

再作答，而第 2 種方式只看問題（圖 8）是只有出現回題讓使用者作答，最後一種方式（圖 9）則是將畫面切割成 2 個部分，同時出現文本以及作答的試題說使用者回答。



圖 8 線上科學語言閱讀理解測驗答題設計（中年級）只看問題



圖 9 線上科學語言閱讀理解測驗答題設計（中年級）回答問題

(4) 回饋設計

下圖 10 為線上科學語言閱讀理解測驗的回饋設計，包括使用者基本資料、

答對題數、百分等級、作答時間以及作答後的建議結果等 5 個部分。



圖 10 線上科學語言閱讀理解測驗回饋設計（中年級）

（二）描述性資料

下表 1 為線上科學語言閱讀理解測驗作答之版本、作答人數、試題數、平

均數、變異數、偏態、峰度等描述性統計資料。

表 1 線上科學語言閱讀理解測驗描述性統計量數一覽表

| Ver. | N | 4th | 6th | Item | Mean | Variance | Skew | Kurtosis | Min | Max |
|------|-----|-----|-----|------|-------|----------|------|----------|-----|-----|
| 中 | 219 | 93 | 126 | 25 | 12.95 | 24.26 | .15 | -.65 | 3 | 25 |
| 中 | 93 | 93 | 0 | 25 | 12.97 | 17.97 | .20 | -.74 | 5 | 23 |
| 中 | 126 | 0 | 126 | 25 | 12.94 | 28.73 | .13 | -.76 | 3 | 25 |
| 高 | 223 | 97 | 126 | 26 | 10.77 | 20.07 | .49 | -.18 | 1 | 25 |
| 高 | 97 | 97 | 0 | 26 | 8.79 | 10.12 | .36 | -.24 | 1 | 18 |
| 高 | 126 | 0 | 126 | 26 | 11.98 | 22.33 | .21 | -.54 | 2 | 25 |

由上表 1 資料中可以得知、中年級版本有 25 題、高年級版本有 26 與峰度 (-.76-.24) 係數的資料，並未出現極端的資料，並沒有受試者全部對錯，而中年級版本高年級學生有全部答對，其餘在中年級版本中的中年級以及高年級版本中的中年級以及高年級的受試者並沒有全部答對的情形。

（三）信度考驗

線上科學語言閱讀理解測驗的信度考驗主要分為內部一致性係數 (Alpha) 以及再測信度等 2 種，說明如下。

1. 內部一致性係數

中年級版本部分，93 位四年級、131 位六年級，合計 224 位受試者，其內部

一致性係數為 .79，其中平均答對率 .52，平均試題與總分之相關為 .41，平均的二系列相關為 .52。若單以四年級的 93 位受試者，其內部一致性係數為 .70，其中平均答對率 .52，平均試題與總分之率 .52，平均試題與總分之相關為 .35，平均的二系列相關為 .45。若單以六年級的 131 位受試者，其內部一致性係數為 .83，其中平均答對率 .52，平均試題與總分之相關為 .44，平均的二系列相關為 .56。

高年級版本部分，97 位四年級、158 位高年級，合計 255 位受試者，其內部一致性係數為 .74，其中平均答對率為 .41，平均試題與總分之相關為 .36，平均的二系列相關為 .47。若單以四年級的 97 位受試者，其內部一致性係數為 .49，其中平均答對率 .34，平均試題與總分之相關為 .27，平均的二系列相關為 .36。若單以六年級的 158 位受試者，其內部一致

性係數為 .76，其中平均答對率 .46，平均試題與總分之相關為 .38，平均的二系列相關為 .49。

2. 再測信度

重測信度的計算是以第 1 次施測後大約間隔 3 週後重測的結果，在中年級的線上科學閱讀理解部分的重測信度為 .49 ($p<.01$)，而高年級的線上科學閱讀理解部分的重測信度則為 .64 ($p<.01$)，達中等程度並且皆達統計上的顯著水準，亦即無論是中年級或者是高年級的線上科學閱讀理解測驗皆達一定程度的穩定性。

(四) 效度考驗

線上科學閱讀理解測驗的效度資料方面，主要以與中文理解測驗相關、科學閱讀理解測驗紙筆版、年級差異以及性別差異等 4 個方面說明如下表 2。

表 2 線上科學閱讀理解測驗之效標關連效度資料一覽表

| 線上科學閱讀理解測驗 | 中文閱讀理解測驗 | 科學閱讀理解測驗紙筆版 |
|------------|----------|-------------|
| 中年級版本 | .64** | .65** |
| 高年級版本 | .65** | .76** |

** $p<.01$

以下進行效標關聯效度，其中的效標分別有 2，其一為中文閱讀理解測驗（林寶貴 & 錡寶香，1999），另一則為紙筆版本的科學語言閱讀理解測驗（Wang, et al., 2011a）。

1. 中文閱讀理解測驗相關

線上科學閱讀理解測驗中年級版本與中文閱讀理解測驗的相關達 .64 ($p<.01$)，高年級版本與中文閱讀理解測驗的相關達 .65 ($p<.01$)（如上表 2 所示），亦即線上科學閱讀理解測驗與中

文閱讀理解測驗達到中上程度的相關，並且達統計上的顯著水準，表示線上科學閱讀理解測驗具有測量中文閱讀理解能力的效能。

2. 科學閱讀理解測驗紙筆版本之相關

線上科學閱讀理解測驗中年級版本與科學閱讀理解測驗紙筆版中年級版本之相關達 .65 ($p<.01$)，高年級版本與閱讀紙本高年級版本之間相關達 .76 ($p<.01$)（如上表 2 所示），亦即科學閱讀理解測驗線上版與紙筆版之間達中

上程度的相關，且達統計上的顯著水準，表示科學閱讀理解測驗線上版與紙筆版有其一致性。

進一步探討線上科學閱讀理解測驗各面向國小學童表現（如下表3），由下表3可知國小學童在線上科學閱讀理解測驗各面向表現能力，在中年級版本部分以科學語彙的表現較佳，其次為能形成解釋、能蒐集資料以及能辨識主題的能力

表現，與 Wang 等人（2011a）利用紙筆科學閱讀理解測驗的版本其表現次序完全一次，只是國小學童在線上科學閱讀理解能力上的表現還是比紙筆版本略差。至於在高年級的版本部分，則以科學語彙表現最佳，其次為能蒐集證據、能形成解釋以及能辨識主題，Wang 等人（Wang, et al., 2011a）紙筆版本相較國小學童在科學語言閱讀理解各向度上的表現一致。

表 3 線上科學閱讀理解能力測驗各面向之平均數、標準差、範圍等描述統計量數

| | MORCST | | | HORCST | | |
|-------|--------|------|-----------|--------|------|-----------|
| | Mean | SD | Range | Mean | SD | Range |
| 科學語彙 | 0.65 | 0.22 | 0.24-0.99 | 0.47 | 0.15 | 0.22-0.97 |
| 能辨識主題 | 0.48 | 0.17 | 0.23-0.99 | 0.41 | 0.17 | 0.22-0.98 |
| 能蒐集證據 | 0.50 | 0.19 | 0.25-0.99 | 0.44 | 0.17 | 0.23-0.96 |
| 能形成解釋 | 0.51 | 0.19 | 0.25-0.99 | 0.43 | 0.16 | 0.23-0.97 |
| 總測驗 | 0.53 | 0.19 | 0.24-0.99 | 0.43 | 0.16 | 0.23-0.97 |

3. 年級向度之比較分析

進行年級間差異分析時發現，線上科學閱讀理解測驗中年級版本中，高年級與中年級的表現並無差異（ $t=.05$ ， $df=207$ ， $p=.96$ ），由其年級平均數（4th=13.13，6th=13.16）可以得知高年級雖略高於中年級，但是在統計上並無差異，亦即高年級雖然具有經驗上的優勢，但面對中年級的學習內容並未表現出年級上的優勢，而只是與中年級的表現一致而已，也突顯出科學閱讀理解其科學概念有其特殊性。

至於線上科學閱讀理解測驗中高年級版本中，高年級的表現與中年級的表現達顯著性的差異（ $t=6.38$ ， $df=207$ ， $p<.001$ ），由其年級的平均數（4th=8.91，6th=12.59）中可以得知，高年級優於中

年級在科學閱讀理解能力上的表現。

綜上年級的差異比較得知，科學語言閱讀理解中的科學術語有其特殊性，由高年級的版本中高年級的表現優於中年級，亦即科學語言若未曾接觸過、未學習過即不容易理解，反之，以中年級版本高年級與中年級的版本而言，科學語言若已學習，則不會因為年級的差異而有所差異。

4. 性別向度之比較分析

進行性別間的差異分析，無論是線上科學閱讀理解中年級版本（ $t=3.18$ ， $df=207$ ， $p=.002$ ）或者是高年級版本（ $t=3.87$ ， $df=207$ ， $p<.001$ ），其性別差異皆達統計上的顯著水準。由其平均數上可以得知皆為女學生優於男學生（中年級，女=14.27>男=12.16，高年級，女

=12.27> 男 =9.93)。

(五) 測驗特徵

下圖 11 為線上科學閱讀理解測驗中年級版本以及高年級版本 3PL 試題反應理論的測驗訊息圖，由圖中可以得知中年級版本的線上科學閱讀理解測驗對於中上能力者提供較佳的訊息，亦即較適合中上程度的中年級學生，高年級版本線上科學閱讀理解測驗對於中上能力者提供較佳的訊息，亦即較適合中上程度的高年級學生。比較中年級與高年級版本的線上科學語言閱讀理解測驗訊息中可以發現，線上科學閱讀理解測驗高年級版本比中年級版本較難，亦即高年級版本比中年級版提供能力高者較高的訊息，亦即具有年級上的差異。

由下圖 11 可以發現高年級版本的線上測驗 (HORCST) 比高年級版本的紙筆測驗 (HPRCST) 對於能力高者提供較豐富的訊息，亦即國小 6 年級學童的反應結果呈現線上版本較紙筆版本困難；另外，中年級版本的線上測驗 (MORCST) 比紙筆測驗 (MPRCST) 對於能力較高者提供較豐富的訊息，亦即國小 4 年級學童的反應結果呈現線上版本較紙筆版本困難；綜上所述本研究的研究發現為國小學童無論是 4 年級或者是 6 年級皆呈現出線上版本較紙筆版本困難。

至於紙筆高年級科學語言閱讀理解測驗較紙筆中年級版本困難，與線上高年級版本與中年級版本的結論相同，具有年級上的差異。

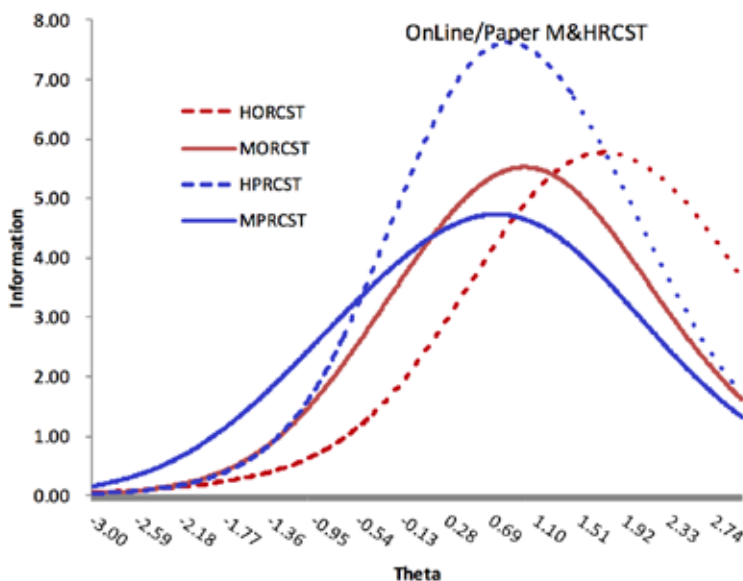


圖 11 科學語言閱讀理解線上與紙筆測驗訊息函數比較圖

綜上所述，線上科學閱讀理解測驗在效度方面具有相當不錯的效度，從與中文閱讀理解測驗、紙筆版本的科學閱

讀理解測驗、年級差異以及性別差異中具有一定的效能。

五、結論與建議

語言在科學中是一個重要的傳達的媒介，科學發明中部分的重要發明是思維的深入語延展、要表徵研究或發現的思維，此時為傳達這些重要的發現或其中的思維就必須介入語言的表徵。Carnap (2002) 指出 " 知識的傳達只有用符號表徵出來，或是用語詞或其他符號的命題，才會存在 "，科學的知識體系藉由科學語言的穿針引線，有條不紊而一目了然，科學語言就好比是一張大綱，科學的基本原理就像是大綱中的繩索，網中的結點猶如科學的基本概念，縱橫交錯的細繩形成各種定律與命題，因此科學語言在科學知識體系中學習具有其重要性。科學語言在國小學童學習科學的歷程之中，扮演著相當重要的角色，誠如陳世文與楊文金 (2006) 所提出科學文本的術語經常具有某種特定的科學意涵或其科學上所特定的定義，如果學習者缺少了這些科學語言，極有可能無法建構系統性的科學概念，而造成國小學童的學習上形成片面或零碎的科學知識，因此科學語言的閱讀理解即是國小學童科學學習上的影響因素。

本研究探討不同試題表徵下 (線上版本與紙筆版本) 國小學童科學語言閱讀理解能力表現中可以發現，本研究建置網際網路上的線上科學語言閱讀理解測驗工具，可以線上收集國小學童在科學文本閱讀後的反應資料，具有作答說明、作答時具有重讀科學文本或只有出現題目的作答方式、作答完畢後有即時的回饋，並且測驗系統利用 Flash 配合 HTML 來製作，畫面的設計也偏向國小學童的學習式樣，作答結束之後即將作答結果完整地傳回系統伺服器，避免造成日後若紙筆作答

後的輸入可能人工疏漏之處，種種的優勢讓本研究在資料的收集上特別地方便、即時以及迅速，對於資料的收集提供不少優勢，誠如陳新豐 (2002a) 所提之資源網路化可以使評量工具更具經濟效益，而且本研究所建置之電腦化測驗為線上測驗，也達成了作答線上化、結果自動化、應用多元化以及資源網路化的線上測驗的目標 (Hricko & Howell, 2006; 陳柏熹, 2006; 陳新豐, 2002b)。

線上科學語言閱讀理解評量工具在中年級的版本之信度為 .79，高年級版本之信度為 .74，其內容具有穩定的程度 (Hopkins, Stanley, & Hopkins, 1990; Popham, 1999)，表示線上科學語言閱讀評量工具的內容與紙筆版本的科學語言理解測評量工具同時具有內容的一致性程度 (Wang et al., 2011a)。至於在重測信度方面雖然中年級版本的重測信度為 .49，高年級版本的重測信度為 .64，只達中等程度，但仍達統計上的顯著水準，亦即時間上達一定程度的穩定性。在效度部分，線上版本的科學閱讀理解評量工具與紙筆版、中文閱讀理解測驗的相關達中上程度的相關，表示線上科學語言閱讀理解測驗工具與紙筆版本、中文閱讀理解測驗有其一致性的程度，至於國小學童在科學語言閱讀理解各面向的能力表現上，中年級與高年級的版本相同一致，綜合言之線上科學語言閱讀理解測驗與紙筆版本內容的正確性大致相同。

線上科學語言閱讀理解測驗在年級差異部分，中年級版本並無差異而高年級版本達顯著性的差異，代表科學語言有其特殊性，未學習的內容一定會比學習過的表現還要差 (高年級版本，高年級

優於中年級），至於學習的內容則仍然具有一定程度的保留（中年級版本，中年級的表現與高年級的表現並無差異）。在性別上的分析結果發現，無論是中年級或者是高年級的版本，女學生皆優於男學生（邱美菁，2011；謝地利，2007），亦即表示科學文本的閱讀仍然是以女學生的表現較佳。

最後進行線上科學語言閱讀理解測驗的試題反應理論參數估計，並繪製測驗訊息函數比較發現無論是紙筆版本或者是線上版本高年級版本比中年級版本提供能力高者較高的訊息，亦即本測驗在年級上具有其區辨性，並且呈現出線上版本較紙筆版本困難的傾向，與陳新豐（2005）所進行的研究結果大致相符。

綜上所述，本研究成功地建置線上科

學語言閱讀理解測驗，設計不同的作答方式，除了提供作答說明、線上科學文本以及即時的作答回饋。並比較紙筆版本的科學語言閱讀理解測驗工具與線上版的特徵，建立線上版本的內部一致性信度、重測信度、二系列相關、效標關聯效度、年級差異以及性別差異分析等測驗訊息，最後以試題反應理論的測驗訊息曲線來說明比較線上科學語言閱讀理解測驗工具的特徵，提供日後進行科學語言、閱讀理解相關研究的參考。

致謝

本論文為國科會編號 NSC100-2511-S-153-007- 之計畫，由於國科會的支持，使本計畫得以順利進行，特此致上感謝之意。

參考文獻

- 林寶貴、錡寶香（1999）。中文閱讀理解測驗指導手冊。台北市：國立台灣師範大學特殊教育中心。
- 邱美菁（2011）。臺中市國小學童閱讀行為、閱讀理解能力對國語和數學學業成就之影響。東海大學碩士論文，未出版，台中市。
- 陳世文、楊文金（2006）。以系統功能語言學探討學生對不同科學文本的閱讀理解。師大學報：科學教育類，51（1，2），107-124。
- 陳柏熹（2006）。國家考試電腦化測驗相關問題探討。國家菁英，2（2），125-138。
- 陳新豐（2002a）。現代測驗的新趨勢 -- 電腦化適性測驗。菁莪季刊，14（3），16-28。
- 陳新豐（2002b）。線上題庫與適性測驗整合系統之發展研究。國立政治大學博士論文，未出版，台北市。
- 陳新豐（2005）。傳統紙筆測驗與線上電腦化測驗試題參數估計差異之比較。教育研究與發展期刊，1（3），123-145。
- 陳新豐（2007）。臺灣學位電腦化測驗研究的回顧與展望。教育研究與發展期刊，3（4），217-248。
- 陳新豐（2010，5月）。電腦化適性測驗及其建置流程。論文發表於國家考試 e 化策略研討會，台北市。

- 陳新豐、方金雅、王靜如、林曉雯（2012，6月）。不同試題表徵科學語言閱讀理解測驗之發展。論文發表於第一屆「閱讀評量與教學」理論與實務研討會，高雄市。
- 謝地利（2007）。國小學童課外閱讀情況與科學文章閱讀理解及科學創造力之調查研究。國立屏東教育大學碩士論文，未出版，屏東市。
- Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 367-407). New York: Macmillan.
- Carnap, R. (2002). *The Logical Syntax of Language* (A. Smeaton, Trans.). Oprn Court.
- Englert, C.S., Mando, M., Zhao, Y.(2004). *I Can Do It Better on the Computer: The Effects of Technology-enabled Scaffolding on Young Writers' Composition*. *Journal of Special Education Technology*, 19(1), 5-21.
- Fang, Z. (2006). *The Language Demand of Science Reading In Middle School*. *International Journal of Science Education*, 28(5), 491-520.
- FCRR. (2011). *2011-2012 PMRN User's Guides*. Tallahassee, FL: Florida Center for Reading Research, FCRR.
- Halliday, M. A. K. (1993). The Analysis of Scientific Texts in English and Chinese. In M. A. K. Halliday & J. R. Martin (Eds.), *Writing Science: Literacy and Discursive Power* (pp. 304). Pittsburgh: University of Pittsburgh Press.
- Hoover, W. A., & Gough, P. B. (1990). *The Simple View of Reading*. *reading and writing: An Interdisciplinary Journal*, 2, 127-160.
- Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). *Educational and Psychological Measurement and Evaluation* (7th ed.): Prentice Hall.
- Hricko, M., & Howell, S. L. (2006). *Online assessment and measurement : foundations and challenges*. Hershey, PA: Information Science Pub.
- Lin, S.-W., Wang, J.-R., Kao, H.-L., & Chen, S.-F. (2009). *Development of an instrument for measuring elementary students' listening comprehension on science*. Paper presented at the International Conference of East-Asian Science Education, Nation Taipei University of Education, Taipei City, Taiwan.
- Love, N. (2009). Science, language and linguistic culture. *Language & Communication*, 29(1), 26.
- Osborne, J. (2010). *Arguing to Learn in Science: The Role of Collaborative, Critical Discourse*. *Science*, 328, 463-466.
- Popham, W. J. (1999). *Modern Educational Measurement: Practical Guidelines for Educational Leaders* (3 ed.). Allyn & Bacon.
- Schellings, G., Aarnoutse, C., & Leeuwe, J. v. (2006). Third-grader's think-aloud protocols: Types of reading activities in reading an expository text. *Learning and Instruction*, 16(6), 549-568.
- Wang, J.-R., Chen, S.-F., Tsay, R.-F., Chou, C.-T., Lin, S.-W., & Kao, H.-L. (2011a). *Developing a Test for Assessing Elementary Students' Comprehension of Science Texts*. *International Journal of Science and Mathematics Education*, 1-19. doi: 10.1007/s10763-011-9307-y
- Wang, J.-R., Chen, S.-F., Tsay, R.-F., Chou, C.-T., Lin, S.-W., & Kao, H.-L. (2011b). *Development of an Instrument for Assessing Elementary School Students' Written Expression in Science*. *The Asia-Pacific Education Researcher*, 20(2), 276-290.
- Wang, J.-R., Lin, S.-W., Kao, H.-L., & Chen, S.-F. (2009). *Development of an instrument for assessing elementary students' scientific literacy on writing task*. Paper presented at the International Conference of East-Asian Science Education, Nation Taipei University of Education, Taipei City, Taiwan.
- Weiss, D. J. (2010). Operational CAT Testing Programs Retrieved 07/18, 2010, from <http://www.psych.umn.edu/psylabs/catcentral/operationalcatprograms.htm>

- Yang, W.-G., & Chen, S.-W. (2008). A Comparison of the Discourses of Science Texts in English and Mandarin on Newton's First Law of Motion. *Journal of National Taiwan Normal University: Mathematics & Science Education*, 53(1), 113-137.
- Yore, L. D. (2011). *Promoting scientific literacy from viewpoints of policy, curriculum reform, and evidence-based practice*. Paper presented at the International Conference of Promoting Scientific Literacy from Viewpoints of policy, Curriculum and Practice, Taipei, Taiwan.

如何運動才能改善大腦？數位穿戴科技應用於教育的可能性探討

吳忠霖

新北市國教健體輔導團輔導團員、新北市武林國小教學組長

陳宇虹

新北市武林國小教師

一、源起

近來世界各國掀起一股「教改熱」，但他們不是從課業上著手，而是「返樸歸真」地增加學生的身體運動量；身體活動量一直是全球重視的議題，這個趨勢之所以可以歷久不衰，乃是由於國際間研究不斷推陳證實，運動不單單可以促進健康、提升智力，更可以達到於防高血壓或是老人痴呆等問題（Andersen, Harro, & Sardinha, 2006），若是盡早提倡於孩童之間，更可以降低其將來心血管危險因子、增加骨量與骨密度，甚至是促進精神健康等等的優點（Wang, Suominen, & Nicholson, 2005）。美國科學研究更證實，規律的身體活動可以促進腦部的運作效率，若透過運動與學習的理解過程訓練，更進而能夠提升學習效率等（Chaddock, Hillman, Buck, & Cohen, 2011；Donnelly, & Lambourne, 2011）。運動對於孩童的發展相當重要，原因在於腦部發育，其腦部不論是在結構或是認知，功能等方面成長都相當快速，而運動的過程當中，腦神經不斷受到的刺激進而增加了突觸修剪（synaptic pruning），達到強化整體神經傳導上的效率（Casey, Tottenham, Liston, & Durston, 2005），然而，在我國的調查資料卻顯示，在腦部發育重要的學齡時期，國內學童有將近六成的比例是不熱愛運動，甚至未養成規律運動的習慣，

與國際相比較下，台灣孩童無論是在身體質量指數或心肺耐力等的表現，均落後鄰近地區的中國大陸、日本以及韓國等地學生（彭臺臨，2007；崑山科技大學，九十五學年度學生體適能檢測報告）。

台灣近年來對於孩童體育的養成及規劃，不僅建設完整的體適能檢測、促進學生身體活動的快活計畫，甚至為了宣倡體適能運動，也計畫將體能表現列入升學評分中的重要標準，如此用心的經營，無非是希望孩童能夠自主的養成運動習慣，並獲得更好的全人發展，近年，國家發展與科技進步，以「記錄」協助檢測的電子工具也日趨成熟，除了可以利用「累積」與「分析」等正面回饋方式鼓勵運動習慣養成之外，更可以映照現有孩童運動規劃的成果，進而找出台灣學齡孩童的運動發展弱勢，達到自我重視及永續成長的運動觀念養成。

二、現況與問題評析

由於課業壓力以及科技的發達，現代的孩童大多處於坐式生活型態，因此，第二型糖尿病、高血脂以及心血管等疾病的患病年齡更是逐年下降，除此之外，許多研究也已證實，活動量較低的孩童，不論在學業表現、智力表現、認知功能、大腦結構與功能也會相對地較差（王駿

濠、張哲千、梁衍明、邱文聲、洪蘭、曾志朗與阮啟弘，2012），鑒於這些發現，國際間各國政府也紛紛訂立與修改健康政策的建議指數，例如：歐盟各成員國的兒童與學生，每週至少要有 180 分鐘或以上身體活動的可能性；美國在 2000 年提出 2010 健康國民的新政策（Healthy People 2010），建議兒童及青少年每天至少有 30 分鐘的中等費力身體活動，每週累積身體活動 150 分鐘，2006 年則宣佈他們的兒童青少年的身體活動量要改為每週累積 420 分鐘等，台灣也以心肺耐力（800/1600 公尺跑走）、肌力（立定跳遠）、肌耐力（一分鐘屈膝仰臥起坐）、柔軟度（坐姿體前彎）以及身體組成（BMI）等測量，配合體適能 333 計畫（每周 3 次，每次最少 30 分鐘，每分鐘心跳 130 以上）來進行運動推廣與量測，近來更構思在十二年國教加入每週至少 150 分鐘身體活動量和體育課由 2 節增加為 3 節。

綜觀國際及國內的體適能計畫，多以提出運動時間的長短為主要建議方式，對於運動強度，多以心跳程度作為費力指標，對於孩童而言，這類訊息的理解程度不僅無法直覺聯想，對於運動習慣的養成，更是需要成年人的悉心維護才較有可能長期維持，除此之外，對於體適能計畫的實施結果的探討，也多採取主觀的自我報告法，例如：三日身體活動紀錄量表、七日身體活動回憶紀錄法等，雖然其優點是使用方便、快速填答、研究成本低以及了解身體活動種類等，然而實際落實於「運動習慣」的養成上，不但缺乏趣味性，更是沒有成就感，甚至許多學童在回憶上有困難，因此更添困擾，也由於這類型運動數據的紀錄及

常模分析不易，所以也容易造成政策雖然嚴謹安排，但是基層實施與評鑑卻仍勞師動眾的非高效率行為。

三、實施理念

有「網路女王」之稱的瑪麗·米克（Mary Meeker），今年發表趨勢報告時，特別宣布：在經過智慧型手機與平板電腦的兩大循環週期後，個人電腦產業正式邁入第 3 波循環——穿戴式行動裝置時代。市場討論沸騰，但頭上掛個探照燈式的眼鏡、腕上戴著像中學時代擁有的電子錶，穿戴式裝置可能像筆電、手機、平板電腦一樣，成為下一個主流裝置？當前的穿戴裝置平台，就屬頭戴式裝置和智慧錶最為熱門。前者以 Google 推出的眼鏡（Glass）為主，但仍停留在測試版，還沒有太多相關軟體應用消息。至於智慧錶，市場上已有三星 Galaxy Gear、高通 Toq、以及靠群眾募資起家的 Pebble、並開始出現相關 App 服務，並鎖定管理個人身體適能、健康紀錄的 App，看起來最能與貼身的穿戴裝置結合。

一般的電子計步器僅限於步數的計算以及推導的預估能量消耗，無法有效辨識出活動強度（Troost, Wong, Pfeiffer, & Zheng, 2012）；而研究室等級的身體活動量計，多由計算時間內身體活動產生的加速度積分數據（counts），來判定除了步數之外的活動強度程度，然而，其單只造價從萬元到數萬元不等，因此，無法有效推廣至學校或是大量樣本之採集，有鑑於此，新北市國教健體輔導團預計結合台灣師範大學研發設計之隨時記錄日常生活運動強度之量測器（Kang, Choi, Lee, & Tack, 2012），致力推展全市運動 3Q「零時體育」計畫，希望讓孩童

可以直接認知玩樂同時也正是身體活動的發生。此具有判定活動強度之身體活動量計，藉由計算使用者步伐衝擊地面產生的加速度數值，定義出使用者目前活動屬於一般步行、輕微慢跑或是高速衝刺等不同活動強度之區別。除此之外，並以活動有趣手錶式介面，讓小朋友隨時能夠自主地觀察到每天的任何時段身體活動強度、卡路里消耗、步數計算等等，藉由傳輸、分享以及競賽等的方式，計畫在網路雲端與全台各地的學生進行互動，更可以配合生動的台灣地圖互動介面軟體，寓教於樂，不僅讓孩童養成了運動習慣以及觀察紀錄的專注力，對父母來說更獲得了指標性的生活型態資訊，對政府教育以及健康單位來說，更是獲得了孩童健康生活習性的基礎分布資料用以佐證行政規劃的落實，可說是一舉三得且持續長遠的健康計畫。

獨立架構資訊處理平台，在孩童上傳運動資訊的同時，亦可即時更新日常生活型態的常模數據，且藉由帳號權限等的設計，讓學校單位得以獲得該校學童的體適能分析情況，而縣市政府也可以電子取代人工的方式，隨時獲得該區域學齡兒童的體育適應能力數據，用以規劃、安排與檢核各計畫實施的成果。除此之外，若更進一步能長期追蹤學齡孩童學業成績表現，並且以統計分析的方式找出學童身體活動量與學業成績之相關性，如此一來，以穿戴科技的輔助，促成運動改善大腦的實施方針，不僅開啟體育運動數位學習與評量另一扇窗，也可能成為未來體適能教育的全新領域以及方向。

四、計畫目標

學習過程中的增強與立即、多元的回饋將促發腦內多巴胺、血清素與正腎上腺素的分泌，可以鞏固學習的聯結。亦即身體活動量錶運用於教學中，教師可以積極的回應學生的需求與問題，在正向的學習氣氛中，透過師生間、同儕間與自我的回饋方式，協助學生澄清概念。分子生物學家麥迪納博士在《大腦當家》一書中大腦守則一即說明運動增強腦力，運動刺激大腦一個最重要的生長因素大腦神經生長因子（Brain Derived Neurotrophic Factor, BDNF），它幫助大腦中某些神經元的生長，這種蛋白質使已經存在的細胞保持年輕和健康，使他們更願意和別的神經元相連接，對BDNF最敏感的細胞是海馬迴的細胞，運動會增加海馬迴細胞內BDNF的濃度。而運動可以改善大腦的證據，已在科學研究中，以認知、心理等面向獲得證實（張育愷、林珈余，2010），然而，實際運用在教育層面上，綜觀教育界可以告訴孩童，究竟要多少時間或是強度的運動才能達到這樣效果的數據和器材卻仍是寥寥可數（古博文、陳俐蓉，2013）。因此，為了有效將「理論」轉譯成為「價值」，本計畫主要在於研究兒童體適能的現況，尤重於日常生活型態部分的紀錄與分析，從海量身體活動數據採集到使用者能力分析，一併建立起運動與體能、智能發展的相關性常模，除了可以有效建議孩童體能進步的標準外，再藉此研究結果設計體適能相關教案與推廣活動，以達到「有效運動、改善大腦與體適能」的貢獻，更能同步促進國家孩童的大腦發展水平，

茲將目標條列如下：

（一）常模建立與環境因子分析

基礎研究的設計，在於探索各微小環節的相關與差異，本研究雖以養成學齡孩童運動習慣為出發，但更希望藉由基礎的研究，建立台灣地區使用之學齡孩童身體活動量常模，此常模以記錄生活型態，例如上下學時間的身體活動情形、體育課的身體活動情形，甚至是課中或是課間休息的身體活動情形等等，運用此常模便可以延伸了解年齡差異、城鄉差距、性別差異等細項，找尋環境因子對於孩童身體活動量上造成的影響，如此，用以提供教育單位或是健康單位更準確的政策執行參考依據，除此之外，也進一步建立身體活動量與環境因子、身體活動量與學習效率之關係。透過本研究提案，也將可建立與釐清以下項目：

1. 建立台灣地區各學齡孩童之生活型態式身體活動量常模：經由本研究使用身體活動量計之紀錄，將可了解孩童於課堂、課間時間甚至是體育課時間之身體活動情形，根據此海量數據分析，便可以獲得更勝以往「柔軟度」、「心肺能力」等專項式型態的體適能常模，而得到「生活型態式」的體適能常模。
2. 釐清城鄉環境差異是否促成身體活動量的不同？差異程度有多少？造成差異的時間點與情形各是如何？此項目的釐清，乃是有鑑於已有研究（張樹立，2004）顯示新北市城鄉國小學童身體活動量與健康體適能之差異及其相關性，其結果顯示鄉村男女學童的身體活動量和健康體適能均高於城市男女學童。筆者相信，更多類似的結果或許也在各縣市發生，因此，如何有效地解釋造

成差異的關鍵轉折，勢必要從差異的時間點及程度進行第一步客觀的了解，方能抽絲剝繭的推敲出其中原因。

3. 釐清升學壓力與身體活動量之間的關連性？影響程度有多少？造成差異的時間點與情形各是如何？
4. 性別在身體活動量上的差異是否有所不同？造成差異的可能關鍵是什麼？造成差異的時間點與情形各是如何？
5. 年齡在身體活動量上的改變是否有可循的趨勢？
6. 不同身體質量指數（BMI）族群的身體活動量差異？是否可指出關鍵的身體活動量警示指標？

（二）常模建立與環境因子分析

1. 體育課程活動究竟需要多少的活動量才算足夠？本提案預計提供目前常見之孩童伸展操、跳繩、籃球以及其他項目運動之活動參考值，藉以讓授課老師或是學校單位了解其課程內容，並分析後依需求調整課程之安排。
2. 本提案所使用之資訊處理平台具有分析「每一分鐘」活動情形之能力，藉此科技發展，將可以分析孩童的上課期間、下課期間、放學期間甚至是回家後的身體活動情形，在了解容易發生活動量不足的時間點或是年齡後，便可以提供學校與教育單位課程安排以及回饋策略的布置。

五、具體作法

筆者以運動感測器結合雲端數位學習與評量教學策略，實際運用於任教學校健體領域教學及高關懷學生的補救教學，選擇以取得容易，時間、空間彈性大，變化多的「跳繩」為教學的主軸，設計

發想與跳躍相關學習課程，其架構如圖 1 所示。藉由手錶式或腰掛式運動感測器，讓小朋友隨時能夠自主地觀察到每一次下課時間、每一堂體育課程、甚至是每天的任何時段身體活動強度，並且透由傳輸、分享以及競賽等方式，在網路雲端與其他學生進行互動，更可以配合生動的台灣地圖互動介面軟體，寓教於樂。

課程實施第一節課以新北市國教健體輔導團自編運動改造大腦簡報及影片，結合「幸福聯絡簿」探討運動與大腦的節目，引發學生內在深層體認並啟發學生尋求自我的聰明力，輔以「運動感測器」所量測之資料，可以立即的換算成步數顯示於台灣地圖上，學童可以獲得多元且立即的回饋，包含教師、家長及同儕鼓勵。

基於新鮮的刺激是抓住學童注意力最好的工具，教學團隊在第二至四節課分別以水管軟管、彈性繩索、跳繩等多元的教學媒材，結合課程學習單，設計發想跳躍的遊戲與教學策略，提供學生

反覆練習的機會。並於每節課結束前 5 分鐘，以今日所學、所見的動作，分組創意編排二個八拍舞蹈，成果影片隨後上傳至 youtube 及學校網站 (<http://youtu.be/8McVhUY86Nc>)。第五節課則分組呈現舞蹈演出，強調人際互動與團隊合作的重要性，提高學生解決問題的能力，培養創新精神，增強自信，形成樂觀開朗的生活態度。

課程實施後，藉由雲端數位學習與評量，提昇了孩子運動的意願，配合「零時體育」課程實施，教學現場教師發現孩子不僅在學習的專注力及記憶力都有顯著的進步，學習情緒和意願也都提昇了，而弱勢孩子經過一連串的數位學習，不僅臉上多了更多笑容，學習意願明顯提昇，也對自己更有了自信心（圖 2）。團隊教師相信藉由雲端數位學習與評量，大大提高了孩子學習興趣，更讓孩童保持持續性運動習慣和熱情。

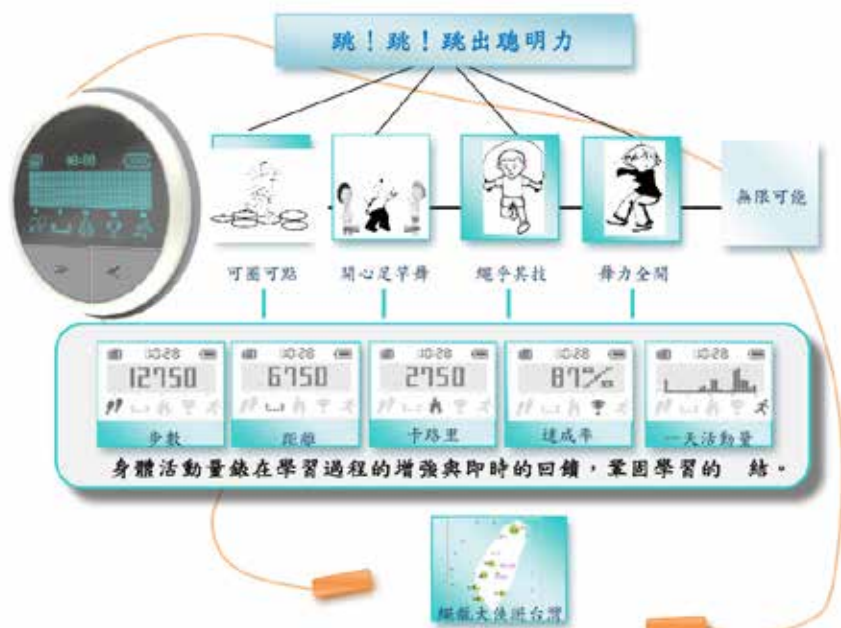


圖 1 身體活動量錶與健體領域教學架構圖



圖 2 聰明學習靠運動

六、未來展望

（一）自我量化與體適能穿戴科技宣導 本階段目的在於宣導自我量化以

體適能穿戴科技的功用與使用方式，讓孩童、家長、師長、教育端了解唯有讓孩童自我認知並養成運動習慣，才得以使體適能情形提升；除此之外，本階段也同時進行身體活動量計的原理與應用進行科普推廣，讓台灣科技應用的冰冷感覺，普及到隨處可見的生活互動。

1. 器材使用與教育訓練：聚集學校老師以及相關施測人士，提供器材硬體之使用、配戴以及保養教學，並輔導如何將資料上傳平台、基本資訊設定與編輯以及個人表現評估與分析等功能（依權限不同，僅學校單位可閱讀該校內學童之比較與報表，僅政府單位可閱讀區域性學校比較與報表）。
2. 實施與輔導：定時前往實施學校進行問題處理與實施現況了解，改良系統與硬體之設定。

（二）指標性實施學校遴選與認養

所謂指標性學校，乃指其學校所在位置（靠山、靠海、都市、鄉村等）與學校風氣（升學、運動、飲食等）特殊之單位，由於該因子具有影響孩童身體活動量可能性，因此期望將優先進行該校之實施，而本階段研究結果，將用以建立全域不同區域類似學校之比較依據。

（三）常模更新

根據本提案所使用之獨立資訊處理平台，不同年齡層、不同性別、不同BMI甚至不同區域的學童身體活動量常

模，可以學童上傳資訊後的短時間內進行常模更新，提供表現落點、建議數值等參考數值。

（四）各層級報表檢視

鑒於各學校及政府單位可能有其活動上安排的需求，例如學校舉辦之運動會、政府提倡之快活計畫等，本提案也可針對不同時間點的孩童身體活動量進行分析，提供各單位活動的實際狀況與影響程度報表，以利下一學期、年度的活動調整與安排。

綜合以上所知，運動本身除了對身體機能及動作發展產生許多正面效果之外，運動與學習有很大的關聯性。運動會調節多巴胺、血清素、正腎上腺素的產生（運動改造大腦一書，2011）。其中多巴胺讓孩子有好情緒，血清素和記憶與情緒有關，正腎上腺素和注意力有直接的相關。運動可以讓學習效果變好，同時也提升孩子的自信心；身體不但變得更健康，抵抗力也變強了。因此，筆者提倡利用「運動感測器」結合雲端數位學習與評量，玩出運動的樂趣，讓全家培養一起運動的好習慣，例如：跑步、游泳、打球等活動，或是帶孩子至戶外活動、公園玩耍。在日常生活中多走路、多爬樓梯，能動就多動，既可以體驗到運動的好處，也可以促進親子之間的感情，讓「運動感測器」穿戴科技成為大家的健康守護神的神奇法寶。

參考文獻

- 王駿濠、張哲千、梁衍明、邱文聲、洪蘭、曾志朗、阮啟弘（2012）。運動對孩童認知功能及學業表現的影響。《教育科學研究期刊》，57(2)，65-94：文獻回顧與展望。
- 古博文、陳俐蓉（2013）。客觀測量身體活動量：當前議題與未來方向。《身體活動與運動科學學刊》，2（1）。
- 張育愷、林珈余（2010）。身體活動對孩童認知表現的影響。《中華體育季刊》，24（2），83-92。
- 張樹立（2004）。台北縣城鄉國小學童身體活動量與健康體適能之比較研究。未出版碩士論文，國立臺北師範學院，臺北市。
- 彭臺臨（2007）。臺灣地區國民運動行為與其政策意涵之研究。行政院體育委員會。崑山科技大學。九十五學年度學生體適能檢測報告。
- Andersen, L. B., Harro, M., Sardinha, L. B., : Physical activity and clustered cardiovascular risk in children: a cross-sectional study (The European Youth Heart Study). *Lancet* 2006;368:299-304.
- Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing: What have we learned about cognitive development? *Trends in Cognitive Science*, 9(3), 104-110.
- Chaddock, L., Hillman, C. H., Buck, S. M., & Cohen, N. J. (2011). Aerobic fitness and executive control of relational memory in preadolescent children. *Medicine & Science in Sport & Exercise*, 43(2), 344-349.
- Donnelly, J. E., & Lambourne, K. (2011). Classroom-based physical activity, cognition, and academic achievement. *Preventive Medicine*, 52(1), S36-S42.
- Kang, D. W., Choi, J. S., Lee, J. W., & Tack, G. R. (2012). Prediction of energy consumption according to physical activity Intensity in daily life using accelerometer. *International Journal of Precision Engineering Manufacturing*, 13(4), 617-621.
- Trost, S. G., Wong, W. K., Pfeiffer, K. A., & Zheng, Y. (2012). Artificial neural networks to predict activity type and energy expenditure in youth. *Medicine and Science in Sports and Exercise*, 44(9), 1801-1809.
- Wang, Q. J., Suominen, H., Nicholson, P. H. F. (2005). Influence of physical activity and maturation status on bone mass and geometry in early pubertal girls. *Scand Journal Medicine Science Sports*, 15, 100-6.
- Welk, G. J., Schaben, J. A., & Morrow Jr., J.R. (2004). Reliability of accelerometry-based activity monitor: A generalizability study. *Medicine & Science in Sports & Exercise*, 36, 1637-1645.
- Westerterp, K. R. (1999). Physical activity assessment with accelerometers. *International Journal of Obesity*. 23(3), S45-49.

成就目標與課室目標對教師教學之啓發

吳中勤

國立成功大學教育研究所博士生

一、前言

在接受教育的過程中，學習成就一直是學生、家長、教師乃至於社會大眾所普遍關注的焦點，並普遍以外在表現，如：分數，來解釋學生的學習成就，忽略從事學習活動時，學生複雜的心理歷程，可能造成學生不適當的負向情緒隨學習活動而增加（Pekrun, Elliot, & Maier, 2006, 2009），並可能導致較差的學業表現（Elliot, McGregor, & Gable, 1999；Harackiewicz, Barron, Tauer, & Elliot, 2002）。探討學生目標動機的成就目標理論，有助於教師了解不同學習者從事學習活動背後的不同原因，提供教師教學之參考，裨益提升學生學習動機與學習成效。

成就目標的研究提供教師了解學生追求學業成就背後原因的重要依據，近三十年來受到國內外研究的重視（程炳林，2003；Linnenbrink & Pintrich, 2002；Murayama, Elliot, & Yamagata, 2011），但國內的研究卻相對較少。學生追求學業成就的原因可能是為了要努力的學會（趨向精熟）或避免學不會課程內容（逃避精熟）；亦或想要贏過別人（趨向表現）或不想要輸給別人（逃避表現），以致於因為不同原因而從是學習活動的個人可能會有不同的學習行為與結果（Elliot & Thrash, 2010；Elliot, Shell, Henry, & Maier, 2005；Harackiewicz, Barron, Pintrich, Elliot, & Thrash, 2002；Pekrun et al., 2009）。

教師教學或進行課程設計主要目的之一便是在於誘發、維持並提升學生的學習目標動機，以期學生以更主動、積極的態度求取知識、追求學業成就。然而，隨著年齡的增長學習者的趨向精熟與趨向表現目標皆呈現顯著的下降（林宴瑛、程炳林，2007；Shim, Ryan, & Anderson, 2008），而逃避表現目標則是顯著的上升（Fryer & Elliot, 2007）。因此，教師必須透過不同的教學或課程設計型塑不同的課室目標結構，進一步維持獲提升學生所抱持的不同目標動機（Church, Elliot, & Gable, 2001；Linnenbrink, 2005）。本研究之主要目的便在於透過成就目標理論及學生知覺到教師所傳遞出之課室目標結構之探討，植基於研究發現與學者之論點，提供教師教學之參考。

二、成就目標理論之意涵

成就目標理論（achievement goal theory）又稱為目標導向理論（goal orientation theory），主要源自於成就動機理論（Maehr & Zusho, 2009）。成就動機構念最早由 James 於 1890 年提出，以能力（competence）為其核心概念，成為當代成就目標測量中，以能力作為架構其理論模型的依據。近三十多年來，成就動機的研究焦點轉向學生的學業與學習目標後（Urduan & Maehr, 1995；Elliot, Murayama, & Pekrun, 2011），使

得成就目標的測量備受矚目，成為當代成就動機中最具影響力的一個研究議題（Elliot, 1999；Pintrich, 2000a；Pintrich, 2000b）。

成就目標中，所謂目標的內涵不僅止於具體與明確之成就目標的設定（Locke & Latham, 1990），而是更進一步探討從事相同學習活動背後的不同原因（Meece, Beverly, & Askew, 2009；Schunk, Pintrich, & Meece, 2008），例如：目標同樣是考試考一百分的兩位學生，背後的原因可能是想要精熟學習內容以發展能力，也可能是為了贏過別人來展現能力（Urdu, 1997；Urdu & Maehr, 1995）。在成就情境中，因個人所抱持的成就目標不同，對行為結果會產生不同的認知知覺與情緒反應（Dweck, 1986），因此，成就目標的測量成為了解學生學業成就歸因、成就情緒與成就行為的重要依據。成就目標理論有其不同的理論主張，茲將之分述如下。

（一）二向度成就目標理論

最早對成就目標構念的定義，是將成就目標視為成功達到優秀（excellence）的標準，這些優秀的標準包含了勝過他人、達到或超過對自我的要求以及達到標準時的情感反應，例如：驕傲或痛苦（McClelland, Atkinson, Clark, & Lowell, 1976），形成日後成就目標相關研究的基礎。成就目標相關的實證研究，最早在1970年代晚期以及1980年代初期，由Ames、Dweck、Maehr與Nicholls等人，各自進行相關研究後，分別針對成就目標提出各自的看法，為成就目標之發展立下了重要的根基，其中又以Dweck與

Nicholls二人對成就目標的概念化二分影響最為深遠（Elliot, 2005）。

最初，Dweck與Nicholls所提出的成就目標，在對能力的定義上，是全然二分且互斥的概念。Dweck（1986）將成就目標分成表現目標與學習目標；Nicholls（1979, 1984）將成就目標區分自我投入（ego involvement）與工作投入（task involvement）。持「表現」與「自我投入」目標的學習者，學習的目的是要勝過他人或投入較少的努力而成功，藉此證明自己是有能力的；相對的，持「學習」與「工作投入」目標的學習者，其學習焦點則放在發展新的能力與學習任務的精熟（Ames & Archer, 1987）。雖然Dweck與Nicholls對成就目標給予不同的分類，但其中卻有許多相似之處，因此分別被以精熟目標與表現目標稱之（Schunck et al., 2008），這樣的二分法也成為往後研究成就動機的基本架構，稱之為二向度成就目標模型或基準目標理論（Pintrich, 2000a）。在基準目標理論中，不論精熟或表現目標，從事學習活動的目的都被認為是努力的發展或證明自己的能力，因此，在二分成就目標對能力的定價或動機傾向上，皆被視為是「趨向」目標（Ames, 1992）。

（二）三向度成就目標理論

Elliot與Harackiewicz（1996）認為，將精熟與表現成就目標皆視為趨向目標的觀點，並未考慮到在表現目標導向中，趨向與逃避是獨立的動機傾向。二向度成就目標理論假定，持精熟目標的學習者會產生精熟的動機組型或適應性的學習結果（Senko, Durik, & Elliot, 2008），例如：

偏好具有適當挑戰的任務、對學習抱持著正向的情感、面對失敗時的堅持；反之，抱持表現目標的學習者卻會產生不適應性的動機組型或學習結果，像是偏好簡單任務、將失敗歸因於缺乏能力或逃避失敗的情境等（Senko, Durik, & Harackiewicz, 2008；Wolters, Yu, & Pintrich, 1996）。但持趨向表現目標者的行為目的，在於成功結果的追求，以獲得能力的正向評斷，因此，如同精熟目標，也會產生適應性的學習結果（Pintrich, 2000b）；逃避表現目標者的行為在於對負向結果的逃避，以避免能力的負向評斷（Köller, 2000; Meece et al., 2009），行為結果卻會產生較為負向或不適應的學習組型，因此，成就目標理論學者將表現目標進一步區分為趨向表現與逃避表現目標。

（三）四向度成就目標理論

在實際生活中，三分成就目標理論中對精熟目標的理解卻過於簡單。以數學概念的學習為例，想要學會某一個數學概念是屬於精熟目標，但同樣抱持此目標的人，卻可能有不同的動機傾向，例如：有些人努力想要學會（趨向精熟），而有人卻是努力避免學不會某一個數學概念（逃避精熟）、不理解或學不會教材內容，甚至避免忘記已經學會的……等，這種類型的學習者可以說是完美主義者，會努力避免犯任何錯誤，行為目的在於逃避任何負向結果發生的可能。因此，Elliot（1999）主張將精熟目標進一步區分為趨向精熟與逃避精熟。

Elliot 與 MacGregor（2001）依據 Elliot（1999）四向度成就目標模型的理論構想，分別根據能力的定義（精熟與表現）與

定價（valence）（趨向與逃避）為經緯，將成就目標區分成趨向精熟、逃避精熟、趨向表現與逃避表現等四向度，相關研究結果亦顯示四向度成就目標模型，較成就目標之其它分類模型與資料的適配度更佳（程炳林，2003；Elliot & MacGregor, 2001），提供四向度成就目標之理論區分的證據。

三、學生知覺到的課室目標結構

根據社會認知理論中的三元交互決定論，個人成就目標為個人、環境與行為交互作用下的產物（Schunk et al., 2008），Ames 與 Archer（1988）研究指出，學生知覺到教師所傳達出的課室目標結構，亦即教師所抱持的某種教學要求或情境限制（Ames, 1992），會影響學習者學習策略的使用與動機，同時也會影響個人所採用的成就目標，意謂著學習者個人的成就目標會因教師教學與課程設計而有所不同。Wolters（2004）認為，課室目標結構是指學生知覺到教師所強調的成就目標類型，因此，如同個人成就目標，可分成二向度、三向度與四向度課室目標結構。

Lau 與 Nie（2008）指出課室表現目標結構增強了個人逃避表現對課室活動參與學業退縮（withdraw）間的負向關係。並且，課室表現目標結構與個人逃避表現目標對學生的學習結果皆呈現不良的影響，反之，課室精熟目標結構與個人精熟目標顯示出適應性的關係。相關研究更進一步指出，知覺到趨向表現課室目標結構的學生較傾向抱持趨向表現與逃避表現目標（Wolters, 2004），較缺乏內在動機（Murayama & Elliot, 2009），

但抱持著較高的自我效能（Murayama & Elliot, 2009）；而高精熟目標結構的班級中學生較少抱持趨向表現目標（Wolters, 2004），並具有較高的內在動機。然而，林宴瑛（2012）的研究卻指出趨向精熟與趨向表現課室目標結構皆有助於學生正向情緒的產生，上述研究結果似乎意謂著精熟目標與趨向表現目標都可能產生適應性的學習結果，因此，精熟與趨向表現兼具的課室目標結構可能對學生學習表現（Barron & Harackiewicz, 2001；Pintrich, 2000a, 2000b）及學習情緒會有最佳的效果（林宴瑛，2012）。

四、型塑課室目標結構的動機策略

教師可使用不同的動機策略來提生學生的目標動機（Schunk et al., 2008），Epstein（1989）提出 T.A.R.G.E.T 此六個動機策略來型塑不同之課室目標結構，此六個向度分別為學習任務的安排（task）、權力的運用（authority）、對學生的認同（recognition）、分組進行學習（group）、學習成效的評量（evaluation）與時間的分配（time）。Ames（1992）在綜合了許多學者後指出，教師可依據學習任務與活動設計、評量與酬賞的使用，以及權力與責任的分配等動機策略，來型塑不同的課室目標結構。

學習活動的安排或設計，是指學習任務是否允許以不同方式完成；在權力的運用或責任的分配上，則是指給予學生自主決定的空間並為自己的決定負責；教師的認同則是指教師是否適時的使用正式或非正式的酬賞，激勵學生繼續學習；分組學習目的在於透過與他人合作取代競爭；學習成效的評量若以公開的成績排名，

可能增強學生趨向表現或逃避表現目標；學習時間的分配涉及了是否考量學習者在學習上的個別差異，允許學習者可根據自己的學習速度來分配學習或完成任務的時間（Schunk et al., 2008）。

五、建議

（一）持續關注學生成就目標

由於逃避動機有害學生的學業表現，因此教師不僅應在學期開始時，進行簡單的問卷調查以了解學生的成就目標，以便給予逃避傾向的學生較多的關注；另一方面，由於學習者的趨向精熟與趨向表現目標會隨時間而下降，而逃避表現目標可能隨時間而增加，因此，教師也應該注意趨向精熟與趨向表現的轉變，例如：在學期中再次進行相同的問卷調查，了解學生成就目標的變化情形，以避免學生的目標動機從趨向轉變為逃避。

（二）適當提供表現機會

教師應給予學生自主決定完成學習任務的時間與表現的機會，例如：鼓勵學生擬定學習計畫再由教師針對個別計畫給予建議，並讓學生有與同學分享其優良學習計畫的機會，以收見賢思齊之效；此外，也應適時給予正式或非正式的獎勵，但不應鼓勵競爭的行為，讓學生能夠有機會展現自己精熟學習的成果，並且透過分組學習的方式，讓小組成員有合作學習與發表的機會，例如：於學期中或學期末舉辦學習成果發表活動，增加學生的自我效能感與趨向精熟的目標動機。

(三) 規劃多元學習任務與建立多元評量機制

由於知覺到的趨向精熟、逃避精熟與逃避表現課室目標結構可能隨時間呈現下降的趨勢（林宴瑛、程炳林，2007）。除了設計具體與稍為困難的目標能夠有助於提升學生的表現之外（Locke & Latham, 1990），教師亦可設計多元的學習任務以減少學生因為從事相同的學習任務而容易產生的社會比較。此外，教師應採多元的評量標準，避免以單一的表现標準（如：考試分數），來評量學

生的學習成就，營造趨向精熟的課室目標結構，進一步提升個人趨向精熟目標。舉例而言，教師應允許學生以紙筆、口頭報告、音樂或美術作品等不同的方式，或以小組或個人為單位來完成學習任務，並同時兼採總結性（如：期中、期末測驗或小考）與形成性（如：檔案評量）的評量方式來評估學生的成就，讓學生不僅有發展其多元智能的機會，也能夠增加其精熟學習任務的動機，並同時避免紙筆測驗的挫折造成逃避學習傾向的增加。

參考文獻

- 林宴瑛（2012）。從人境互動觀點探討目標結構、目標導向、情緒調整對考試情緒與因應策略之效果：條件化間接效果暨調節效果分析（未出版博士論文）。國立台灣師範大學，台北。
- 程炳林（2010）。四向度目標導向模式之研究。師大學報，48（1），15-40。
- 林宴瑛、程炳林（2007）。個人目標導向、課室目標結構與自我調整學習策略之潛在改變量分析。教育心理學報，39（2），173-194。
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261-271.
- Ames, C., & Archer, J. (1988). Achievement goals in the classroom students' learning strategies and motivation processes. *Journal of Educational Psychology*, 80(3), 260-267.
- Barron, K. E., & Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology*, 80(5), 706-722. doi: 10.1037//0022-3514.80.5.706
- Church, M. A., Elliot, A. J., & Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology*, 93(1), 43-54. doi: 10.1037//0022-0663.93.1.43
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41(10), 1040-1048.
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34(3), 169-189.
- Elliot, A. J. (2005). *A conceptual history of the achievement goal construct*. In A. J. Elliot, & C. S. Dweck (Eds.). *Handbook of competence and motivation*(pp.52-72). New York: Guilford.
- Elliot, A. J., & Dweck, C. S. (2005). *Competence and motivation: Competence as the core of achievement motivation*. In A. J. Elliot & C. S. Dweck (Eds.). *Handbook of competence and motivation* (pp.3-14). New York: Guilford.
- Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, 70（3），461-475.

- Elliot, A. J., & McGregor, H. A. (2001). A 2×2 achievement goal framework. *Journal of Personality and Social Psychology, 80* (3), 501-519.
- Elliot, A. J., McGregor, H. A., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology, 91*(3), 549-563. doi: 10.1037/0022-0663.91.3.549
- Elliot, A. J., & Thrash, T. M. (2010). Approach and avoidance temperament as basic dimensions of personality. *Journal of Personality, 78*(3), 865-906. doi: 10.1111/j.1467-6494.2010.00636.x
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist, 34*(3), 169-189. doi: 10.1207/s15326985ep3403_3
- Elliot, A. J., Shell, M. M., Henry, K. B., & Maier, M. A. (2005). Achievement goals, performance contingencies, and performance attainment: An experimental test. *Journal of Educational Psychology, 97*(4), 630-640. doi: 10.1037/0022-0663.97.4.630
- Epstein, J. (1989). Family structures and student motivation: A developmental perspective. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (Vol. 3, pp. 259-295). San Diego: Academic Press.
- Fryer, James W., & Elliot, Andrew J. (2007). Stability and change in achievement goals. *Journal of Educational Psychology, 99*(4), 700-714. doi: 10.1037/0022-0663.99.4.700
- Harackiewicz, Judith M., Barron, Kenneth E., Pintrich, Paul R., Elliot, Andrew J., & Thrash, Todd M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology, 94*(3), 638-645. doi: 10.1037//0022-0663.94.3.638
- Harackiewicz, Judith M., Barron, Kenneth E., Tauer, John M., & Elliot, Andrew J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology, 94*(3), 562-575. doi: 10.1037//0022-0663.94.3.562
- Köller (2000). Goal orientations: Their impact on academic learning and their development during early adolescence. In J. Heckhausen (Eds.). *Motivational psychology of human development* (pp.129-142). Amsterdam: Elsevier Science.
- Lau, Shun, & Nie, Youyan. (2008). Interplay between personal goals and classroom goal structures in predicting student outcomes: A multilevel analysis of person-context interactions. *Journal of Educational Psychology, 100*(1), 15-29. doi: 10.1037/0022-0663.100.1.15
- Linnenbrink, E. A. (2005). The dilemma of performance-approach goals: The use of multiple goal contexts to promote students' motivation and learning. *Journal of Educational Psychology, 97*(2), 197-213. doi: 10.1037/0022-0663.97.2.197
- Linnenbrink, E. A., & Pintrich, P. R. (2002). Achievement goal theory and affect: An asymmetrical bidirectional model. *Educational Psychologist, 37*(2), 69-78. doi: 10.1207/s15326985ep3702_2
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Murayama, K., & Elliot, A. J. (2009). The joint influence of personal achievement goals and classroom goal structures on achievement-relevant outcomes. *Journal of Educational Psychology, 101*(2), 432-447.
- Maehr, M. L., & Zusho, A. (2009). *Achievement goal theory: The past, present, and the future*. In K. R. Wentzel & A. Wigfield (Eds.) . *Handbook of motivation at school* (pp.77-104). New York: Routledge.
- Meece, J. L., Beverly, B. B., Askew, K. (2009). *Gender and motivation*. In K. R. Wentzel & A. Wigfield, (Eds.) . *Handbook of motivation at school*(pp.417-431). New York: Routledge.
- Murayama, K., Elliot, A. J., & Yamagata, S. (2011). Separation of performance-approach and performance-

- avoidance achievement goals: A broader analysis. *Journal of Educational Psychology*, 103(1), 238-256. doi: 10.1037/a0021948
- McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1976). *The achievement motive* (pp.107-138). New York: John Wiley & Sons.
- Nicholls, J. G. (1979). Quality and equality in intellectual development. *American Psychologist*, 34, 1071-1084.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91 (3) , 328-346.
- Pekrun, R., Elliot, Andrew J., & Maier, Markus A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*, 98(3), 583-597. doi: 10.1037/0022-0663.98.3.583
- Pekrun, R., Elliot, Andrew J., & Maier, Markus A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology*, 101(1), 115-135. doi: 10.1037/a0013383
- Pintrich, P. R. (2000a). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology*, 25(1), 92-104. doi: 10.1006/ceps.1999.1017
- Pintrich, P. R. (2000b). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, 92(3), 544-555. doi: 10.1037/0022-0663.92.3.544
- Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education: Theory, research, and applications*. Upper Saddle River, N.J.: Pearson/Merrill Prentice Hall.
- Senko, C., Durik, A. M., & Harackiewicz, J. M. (2008). Historical perspectives and new directions in achievement goal theory: Understanding the effects of mastery and performance-approach goals. In Shah, J. Y & Gardner, W. L. (Eds.). *Handbook of motivation science* (pp.100-113). New York: Guilford Press.
- Shim, S. S., Ryan, A. M., & Anderson, C. J. (2008). Achievement goals and achievement during early adolescence: Examining time-varying predictor and outcome variables in growth-curve analysis. *Journal of Educational Psychology*, 100(3), 655-671. doi: 10.1037/0022-0663.100.3.655
- Urdu, T. C. (1997). *Achievement goal theory: Past results, future directions*. In M. L. Maehr & P. R. Pintrich, (Eds.). *Advances in motivation and achievement* (vol. 10). London: Jai Press.
- Urdu, T. C, & Maehr, M. L. (1995) .Beyond a two-goal theory of motivation and achievement: A case for social goals. *Review of Educational Research*, 65(3), 213-243.
- Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology*, 96(2), 236-250. doi: 10.1037/0022-0663.96.2.236

活化教學的動能：教師專業學習社群觀點

賴協志

國家教育研究院教育制度及政策研究中心助理研究員

一、前言

處在少子女化、強調競爭力與教育品質的新時代，教師應體會活化教學能力及促進專業發展的重要，除了重視個人知識學習外，應積極參與專業學習社群，藉由溝通分享及團隊學習的力量，持續精進與成長，進而提升教與學的效能。

推動「活化教學」的目的，在於打破老師傳統僵化教學模式的限制，以更生動活潑的教學方法啟發學生對知識的熱愛與追求，而學生在此過程中所扮演的角色並非被動的知識接收者，而是主動的知識探求者（林騰蛟，2013）。活化的教學，意謂著融合各種不同的教學策略，然後依據不同情境隨時做調整，是無法一式多用、或是「定型挪移」。也就是說，知識是「活的」、必須能夠「活用」、教學方式更需要「活化」。而學習社群的建構是活化教學的重要策略之一（李宜芳，2013）。

專業學習社群的建構，係發展教師專業學習與反思的機制（Munby, Russell, & Martin, 2001）。透過參與社群活動的合作與對話，裨益其發展教師專業的能力（吳明清，2002）。教師專業學習社群成立的目的，除了分享知識和聯絡情感之外，最重要的在於豐富教學知識、強化理論與實務的連接，著重於教師教學的反思，將所學落實到實際教學上（吳清山，2012）。教師專業學習社群成立目

的是社群成員一同分享與探討教學的問題與困境，並提供改進或精進教師教學，進而提升學生學習表現。其運作過程強調社群成員中的知識分享，集體的知識創造（鮑慧門，2013）。

在教師專業學習社群裡，一群教師們抱持共同的信念、願景或目標，為致力於促進學生獲得最佳的學習成效，而努力不懈地以合作方式共同進行探究和問題解決，其不僅重視教師專業知能的成長或個別興趣的追求，更強調學生學習成效的提升（張德銳、王淑珍，2010）。學習社群是有效的專業成長方式，教師們形成一個成長團體，為改進教學、促進學生學習彼此切磋成長：成員們定期聚會，最好是一星期數次，共同討論課程與學習目標、教學計畫、解決教學上遭遇到的問題。整體來說，研究社群已逐漸釐清從「教師專業發展」到「教學改進」的連結關係（張景媛、鄭章華、范德鑫、林靜君，2012）。教師專業學習社群的推動，能協助教師發展教學特色及增進教學能力，並藉由團隊學習及共同行動研究的結合，提升教師教學效能及持續改善教學品質（Huffman & Hipp, 2003; Roberts & Pruitt, 2003）。本文主要在探討教師專業學習社群的意義及內涵，並且從 APPLE 觀點分析教師專業學習社群運作的要素。

二、教師專業學習社群的意義與內涵

教師專業學習社群理念可回溯至 Judith Little 於 1981 年發表的《組織的權力》(The power of organizational setting)；Little 認為教師教學的改善與學校學習風氣的提升緊密相關，並提出教師可透過協力合作模式，激發教育改革動能 (Fullan, 2006)。最早提出專業學習社群是 Shirley. M. Hord (1997) 所著的《專業學習社群：持續探究與改進的社群》(Professional learning communities: Communities of continuous inquiry and improvement) 一書。隨著學習型組織概念的倡導，專業學習社群的理念漸漸開展與應用於教育情境中 (Ostmeyer, 2003)。

所謂教師專業學習社群，Dufor 與 Eaker (1998) 認為其係指一群教育人員，他們建立一個能增進互相合作、情緒支持和個人成長的環境，在這個環境中，達成個人無法達成的目標。Hord (2004) 強調專業學習社群係在共有的目標與願景下，社群內領導者有效帶領與支持成員參與社群，在過程中採取合作的方式，成員也持續分享實務經驗，並且在外在環境與人員的支持下，使社群有更良好的運作，且成員也非常時時注意社群的運作情況與結果，以確保專業學習社群的品質。

教育部 (2009) 認為教師專業學習社群是指一群志同道合的教育工作者所組成，持有共同的信念、願景或目標，為致力於促進學生獲得更加的學習成效，而努力不懈地以合作方式共同進行探究和問題解決。簡言之，教師專業學習社群必須關注於學生學習成效的提升，不僅

止於教師專業知能的成長或個別興趣的追求而已。吳清山與林天祐 (2012) 指出學習共同體就是學習社群，係指地方、家長以及教師通力合作，共同關注到每一個學生的學習，讓學生在充滿學習氛圍的情境中學習，以發揮最大的學習效果。鮑慧門 (2013) 認為教師專業學習社群是指由教師所組成，基於專業的共同信念、願景或目標，以互動對話、自我省思、探索學習、協同合作及分享討論等方式，來促進學生學習成效，並精進自身的專業素養。

有關教師專業學習社群的內涵，Hord (1997) 提出專業學習社群有五項內涵，分別是：1. 支持性及分享式領導；2. 共同的價值與願景；3. 合作學習與應用；4. 分享個人實務；5. 支持性的環境。Roberts 與 Pruitt (2003) 指出學習社群的內涵有：1. 學校中的領導權分享；2. 學校中的合作；3. 學校的共同願景；4. 學校的結構化／社會支持；5. 學校的共同實務。Stoll 與 Louis (2007) 則認為專業學習社群的內涵在於：1. 專業學習；2. 凝聚力的團體；3. 集體知識；4. 學校領導者、教師及學生間的人際關懷倫理。

Hipp, Huffman, Pankake 與 Olivier (2008) 認為專業學習社群的內涵包括：1. 共享與支持領導；2. 共同價值與願景；3. 合作學習與應用；4. 分享個人實踐；5. 支持條件 (關係與結構)。陳佩英 (2009) 認為專業學習社群的內涵有：1. 支援性和共享領導；2. 共享價值觀與願景；3. 共同學習和應用學習成果；4. 共享實務經驗；5. 內部與外部支援。教育部 (2009) 指出教師專業學習社群具有不同於一般團體的特徵，包括：1. 共同願景、價值

觀與目標；2. 協同合作：聚焦於學習；3. 共同探究學習；4. 分享教學實務；5. 實踐檢驗：有行動力，從做中學；6. 持續改進；7. 檢視結果。

綜合上述，教師專業學習社群係指一群秉持共同願景、信念、價值觀的學校教育工作者，能相互合作且發揮團隊力量，具體展現實際行動，透過學習、研討、對話、分享及省思教與學的實務經驗，以增進教師教學專業知能，進而能有效提升學生學習成效及達成教學目標所組成的專業成長團體。其內涵主要包括：1. 展現專業及支持性的領導；2. 秉持共同願景、信念與價值觀；3. 共享教學實務；4. 相互合作與共同學習；5. 關注教與學效能的提升；6. 學校內外部資源的支援。

三、從APPLE觀點談教師專業學習社群運作的要素

教師專業學習社群的有效運作，是活化教師教學及促進專業成長的重要動能。王巧萍（2011）的研究發現教師專業學習社群之運作成效來自於教學方案研發與教學啟發，有利於解決教學現場實務問題，減少教師孤立不安感，增進教師對課程與教學的關心。許詩珮（2010）的研究指出在教師專業學習社群的運作模式與過程中，教師的專業能力逐漸產生改變。鄭傑文（2013）的研究發現精進教學計畫學校之教師學習社群發展愈好，其教師專業發展成效亦佳。吳明芬（2013）的研究指出教師專業學習社群與教師教學效能具有正相關，教師專業學習社群對教師教學效能具有預測力。鍾昫珊（2013）的研究發現參與教師專業學習社群教師的教學效能有高精度水準，在多元教學

策略表現水準較高；教師在教師專業學習社群表現，有助於教師教學效能表現水準。依據專家學者（丁一顧，2013；吳清山，2012；許德田，2012；張德銳、王淑珍，2010；Hord & Hirsh, 2008; Huffman & Hipp, 2003; Roberts & Pruitt, 2003）的看法，分別從行動（Action）、專業（Profession）、熱情（Passion）、學習（Learning）、效能（Effectiveness）等面向，即APPLE觀點，探討教師專業學習社群運作的要素，並敘述如下：

（一）行動：具體行動且落實理念

專業學習社群是行動取向的，願意嘗試與創新，社群成員需將其目標與理想轉換成實際的行動。共同學習不是只有討論，而是一起去做，付諸實踐，從做中學，並且針對結果和影響結果的因素進行反省，發展新的理論、嘗試新的試驗、檢驗其成效（教育部，2009）。學校的專業學習社群宜規劃多元型態的運作模式，符應不同社群需求，包括：1. 協同備課、觀課與評課；2. 教學觀察、討論與回饋；3. 同儕討論、對話與省思；4. 建立E化專業歷程檔案分享平臺；5. 案例分析討論與資源運用分享；6. 主題經驗分享；7. 主題探討（含專書、影帶）；8. 課程發展與評鑑；9. 教學媒材研發；10. 教學方法與多元評量創新；11. 行動研究；12. 標竿楷模學習，觀摩優質社群運作經驗；13. 新進夥伴輔導；14. 專題講座與研討（林志成，2013）。教師專業學習社群提供了教師各種不同之成長途徑以促進教師專業成長；由於各學科領域的內涵及教學方式殊異，教師可依學校發展或教師需求募集校內同領域、跨領域教師或校外夥伴學

校教師，成立教師專業學習社群，增進教師同儕的專業成長；其參與類型可包括「年級/年段」、「學科/領域/學群」、「學校任務」、「專業發展主題」等四類，而這四類型可依教師所需，以校內或跨校方式進行（葉素枝，2011）。由上可知，專業學習社群係屬行動取向，行動力的具體展現，是落實教師專業學習社群理念及促進教師專業成長的關鍵要素。在行動過程中，其運作方式十分多元且兼具彈性，在實際推動時，需結合教學目標及實際需求，擇取適當的運作模式或成長途徑，方能達到事半功倍之效。

（二）專業：省思對話與專業發展

在教師專業學習社群的動態發展歷程中，開啟專業對話、回應社群需求及促進專業發展是必須受到重視的（Huffman & Hipp, 2003）。省思對話是專業學習社群的重要特徵；專業學習社群內的教師在彼此互動過程中，對於工作中有困難或疑惑的地方會提出來一起討論，透過解決意見的提供與交換，讓教師進行反省或批判性思考，以獲得解決的方法，藉此專業的對話使教師不斷改善缺點（張德銳、王淑珍，2010）。「教育實踐的省思與對話」是學校專業學習社群的性質與要素之一，透過對話可以協助社群成員達成不同層面的專業發展，重視教師聲音和觀點的表達，讓教師獲得應有的專業尊嚴、建立自信心，發展出相互支持的集體專業意識，再發展出更有效教學策略（許德田，2012）。教師專業學習社群是藉由經驗的分享、知能的傳承、觀摩學習，持續性的引領成員專業發展、建構組織的專業形象，進而促進教師專業發展（蔡

玫芳、黃惠貞，2012）。據此了解，省思對話及共同研討是促進教師專業成長、提升社群運作成效的關鍵要素，能讓教師反思及分享實務經驗，並獲得解決問題的有效良方。

（三）熱情：有使命感且積極參與

理論上教師團體應可以成為一個專業社群，因其具有共同的專業目標、專業活動，以及專業規範，因此，社群成員對所屬的社群不僅具有強烈的歸屬感，也會有榮辱與共的情懷，以及致力社群發展的使命感（吳明清，2002）。社群參與係基於個體自願性；當社群成員的自我認知及社群認知相重疊時，他們會表現出熱情，積極且主動參與社群，並藉著在社群中參與各項活動、協助他人解決問題，來提升自我價值，達成個人及社群目標（Bhattacharya & Sen, 2003; Roberts & Pruitt, 2003）。若學校擁有高度參與的學習社群，則成員對社群會有向心力及使命感，在忠誠、專業或產出方面均有良好表現，能提升教師個人信心與專業表現、促進團體成員互相砥礪且對學校組織文化有正面影響（何文純，2006；Huffman & Hipp, 2003）。依此而言，社群成員能保持高度活力與熱情，持續且主動參與社群活動，且對社群有高度向心力及使命感，將使社群的推展及校務整體的運作更加順遂。

（四）學習：共同合作及團隊學習

專業學習社群的核心機制在於學習，強調運作的內容；在專業學習社群中，社群成員共同探究知識和進行實踐（Hord & Hirsh, 2008）。教師專業發展的學習能

否有效果，取決於教師們各自的能力與努力，主要是能否有效學習並且在課堂上實施出來（Leithwood, Aitken, & Jantzi, 2001）。當社群成員在協商找出一致性的結論時，難免會有衝突及磨合，之後便逐漸凝聚一股團結力量，彼此之間更加了解，也更容易相互合作與學習，當團結一致的信念在社群中形成時，該社群便開始有效能及效率地運作（Diaz-Maggioli, 2004）。學校教師必須持續學習才能增進變革能量，學校行政人員和決策者也都必須持續學習，才能提升教育變革方案的品質與效率，主要目標是讓「教學專業」轉化為「學習專業」（吳明清，2005）。

學校在安排研習和進修活動時，依據教師整體需求，採取年級、領域或專業社群的型態進行。而在規劃活動主題和講師時，也可由年級、領域或專業社群的教師自主編選（潘慧玲、高嘉卿，2012）。由此了解，共同合作及團隊學習可說是專業學習社群運作的核心機制，在社群裡，結合成員的能力與努力，發揮團隊合作精神及團結力量，更能展現教師教學專業與學習專業。

（五）效能：能提升教與學的效能

教師專業學習社群的有效運作，能提高教師教學技巧與效能，進而提升學生學習成效（Bunker, 2008; Waters, 2009）。教學效能之提升，在教育改革中扮演極重要之角色；一位專業的教師勢必擁有良好的教學效能，才能在師生互動中將專業的教育理念與創意觀念直接或間接實踐於教室教學與學生學習當中（羅仁駿、蔡慧敏、許惠英、黃寶月，2010）。教師專業學習社群的有效運作，

促進了教學知識的交流，提高教學問題解決能力，有助於促進班級教師教學效能的提升；尤其，藉由彼此的教學觀察與課程內容的觀點分享，更對學生學習成效提供了直接的促進影響效果（鮑慧門，2013）。專業學習社群運作與推展其實在於關注學生學習，因為，唯有關注學習學習，學生學習成效才有改善的可能，也唯有學生學習有改變，教師專業與自主才能為社會及家長所認同（丁一顧，2013）。依此而言，教師在組成或參與專業學習社群的過程中，應以增進教學知能、促進專業成長及提升學習成效為努力方向，並將所學用於實際教學及班級經營之中，發揮社群的影響力，使學生能真正受惠。

四、結語

為了活化教學動能、精進教學能力及促進專業成長，教師應了解專業學習社群的理念與內涵，並且有效掌握運作要素，包括：1. 行動（Action）：具體行動且落實理念；2. 專業（Profession）：省思對話與專業發展；3. 熱情（Passion）：有使命感且積極參與；4. 學習（Learning）：共同合作及團隊學習；5. 效能（Effectiveness）：能提升教與學的效能。在專業學習社群的運作過程中，教師應採取具體有效的行動，主動成立或積極參與專業學習社群，例如：組成或參加讀書會、工作坊、團體研討或發表會議、各學年或學習領域成長團體、學習分享平臺等，並且充分展現熱情與活力，透過專業對話、團隊學習及互助合作，幫助教師反思班級經營與教學歷程，解決教學實務上遇到的難題，以強化教師專業知能，精進教師教學能力，進而提升教師教學效能及學生學習成效。

參考文獻

- 丁一顧（2013）。從專業社群觀點談學校教師會轉型之芻議。臺灣教育評論，2（6），109-112。
- 王巧萍（2011）。臺中市一所國小發展教師專業學習社群之個案研究—以曙光學習社群為例。國立臺中教育大學課程與教學研究所碩士論文，未出版，臺中市。
- 何文純（2006）。國民小學社會資本與學習社群關係之研究。國立臺北教育大學教育政策與管理研究所碩士論文，未出版，臺北市。
- 吳明芬（2013）。雲嘉地區國小教師專業學習社群與教師教學效能關係之研究。國立嘉義大學教育行政與政策發展研究所碩士論文，未出版，嘉義縣。
- 吳明清（2002）。促進教師專業發展的策略。理論與政策，16（1），99-114。
- 吳明清（2005）。教改辛苦為誰忙—評介 Michael Fullan 教育變革的意義。當代教育研究，13（1），265-272。
- 吳清山（2012）。教師專業學習社群與學生學習。教育人力與專業發展，29（1），1-4。
- 吳清山、林天祐（2012）。學習共同體。教育研究，224，111-112。
- 李宜芳（2013）。活化教學的策略與應用—建構 Facebook 班級學習社群。新北市教育，7，62-64。
- 林志成（2013）。教育專業發展的潛在危機與理想藍圖。教育研究，228，27-43。
- 林騰蛟（2013）。活化教學 讓學習更精彩。新北市教育，7，4。
- 張景媛、鄭章華、范德鑫、林靜君（2012）。「教師學習社群」發展對話式形成性評量實務及其對學習成效之影響。教育心理學報，43（3），172-188。
- 張德銳、王淑珍（2010）。教師專業學習社群在教學輔導教師制度中的發展與實踐。臺北市立教育大學學報，41（1），61-90。
- 教育部（2009）。中小學教師專業學習社群手冊。臺北市：教育部。
- 許詩珮（2010）。小學教師專業學習社群發展之研究—以同德國小閱讀教學教師專業學習社群為例。國立新竹教育大學人資處語文教學碩士班論文，未出版，新竹市。
- 許德田（2012）。邁向專業學習社群的學校微革命。臺灣教育，676，10-19。
- 陳佩英（2009）。一起學習、一起領導：專業學習社群的建構與實踐。中等教育，60（3），68-88。
- 葉素枝（2011）。基隆市國民小學教師之同僚關係知覺與專業學習社群參與之研究。淡江大學教育政策與領導研究所碩士在職專班論文，未出版，新北市。
- 潘慧玲、高嘉卿（2012）。臺北市國民小學試辦教師專業發展評鑑之成效分析：理論導向評鑑取徑之應用。教育政策論壇，15（3），133-166。
- 蔡玫芳、黃惠貞（2012）。建構教師專業社群—以板橋高中導師專業發展社群為例。中等教育，63（2），177-185。
- 鄭傑文（2013）。教師學習社群與教師專業發展之相關性研究—以基隆市精進教學計畫學校教師社群成員為例。國立臺灣海洋大學教育研究所碩士論文，未出版，基隆市。

- 鮑慧門 (2013)。知識管理應用於教師專業學習社群之探究。《學校行政》，83，134-155。
- 鍾昫珊 (2013)。屏東縣偏遠地區國小教師專業學習社群、組織承諾與教學效能之研究。國立屏東教育大學教育行政研究所碩士論文，未出版，屏東縣。
- 羅仁駿、蔡慧敏、許惠英、黃寶月 (2010)。臺北市國小體育教師專業社群互動、教學輔導創新與教學效能之關聯性。99年度大專體育學術專刊，233-239。
- Bhattacharya, C. B & Sen, S. (2003). Consumer-company identification: A framework for understanding consumers' relationship with companies. *Journal of Marketing*, 67(2), 76-88.
- Bunker, V. J. (2008). *Professional learning communities, teacher collaboration, and student achievement in an era of standards based reform*. Unpublished doctoral dissertation, Lewis and Clark College, Oregon.
- Dufour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. Bloomington, In National Educational Service.
- Fullan, M. (2006). Leading professional learning. *School Administrator*, 63 (10), 10-14.
- Hord, S. M. (1997). *Professional learning communities: Communities of continuous inquiry and improvement*. Austin, TX: Southwest Educational Development Laboratory.
- Hord, S. M. (2004). Professional learning communities: An overview. In S. M. Hord (Ed.), *Learning together, leading together: Changing school through professional learning communities* (pp. 5-14). NY: Teachers College.
- Hord, S. H. & Hirsh, S. A. (2008). Making the promise a reality. In A. M. Blankstein, P. D. Houston & R. W. Cole (Eds.). *Sustaining professional learning communities* (pp.23-40). Thousand Oaks, CA: Corwin.
- Hipp, K. K., Huffman, J. B., Pankake, A. M., & Olivier, D. F.(2008). Sustaining professional learning communities: Case studies. *Journal of Educational Change*, 9(2), 173-195.
- Huffman, J. B., & Hipp, K. K. (2003). Professional learning community organizer. In J. B. Huffman & K. K. Hipp. (Eds). *Professional learning communities: Initiation to implementation*. Lanham, MD: Scarecrow.
- Leithwood, K. A., Aitken, R., & Jantzi, D. (2001). *Making schools smarter: A system for monitoring school and district progress*. Thousand Oaks, CA: Corwin.
- Munby, H., Russell, T., & Martin, A. K. (2001). Teachers' knowledge and how it develops. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.)(pp. 877-904). New York, NY: American Educational Research Association.
- Ostmeyer, C. J. (2003). *Professional learning community characteristics: A study from a district perspective*. Unpublished doctoral dissertation, Texas A&M University, College Station, Texas.
- Roberts, M. S. & Pruitt, Z. E. (2003). *School as professional learning communities: Collaborative activities and strategies for professional development*. Thousand Oaks, CA: Corwin.
- Stoll, L., & Louis, K. S. (2007). Professional learning communities: Elaborating new approaches. In L. Stoll & K. S. Louis (Eds.), *Professional learning communities: Divergence, depth and dilemmas* (pp. 1-13). Maidenhead, UK: Open University.
- Waters, C. D. (2009). *Case study: Teacher perceptions of professional learning communities in an elementary school setting*. Unpublished doctoral dissertation, Walden University, Minnesota.

論權重與似真值在大型成就評量資料分析的意義、重要性與影響

葛湘璋

致理技術學院國際貿易學系副教授

一、前言

近年來許多國家關心教育現象，政府機構陸續委託專業學術單位對學生、教師及學校單位進行大規模的國內或國際調查。如美國有全國教育進展評量（National Assessment of Educational Progress，簡稱 NAEP），德國有國家長期追蹤調查（National Educational Panel Study，簡稱 NEPS），台灣有台灣教育長期追蹤資料庫（Taiwan Educational Panel Survey，簡稱 NEPS）。著名的國際大型成就評量（large-scale assessment，簡稱 LSA），有國際數學與科學教育成就趨勢調查（Trends in International Mathematics and Science Study，簡稱 TIMSS），全球學生閱讀能力進展研究（Progress in Reading Literacy Study，簡稱 PIRLS），學生基礎素養國際研究計畫（Programme for International Student Assessment，簡稱 PISA），國際公民教育與素養研究（International Civic and Citizenship Education Study，簡稱 ICCS）。而因為 LSA 資料庫包含豐富可貴的資訊，近年來使用 LSA 資料庫的研究越來越盛行。如 TIMSS 的相關論文自 1995 年至 1999 年的 340 篇增為 2005 年至 2009 年的 851 篇（Rutkowski, Gonzalez, Joncas, & von Davier, 2010）。而 TEPS 相關的期刊論文、研討會論文及碩博士論文共 155 篇（中央研究院調查研究專題中心，2012）。

LSA 的主要目的為了解學生的學習成就與家庭背景、學校環境、教師等影響因素的關係，並探討學生的學習特色、學習行為與教育相關議題，以做為教育政策擬定者改進教育參考。而國際 LSA 更進一步進行國際比較或國際教育的交流合作，提供跨國觀摩教育現象的機會，在拓廣教育視野、制定教育決策有其獨特之參考價值。

LSA 一般均採複合調查設計（complex survey design），在資料整理與建立模式比一般資料分析複雜。在 LSA 中，受測學生非隨機取樣，且學生只回答測驗題庫中的部分題目。這種形式的設計在資料分析時需使用抽樣權重（sampling weight），因此，LSA 資料的變異數計算須同時考慮差補變異數（imputation variance）與抽樣變異數（sampling variance）。一般教科書所列之變異數的計算公式僅適用於隨機取樣及學生回答了所有試題的情形，不適用 LSA 的資料（Olson, Martin, & Mullis, 2008; OECD, 2009）。再者，國際 LSA 資料庫在學生的成就以似真值（plausible values, 簡稱 PVs）來估計學生能力，如果直接將 PVs 取其平均或只使用使單個 PV 會低估能力分配的變異，所得的統計量的標準誤有偏差（Willms & Smith, 2005; NCES, 2001）。因此在分析 LSA 時，不

宜將 PVs 取其平均數或只取單個 PV 進行資料分析。由於 LSA 在學生抽樣、試題抽樣、及以 PVs 來估計學生能力的特性，必須採用合適的資料分析方法，才能獲得學生成就的不偏母體估計值。

如前所述，在使用 LSA 資料分析時，須將 PVs 與權重列入考慮，但因計算過程繁雜，須視所使用的軟體是否具使用權重與計算 PVs 合併的功能，因此大部分已發表的論文或研討會論文仍採不正確或簡便的統計方法求平均數、變異數或其他的分析（Carstens & Hastedt, 2010）。研究者常用的錯誤分析方式有僅使用一個 PV 或將 PVs 取平均當依變數，不採用權重或選擇錯誤的權重。本文從實務應用的觀點舉例說明不使用權重對資料的分析影響及可能獲得錯誤的結論，同時，也提出在 LSA 資料分析使用 PVs 的討論及舉例說明變異數的計算方式，以提供使用國際 LSA 資料庫的研究者在次級分析（secondary analysis）時參考。

二、權重

LSA 採複合調查設計，子樣本（subsample）有時未反應母體子群的真正比例。權重主要是考慮到被選取學校/教師/學生的機率並非完全相同，每個受測單位的回應需調整到能反映出學生在母體中被抽中的真正比例，這種調整即是抽樣權重（sampling weight）（Rutkowski et al., 2010），譬如在 LSA 調查設計時，抽樣考慮男女學生的比例需近似相等，然而如果母體中男女學生比例有差異，此時有使用權重及未使用權重的結果會有很大的差異。因此，資料分析需依研究目的，選擇適當的權重類型，以獲得正確結論。

茲舉例說明權重對估計值的影響如下。假設一個班級有 24 個學生，16 位男生，8 個女生。如果要從此班隨機抽取六位學生進行調查以研究學生每週上網時間。依照母體比例，應抽 4 位男生，2 位女生。然而如果抽樣要男女生人數相等（亦即要抽取三位男生，三位女生），同時也要考慮到母體男女生真正比例的情形（母體女生比例小於男生），此時被抽取的學生的比例必須能反應真正的母群體男女生比例，因而在估計時必須要調整，此即所謂的抽樣權重。有採取及未採取抽樣權重的估計值會因子樣本（subsample）的比例不同而不同。以上例而言，女生的權重為 2.667，男生的權重為 5.333。假設此三位女生每週各花 2、3.5、1.5 小時上網，此三位男生每週各花 0.5、1、1.2 小時上網，如果沒有採取權重，則此六位學生的平均數為 $(2+3.5+1.5+0.5+1+1.2)/6=1.617$ 小時，而如果採取抽樣權重，則平均數為 $(2.667*(2+3.5+1.5)+5.333*(0.5+1+1.2))/24=1.378$ ，比未加權平均數少了 14.8%。另舉一例，如果此班有 24 位學生，10 位男生，14 位女生，要從中抽取六位學生，若依母體比例應抽取 2.5 位男生 3.5 位女生。若要抽取相等人數的男女生，即男生 3 位，女生 3 位，此時男女生權重各為 3.333 及 4.667。此時採抽樣權重所得平均數為 $(4.667*(2+3.5+1.5)+3.333*(0.5+1+1.2))/24=1.736$ ，比未加權的平均數高 7.39%。表 1 列出此兩例的比較。由此表可看出，使用權重與未使用權重，子樣本對估計值的影響。未使用權重可能高估或低估母體估計值，使用權重時，樣本數較小的子樣本會比未使用權重時對估計值的影響較大。

表 1 加權與未加權平均數例題比較

| 母體數 | | 應抽樣本數 | | 權重 | | 上網時數 | | | | | | 平均數 | | |
|-----|----|-------|-----|-------|-------|------|-----|-----|-----|-----|-----|-------|-------|--------|
| 男 | 女 | 男 | 女 | 男 | 女 | 男 1 | 男 2 | 男 3 | 女 1 | 女 2 | 女 3 | 未加權 | 加權 | 增減比例 |
| 16 | 8 | 4 | 2 | 5.333 | 2.667 | 0.5 | 1 | 1.2 | 2 | 3.5 | 1.5 | 1.617 | 1.378 | -14.8% |
| 10 | 14 | 2.5 | 3.5 | 3.333 | 4.667 | 0.5 | 1 | 1.2 | 2 | 3.5 | 1.5 | 1.617 | 1.736 | 7.4% |

以 TIMSS 為例，TIMSS 使用二階段抽樣 (Martin, Mullis, & Kennedy, 2007; Olson et al., 2008; Joncas, 2008; OECD, 2009)，即先抽學校再抽班級，各年級抽樣人數有 150 所學校及 5000 位學生。各抽樣單位 (學生 / 班級 / 學校) 並非在相同機率下被抽，抽樣單位需適當加權，以代表母群體中子群的真正比例。TIMSS 採多種權重，分別為學生總數權重 (Total Student Weight)、學生樣本權重 (Student House Weight)、學生等數權重 (Student Senate Weight)、整體及學科教師權重 (Overall and Subject Teacher Weight) 及學校權重 (School Weight)。Rutkowski et al. (2010)、Foy 與 Olson (2008)、Williams, Ferraro, Roey, Brenwald, Kastberg, Jocelyn, Smith 與 Stearns (2009) 在 TIMSS 資料分析時，對權重的選擇提出建議以作為研究者在使用 TIMSS 資料分析時參考。

茲將 TIMSS 權重的使用原則綜合歸納如下，表 2 為 TIMSS 的權重名稱、適用時機及分析層次之綜合整理，其他國際 LSA 資料庫的使用者在資料分析前亦須參考官方技術報告及相關文獻對權重類型與使用方式的說明。

(一) 對樣本數敏感的分析 (如卡方檢定) 採用正規化權重 (normalized weights)，在 TIMSS 稱為 House weight。House Weight 是 Total Student

Weight 的線性轉換，總合等於樣本數。

(二) 在做跨國分析時，合併多國資料後，必須注意到母體人數會影響到結果。例如美國八年級生有三百多萬人，賽浦勒斯八年級生有 9,237，馬爾他的八年級生只有 4,943 人 (Olson et al., 2008)。在做國際數學成就的比較分析時，如果採用 Total Student Weight，結果會受到美國的大母體的影響。TIMSS 提供 Student Senate Weight 作為跨國比較。Student Senate Weight 是 Total Student Weight 的轉換，每個國家的權重學生樣本數 (weighted student sample size) 為 500，因此在跨國比較時，不論母群體的大小，每個國家的貢獻是相等的。在作跨國比較分析時，應採用 Student Senate Weight，以避免母體數差異太大造成研究結果偏誤。

(三) 如果要將教師變數當為學生特質的分析時，適合使用 Overall Teacher Weight。如果是做數學成就分析時，則使用數學教師權重 (Mathematics Teacher Weight)，如果是做科學成就分析時，則使用科學教師權重 (Science Teacher Weight)。但研究者須注意到，TIMSS 的教師並非隨機取樣，而是被抽選到的學生的教師回答教師問卷，因此在分析教師變數時，須採用教師 - 學生資料進行分析。

(四) 如果研究興趣是在學校層次的特質，則使用學校層級的權重。

多層次的統計軟體如 HLM (Raudenbush, Bryk, Cheong, & Congdon, 2004) 及 Mplus (Muthen & Muthen, 1998-2010) 能讓使用者指定各層次的權重並將似真值列入模式參數計算的選項，而 TIMSS 計畫的主持單位國際教育成就評比學會 (International Association for the Evaluation of Educational Achievement,

簡稱 IEA) 發展出 TIMSS 資料的 IEA 國際資料庫分析軟體 (IEA International Database Analyzer Software, 簡稱 IEA IDB Analyzer) 也提供了 TIMSS 研究者資料分析使用權重與 PVs 的選擇。在資料分析時，研究者需參考軟體使用手冊，官方的使用手冊及技術報告，配合研究目的與問題，選擇各層次合宜的權重。

表 2 TIMSS 的權重名稱、適用時機及分析層次

| 權重 | 變數名稱 | 說明 | 適用時機 | 分析層次 |
|----------------------------|--------|--------------------|-------------|---------|
| Total Student Weight | TOTWGT | 加總等於國家受測年級母群體 | 單一國家研究 | 學生 |
| Student Senate Weight | SENWGT | 每個國家權重學生樣本數加總為 500 | 合併多國資料，跨國研究 | 學生 |
| Student House Weight | HOUWGT | 加總為每個國家的學生樣本數 | 樣本數敏感的顯著性檢定 | 學生 |
| School-level Weight | SCHWGT | 學校層級的權重 | 學校 | 學校 |
| Overall Teacher Weight | TCHWGT | 整體教師的權重 | 教師及學生 | 學生 / 教師 |
| Mathematics Teacher Weight | MATWGT | 數學教師權重 | 數學教師及學生 | 學生 / 教師 |
| Science Teacher Weight | SCIWGT | 科學教師權重 | 科學教師及學生 | 學生 / 教師 |

三、似真值

在測驗施測時，測量誤差主要來自於測驗內容題項、學生的身心狀況及測驗情境 (OECD, 2005)。而在大型成就評量時，通常會從題庫中抽取題目編製成不同版本的題本，每個版本的題本雖然題目不同，但所要測量的特質與內涵是相同。由於題庫測驗題數相當多，每個學生只回答其中部分題目，學生的能力無法精確得知，使用不同的題本，有些學生的能力可得到良好估計，有些則否。因此，一般國際 LSA 以 PVs 來估計學生成就而不以一個測驗總分來代表學生的測

驗結果。以 TIMSS 而言，TIMSS 建立試題庫，採用複合輪換題本設計 (complex rotated booklet design) 以確保每題試題都能充分使用及題本試題數足以測量出學生的能力，並以五個 PVs 來估算學生數學成就分數 (Foy & Olson, 2008)。同時 TIMSS 將測驗時間限制於 90 分鐘，在複合輪換題本設計及時間限制下，傳統估計學生能力方式所得的母體參數變異數是有偏誤的 (Mislevy, Beaton, Kaplan, & Sheehan, 1992)。亦即從一個測驗題庫隨機抽取部分測驗題來估計學生的能

力，有其不確定性（uncertainty）存在，估計的準確度會受影響。在成就測驗中，研究者通常希望能得到一個似乎學生完成整個測驗題庫的分數，這與插補遺漏（imputing missing）的觀念相似，因此，就有一個 PV 分數（假設學生完成了所有測驗題目）及其測量誤差（Willms & Smith, 2005）。PVs 所代表的意義可說是學生能力的各種可能值的範圍及其機率，而非學生個人成績，學生個人的 PVs 則從此機率分配中隨機產生（von Davier, Gonzalez, & Mislevy, 2009）。

在 TIMSS 中，學生在有限的時間裡回答了每個領域的部分試題，學生的 PVs 隨機從所估計的能力分配所得，而這能力分配主要從有相似的試題回答方式（item response pattern）及相似背景資料的受試者所估計而得。在使用 PVs 時，同時也必須計算由於測驗缺乏精確性所造成的不確定性，如果測驗是精確的，那麼測驗誤差為零，而每個 PV 分析所得的統計數值應相等。然而，完全精確無誤的測驗幾乎是不可能，因此會有測量變異，通常稱為差補變異數（imputation variance）。LSA 研究中，和 PVs 有關的資料分析須同時考慮到差補變異與抽樣變異，由於計算繁雜，一般研究者在使用 PVs 分析資料時，通常會採取兩種簡便的方式，一種為只使用一個 PV 作分析，因為和力量測有關的不確定性被忽略（如抽樣誤差），即標準誤的估計值沒有考慮抽樣變異，因此這種分析方式所得的統計量的標準誤會被低估。第二種方式是取 PVs 的平均值作分析，這種採用平均 PVs 的方式比只採用一個 PV 的分析方式更嚴重的低估了標準誤（von Davier et al.,

2009），這兩種分析方式均會使推論不正確。

因此，在和測驗有關的資料分析中，若使用到 PVs，必須對每個變數依公式估計其參數的變異。例如，在迴歸分析時，因 TIMSS 有五個 PVs，因此必須對每個 PV 作個別分析，重複五次，然後再依公式求其合併後之係數與標準誤（Willms & Smith, 2005; Foy & Olson, 2008）。即若為母體統計數值，是第 i 個（ $i=1, \dots, k$ ）PV 的分析所得的統計數值，則

$$\theta = \frac{1}{k} \sum_{i=1}^k \theta_i, \quad i=1,2,3, \dots, k, \quad (1)$$

設 B_k 為差補變異數

$$B_k = \frac{1}{k-1} \sum_{i=1}^k (\theta_i - \theta)^2, \quad (2)$$

抽樣變異數（sampling variance）則為

$$U = \frac{1}{k} \sum_{i=1}^k u_i, \quad (3)$$

u_i 為第 i 個 PV 分析的誤差變異數（error variance），再將抽樣變異數與插補變異數相加，得合併後得係數之變異數，即

$$V = U + (1 + \frac{1}{k})B_k, \quad (4)$$

例如，TIMSS 有五個 PVs，因此

$$V = U + 1.2B_k \quad (5)$$

茲舉例說明如下：假設在以 TIMSS 資料作迴歸分析，以 PV 為依變數，截距、變數一、變數二之迴歸係數分別為 g_{00} 、 g_{10} 、 g_{20} 。對每個 PV 作個別分析，重複作五次。再依公式計算每個係數的 θ 、 B_k 、 U 、 V ，再以合併後的係數及標準誤計算最後的 t_i 值為 $\frac{\theta_i}{\sqrt{V_i}}$ 。表 3 為五次迴歸分析之迴歸係數及標準誤、及 θ 、 B_k 、 U 、 V 、 t 值的應用實例。

表 3 五次迴歸分析之迴歸係數及標準誤、及 θ 、 B_k 、 U 、 V 、 t 值

| | g00 | SE(g00) | g10 | SE(g10) | g20 | SE(g20) |
|----------|--------|---------|------|---------|-------|---------|
| pv1 | 531.12 | 6.88 | 5.44 | 2.73 | 18.30 | 1.57 |
| pv2 | 535.47 | 7.11 | 2.47 | 2.79 | 18.09 | 1.56 |
| pv3 | 530.21 | 7.04 | 5.12 | 3.03 | 19.06 | 1.60 |
| pv4 | 533.16 | 7.07 | 5.44 | 2.84 | 18.44 | 1.63 |
| pv5 | 533.28 | 7.35 | 3.90 | 2.94 | 17.95 | 1.65 |
| θ | 532.65 | 7.09 | 4.47 | 2.87 | 18.37 | 1.60 |
| B_k | | 50.30 | | 8.24 | | 2.58 |
| U | | 4.23 | | 1.66 | | 0.19 |
| V | | 55.37 | | 10.24 | | 2.80 |
| t | 71.58 | | 1.40 | | 10.98 | |

四、結論與討論

LSA 計劃的專責單位在完成調查後釋出豐富的資料供教育專家學者及研究者進行二次分析。然而由於 LSA 的調查設計不同於一般隨機取樣調查，在學生抽樣、試題抽樣及學生能力估算的複雜性，使用一般統計公式及分析技巧，所得的結果通常是有偏誤。本研究舉例說明使用 / 不使用權重在估計值的差異及整合 PVs 標準誤的計算。同時，也指出在 LSA 調查設計，由於學生抽樣及試題抽樣的複雜性，學生在有限的測驗時間及試題限制下，如只選用一個 PV，因為沒有考慮到抽樣的誤差，因此標準誤將被低估且能力分配測量（measurement of proficiency distribution）的不確定性將被忽略。將 PVs 取平均比只用單個 PV 更低估了標準誤，因此，在有 PVs 的 LSA 資料庫中，最不宜將 PVs 取平均做二次分析。單個 PV 及 PVs 的平均適用於每個學生均回答了所有試題的情形，不適用於如 TIMSS 採用複合輪換題本設計，亦即 LSA 資料分析時，PVs 的處理須同時考慮差補變異與抽樣變異，以獲得不偏的母體估計值。使用綜合 PV（每個 PV 作個別分析，然

後再依公式求其合併後之係數與標準誤）比使用 PV 取平均或僅使用單個 PV 有較大的標準誤，小標準誤可能導致不同的結果。在 Carstens 與 Hastedt（2010）的研究中以 TIMSS 資料分析也發現，使用單個 PV 或使用 5 個 PVs 的平均來做國家排比時，排比順序會改變，使用不適當的分析方法會使原來未具顯著差異變為顯著不同。研究者若將 PVs 當依變數，可考慮資料分析時以單個 PV 做初步的探索性分析，最後仍應以對每個 PV 作個別分析，然後以合併後之係數與標準誤作最後結論。

LSA 提供了豐富的教育資訊，一個 LSA 資料庫的產生，通常是教育相關領域的政府單位、專家學者、學術單位，實務工作者及研究團隊投注多年心力、時間與資源共同努力的結果，譬如 TEPS 從 2001 年開始至今已執行近 12 年，TIMSS 從 1995 年執行至今近 18 年，TIMSS 資料庫長久以來，提供國內 / 國際教育專家學者研究國內 / 國際學生數學及科學成就及相關教育議題的資訊，是數學及科學教育研究的可貴來源，TIMSS

2011 資料亦於今年（2013）二月釋出，這樣的資料庫是研究者探討教育相關議題的次級分析的可貴來源。然而 LSA 在試題設計、抽樣設計、能力分數的估算不同於一般橫斷資料，因此不宜以一般的統計技巧做資料分析。研究者在使用 LSA 資料須了解資料庫變數的特性外，更應參考官方使用手冊如抽樣、權重、學生

能力估算及資料分析應注意事項等議題，配合研究目的，選用正確的分析方式，才能獲得正確的母體估計值與做出正確的結論。本研究提出次級分析時須注意的議題，期能使 LSA 研究者瞭解權重及 PVs 在 LSA 資料分析的重要及錯誤的分析方式可能導致錯誤的結論，提供研究者在進行次級資料分析時參考。

參考文獻

- 中央研究院調查研究專題中心（2012）。台灣教育長期追蹤資料庫資料使用工作坊。台灣教育長期追蹤資料庫資料使用工作坊會議資料，中央研究院調查研究專題中心。
- Carstens, R. & Hastedt, D. (2010). *The effect of not using plausible values when they should be: An illustration using TIMSS 2007 grade 8 mathematics data*. Retrieved August 10, 2011, from http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2010/Papers/IRC2010_Carstens_Hastedt.pdf
- Foy, P., & Olson, J. F. (Eds.) (2008). *TIMSS 2007 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Joncas, J. (2008). TIMSS 2007 Sampling weights and participation rates. Chapter 9 in J. F. Olson, M. O. Martin, and I. V. S. Mullis (Eds.), *TIMSS 2007 technical report* (pp.153-192). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M., Mullis, I., & Kennedy, A. (Eds.). (2007). *PIRLS 2006 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mislevy, R., Beaton, A., Kaplan, B., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Education Measurement*, 29, 133-161.
- Muthen, L., & Muthen, B. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthen & Muthen.
- NCES(2001). *User's guide for the Third International Mathematics and Science Study(TIMSS)*. NCES 2001-065 by B. Chaney, L. Jocelyn, D. Levine, T. Mule, L. Rizzo, S. Roey, K. Rust. & T. Willians Washington, DC: Government Printing Office.
- Olson, J., Martin, M., & Mullis, I. (Eds.). (2008). *TIMSS 2007 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Organisation for Economic Co-operation and Development. (2005). *PISA 2003 data analysis manual*. Retrieved May 8, 2010 from www.pisa.oecd.org/dataoecd/53/22/35014883.pdf
- Organisation for Economic Co-operation and Development. (2009). *PISA 2006 technical report*. Paris: Author.
- Raudenbush, S., Bryk, A., Cheong, Y., & Congdon, R. (2004). *HLM 6 [Manual]*. Lincolnwood, IL: Scientific Software International.
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, 39, 142-151. Retrieved August 10, 2010, from <http://edr.sagepub.com/content/39/2/142>.
- von Davier, M., Gonzalez, E., & Mislevy, R. (2009). Plausible values: What are they and why do we need them?

IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, 2, 9–36.

Williams, T., Ferraro, D., Roey, S., Brenwald, S., Kastberg, D., Jocelyn, L., Smith, C., & Stearns, P. (2009). *TIMSS 2007 U.S. technical report and user guide (NCES 2009-012)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Willms, J., & Smith, T.(2005) . *A manual for conducting analysis with data from TIMSS and PISA*. from www.unb.ca/crisp/pdf/manual_TIMSS_PISA_2005-0503.pdf

研習訊息 (103年6月至7月)

| 序號 | 期別 | 院區 | 研習班別 | 日期 | 承辦人 |
|----|------|-------|-------------------------------|-----------------|------------|
| 1 | 141 | 三峽總院區 | 103年國小主任儲訓班 A1 | 103.06.09-07.11 | 林巧涵 |
| 2 | 141 | 三峽總院區 | 103年國小主任儲訓班 A2 | 103.06.09-07.11 | 王美芸 張瑜芬 |
| 3 | 141 | 三峽總院區 | 103年國小主任儲訓班 B1 | 103.06.09-07.11 | 翁秀惠 |
| 4 | 141 | 三峽總院區 | 103年國小主任儲訓班 B2 | 103.06.09-07.11 | 吳翊菁 許茹蘭 |
| 5 | | 三峽總院區 | 辭典編輯培訓工作坊 | 103.06.30-07.04 | 林韋均 |
| 6 | | 三峽總院區 | 國民中學跨領域閱讀教學研習 | 103.07.07 | 黃玉萍 萬寧 |
| 7 | 3033 | 三峽總院區 | 事務管理研習班 | 103.07.21-07.22 | 吳翊菁 王美芸 |
| 8 | 3035 | 三峽總院區 | 不動產管理研習班 | 103.07.24-07.25 | 王美芸 吳翊菁 |
| 9 | 3036 | 三峽總院區 | 中小學校長在職專業研習 - 國小 7 | 103.07.22-07.25 | 張瑜芬 |
| 10 | 3037 | 三峽總院區 | 幼兒園園長專業研習班 -1 | 103.07.23-07.25 | 吳翊菁 王美芸 |
| 11 | 3038 | 三峽總院區 | 103學年度中央課程與教學輔導諮詢 教師團隊期初會議 | 103.07.28-08.01 | 張瑜芬 林巧涵 |
| 12 | 3039 | 三峽總院區 | 境外學校性別平等教育研習班 | 103.07.28-08.01 | 許茹蘭 |
| 13 | 3040 | 三峽總院區 | 幼兒園園長專業研習班 -2 | 103.07.28-07.30 | 吳翊菁 王美芸 |
| 14 | | 三峽總院區 | 溫世仁文教基金會 -1 | 103.07.28-08.02 | |

| 序號 | 期別 | 院區 | 研習班別 | 日期 | 承辦人 |
|----|------|------|---------------------|-----------------|-----|
| 15 | | 臺中院區 | 閩南語語言能力認證命題研習班 | 103.06.05-06.07 | 郭益豪 |
| 16 | 3521 | 臺中院區 | 海洋教育 | 103.06.09-06.10 | 林例怡 |
| 17 | 3519 | 臺中院區 | 與媒體互動研習班 | 103.06.18-06.19 | 郭益豪 |
| 18 | 3503 | 臺中院區 | 行政中立研習班 | 103.06.23-06.24 | 林例怡 |
| 19 | | 臺中院區 | 國民中學教師閱讀教學增能計畫教師研習 | 103.07.07-07.08 | 吳于嫻 |
| 20 | 3522 | 臺中院區 | 全國高級中等學校誠信研習營第一梯 | 103.07.09-07.11 | 羅彩紅 |
| 21 | 3531 | 臺中院區 | 12年國民基本教育境外學校教學研習班 | 103.07.14-07.18 | 郭益豪 |
| 22 | 3703 | 臺中院區 | MOOCs 課程研擬研習營 | 103.07.14-07.16 | 吳于嫻 |
| 23 | 3512 | 臺中院區 | 中階人員管理發展訓練研習營 | 103.07.24-07.25 | 林例怡 |
| 24 | 3533 | 臺中院區 | 103年度新進學校護理人員職前訓練研習 | 103.07.26-07.27 | 羅彩紅 |
| 25 | 3523 | 臺中院區 | 103年度新進學校護理人員職前訓練研習 | 103.07.28-08.01 | 羅彩紅 |

院務花絮



103.04.16 本院辦理 103 年度國民中學校長儲訓班，蔣部長偉寧親臨勉勵並與師傅校長及本中心洪啟昌主任合影。



103.04.17 本院辦理 103 年度國民中小學校長儲訓班結業典禮，本院柯院長華葳感謝師傅校長的辛勞及教導。



103.04.17 本院辦理 103 年度國民中小學校長儲訓班結業典禮，學員表演節目：歡慶豐收 (擊鼓 - 節令鼓)。



103.04.23 「第 3024 期國民小學校長在職專業研習班」，學員結業式後與本中心洪啟昌主任合影。



103.05.07「第 3031 期國民小學校長在職專業研習班」，圓桌對話 - 重大學校行政議題探討：課程與教學領導。



103.05.08「第 3031 期國民小學校長在職專業研習班」，文化參訪課程 - 安排新北市李梅樹紀念館導覽。



103.05.09「第3031期國民小學校長在職專業研習班」結業式，本院潘文忠副院長、本中心洪啟昌主任與全體學員合影。



103.05.14「第3034期國民中學校長在職專業研習班」，圓桌對話-重大學校行政議題探討：教師工會互動實務。



103.05.15「第3034期國民中學校長在職專業研習班」，文化參訪課程 - 捏陶體驗課程。



103.05.15「第3034期國民中學校長在職專業研習班」結業式，本院潘文忠副院長與全體學員合影。



103.05.16 「第 3034 期國民中學校長在職專業研習班」結業式，本院潘文忠副院長頒發研習證書予學員代表。



103.05.19 本院人事室辦理「103 年度教育研習」，邀請本院曾世杰副院長專題演講，講題「故事會說話：您自己怎樣看男男女女的事」。



103.05.29 本院人事室辦理「本院員工自我成長研習營 - 感動服務、效能提升」，課程之一：「團體動力學」。



103.05.29 本院人事室辦理「本院員工自我成長研習營 - 感動服務、效能提升」，邀請新北市龍埔國小林瑞昌校長授課：「分享設校過程中的幾個故事」。

教育人力與專業發展

educators and professional development

雙月刊

1984年3月31日 創刊

2014年6月15日 出刊

發行人 | 柯華蕙

總編輯 | 潘文忠

編輯委員 | 丁一顧 臺北市立大學總務長
李俊仁 國家教育研究院測驗及評量研究中心主任
李麗玲 國家教育研究院教科書發展中心助理研究員
何思謎 國家教育研究院教科書發展中心副研究員
林信志 國家教育研究院教育制度及政策研究中心助理研究員
范信賢 國家教育研究院課程及教學研究中心主任
孫志麟 國立臺北教育大學教育經營與管理學系教授
秦葆琦 國家教育研究院課程及教學研究中心副研究員
張鈿富 淡江大學教育學院院長
郭昭佑 國立政治大學附屬高級中學校長
蔡進雄 國家教育研究院教育制度及政策研究中心主任
賴協志 國家教育研究院教育制度及政策研究中心助理研究員
謝名娟 國家教育研究院測驗及評量研究中心副研究員

執行編輯 | 洪啓昌

輪值主編 | 謝進昌、謝名娟

發行機構 | 國家教育研究院
地址 | 23703 新北市三峽區三樹路2號
電話 | (02) 8671-1151
傳真 | (02) 8671-1146

期刊網址 | <http://study.naer.edu.tw>
GPN | 4810100028
ISSN | 2306-5966