

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。

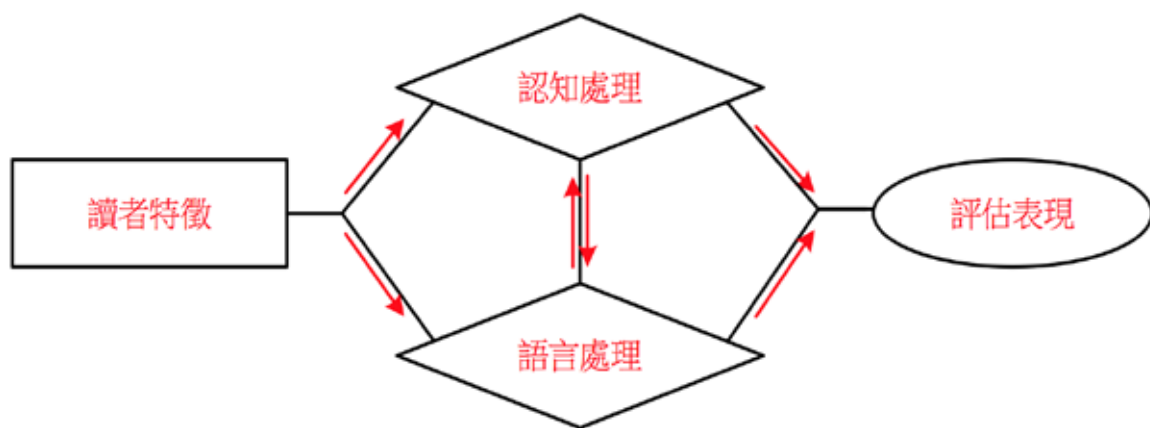


圖 1 閱讀理解的過程 (FCRR, 2011)

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。

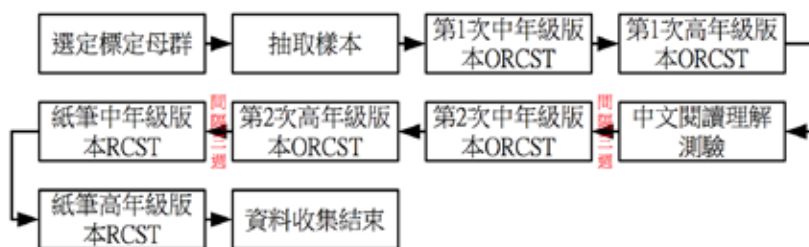


圖 2 研究流程圖

(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、銓寶香（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。

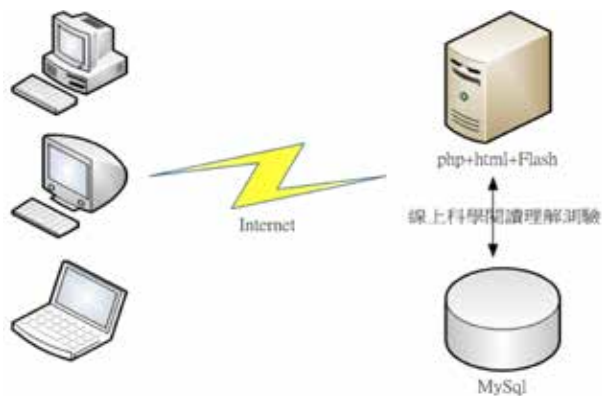


圖 3 線上科學語言閱讀理解測驗系統架構圖

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。

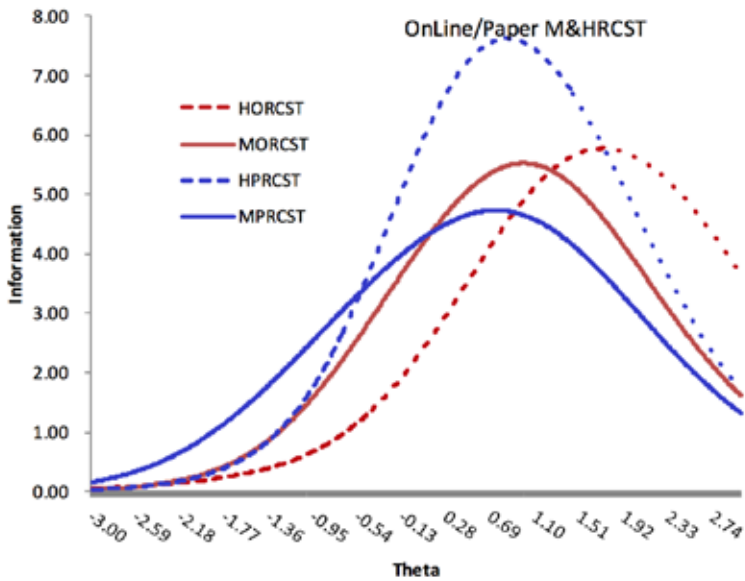


圖 11 科學語言閱讀理解線上與紙筆測驗訊息函數比較圖

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.). Opre 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 sientific literacy from viewpoints of policy, curriculum reform, and evidence-based practice*. Paper presented at the International Coference of Promoting Scientific Literacy from Viewpoints of policy, Curriculum and Practice, Taipei, Taiwan.