

青少年線上閱讀素養評量工具之發展

張貴琳*

本研究兼顧閱讀素養和資訊與通訊科技媒介特色，針對線上閱讀素養的可能內涵，發展兩個青少年線上閱讀素養評量工具，並針對可行性樣本，提出初步效度的檢驗，同時描述樣本的表現特徵。研究者依據 PISA 電子閱讀評量架構，並參酌 Leu 等人（2004）線上閱讀理解之內涵，進行評量工具的發展，並分別針對 601 位八年級學生和 618 位十年級學生進行施測。使用 Rasch IRT 模式進行題目參數與表現估計。結果顯示，兩個評量工具有良好的試題品質，難度適中，鑑別度良好，適合青少年學生。在內容效度、實質效度、結構效度、解釋性等層面有良好的效度證據支持。線上閱讀素養包含 ICT 成份，有別於離線閱讀。中學生的線上閱讀素養表現存在顯著的性別差異。

關鍵詞：資訊與通訊科技、線上閱讀素養、Rasch IRT 模式

* 張貴琳：國立台灣師範大學華語文與科技中心博士後研究員
happylin0432@yahoo.com.tw

The Development of Adolescent Online Reading Literacy Assessments

Kuei-Lin Chang*

By adopting the constructs of reading literacy and with the use of information and communication technology, the purpose of this study is to develop two adolescent online reading literacy assessments (ORLA) and examine the preliminary validity of ORLA. In addition, samples of online reading literacy performance are described. The design of ORLA is based on PISA electronic reading assessment framework and Leu et al. (2004)'s online reading comprehension definition. A total of 601 eighth graders from junior high schools and 618 tenth graders from senior high schools in Taiwan City participated in this study. This study uses Rasch IRT model to calibrate the item parameter and scale scores. The results indicated: the online reading literacy assessments had adequate difficulty level, reasonable validity evidences, and students' online reading literacy performance contains ICT element which is different from offline reading; the gender effect exists in the online reading environment by grades.

Keywords: *information and communication technology, online reading literacy, Rasch IRT model*

* Kuei-Lin Chang: Postdoctoral Fellow, Center of Learning Technology for Chinese, National Taiwan Normal University

青少年線上閱讀素養評量工具之發展

張貴琳

壹、緒論

閱讀與閱讀素養（reading literacy）的內涵隨社會、經濟與文化的改變而轉化，學習的定義，特別是終身學習（lifelong learning）的理念更實質擴展了閱讀素養的社會功能層面，素養不再只是孩童早年在學校所習得的能力，而是知識、技能及策略的實際運作效能，個人透過不同生活情境、與同儕互動、和參與社群建立素養。國際學生能力評量計畫（Programme for International Student Assessment, PISA）對閱讀素養的定義為個體理解、運用、省思和投入書面文本（written text），以達成個人目標、發展個人知識和潛能、有效參與社會的能力（Organization for Economic Cooperation and Development [OECD], 2010a）。

數位時代閱讀素養的運作並不單指從靜態紙本資料獲取知識，而是指更廣義的心智模式和能力建立，強調適應網路新科技變化的學習能力（Leu, 2007）。Leu、Kinzer、Coiro 和 Cammack（2004）、Coiro 與 Dobler（2007）從認知、社會和語言的新素養（new literacy）觀點探討學生的閱讀素養發展，認為其受新科技影響而產生變化，因此線上閱讀素養（online reading literacy）是很重要的探討議題。而目前有關網際網路或其他資訊與通訊科技（information and communication technology, ICT）ICT 工具對閱讀理解本質影響的探討纔剛起步，缺乏明晰的理論架構及實徵證據，亟需研究投入（Coiro, Knobel, Lankshear, & Leu, 2007；Leu, Zawilinski, et al., 2008）。

線上閱讀素養側重問題導向的探究歷程或學習作業，通常指個體能使用網際網路或其他 ICT 工具以確認重要的問題（identify important question）、找尋資訊（locate information）、分析資訊的有用性（analyze the usefulness of information）、統整多元文本資訊（integrate multiple-text information），及溝通資訊（Cheung & Sit, 2008；Leu, Coiro, et al., 2008；Leu, Zawilinski, et al., 2008）。國際閱讀協會（International Reading Association, IRA）提出拓展閱讀的定義，使其包含有效的資訊擷取、資訊的評鑑、使用與溝通資訊等實質內涵（IRA, 2002）。

IRA 認為線上閱讀理解有其獨特本質，因此提倡評量工具的發展，並指出評量內涵必須包含資訊與通訊科技（information and communication technology, ICT）的運用（IRA, 2002；Leu, 2007）。現階段 PISA 亦初步於 22 個國家測試電子閱讀評量（electronic reading assessment, ERA），並從事後續研發工作（OECD, 2007, 2010a）；再者，許多閱讀課程標準與評量尚未包括線上閱讀理解，PISA 國際成人能力評量計畫（Programme for International Assessment of Adult Competencies, PIAAC）才剛開始嘗試；國家教育進展評量（National Assessment of Educational Progress, NAEP）的閱讀評量預計 2019 年才會加入線上閱讀理解和策略（Leu et al., 2009）。由此可知，線上閱讀素養評量的發展成為現階段科技融入閱讀的主要研究議題之一。

性別差異是臺灣學生閱讀表現探討的重要議題之一。在國際閱讀素養進展研究（Progress in International Reading Literacy Survey, PIRLS）的整體閱讀表現中，女生的表現皆略優於男生（柯華葳、詹益綾、張建妤、游婷雅，2008；Mullis, Martin, Kennedy, & Foy, 2007）；PISA 2006 閱讀素養調查的資料也顯示，臺灣女生的閱讀素養表現優於男生（林煥祥、劉聖忠、林素微、李暉，2008）；PISA 2009 閱讀素養調查結果指出，男女學生的閱讀素養表現差異更是明顯（OECD, 2010b）。2009 年 PISA 數位閱讀評量（digital reading assessment）結果亦指出，各國女生的表現優於男生，在不同閱讀歷程（aspect）也有差別（OECD, 2011）。

由於數位化紀元的青少年經常以電子化媒介作為閱讀的途徑，而臺灣地區目前缺乏適當的青少年線上閱讀素養評量工具，現有的評量設計與探討議題多著重於資訊檢索與搜尋層面，很少有研究將網際網路的使用視為一項閱讀議題進行討論，了解學生在開放網路資訊空間的閱讀理解特徵。因此，本研究欲兼顧閱讀素養和資訊與通訊科技媒介特色，針對線上閱讀素養的可能內涵，以青少年為對象，嘗試發展線上閱讀素養評量工具，並依 Messick（1989）心理與教育測驗的效度驗證標準，檢驗線上閱讀素養評量的初步效度，同時描述性別差異。其中，青少年以國中階段和高中階段學生為樣本，各發展一個評量工具。

貳、文獻探討

本研究線上閱讀素養評量工具主要是依據新素養、PISA 閱讀素養、Leu 等人

(2004) 提出的線上閱讀素養之理念和內涵，及 PISA 電子閱讀評量的測量取向和特色進行建構。評量系統的開發初步著重在測驗設計層面，包含線上文本的選用與評量試題的編擬兩部分，採半開放式的網際網路環境，有別於只是將紙本材料 (print) 放置於電腦上的線性閱讀文本。以下將針對這些概念進行介紹。首先，從新素養理論出發，提及線上閱讀素養的定義，並對比線上閱讀與離線閱讀 (offline reading) 的異同；其次，論述電子閱讀評量的測量取向，及評量試題或作業 (tasks) 的特色；最後，說明目前各國男女學生電子閱讀表現的調查結果，以作為性別差異檢驗的文獻基礎。

一、線上閱讀素養

線上閱讀素養立基於新素養理論。不同學者曾對新素養提出不同的定義。有人認為它來自社會行為研究，有人認為它強調新論述 (discourse) 或新科技的取得，也有人認為它是讀者對新科技所創造符號和文化脈絡的覺察，其定義將隨科技進展不斷更新 (Asselin & Moayeri, 2010)。新素養的同義詞很多，如網路素養 (Internet literacy)、數位素養 (digital literacy)、新媒體素養 (media literacy)、資訊與通訊科技素養 (ICT literacy)、電腦素養 (computer literacy) 等，全球有將近六分之一的人使用網際網路，使用率持續呈現指數成長，預計 2012 年網際網路將成為無所不在 (ubiquitous) 的科技工具，也就是 21 世紀資訊、溝通與閱讀理解的主要媒介工具 (Coiro et al., 2007; Leu, Coiro, et al., 2008)。ICT 的進展使新素養構念持續更異，內涵成分目前還不夠清晰穩定，統整目前文獻新素養的內涵包括 ICT、心智模式、文化互動、及社會責任。新素養是新科技環境中處理事務的潛能，也是政治、經濟及全球社群中個人參與能力的核心，本質上是複雜而統整的多型態 (multi-modality) 和多面向運作 (Leu, Coiro, et al., 2008; Leu, Zawilinski, et al., 2008)。由上可知，ICT 素養是線上閱讀素養的核心面向。

(一) 線上閱讀素養的定義

根據傳統上對素養 (literacy) 一詞的定義，素養意指讀寫能力，是用以獲取資訊與溝通資訊的工具，當中包含閱讀與溝通面向。閱讀素養包括一組廣泛的認知能力，從簡單的解碼、字彙知識、文法與組織結構，到對世界的認知，偏重目的性和功能性。PISA 2009 將閱讀素養定義為個體能理解、運用、省思和投入書面文本資訊，以達成個人目標、發展個人知識和潛能，並有效參與社會；其中，書面文本包含數位化文本

(洪碧霞, 2008; OECD, 2007, 2010a)。ICT 素養是線上閱讀素養的必要成份，但線上閱讀素養並不全是 ICT 素養，線上閱讀素養中還包含重要的閱讀成份，範圍比 ICT 素養還廣泛。線上閱讀的同義詞包括數位閱讀、電子閱讀、超文本閱讀 (hypertext reading) 等，Leu 等人 (2004) 指出，線上閱讀素養側重問題導向的探究歷程或學習作業，通常指個體能使用網際網路或其他 ICT 工具以確認重要的問題、找尋資訊、分析資訊的有用性、統整多元文本資訊及溝通資訊。換句話說，線上閱讀素養是個體所具備有效使用快速變化溝通科技以解決個人和職涯問題的技巧、策略及氣質 (Cheung & Sit, 2008; Leu et al., 2004; Leu, Coiro, et al., 2008; Leu, Zawilinski, et al., 2008)。PISA 2009 ERA 依據 PISA 閱讀素養的定義與架構，及 Leu 等人 (2004) 的線上閱讀理解歷程，提出電子閱讀的內涵，包含線上資訊的使用擷取 (assess and retrieve)、統整解釋 (integrate and interpret)、反思與評鑑 (reflect and evaluate) 歷程 (Cheung & Sit, 2008; OECD, 2010b)。這些歷程在線上閱讀與離線閱讀中可能是不同的型態。

(二) 線上與離線閱讀的差異

Leu (2007) 採用線上與離線閱讀進行構念探討，主要基於下列四項理由：(1)由閱讀而言，目的在於資訊的理解。(2)網路是目前最大的資訊來源。(3)線上與離線兩詞互斥，對照較為準確。(4)目前部份證據顯示學生線上與離線閱讀表現有所差異。換言之，線上與離線閱讀的區別是必要的，假設兩人面對相同的問題，解決問題時不太可能閱讀相同的資料，而且精熟的紙本讀者線上閱讀效能未必良好。

就文本形式而言，兩者最大差異在於閱讀範圍邊界的固定 (bounded) 或動態 (dynamic) 本質。如圖 1 所示，線上閱讀包含航網工具 (navigation tool) 和多元文本運用，閱讀歷程通常是非線性的。航網工具的形式眾多，包含捲軸、按鈕、表單等，以超連結的使用最多。廣泛的超文本涵蓋航網工具在內，具有多樣化的文本型態，如部落格的評論交流、電子郵件的往返、多媒體超連結等互動式文本。讀者需自行擇取文本，而每個人建構的文本路徑都不相同 (Leu, 2007; Leu, Coiro, et al., 2008)。精熟的線上讀者必須熟悉航網工具的使用，同時建構多重頁面的心理表徵路徑，透過循環、統整的動態歷程進行線上資訊的搜尋與閱讀 (Cheung & Sit, 2008; Henry, 2005; OECD, 2010a, 2011)。

就策略層面而言，線上閱讀除了內涵包含電腦化閱讀文本豐富、動態的特質外，還需要搜尋引擎和網頁連結往返等 ICT 技能的應用。線上閱讀的主要目的多為解決資

訊性的問題，面對的是資訊無限的空間，需要使用多媒體導航工具。閱讀歷程中持續尋找、批判、評鑑資訊的反覆歷程尤其重要。讀者首先使用搜尋引擎找尋資訊、從諸多搜尋結果中評鑑資訊的參考價值、運用網頁連結獲取相關訊息、持續的往返關聯網頁以尋求問題的答案並對照檢核資訊的精確性，這些都是明顯不同於離線閱讀的嶄新閱讀策略（Leu, Coiro, et al., 2008；Leu, Zawilinski, et al., 2008）。

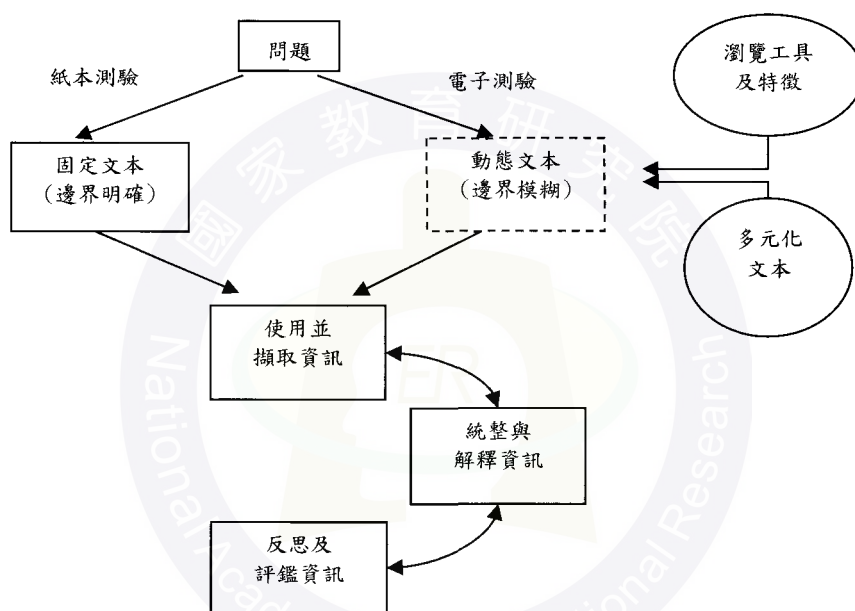


圖 1 PISA 紙本與 ERA 概念之關係（Cheung & Sit, 2008）

就閱讀歷程而言，紙本與電子閱讀的擷取與檢索訊息歷程均涉及尋找、選擇和收集資訊。在某些情境中，讀者會尋找特定的文本資訊，有時則需要同時擷取多個資訊。紙本閱讀是在固定的資訊空間找尋資訊，電子閱讀是在開放的資訊空間找尋資訊，需要使用目錄和頁碼等搜尋工具。紙本閱讀的統整與解釋訊息歷程涉及文本內部的統整，讀者需了解文本各部分的關係、找出適切的關係、確認需要推論的關係等，文本結構的基模固定；電子閱讀也涉及這些內涵，但由於數位化文本的結構沒有固定的基模，跨不同網頁進行統整更形複雜。紙本閱讀的反思與評鑑訊息歷程涉及利用文本外的知識、想法和價值，連結文本與個人的知識或經驗，讀者需從文本抽離，進行客觀

思考，以評鑑文本的品質和適切性，文本風格通常扮演重要角色；電子閱讀由於資訊來源的變化性高，故較著重精確性和可靠性的判斷。簡言之，讀者在進行電子閱讀時，需要有效使用航網工具瀏覽大量的文本資訊，並立即評鑑其真實性、可信度和關連性，這是紙本與電子閱讀最不同的地方（OECD, 2010a, 2011）。為突顯出線上閱讀的特性，使其有別於離線閱讀或電腦化線性文本閱讀，本研究現階段先參酌電子閱讀評量的系統環境進行模擬，使評量工具較貼近實際的情況。

（三）電子閱讀評量

1.測量取向

線上閱讀依據數位化環境（digital environment）的真實性分為兩種不同的測量取向（measurement approach）：(1)在真實的無限網路空間中進行質性研究，採用認知歷程分析的方法收集學生的線上閱讀理解策略和行為組型；(2)在虛擬的有限網路空間中收集學生的線上閱讀表現，採半開放式的網際網路環境，不以關鍵字搜尋（search）作為起點，而是模擬一個線上閱讀或問題解決情境，由受試者根據問題，搜尋（locate）和瀏覽（browse）解題的相關網頁，惟其無法充分代表線上閱讀的真實複雜度和挑戰性（Leu, 2007）。PISA 2009 電子閱讀評量為國際首次的大規模數位化閱讀評量，使用 USB 隨身碟存取測驗題本和作答資料（內含瀏覽和輔助程式軟體），數位化環境係模擬真實的線上閱讀情境，文本內容包含線上常用的資源，有專題或搜尋網頁、電郵通訊、網誌、線上交易或其他互動平台，實施時間為 40 分鐘（香港中文大學，2011；Ho, 2011），屬於第二種取向。本研究線上閱讀素養評量依此進行模擬。

2.試題（作業）特色

電子閱讀評量試題或作業一般會要求學生辨識重要的問題，尋找與這項閱讀作業所預設之提取路徑（access paths）對應的訊息，其內涵同時涵蓋文本處理歷程（text processing）與航網歷程（navigation）兩個成份（Mendelovits, Lumley, & Searle, 2009；Sit & Cheung, 2009）。PISA 將航網歷程定義為讀者選擇與安排超文本資訊的過程，由於線上閱讀時，讀者是利用航網工具建構自己的閱讀文本，故航網歷程也是認知歷程的一部分。評量試題或作業依不同的成份比重，可分為四種類型：低文本處理歷程和低航網歷程、高文本處理歷程和低航網歷程、低文本處理歷程和高航網歷程、高文本處理歷程和高航網歷程（OECD, 2010a, 2011），如圖 2。本評量在試題設計上嘗試這樣的分類。

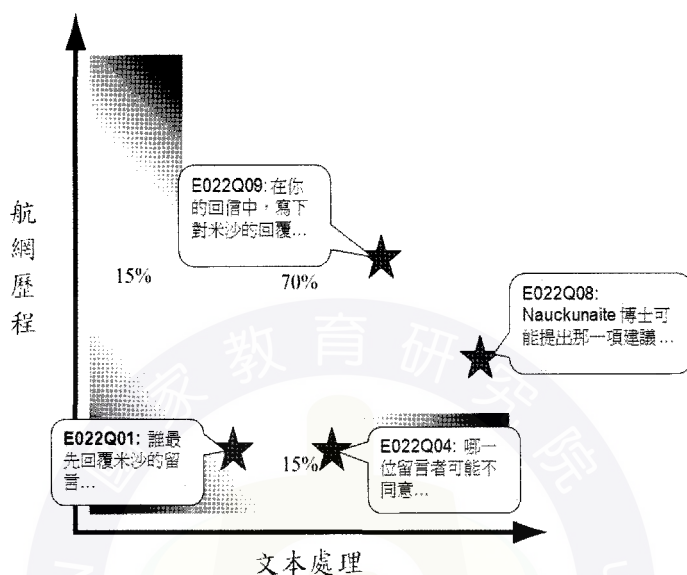


圖 2 試題中航網歷程與文本處理歷程的關係 (OECD, 2010a)

電子閱讀評量試題或作業的難度受到四個變項所影響：(1)文本特徵 (characteristics of text)：包括讀者對文本內容的熟悉度、文本結構的複雜性、文本的資訊量等；(2)航網歷程的複雜度 (complexity of navigation)：包含所需網頁資訊的可見度、讀者對航網工具的熟悉度、航網工具屬性的一致性、網頁結構的配置等；一般而言，航網工具的操作步驟數愈多、形式愈複雜，題目就愈困難；(3)作業要求的明確度 (explicitness of task demands)：例如，題目敘述與文本是否使用相同的用語、是否有清晰的說明等；(4)反應的本質 (nature of response)：意指完成試題所需的心理歷程。搜尋能力雖是線上閱讀的必要條件，但 PISA 2009 數位閱讀評量調查結果指出，難度最高的試題並不是航網複雜度最高的試題，而是評鑑線上文本內容可靠性的試題，難度最低的試題則是擷取單一文本中特定資訊的試題 (OECD, 2010a, 2011)。

二、線上閱讀素養的性別差異

參與 PISA 2009 閱讀素養評量各國女生的整體閱讀素養表現與不同閱讀歷程表現皆優於男生，以 OECD 平均而言，平均閱讀表現差距為 39 分，超過半個閱讀精熟水

準，其中以統整與解釋歷程的差距較小（36 分），反思與評鑑歷程的差距較大（44 分）。大部分學生位於精熟水準 3，有較多比例的男生位於水準 2 和水準 3，較多比例的女生位於水準 3 和水準 4（OECD, 2010a）。在 ERA 方面，各國女生的數位閱讀表現亦皆優於男生，以 OECD 平均而言，平均差異分數為 24 分，比紙本閱讀小，女生亦多位於水準 3 和水準 4，男生多位於水準 2 和水準 3。大部分學生位於水準 3，能在數個網頁中尋找到特定的目標資訊，且當作業沒有明確的作業和指示時，也能自行整合資訊。另一方面，女生的航網優於男生，但其數位閱讀表現不如紙本閱讀表現（OECD, 2011）。

目前在效度研究方面，Leu（2007）發展線上閱讀理解評量，發現七年級學生線上與離線閱讀理解表現的相關未達顯著。Coiro（2007）探討離線閱讀對線上閱讀的解釋量，發現 118 位七年級學生的離線閱讀理解與先備知識能解釋線上閱讀理解的大量變異，但學生線上閱讀的獨立變異仍佔很大比率（增加 15.4% 解釋量），顯示線上閱讀理解同時需要許多新的技巧和策略。Mendelovits 等人（2009）針對電子閱讀評量進行初步構念效度的探討，發現專家所評定的文本處理歷程難度與試題難度的相關高於航網歷程難度與試題難度的相關（ $0.63 > 0.46$ ）。由此可見，電子閱讀評量試題的難度與文本處理歷程有較為密切的關聯。

從上述研究動機與文獻分析可知，目前臺灣地區缺乏以青少年母群為對象、用以了解青少年網路閱讀特徵的線上閱讀素養調查工具，加上臺灣已於 2012 年接受 PISA 電子閱讀評量，故以這樣的評量取向和特色作為本土化評量工具開發與實證研究探討的起步頗值得嘗試。再者，由於目前以真實測量取向作為評量工具發展的實徵研究結果指出，線上閱讀與離線閱讀有明顯差異，本研究嘗試以半開放式的數位化環境建置評量工具，效度證據需要進一步檢驗，以提高評量工具的後續應用價值及相關領域研究的基礎。簡言之，本研究之目的有二：(1)針對線上閱讀素養的可能內涵，各以八年級和十年級學生為對象，發展適用的線上閱讀素養評量工具；(2)檢驗八年級和十年級線上閱讀素養評量的初步效度，其中包含性別差異的分析。待答問題包含：

1. 八年級和十年級線上閱讀素養評量工具的題目計量特徵為何？
2. 八年級和十年級學生線上閱讀素養評量工具的初步效度為何？
3. 八年級和十年級男女學生的線上閱讀素養表現差異為何？

參、研究方法

一、研究樣本

由於線上閱讀素養評量仍在發展的初步階段，本研究先以便利取樣來檢驗其付諸實行的可行性。八年級和十年級皆包含預試樣本與正式樣本，預試樣本用以修正試題，正式樣本用以檢驗初步效度。八年級預試樣本取自臺灣某地區3所國中、7個班級，共216位學生，其中有115位男生，101位女生；正式樣本取自5所國中、20個班級，共601位學生，包含317位男生，284位女生。十年級預試樣本取自2所高中、5個班級，共199位學生，其中有110位男生，89位女生；正式樣本取自4所高中、16個班級，共618位學生，包含307位男生，311位女生。學校皆位於市中心邊陲。

二、線上閱讀素養評量（Online Reading Literacy Assessment, ORLA）

（一）評量架構、素材與細目

線上閱讀素養評量以 PISA 2009 電子閱讀評量架構為命題依據，包含訊息的使用擷取、統整與解釋、反思與評鑑三個層面，同時參酌 Leu 等人（2004）線上閱讀理解構念。評量素材取自國家科學委員會的數位典藏資源，呈現臺灣文化多樣性脈絡，具有相當高的學術性與教育性。網頁配置清晰，適合青少年閱讀。由於閱讀素養著重生活取材，有別於成就測量，故取材範圍相當廣泛，包含學生熟悉和不熟悉的題材，且閱讀素養表現較好者，更經常閱讀小說類文體。本研究在取材上，考量這些因素，在高中階段選用小說類文體作為主要題材。上述材料經研究者適度編修，包含短文本和長文本。

八年級版本包含時光隧道單元與電影欣賞單元，前者主題為史前動物，取自數位博物館網站，內容包含生物化石、食性、演化等；後者主題為臺灣電影，取自國家電影資料館網站，內容包含臺灣電影史、影評論述、放映資訊等。預試版本有 15 題，為單式，經預試後修正 3 個試題，由於作答時間不足，正式施測分為甲乙兩式。正式版本有 17 題，題目形式包含首頁資料判讀、搜尋結果列表、電子郵件、線上討論版、網路訂票等，各單元有 8-9 題，其中，擷取與檢索歷程有 5 題，統整與解釋歷程有 7 題，反思與評鑑歷程有 5 題。各式有 12 題，其中 10 題為選擇題，2 題為建構反應題，

有 7 題為共同試題。網站具多樣化的視覺特徵和屬性，如文字、圖片、動畫，多為短文本，網頁分層如圖 3。

十年級版本包含兩個臺灣文學作家單元，一為鍾理和數位博物館，一為賴和紀念館，內容皆包含作家生平介紹、著名作品、文學貢獻等。預試版本有 18 題，為單式，經預試後修正 4 個試題，受限於作答時間，亦分為兩式。正式版本有 18 題，題目形式包含首頁資料判讀、單一和多元文本閱讀，各單元有 9 題。丙丁式各有 13 題，其中 12 題為選擇題，1 題為建構反應題，有 8 題為共同試題。

網站分層較為複雜，多為長文本。試題設計偏重統整與解釋層面（10 題）。

在試題類型方面，共有四種型式。類型一為低文本處理和低航網歷程，類型二為低文本處理和高航網歷程，類型三為高文本處理和低航網歷程，類型四為高文本處理和高航網歷程。類型的區分係考量文本特徵、航網歷程的複雜度、作業要求的明確度等三個變項。在文本特徵方面，現階段主要是針對文本的長短做控制，在航網歷程的複雜度方面，主要是針對找到答題資訊所需點擊的網頁個數、跨網頁的個數（單一和多元文本）、線上情境等做控制，在作業要求的明確度方面，主要是針對題目敘述是否有明確指定文本、題目與文本是否使用相同的用語做控制。不同線上閱讀歷程（擷取與檢索、統整與解釋、省思與評鑑）的運作都受到這些因素所影響。其中，類型一和二的文本處理成份較重，較貼近紙本閱讀情境，類型三和類型的航網成份較重，較貼近線上閱讀情境，兩者各占一半比例。四種試題類型的示例見附錄（八年級版本）。試題雙向細目表如下。

表 1 ORLA 試題雙向細目表

文本素材	低文本處理 低航網	低文本處理 高航網	高文本處理低 航網	高文本處理 高航網
時光隧道單元	3	4	1	1
電影欣賞單元	2	2	3	1
作家鍾理和單元	2	2	3	2
作家賴和單元	2	2	3	2

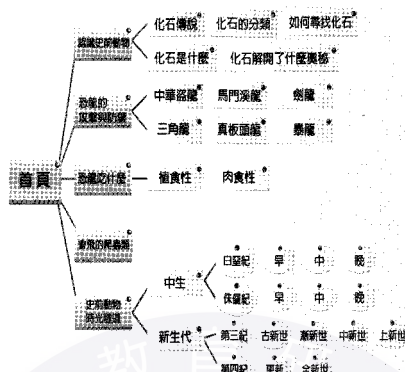


圖 3 網頁分層示例

(二) 實施與計分

本研究自行研發一套線上閱讀素養評量系統，該系統架構在Internet/WWW，建置線上閱讀素養評量工具，提供國高中學生及測驗管理者兩種不同的角色進行施測及管理。受試者可透過網頁瀏覽器(如Internet Explorer)直接在線上進行施測。測驗管理者負責管理維護受試者帳號、監控受試者的評量狀況、匯入匯出測驗資料等。系統架構採用三層式軟體架構設計，分別為使用者介面層(Presentation Tier)、商業邏輯層(Business Logic Tier)及資料服務層(Data Services Tier)。在使用者端採用薄客戶端的方式設計，使用一般的網頁瀏覽器(Browser)進行施測以減輕前端之需求及負荷。試題採用Flash動畫製作呈現於網頁上，受試者以網頁瀏覽器搭配廣及普及的Flash Player元件檢視試題。伺服器端採用Windows Internet Information Server(IIS 6.0)作為網站伺服器，以Microsoft.NET為軟體平台，搭配ASP.NET程式技術，進行測驗系統施測和計分之邏輯判斷及統計分析。後端資料庫則以關聯式資料庫(Microsoft SQL Server)作為整個系統之資料庫伺服器(Database Server)，儲存受試者資料、試題內容、試題參數、答題反應等相關資料。當受試者進入系統首頁後，先輸入個人基本資料(如學校、班級及學號等)，系統會顯示其姓名及性別，以便受試者確認自己的帳號。在確認無誤後，系統顯示一段施測說明，並讓受試者進行作答練習。接著，系統顯示題目並等待受試者反應。當受試者送出答案後，系統會自動將受試者的反應記錄到資料庫中。最後，在受試者確定作答完畢後，系統自動呈現測驗結果報告。

系統環境為半開放式的網際網路資訊空間，模擬一個線上閱讀或問題解決情境，

由受試者根據問題需求，搜尋和瀏覽解題的相關網頁進行作答。創新的評量系統介面分為「網頁瀏覽區」與「問題作答區」兩大部分，網頁瀏覽區包含閱讀材料，問題作答區包含評量試題，作答時由學生依題目敘述到網頁瀏覽區閱讀材料，並可使用網頁上預設的超連結按鍵連結到不同的網頁，並隨時回到問題作答區選擇正確答案或輸入文字（50 字內），按確定鍵後即不能再回上一題。施測時間為40-45分鐘，每人配置一臺電腦，在電腦教室進行團測，電腦的硬體規格皆一致，中隔一個座位，以防止作弊行為。學生在正規的電腦課程已有2-3年使用ICT的經驗，故具備基本的ICT能力，例如，鍵盤輸入、滑鼠操作、打字、使用按鈕、操作捲軸、使用下拉式選單、超連結等，在測驗前5分鐘使用帳號密碼登入，共同閱讀測驗說明，並由施測人員示範作答，施測人員皆為研究者。從實際施測情形來看，線上閱讀以基礎搜尋能力（航網）為必要條件，學生在解決較複雜的問題時，若無法即時找到關鍵網頁，將浪費太多作答時間，貼近一般線上閱讀的實際情境，頗適合用來檢測。

計分方面，選擇題採二元計分（0、1），建構反應題採多元計分（0-3）。後者以人工閱卷，由研究者草擬評分規範，並針對預試結果建立實際作答樣本，以供正式題本使用。評分規範的內容依不同得分點各提供一組編碼原則和數個參考答案。八年級甲式滿分為15分，乙式滿分為13分；十年級丙和丁式滿分皆為14分。皆以多元計分Rasch單向度IRT模式和同時估計法進行全測驗的題目參數估計和受試者能力值估計，理論上，兩者值介於-3-3之間，能力估計值的平均數為0。八年級另以多向度IRT模式進行不同閱讀歷程的能力值估計。

三、效標工具

本研究使用的效標工具有PISA 2006閱讀素養測驗、臺灣地區國中生基本學力測驗、在校各科平均表現。前兩者皆為標準化測驗，具有穩定的信效度，在閱讀（語文科）測量內容方面，皆含有閱讀和溝通部分。2006和2009年十年級學生PISA閱讀素養與基測國文成績（含作文）的相關皆在.70以上，為中高度相關。

（一）PISA 閱讀素養測驗

本測驗為八年級正式版本的效標。試題取自 PISA 2006 正式施測國際閱讀試題。研究者依不同層面試題比重分配與國際試題難度 b 參數選擇不同難度層次的試題組成國中學生適用的題本，包括 5 個單元，每單元有 2-5 題，其中 3 個單元為連續文本，2 個單元為非連續文本，共 18 題，平均難度 b 值為 -3.5 ，包含 5 題擷取與檢索歷程的試題，8 題統整與解釋歷程的試題，5 題反思與評鑑歷程的試題。實施時間為 45 分鐘。2 位編碼人員皆為臺灣 PISA 2009 研究小組成員，接受過 1 個月的編碼訓練，其中一位為研究者。評分時依 PISA 2006 國際閱讀試題的計分規準進行編碼，共 1 星期，計分為 0-3，採多元計分 Rasch 單向度和多向度 IRT 模式進行全測驗與不同閱讀歷程的能力值估計。結果顯示，IRT 信度係數為 0.79，各閱讀歷程與總分的相關介於 0.57-0.90 之間，不同閱讀歷程的相關介於 0.62-0.90 之間。受限於學校配合意願，十年級學生未接受此測驗。

（二）國中基本學力測驗

國民中學學生基本學力測驗（The basic competence test for junior high school students, BCTEST）為臺灣地區九年級學生升學進入高中職的參考依據。測驗目的在檢測學生在經過國中階段的學習後，對教育部所訂定之各科能力指標的精熟程度，包含國文（含作文）、英語、數學、自然、社會科。測驗試題主要是根據各科課程綱要之能力指標進行編製，不限定教材版本，為學習能力測驗，偏向學習成就導向。測驗題型均採單一選擇題，包括單題及題組題兩類。各科採用 IRT 量尺分數計算，滿分為 80 分，加上作文級分，總分為 412 分。在信效度上已獲得客觀資訊，測驗分數可供申請入學、甄選入學、登記分發入學使用。相較於各校實施的國文科段考試題，基測國文成績更具有穩定性和參考價值，故本研究以此作為十年級版本同時效度檢驗的依據，同時並以其他各科成績檢驗區辨效度。

四、資料收集與分析

（一）資料收集

本研究於 2010 年 6 月收集八年級和十年級線上閱讀素養評量的預試資料，作為試題增刪和修改的依據。2010 年 12 月收集八年級和十年級上閱讀素養評量的正式資料，

其中，八年級有11個班級的學生接受甲式，計307人，9個班級的學生接受乙式，計294人；十年級有8個班級的學生接受丙式，計311人，8個班級的學生接受丁式，計307人。在建立學生基本資料的同時，一併收集性別資料。在效標變項方面，2011年4月收集八年級PISA閱讀素養測驗資料，5月徵得學校同意取得十年級學生在前一年參加基測的各科成績，6月收集八年級和十年級在校各科成績，共4次；科目皆包含國文、英語、數學、自然、社會。其中，在校各科成績皆以班級為單位轉為 z 分數，並進行平均。

（二）資料分析

本研究以Iteman for Windows軟體進行傳統題目分析，並以ConQuest 2.0軟體進行IRT題目參數估計與受試者能力值估計（並轉換為平均數500，標準差100的量尺分數），同時以SPSS18.0進行Pearson積差相關、描述統計和 t 檢定分析。首先，以 p 、 b 、 d 、 r_{pb} 指標描述線上閱讀素養評量試題的難度、鑑別度和點二列相關係數。其次，以相關矩陣討論線上閱讀素養評量分數的輻合和區別效度。接著，以描述統計呈現線上閱讀素養的評量結果，並以獨立樣本 t 檢定討論線上閱讀素養表現的性別差異。

肆、研究結果

本研究旨在嘗試研發國、高中學生適用的線上閱讀素養評量工具，並針對可行性樣本，提出評量工具的初步效度證據，同時描述學生的整體表現與性別差異。效度驗證依循 Messick（1989）所提出的架構，包含內容層面（content aspect）、實質層面（substantive aspect）、結構層面（structural aspect）、解釋性（interpretability）等。在內容層面，包含內容相關性（content-relevant）、代表性（representatives）與技術品質（technical quality）。

一、ORLA 題目計量特徵

就 CTT 而言，八年級各式試題的平均難度適中（ $P=0.48$ 、 0.54 ），平均鑑別度良好（ $D=0.42$ ），平均點二系列相關適中（ $r_{pb}=0.36$ 、 0.38 ）；十年級各式試題的平均難度為中間偏易（ $P=0.63$ 、 0.57 ），平均鑑別度良好（ $D=0.44$ 、 0.46 ），平均點二系列相

關良好($r_{pb}=0.50、0.55$)。以 IRT 進行試題等化後，八年級試題的平均難度 b 值為 0.08，各題介於-1.36-2.10 之間；十年級試題的平均難度 b 值為-0.19，各題介於-1.74-2.68 之間。以八年級來說，擷取與檢索歷程的平均難度為-0.32，統整與解釋歷程的平均難度為-0.08，反思與評鑑歷程的平均難度為 0.75。由上可知，線上閱讀素養評量的難度適中，可區分不同程度的學生，對高中學生來說較為容易，而其中，反思與評鑑歷程的試題對國中學生來說比較困難。

表 2 線上閱讀素養評量的題目參數摘要

年級	式別	人數	題數	難度(P)	鑑別度(D)	點二系列相關(r_{pb})	IRT 難度(b)
8	甲	307	12	0.48	0.42	0.38	0.08
	乙	294	12	0.54	0.42	0.36	
10	丙	311	13	0.63	0.44	0.50	-0.19
	丁	307	13	0.57	0.46	0.55	

二、ORLA 效度檢驗

(一) 內容層面

效度證據在內容層面上，可從測驗內容是否能充分涵蓋構念的定義範圍，整合專業判斷和量化指標進行評判 (Messick, 1989)。本評量在內容相關性及代表性方面，邀請多位測驗評量領域學者與中文系教授針對各式版本進行內容和形式上的審查，根據原先的測驗藍圖，考量文本特徵、航網歷程的複雜度、作業要求的明確度等因素，判斷每個試題是否準確對應到藍圖中的各細格。在綜合專家學者的意見後，研究者據以修改閱讀材料、題目敘述和作答說明，同時調整試題分類，最後得到如表 1 的題數組成結果。在量化指標方面，採用 IRT 的 Infit 和 Outfit 指標值檢驗各題的模式適配度。結果顯示，八年級各題的 Infit 值和 Outfit 值皆介於 0.84-1.09 之間，十年級各題的 Infit 值和 Outfit 值皆介於 0.87-1.15 之間，都在模式適配度的合理範圍 0.60-1.30 內。由此可知，ORLA 試題與測量模式的吻合程度良好，兩個評量工具都有不錯的技術品質。

圖 4 和圖 5 分別呈現八年級和十年級線上閱讀素養評量的試題圖。此圖為能力估計值與試題難度值的對應（右方數字為題號），理論上兩者介於-3-3 之間。大抵而言，

就八年級而言，能力估計值的平均數以上與以下 2 個標準差以內涵蓋了所有試題，就十年級而言，能力估計值的平均數以下 2 個標準差、以上 3 個標準差以內涵蓋了所有試題。十年級的試題難度分佈較均勻，都沒有天花板效應或地板效應。由此可知，評量試題具有不錯的代表性。

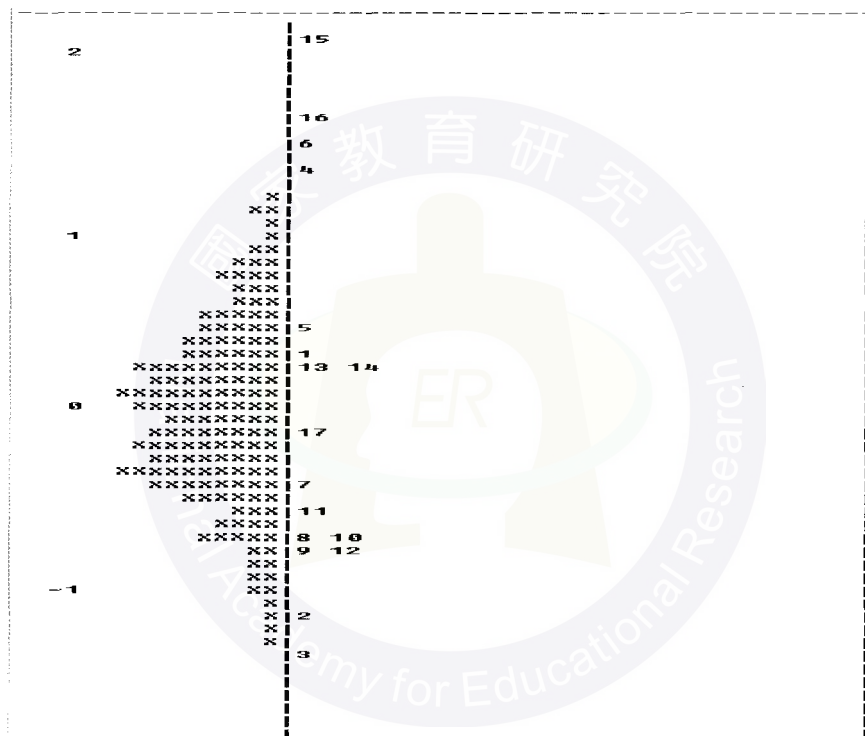


圖 4 八年級線上閱讀素養評量的試題圖

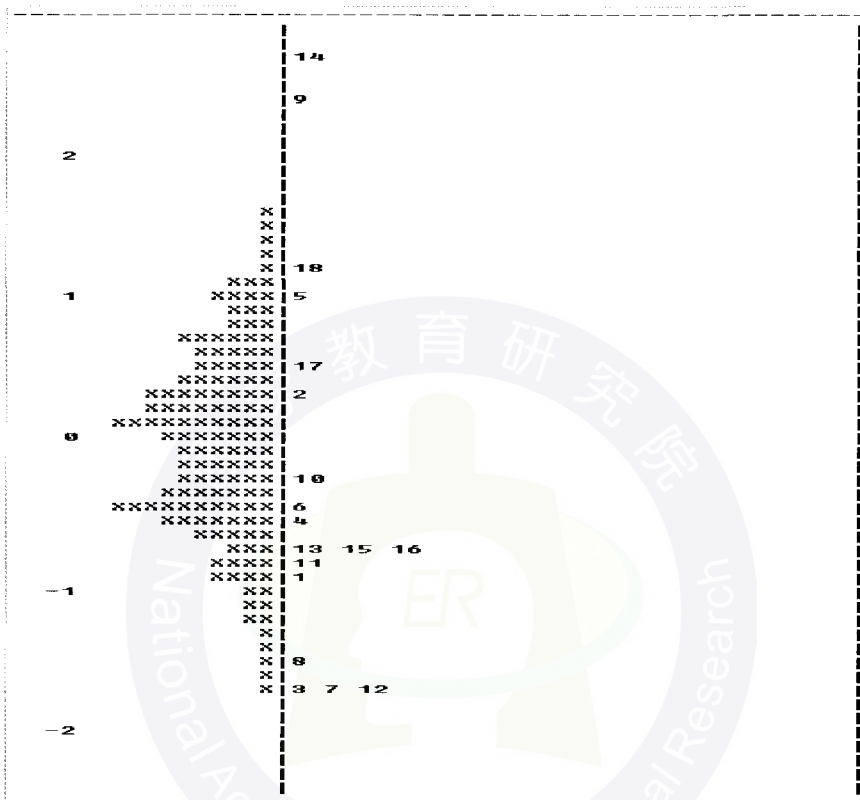


圖 5 十年級線上閱讀素養評量的試題圖

(二) 實質層面

1. 試題難度層次

效度證據就實質層面而言，可檢視作答反應組型是否與文獻理論與假設一致 (Messick, 1989)。本研究從試題難度層次與效標關聯效度係數兩方面進行檢視。在試題難度層次方面，依難度 b 值並參酌試題內容，將八年級不同線上閱讀歷程的試題各分為基礎、精熟和進階三個難度層次。表 3 為八年級甲乙式合併估計後的數據，結果顯示，若將不同線上閱讀歷程分開來看，較簡單的擷取與檢索試題是在指定文本中擷取某個明確訊息的試題，較困難的擷取與檢索試題是在多重文本中尋找某個訊息所在位置的試題。較簡單的統整與解釋試題是整合數個指定網頁中明確訊息的試題，較困

難的統整與解釋試題是往返於多個不確定的網頁中進行資訊的連貫性統整之試題。較簡單的反思與評鑑試題是辨認網路資料來源（如，網址）可靠性的試題，較困難的反思與評鑑試題是針對線上文本內容本身進行批判反思的試題，尤其是線上論壇，而並不是所需航網複雜度較高的試題。由上可知，各閱讀歷程的試題難度層次大致可依指定文本、單一和多元文本、航網複雜度、文本內容的批判程度進行區分。此分類結果與 OECD（2011）PISA 2009 數位閱讀評量各歷程的實徵難度組型一致。

另一方面，若將不同線上閱讀歷程合起來看，統整與解釋的試題有時比擷取與檢索的試題容易，這樣的數據與 PISA 調查結果也有部分謀合。根據 2009 年閱讀素養調查結果，有些擷取資訊的問題比單純理解文義的問題來得困難，這些問題多半是出現在非連續文本（如廣告單）上，原因在於文本中某些細部資訊不易被察覺，但往往卻是答題的關鍵資訊（OECD, 2010b）；且不同線上閱讀歷程對文本特徵、航網歷程的複雜度、作業要求的明確度等因素的控制也有差異。十年級丙丁兩式合併估計後的數據顯示，擷取與檢索試題的難度 b 值介於-0.75~-1.74 之間，統整與解釋試題的難度 b 值介於-1.98-2.68 之間，反思與評鑑試題的難度 b 值介於 1.20-2.34 之間，也是一樣的情形。整體而言，本評量的實徵難度組型與現有的調查結果頗為一致。

表 3 八年級不同線上閱讀歷程試題難度層次的特徵描述

歷程	試題	b	層次	描述
擷取與檢索	2	-1.13	L	擷取指定網頁的單一特定資訊
	10	-0.66	M	
	7	-0.42	M	跨數個指定網頁擷取一到多個資訊
	13	0.28	M	
	1	0.33	H	跨多重網頁以推測資訊的位置
統整與解釋	3	-1.36	L	跨數個指定網頁整合特定的資訊
	9	-0.74	M	
	11	-0.58	M	
	17	-0.08	M	尋找問題關連的網頁以整合資訊
	14	0.24	M	
	5	0.50	M	
	4	1.31	H	在多重不確定網頁中進行動態性統整
反思與評鑑	12	-0.75	L	評鑑網路資料來源的可靠性
	8	-0.68	L	
	6	1.47	M	評鑑線上文本內容的可靠性
	16	1.61	M	
	15	2.10	H	評鑑網路論壇內容的可靠性

L：易，H：難，M：中

2.效標關聯效度係數

八年級線上閱讀素養與 PISA 閱讀素養及在校國文成績皆具有中度相關 ($r=0.52$ 和 0.48)，高於與英語、數學、自然、社會的相關 ($r=0.33-0.47$)。十年級線上閱讀素養與基測國文成績亦具有中度相關 ($r=0.47$)，高於與基測英語、數學、自然、社會的相關 ($r=0.28-0.40$)。若以整個相關矩陣來看，社會科與各科的相關係數都很高，可能原因在於學科內容範疇較為廣泛，涉及綜合性知識的使用，其中閱讀更是基礎，然整體而言，線上閱讀素養評量呈現合理的效度組型。逐題方面，試題與總分的相關均高於與效標的相關，且與國文成績的相關均高於與各科的相關。依逐題效度分析組型，試題可分為兩類：(1)與總分有明顯較高相關的試題，及與各效標均無顯著相關的試題：這類試題的航網成份較重，包含首頁資料判讀的訊息擷取歷程、評鑑網址或搜尋結果可靠性的歷程、多重網頁的資訊統整歷程，足見不同線上閱讀歷程同時含有 ICT 搜尋成份。(2)與各效標皆有顯著相關的試題：試題的文本處理成份較重，平均難度較高，支持 Mendelovits 等人 (2009) 所提出之文本處理是電子閱讀評量主要難度來源的說法。八年級線上閱讀素養評量有 8 題航網成份較重的試題，平均難度 b 值 -0.23 ，9 題文本處理成份較重的試題，平均難度 b 值為 0.37 ；十年級有 8 題航網成份較重的試題，平均難度 b 值為 -0.63 ，10 題文本處理成份較重的試題，平均難度 b 值為 0.22 。

表 4 線上閱讀素養評量分數的效標關聯效度係數

年級	人數	PISA 閱讀	國文	英語	數學	自然	社會
8	601	0.52*	0.48*	0.39*	0.33*	0.41*	0.47*
10	618		0.47*	0.38*	0.28*	0.37*	0.40*

* $p < 0.05$

對照電子閱讀評量所假設的試題分類，各試題可合理歸為低文本處理和低航網歷程、高文本處理和低航網歷程、低文本處理和高航網歷程、高文本處理和高航網歷程等四類。航網歷程又可分為機械性運作與認知性運作。不同類別的試題舉例說明如下：

(1) 低文本處理和低航網歷程：例如，八年級問題 2 明確指示學生到某個指定的網頁中擷取單一明確的訊息，文本處理與航網歷程的複雜度皆低，偏向機械式的航網運作。試題與總分的相關為 0.41 ，與各效標的相關介於 $0.33-0.38$ 之間，皆達 0.05 顯著水準，見附錄。

(2) 高文本處理和低航網歷程：例如，八年級問題 14 要求學生尋找幾個明確與問題有關的長文本網頁，並選出最符合問題所需的資訊，閱讀理解的需求度較高，但只需機械式的航網運作。試題與總分的相關為 0.48，與各效標的相關介於 0.26-0.31 之間，皆達 0.05 顯著水準。

(3) 低文本處理和高航網歷程：例如，八年級問題 12 呈現一個搜尋結果資料網頁，要求學生判斷數個網址來源的可靠性，而各筆資料的內容意義相近。試題與總分的相關為 0.31，與各效標的相關介於 0.05-0.07 之間，未達 0.05 顯著水準，見附錄。

(4) 高文本處理和高航網歷程：例如，八年級問題 4 要求學生根據問題需求，跨多元文本搜尋與問題關連的網頁，同時整合數個資訊，以解決一個小型作業問題。需要較高的認知性航網運作。試題與總分的相關為 0.50，與各效標的相關介於 0.18-0.21 之間，皆達 0.05 顯著水準。

綜上所述，線上閱讀素養評量具備合理的輻合與區辨效度，有別於離線閱讀，研究結果支持 Coiro (2007)、Cheung 與 Sit (2008)，及 Mendelovits 等人 (2009) 的論點，線上閱讀理解僅有一部分可由離線閱讀解釋，電子閱讀同時涵蓋文本處理歷程與航網歷程。

(三) 結構層面

效度證據從結構層面來看，可檢視計分模式是否與構念所假設的能力結構一致 (Messick, 1989)。本研究以單向度 IRT 進行線上閱讀素養表現估計，結果顯示，八年級和十年級線上閱讀素養評量皆符合單向度假定，只測量一個能力因素，Rasch 模式分別可解釋 80.1%與 81.3%的受試者能力變異，大於 Linacre (2007) 所提出的標準 60%。以多向度 IRT 估計來說，八年級不同閱讀歷程與總分的相關介於 0.55-0.82 之間，擷取與檢索歷程、統整與解釋歷程的相關為 0.72，統整與解釋歷程、反思與評鑑歷程的相關為 0.72，擷取與檢索歷程、反思與評鑑歷程的相關為 0.58。線上閱讀素養與 PISA 閱讀素養在擷取與檢索歷程得分的相關係數為 0.39 ($p<0.05$)，在統整與解釋歷程得分的相關係數為 0.49 ($p<0.05$)，在反思與評鑑歷程得分的相關係數為 0.41 ($p<0.05$)，顯示線上閱讀歷程有紙本閱讀歷程所測量的結構，而在數位化環境中，由於資訊的擷取仰賴網航工具，並需立即進行批判性選取，故擷取層面與評鑑層面是紙本和線上媒介的主要差異。另一個解釋方式是，由於各線上閱讀歷程的得分中同時含有文本處理成份和航網成份，其中航網又為必要條件，故使其有別於紙本閱讀。

（四）解釋性

解釋性意指測驗結果可提供的解釋程度（Messick, 1989）。PISA 為了便於追縱各國學生的閱讀素養表現成長，及進行紙本與數位閱讀表現的對照，雖然歷次測驗實施的內容和施測題數都不盡相同，仍皆使用同一組決斷分數作為解釋學生表現水準的依據，且適用於九和十年級。由於本研究之 ORLA 係針對電子閱讀評量進行本土化的開發與實證研究，以其評量架構與試題特色為依據，在總施測題數、題型分配及全測驗量尺估計方法上亦雷同，故以其決斷分數瞭解學生目前的精熟水準特徵有哪些一樣的地方。在測驗結果的解釋上，較高水準的線上閱讀素養表現特徵包含較低水準的線上閱讀素養表現特徵，八年級未達水準 2（407 分以下）的學生能擷取指定網頁中的明確資訊，及整合數個指定網頁的明確資訊；水準 2（407 分以上）的學生能進一步跨數個指定網頁擷取多個資訊、搜尋問題關連的網頁並整合資訊、評鑑網路資料來源的可靠性；水準 3（480 分以上）的學生在擷取訊息方面，能進一步搜尋多重網頁以推論資訊的位置；水準 4（553 分以上）的學生在統整訊息方面，能進一步理解單一文本所隱含的資訊，及不同搜尋結果資料的意義。水準 5（626 分以上）的學生在評鑑訊息方面，能進一步評鑑線上文本內容及網路論壇內容的可靠性。十年級未達水準 2 學生的表現與八年級雷同，水準 2 到水準 4 學生的表現與八年級水準 4 學生雷同，水準 5 學生的表現與八年級水準 5 雷同。不同表現水準皆可根據幾個主要特徵進行區分。

在測驗信度方面，本研究初步檢驗各單元與總分的相關，及建構反應題的評分者一致性。結果顯示，八年級甲乙式評量單元與總分的相關介於 0.77-0.84 之間，內部一致性 α 係數各為 0.58 和 0.60，兩式合併後的 IRT 信度估計值為 0.62。十年級丙丁式評量單元與總分的相關介於 0.76-0.88 之間，內部一致性 α 係數各為 0.57 和 0.59，兩式合併後的 IRT 信度估計值為 0.61。建構反應題由研究者與一位國文科現職教師使用評分規範進行評定，評分者一致性係數分別為 0.85 和 0.81，一致性良好。由於學生在複雜作業所需的答題時間較長，每位學生作答的題數受限，故暫時以群體表現特徵描述為主，因為個體的特徵或診斷描述誤差較大。

三、線上閱讀素養表現

（一）整體閱讀表現

表5呈現全體學生線上閱讀素養評量的原始得分。由表可知，八年級甲式平均得

分為7.19，乙式平均得分為6.57，多數學生答對一半試題。試題經等化後，平均量尺得分為500，標準差為85，女生的平均數為518，男生的平均數為485，女生得分較高。在十年級方面，丙式平均得分為11.06，丁式平均得分為7.63，試題經等化後，平均量尺得分為500，標準差為75，其中女生的平均數為508，男生的平均數為491，女生得分較高。

表 5 線上閱讀素養評量原始得分之描述統計

年級	式別	人數	平均數	標準差
8	甲	307	7.19	2.51
	乙	294	6.57	2.00
10	丙	311	11.06	2.34
	丁	307	7.63	2.57

在精熟水準方面，八年級未達水準2（407分以下）的學生有10.1%，位於水準2（407分以上）的學生有26.5%，位於水準3（480分以上）的學生有37.1%，位於水準4（553分以上）的學生有21.1%，位於水準5（626分以上）的學生有5.2%。八年級未達水準2的男生有12.6%，位於水準2的男生有30.9%，位於水準3的男生有36%，位於水準4的男生有17.7%，位於水準5的男生有2.8%。八年級未達水準2的女生有7.4%，位於水準2的女生有21.5%，位於水準3的女生有38.4%，位於水準4的女生有25%，位於水準5的女生有7.7%。十年級與八年級的比例分布組型相同，多數學生皆位於水準3，其中男生多位於水準2和水準3，女生多位於水準3和水準4，水準5的學生人數明顯較少，如表6。綜上所述，中學生線上閱讀素養的表現適中，且女生表現較好，大部分的學生能擷取數個指定網頁中一到多個明確資訊、搜尋多重網頁以推論資訊的位置、搜尋問題關連的網頁以整合資訊、評鑑網路資料來源的可靠性。各精熟水準的學生人數比例分布接近2009年PISA數位閱讀評量各國調查結果。

若比較全體學生在兩種不同媒介的閱讀素養精熟水準表現，紙本和線上閱讀素養表現達水準3的學生人數皆較多，就水準2以下而言，線上媒介位於水準2的學生人數較多，紙本媒介未達水準2的學生人數較多，皆以男生較多。就水準4以上而言，紙本媒介位於水準4和水準5的學生人數皆較多，女生尤為明顯，女生的線上閱讀不如紙本閱讀。上述結果與2009年PISA數位閱讀評量各國男女學生在不同媒介閱讀表現的組型

相同。男生落後表現的現象值得關注，另一方面，水準5的試題多需反思與評鑑，可見加強青少年線上批判性閱讀是當前的要務。

表 6 八年級男女學生紙本和線上各閱讀精熟水準的比例分布

性別	媒介	未達水準2	水準2	水準3	水準4	水準5
全體	紙本	16.1	21.8	31.3	24.6	6.2
	線上	10.1	26.5	37.1	21.1	5.2
男生	紙本	22.1	23.7	30.3	18.9	5.0
	線上	12.6	30.9	36.0	17.7	2.8
女生	紙本	9.5	19.7	32.4	31.0	7.4
	線上	7.4	21.5	38.4	25.0	7.7

註：紙本媒介使用 PISA 閱讀素養測驗

(二) 不同閱讀歷程表現

在不同線上閱讀歷程表現方面，八年級甲式擷取與檢索試題的平均原始得分率為 0.51，統整與解釋試題的平均得分率為 0.48，反思與評鑑試題的平均得分率為 0.45，其中選擇題的平均得分率為 0.55，建構反應題的平均得分率為 0.15；乙式擷取與檢索訊息的平均得分率為 0.62，統整與解釋訊息為 0.58，反思與評鑑訊息為 0.35，其中選擇題的平均得分率為 0.61，建構反應題為 0.16，建構反應題多為反思與評鑑試題。十年級丙丁式建構反應題的平均得分率在 0.30-0.35 之間，也遠低於選擇題。多向度 IRT 估計結果顯示，八年級學生在擷取與檢索歷程的平均得分為 503，在統整與解釋歷程的平均得分為 501，在反思與評鑑歷程的平均得分為 499；男女學生在擷取與檢索歷程的得分差距較小，在反思與評鑑歷程的得分差距較大。就男生而言，擷取與檢索歷程的表現相對較好，就女生而言，反思與評鑑歷程的表現相對較好。以建構反應題的實際作答樣本來看，大部分學生習慣以自己的意見回答問題，無法有效根據網站資訊做為推論或判斷的依據，另一部分學生偏重複製文本，未能提供網站資訊與自己立論的關連說明。十年級也有同樣的情形。

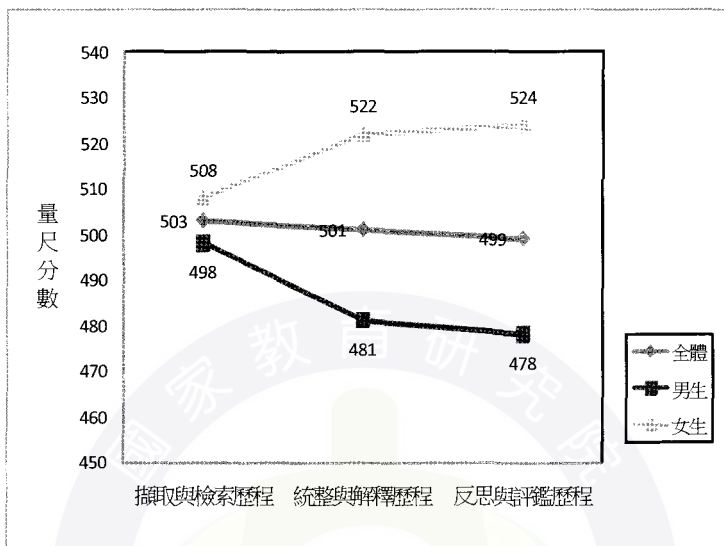


圖 6 八年級男女學生在不同線上閱讀歷程的表現對照

(三) 性別差異

在性別差異檢定方面，青少年女生的線上閱讀素養平均表現優於男生，八年級女生的平均得分為 518，男生的平均得分為 485，十年級女生的平均得分為 508，男生的平均得分為 491。對全測驗進行獨立樣本 t 檢定，結果發現，八年級的 t 檢定值為 4.77 ($p < .01$)，解釋力 η^2 為 0.037；十年級的 t 檢定值為 4.11 ($p < .01$)，解釋力 η^2 為 0.027。由此可見，就線上閱讀素養而言，性別差異的現象依然存在，且國中有更明顯的差異。若以單題表現來看，女生在閱讀成份較重與航網成份較重試題的表現都比較好，尤其是電子郵件與討論交流的問題，此與 2009 年電子閱讀評量的調查結果一致。男生表現略好的題目則在搜尋多重網頁以推論資訊位置的問題。

上述研究結果指出，青少年線上閱讀素養評量工具有不錯的品質，並有初步效度證據支持，全體學生的表現適中，多項發現與 PISA2009 各國數位閱讀評量調查結果相似。

伍、結論與建議

一、討論與結論

本研究旨在嘗試研發兩個中學生線上閱讀素養評量工具，並針對可行性樣本，進行初步效度檢驗，同時描述青少年樣本的閱讀表現。茲針對試題品質、效度檢驗、表現特徵等方面之初步研究發現進行討論，並說明結論：

(1) 在試題品質方面：兩個線上閱讀素養評量的難度適中，鑑別度良好，對高中學生來說較為容易，對國中學生來說比較困難。惟受限於作答時間與測驗內涵的複雜性，尚未獲得良好的內部一致性 α 係數與穩定的信度估計值。由於各式評量的實施時間僅 45 分鐘，某些試題設計涉及較複雜的航網運作，所需的答題時間較長，因此使資料收集的代表性受限。本研究使用題本配置進行較長測驗的表現估計，結果顯示，可得到比較好的估計值穩定性，此可作為後續線上閱讀題庫擴充，及類推性效度證據收集與探討的基礎，包含不同線上閱讀歷程的表現估計。整體而言，線上閱讀素養評量頗適用於青少年學生，本研究參考 PISA 在題目敘述中採用不同程度航網指示的試題設計方法可有效指引學生搜尋關連網頁與閱讀，有別於使用關鍵字作為搜尋起點的 ICT 素養評量，能克服真實線上情境中關鍵字搜尋能力不佳所產生的後續問題，此精神可作為後續以青少年階段閱讀作為焦點之線上評量設計的起點。

(2) 在效度檢驗方面：線上閱讀素養評量在內容層面上，實徵資料與模式的適配度良好，具有不錯的技術品質。不同難度層次的試題與電子閱讀評量所提出的試題分類假設與難度來源大致相符，且測量的特質包含 ICT 航網歷程，有別於離線閱讀，在實質效度上有合理的證據支持，與 Coiro(2007)、Cheung 和 Sit(2008)，及 Mendelovits 等人(2009)、OECD(2011)的論點一致。受試者個體表現可依不同水準試題的特徵進行界定，提供線上資訊擷取、資訊統整及資訊評鑑的測驗結果解釋架構，且較高水準的表現特徵包含較低水準的表現特徵，也與 OECD(2010b) PISA 現有的結果資料不謀而合。整體而言，線上閱讀素養評量在內容層面、實質層面、結構層面、解釋性等方面皆有良好的效度證據支持。目前以半開放式網際網路環境模擬真實的線上環境，在各閱讀歷程中亦可見其展現不同於紙本閱讀的型態，效度係數組型比 OECD(2011)的 PISA 資料合理，顯示如果相關變項控制得當，採用虛擬情境所測量的表現可能與真實情境一致，因此，未來在評量情境的安排上，可考慮在似真性與施測成

本上取得一個較佳平衡。

(3) 在發展特徵與性別差異方面：本評量工具是一個初步嘗試。從該地區中學生的初步表現來看，青少年學生線上閱讀素養的表現適中，對於一個創新評量設計形式來說，學生的表現算是不錯。大部分學生能擷取數個指定網頁中一到多個明確資訊、搜尋多重網頁以推論資訊的位置、搜尋問題關連的網頁以整合資訊、評鑑網路資料來源的可靠性；惟在線上文本的評鑑上表現較不理想，習慣以自己的意見回答問題，無法進行有依據的說理，可能原因在於平日沒有這樣的習慣，多以複製文本的方式收集和組織線上資料，由此可知，資料依據的線上批判思考教學是中學語文數位化課程訓練的要項。另一方面，女生在電子郵件和線上交談的問題上表現較好、男生在搜尋多重網頁的問題上表現較好，性別差異特徵與 OECD 平均各國男女學生的表現相同 (OECD, 2011)。由於 PISA 2009 問卷調查結果指出，女生較常使用電子郵件並從事線上聊天等社會性活動，男生較常參與網路小組論壇及搜尋網路資訊等資訊搜尋活動 (OECD, 2010c; OECD, 2011)。同時，男生也比女生更容易使用多種搜尋策略、點閱不同超連結、往返於網頁之間、花費較少時間瀏覽每個網頁 (Chen, 2009; OECD, 2011; Rahman, Yassin, Ishak, & Amir, 2008)，這些可能都是造成表現不同的原因，值得後續進行背景變項的探討。另一方面，從階段別來看，國中性別差異比高中明顯，皆以男生為落後，尤其是線上文本的評鑑，顯示在進入國中階段後，即宜關注及加強線上文本的教學，特別是針對男生。

(4) 在評量革新方面：本研究線上閱讀素養評量是臺灣地區較早以閱讀科為主所發展的數位化評量，由於線上測驗在技術上與硬體上的門檻較高，本評量結合線上典藏資源，為國內少見。從施測現場來看，學生多專注於作答，顯示這樣的嘗試頗值得被鼓勵。

二、未來研究與教學上的建議

本研究參考 PISA 電子閱讀評量的試題特色與施測方法，進行本土化線上閱讀素養評量工具的開發與實證研究。無論在應用或學術研究層面，線上閱讀素養是目前一個相當重要的議題，且尚在起步階段，仍缺乏明晰的理論架構與實證資料。雖然近來國內外學界亦開始重視電子閱讀，線上閱讀素養因本質上仍有差異且目前受到較少研究者的關注，因此本研究提供可接受的學術理論創新性和應用價值。本研究根據研究限制與上述結論，提出未來研究與應用上的建議。

(1) 擴充線上閱讀評量題庫：本研究線上閱讀素養評量受限於作答時間與測驗內涵的複雜性，尚未獲得良好的內部一致性 α 係數與穩定的信度估計值。使用題本配置進行較長測驗的表現估計後，可得到比較好的估計值穩定性，此可作為後續線上閱讀題庫擴充與信度檢驗的基礎，包含不同線上閱讀歷程的表現估計。

(2) 記錄航網歷程與計分：本研究線上閱讀素養評量未記錄航網歷程並予以計分，故在評量結果的解釋上，無法提供較具有診斷價值的剖面資訊，後續研究可記錄航網歷程，並發展計分標準，以提供線上閱讀的部分給分機制。

(3) 數位化閱讀評量的革新：本文最主要的貢獻在線上評量的革新方面。由於線上測驗在技術上與硬體上的門檻較高，本評量是臺灣地區較早以閱讀科為主所發展的數位化評量，文本素材結合線上典藏資源，為國內少見，值得進一步嘗試。

(4) 增進青少年線上批判性閱讀：本研究青少年樣本的線上文本批判性閱讀表現不盡理想，尤其是建構反應題。未來教學宜強調資訊依據為批判性閱讀的關鍵，需教導學生答案需改寫自網站資訊，選用文本或圖表資訊支持自己的論點，強調資訊依據為評分的關鍵。閱讀內容可從線上網路論壇、多元線上資訊來源著手，以進行線上文本的評鑑。

(5) 加強青少年男生的閱讀表現：男生在紙本和線上媒介的閱讀表現都比較不好，除了針對落後學生的表現特徵規畫合適的教學介入方案，例如，針對跨網頁擷取資訊、搜尋問題關連的網頁以整合數個資訊、評鑑網路資料來源的可靠性等進行教學外，也可從事相關背景變項的分析，以了解落後表現的主要原因。

誌 謝

本文感謝《教育實踐與研究》編輯委員會多位審查委員老師在論文寫作嚴謹度、研究結果意義的闡述等方面給予的評論與建議。另外感謝臺南市立大灣高中教務處資訊組蘇宗立老師於測驗系統及施測樣本的協助。

參考文獻

- 林煥祥、劉聖忠、林素微、李暉 (2008)。臺灣參加**PISA 2006**成果報告。行政院國家科學委員專題研究計畫 (編號: NSC 95-2522-S-026-002), 未出版。
- 柯華葳、詹益綾、張建好、游婷雅 (2008)。臺灣四年級學生閱讀素養**PIRLS2006**報告。行政院國家科學委員專題研究計畫 (編號: NSC 96-MOE-S-008-002), 未出版。
- 香港中文大學 (2011)。全球學生數碼閱讀能力首次揭示: 分析香港學生數碼閱讀與科技應用能力的關係。中大公布學生能力國際評估計畫 (**PISA 2009**) 研究結果新聞稿。取自
http://www.fed.cuhk.edu.hk/~hkpisa/events/2009/files/PISA09_digital_pr_c_28jun.pdf
- 洪碧霞 (2008)。臺灣 15 歲學生閱讀、數學和科學素養調查研究: 教育品質和均等議題 (**PISA 2009**)。行政院國家科學委員會專題研究計畫 (編號: NSC-97-2522-S-024-001), 未出版。
- Asselin, M., & Moayeri, M. (2010). Examining adolescent internet literacy practices: An exploration of research methods. *Journal of Theory and Practice in Education*, 6(2), 191-210.
- Chen, H. Y. (2009). *Online reading comprehension strategies among general and special education elementary and middle school students* (Unpublished doctoral dissertation). University of Michigan State University, MSU.
- Cheung, K. C., & Sit, P. S. (2008). Electronic reading assessment: The PISA approach for the international comparison of reading comprehension. *Journal of Educational Research and Development*, 4(4), 19-39。
- Coiro, J. (2007). *Exploring changes to reading comprehension on the internet: Paradoxes and possibilities for diverse adolescent readers* (Unpublished doctoral dissertation). University of Connecticut, CT.
- Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, 42(2), 214-257.
- Coiro, J., Knobel, M., Lankshear, C., & Leu, D. (2007). *The handbook of research in new literacies*. Mahwah, NJ: Lawrence Erlbaum Associates.

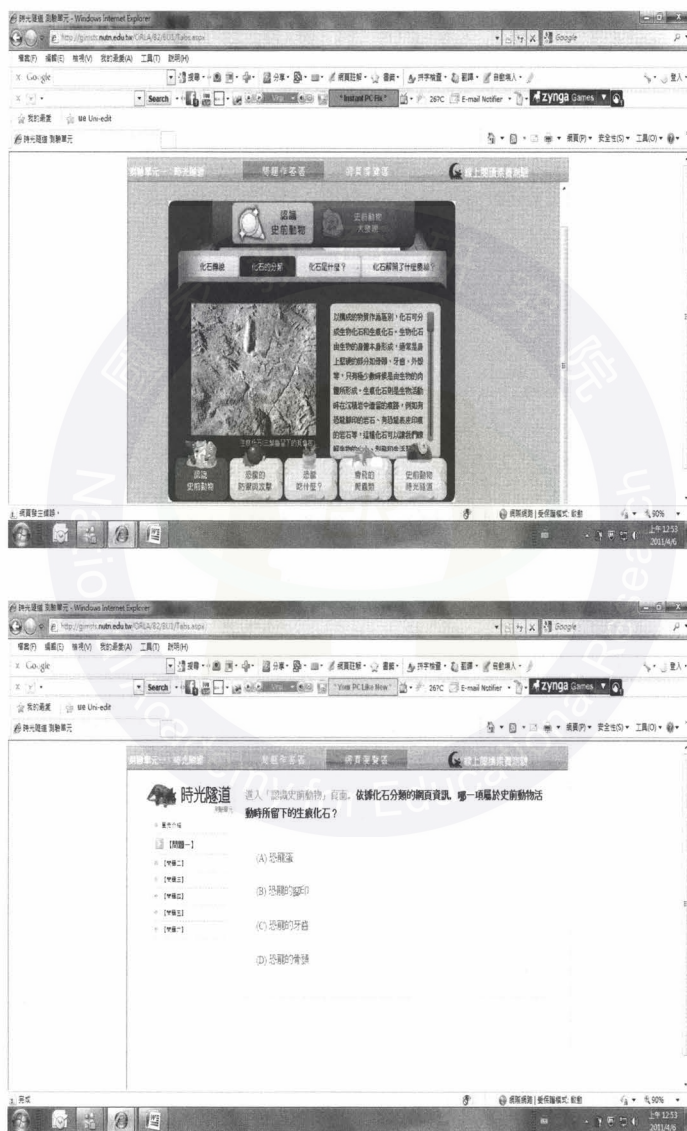
- Henry, L. A. (2005). Information search strategies on the Internet: A critical component of new literacies. *Webology*, 2(1), 1-14.
- Ho, S. C. (2011). *Results from HKPISA 2009: Hong Kong students on line: Digital technologies and reading in PISA 2009*. Retrieved from http://www.fed.cuhk.edu.hk/~hkpisa/events/2009/files/PISA09_digital_ppt_28jun.pdf
- IRA (2002). *Integrating literacy and technology in the curriculum: A position statement*. Newark, DE: Author.
- Leu, D. J. (2007). *Expanding the reading literacy framework of PISA 2009 to include online reading comprehension*. A working paper commissioned by the PISA 2009 reading expert group. Princeton, NJ: Educational Testing Services.
- Leu, D. J., Coiro, J., Castek, J., Hartman, D. K., Henry, L. A., & Reinking, D. (2008). Research on instruction and assessment in the new literacies of online reading comprehension. In C. C. Block, S. Paris, & P. Afflerbach. (Eds.), *Comprehension instruction: Research-based best practices* (pp.61-79). New York, NY: Guilford Press.
- Leu, D. J., Jr., Kinzer, C. K., Coiro, J., & Cammack, D. (2004). Toward a theory of new literacies emerging from the Internet and other information and communication technologies. In R. B. Ruddell, & N. Unrau (Eds.), *Theoretical models and processes of reading* (15th ed., pp. 1568-1611). Newark, DE: International Reading Association.
- Leu, D. J., McVerry, J. G., O'Byrne, L., Zawilinski, L., Castek, J., & Hartman, D. K., (2009). The new literacies of online reading comprehension and the no child left behind: Students who require our assistance the most, actually receive it the least. In L. M. Morrow, R. Rueda, & D. Lapp (Eds.), *Handbook of research on literacy instruction: Issues of diversity, policy, and equity* (pp. 173-194). New York, NY: Guilford Press.
- Leu, D. J., Zawilinski, L., Castek, J., Banerjee, M., Housand, B. C., Liu, Y., & O'Neil, M., et al. (2008). What is new about the new literacies of online reading comprehension? In L. S. Rush, A. J. Eakle, & A. Berger (Eds.), *Secondary school literacy: What research reveals for classroom practice* (pp.61-79). Urbana, IL: National Council of Teachers of English.
- Linacre, J. M. (2007). *Dimensionality: Contrasts and Variances*. Retrieved from <http://www.winsteps.com/winman/principalcomponents.htm>
- Mendelovits, J., Lumley, T., & Searle, D. (2009). *Assessing reading literacy in the digital age*. Paper presented at The PISA Research Conference 2009, Kiel, Germany.

- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *educational measurement* (3rd ed., pp.13-104). New York, NY: Macmillan.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.
- OECD (2007). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris, France: Author.
- OECD (2010a). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris, France: Author.
- OECD (2010b). *PISA 2009 results: What students know and can do*. Paris, France: Author.
- OECD (2010c). *PISA 2009 Results: Learning to Learn (Volume III)*. Paris, France: Author.
- OECD (2011). *PISA 2009 results: Students on line: Reading and using digital information, explores students' use of information technologies to learn (Volume VI)*. Paris, France: Author.
- Rahman, S., Yassin, S. F. M., Ishak, N. M., & Amir, R. (2008). *The use of metacognitive strategies in accessing and studying hypertext materials online*. Paper presented at the 2008 EABR & TLC Conference Proceedings, Rothenburg, Germany.
- Sit, P. S., & Cheung, K. C. (2009). *An examination of the design and implementation logistics of electronic reading assessment in PISA 2009 study: The macao experiences*. Paper presented at The PISA Research Conference, Kiel, Germany.

投稿收件日：2013 年 3 月 6 日

接受日：2013 年 12 月 3 日

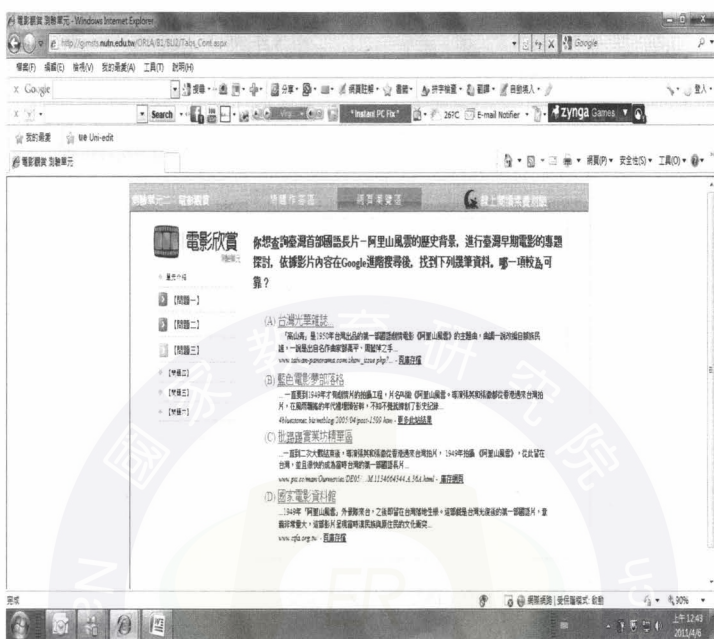
附錄 線上閱讀素養評量樣本試題



類型一：低文本處理和低航網—到指定網頁中擷取單一資訊（問題 2）



類型二：低文本處理和高航網一搜尋多重網頁以擷取單一資訊（問題 1）



類型二：低文本處理和高航網—評鑑搜尋結果資料來源的可靠性（問題 12）



類型三：高文本處理和低航網—到指定網頁中評論某一長文本（問題 16）



類型四：高文本處理和高航網—搜尋多重網頁以統整多項資訊（問題4）