

教育科學研究期刊 第六十卷第三期

2015 年，60 (3)，95-126

doi:10.6209/JORIES.2015.60(3).04



大學生基本素養測驗的發展及信度效度分析

陳柏熹

國立臺灣師範大學
教育心理與輔導學系

黃馨瑩

國立臺灣師範大學
科學教育中心

陳郁欣

國立臺灣師範大學
教育心理與輔導學系

葉泰廷

國立臺灣師範大學
教育心理與輔導學系

蘇少祖

國立臺灣師範大學
教育心理與輔導學系

摘要

本研究目的旨在發展大學生基本素養測驗並進行信度與效度評估。藉由分析國內大專院校的通識教育目標和核心素養，並參考 ATC21S 提出的 21 世紀現代學生需具備的 10 項基本素養，歸納出大學生基本素養測驗的九項素養，分別為：溝通合作、美感素養、科學思辨、資訊素養、終身學習、創新領導、問題解決、公民社會及生涯發展。測驗形式為線上多媒體情境式題型，每個題本均包含九項素養的內容，每項素養皆有二至三個題組。研究對象為全國大專校院一至四年級學生，研究樣本來自 20 校 10,958 名大學生。由效度評估結果可知，大學生基本素養測驗的題組效果不大，可以忽略，並採用部分計分模式來估計，幾乎所有試題與模式都能適配，顯示建構效度良好。試題發展過程均歷經嚴謹修審題程序，取得良好專家效度證據。此外，不同性別和年級的學生在各素養的表現上差異不大，和過去的文獻相符合，具有良好的效標關聯效度。信度證據方面，各素養能力估計誤差約在 .20~.60 logit 之間，單一題本的信度高於 .69，顯示本測驗題數雖少，但信度大致良好。整體而言，大學生基本素養測驗具良好的信度與效度。

關鍵字：信度、效度、素養評量、情境式測驗、線上測驗

通訊作者：黃馨瑩，E-mail: aessaess@gmail.com

收稿日期：2014/10/06；修正日期：2015/02/28；接受日期：2015/03/05。

壹、緒論

一、研究源起

21 世紀的高等教育面臨人才培育的挑戰，大學不只傳授學生專業知識，更需要培育學生的文化性格、職場能力與國際競爭力。歐洲在《波隆納宣言》(Bologna Declaration) 的指引下，建立歐洲高等教育區 (European Higher Education Area)，希望歐洲各國在高等教育上互相合作，以發展出可以比較的品質保證標準，而其中的品質保證標準之一，即是瞭解學生的學習成效。另外，在 2004 年的《魯汶公報》(Leuven/Louvain-la-Neuve Communiqué) (European Association for Quality Assurance [ENQA], 2009a, 2009b) 更明確指出，未來歐洲高等教育政策的改革重點為「以學生為中心的學習」。我國高等教育評鑑基金會自 2011 年起對各大學系所的評鑑，亦將以「品質保證」、「學生學習成效」為主軸 (王保進, 2011)，也在 100 學年度校務評鑑實施計畫中將「學校規劃與評核學生達成基本素養與核心能力的機制」列入評鑑學校是否有盡到確保學生學習之績效責任。由以上可知，「學生學習成果」已逐漸成為評量高等教育的主要指標，高等教育也相當重視學習成果的展示，也就是學生核心能力與基本素養的展示。

學生基本素養的表現已成為國內、外關心的議題，21 世紀技能評量與教學計畫 (Assessment & Teaching of 21st Century Skills [ATC21S], 2010) 提出了 21 世紀學生需具備的 10 項基本素養，包含創意、資訊、生涯、公民、問題解決、批判思考能力等。而國際學生評量計畫 (Programme for International Student Assessment, PISA) 也將數學、閱讀、資訊、科學等素養的評量列為瞭解各國教育成效的主要評估項目。在我國，十二年國民基本教育的配套措施三「學生生涯規劃與國民素養提升」其中一個方案為「提升國民素養實施方案」，預計 2016 年全國抽測學生在語文、科學、數學、數位、教養 (含美學) 等素養的表現情形 (國家教育研究院, 2014)。綜觀以上，評量學生的基本素養表現情形已成為國際一致認同的重要教育目標。

有鑑於高等教育注重學生學習情形，以及國際對學生素養表現的關注，我國各大專院校多根據自己學校的特色，陸續發展學生所需的基本素養與核心能力，基本素養主要倚賴全校性通識課程來培養，核心能力則由院系專業課程來養成，並搭配學校發展的課程地圖，讓學生能依據自己的興趣、職業性向，選擇自己要培養的能力與應該進修的課程，有些大專院校更進一步發展評量測驗，以利瞭解學生在各素養或各核心能力上的發展情形。然而，目前我國各大專院校發展的基本素養評量工具有幾個缺點：(一) 評量皆為自陳式量表，自陳式量表能評估學生在自我概念上的自我感受，比較適合用在態度或人格測量，但不適合評估學生在各素養上的成就表現情形 (Aiken & Groth-Marnat, 2005)。(二) 自陳式量表的題目都是單一敘述句，缺乏實際的生活情境，題目無法和實際的生活事件相互連結，導致作答結果無法反映

出學生在面對日常生活事件時的真實能力表現，也影響測驗的信度和效度。(三) 題目皆為單一題本沒有建立題庫，故題數有限，容易有試題外洩的風險，且試題樣本不足也較難反映出完整的評量架構，在評估學生素養表現時證據會稍顯不足。

為改善上述問題，本研究目的為透過嚴謹的程序發展電腦化大學生基本素養測驗，希望藉由多媒體情境式測驗工具，協助大專院校瞭解校內基本素養教育的培育成效，並使學生能檢核其各項基本素養的學習情形，作為未來選課規劃與生涯發展之參考。本研究發展之大學生基本素養測驗有幾個特點：(一) 包含認知能力與情意態度測驗：坊間大學生的素養評量工具常以自陳式量表為主，但根據國際大型資料庫的研究結果可知，學生對自身能力的感知和自己真正具備多少能力兩者並不一致，故自陳式量表並不能代表學生具備多少能力，大學生基本素養測驗包含認知能力和情意態度測驗，評量學生在基本素養的實際學習狀況與認知能力，並捕捉學生在各素養的自我信心與實踐能力等情意態度。(二) 多媒體情境式命題：以常見的日常生活事件進行情境式命題，希望在貼近學生生活的前提下，較準確地測量學生的表現。(三) 建立題庫：建立完善嚴謹的題庫，並附上說明詳盡的測驗指導手冊，使用者可在不同測驗精準度的前提下，挑選所需的測驗試題數。

本研究將說明大學生基本素養測驗的編製發展情形，並根據 2013 年蒐集而得的結果說明測驗信度與效度，以供未來國內、外有興趣的人員參考。

二、測驗發展流程

(一) 訂定九項素養內涵

本研究先蒐集國內知名大專院校的通識教育目標和核心素養(如表 1 所示)，並參考 ATC21S 中所提出的 21 世紀現代學生需具備的 10 項基本素養，將國內、外重視的素養歸納如表 1 所示，本研究加以統整並評估社會所需人才之趨勢後，總結出大學生基本素養測驗包含九項素養，分別為：溝通合作、美感素養、科學思辨、資訊素養、終身學習、創新領導、問題解決、公民社會、生涯發展，而九項素養的內涵整理如表 2 所示。由於本研究對基本素養的選取是以 ATC21S 為基礎，再加入國內大學共同提及的素養作為選取大學生基本素養測驗的目標，以期未來能符合國際潮流。其中，雖然僅兩校將生涯發展列入，但 ATC21S 將其列為學生未來重要能力，所以在大學生基本素養測驗中亦將之納入。此外，雖然有些學校將國際視野相關概念納入基本素養，但各校的定義及內涵較分歧，舉凡有競爭力、國際觀、學習文化與理解國際等，其中多元文化的內容已納入本測驗的公民社會素養中。其他有關國際視野的定義因各校較不一致且模糊不清，因此暫時不將國際視野納入本測驗中，待未來若能將各校對於國際視野的定義釐清並統一後，再考慮將國際視野納入大學生基本素養測驗中。

表 1
大專院校通識教育目標和核心素養摘錄

素養名稱	國立中央大學	國立臺灣大學	國立成功大學	國立清華大學	國立交通大學
美感素養	<ul style="list-style-type: none"> 美感品味 	<ul style="list-style-type: none"> 藝術與人文涵養 			
溝通合作	<ul style="list-style-type: none"> 良好的人際關係 正直與容忍他人意見的容他性 溝通與表達的能力、團隊合作 	<ul style="list-style-type: none"> 溝通表達與團隊合作 	<ul style="list-style-type: none"> 語文與溝通能力 	<ul style="list-style-type: none"> 自我瞭解與溝通表達 	
創新領導	<ul style="list-style-type: none"> 經營與管理的能力 		<ul style="list-style-type: none"> 創新與領導能力 		
公民社會	<ul style="list-style-type: none"> 履行公民責任 瞭解尊重多元文化 	<ul style="list-style-type: none"> 公民素養 社會關懷 	<ul style="list-style-type: none"> 多元觀點與社會實踐 	<ul style="list-style-type: none"> 公民責任 品德涵養 人文及在地關懷 	
資訊素養	<ul style="list-style-type: none"> 使用電腦與資料搜集及處理的能力 		<ul style="list-style-type: none"> 資訊科技與媒體素養 	<ul style="list-style-type: none"> 資訊素養 	
生涯發展					
終身學習	<ul style="list-style-type: none"> 終身學習與應用知識的意願 				
問題解決	<ul style="list-style-type: none"> 獨立批判與分析、創意思考 選擇最適工作、規劃與組織 解決問題與決策的能力 	<ul style="list-style-type: none"> 獨立思考與創新 	<ul style="list-style-type: none"> 思考與判斷能力 	<ul style="list-style-type: none"> 邏輯推理與批判思考能力 發現及解決問題的能力 跨界多元思考的能力 	
科學思辨	<ul style="list-style-type: none"> 吸收科學知識的能力 		<ul style="list-style-type: none"> 科學思維與反思 		
其他	<ul style="list-style-type: none"> 國際視野、閱讀中英文、通順寫作、數理與邏輯的能力、重承諾與全力投入工作的專業精神..... 	<ul style="list-style-type: none"> 道德思辨與實踐、身心健康管理、人文關懷、國際視野 	<ul style="list-style-type: none"> 國際視野、人文素養 	<ul style="list-style-type: none"> 人文藝術陶冶、國際觀、外語能力 	

(續)

表 1
大專院校通識教育目標和核心素養摘錄

素養名稱	國立中正大學	國立東華大學	國立嘉義大學	輔仁大學	長庚大學
美感素養	<ul style="list-style-type: none"> 人文藝術欣賞能力 美感與藝術欣賞 	<ul style="list-style-type: none"> 情意美感與創造思考的能力 			
溝通合作	<ul style="list-style-type: none"> 溝通表達與團隊合作 	<ul style="list-style-type: none"> 溝通合作與社會實踐的能力 語文表達 		<ul style="list-style-type: none"> 人際溝通行為 合作 團隊合作行為 	<ul style="list-style-type: none"> 語文閱讀、理解與表達應用能力 協調、溝通與團隊合作能力
創新領導			<ul style="list-style-type: none"> 關懷、參與與領導能力 		
公民社會	<ul style="list-style-type: none"> 校園環境薰陶、多元文化 公民素養與社會參與 道德思辨與實踐 	<ul style="list-style-type: none"> 民主與法治之公民責任的能力 在地關懷與全球視野的能力 	<ul style="list-style-type: none"> 在地視野 	<ul style="list-style-type: none"> 服務 	<ul style="list-style-type: none"> 尊重多元文化與國際視野 現代公民民主素養與思辨能力
資訊素養	<ul style="list-style-type: none"> 資訊能力 	<ul style="list-style-type: none"> 資訊之統整應用的能力 	<ul style="list-style-type: none"> 資訊 	<ul style="list-style-type: none"> 資訊能力 	
生涯發展	<ul style="list-style-type: none"> 生命探索與生涯規劃 				
終身學習		<ul style="list-style-type: none"> 卓越之專業智能與終身學習的能力 			<ul style="list-style-type: none"> 終身學習的能力
問題解決	<ul style="list-style-type: none"> 思考判斷能力 思考與創新 問題分析與解決 		<ul style="list-style-type: none"> 獨立思考及創新能力 	<ul style="list-style-type: none"> 問題分析及解決行為 創新行為 	<ul style="list-style-type: none"> 獨立思考與負責任的能力
科學思辨					
其他	<ul style="list-style-type: none"> 培養學術能力、基本數理能力、國際視野、建立學生對人、社會及自然各種層面的基礎認識 	<ul style="list-style-type: none"> 康健自由、敦厚樸實的 身心特質、博雅多元的 識見與人文素養 	<ul style="list-style-type: none"> 博雅知能、全球視野、 自主自律、語文、體適 能 	<ul style="list-style-type: none"> 中文能力、英文能力、 誠信、謙卑、勇氣、卓 越 	<ul style="list-style-type: none"> 跨領域的知識與整合 能力、蒐集資料與口頭 報告的能力

(續)

表 1
大專院校通識教育目標和核心素養摘錄（續）

素養名稱	東吳大學	淡江大學	中山醫學大學	國立臺灣科技大學	國立臺北科技大學
美感素養			• 審美能力	• 藝術創造能力	• 文學與藝術
溝通合作		• 表達能力與人際溝通	• 溝通能力	• 溝通協調能力	
創新領導		• 團隊合作與公民實踐			
公民社會	• 民主法治與公民素養 • 國際視野與社會認知能力			• 公民責任能力 • 多元關懷能力	• 民主與法治
資訊素養	• 資訊與科技運用能力	• 科技應用與資訊處理能力	• 資訊科技應用能力	• 培養科技與人文素養 均衡發展	
生涯發展		• 專業發展與職涯規劃			
終身學習		• 自我瞭解與主動學習		• 終身學習能力	
問題解決	• 思考與判斷能力	• 主動探索與問題解決		• 解決問題能力 • 訓練理性思考與判斷	
科學思辨	• 科學素養		• 邏輯與計算，推理能力	• 自然科學與邏輯推理	
其他	身心均衡、專業實踐、人文精神與健全身心、語文能力、歷史思維能力	學習文化與理解國際、洞察未來與永續發展	專業知能、創造能力、倫理涵養與道德思辨能力	國際競爭能力；塑造全人化的人格、生活價值與人生觀	歷史思維與世界文明、社會經濟與管理、哲學思考與倫理

（續）

表 1
大專院校通識教育目標和核心素養摘錄（續）

素養名稱	國立雲林科技大學	國立高雄應用科技大學	ATC21S
美感素養		<ul style="list-style-type: none"> 高超的欣賞能力 	
溝通合作	<ul style="list-style-type: none"> 合作溝通 		<ul style="list-style-type: none"> 人際溝通 (communication) 團隊合作 (collaboration, team work)
創新領導			
公民社會	<ul style="list-style-type: none"> 公民素養 社會關懷 多元文化 	<ul style="list-style-type: none"> 多元化的適應力 	<ul style="list-style-type: none"> 公民觀點 (citizenship-local and global) 個人及社會責任 (personal and social responsibility)
資訊素養			<ul style="list-style-type: none"> 資訊素養 (information literacy) 資訊通訊科技素養 (Information and communications technology)
生涯發展			<ul style="list-style-type: none"> 生涯規劃 (life and career)
終身學習			<ul style="list-style-type: none"> 學習學習、後設認知 (learning to learn, metacognition)
問題解決	<ul style="list-style-type: none"> 獨立思考 		<ul style="list-style-type: none"> 問題解決 (problem solving) 思考決策 (decision making)
科學思辨		<ul style="list-style-type: none"> 理性的科學精神 敏銳的分析能力 	<ul style="list-style-type: none"> 批判思考 (critical thinking)
其他	認真學習、力行負責、國際視野	健康的自我概念、清晰的價值系統、通達的倫理觀念、睿智的道德情操.....	創造力 (creativity and innovation)

註：學校資訊擷取至各大專院校通識中心及相關網頁，並蒐集至 2011 年 11 月止。

表 2

大學生基本素養測驗的認知層面與情意態度層面的內涵

素養名稱	認知能力層面的內涵	情意態度層面的內涵
創新領導 (Innovative leadership)	能夠帶領團隊達成任務目標，並且引導新思維的產生。對於任務的規劃、實行與結束皆有適當且完善的應對方式。	具備相當的自信、開放性以及道德觀，並且對於與他人互動、帶領團隊和達成團隊任務有較高動機。
問題解決 (Problem solving)	面對問題時能夠分析關鍵的原因，有效的利用資源及經驗提出可能的解決方式，能整合各種解決方案的優缺點，且能發展解決問題的計畫，以及評估方案是否有效並進行改進。	對於分析問題的關鍵原因有足夠的信心，願意在日常生活中使用策略解決問題，並且理性地使用策略。
終身學習 (Lifelong learning)	能知道如何評估自己學習的優、劣勢，在學習中知道如何專注精神及保持彈性，能自我監控學習狀況並促進自我成長，能選擇合適的資源進行學習。	願意監控、提升學習動機、自我效能、自尊等自我概念，積極樂觀的學習，遇到挫折時願意自我調適以面對挑戰。
溝通合作 (Communication and collaboration)	能注意到他人意見，正確理解訊息，並知道如何運用適當的表達與溝通技巧闡述個人見解。在團隊中能進行有效分工並參與團隊討論，達成團隊目標。	喜歡與他人建立正向的互動關係，在團體中願意表達自己的意見也願意聆聽，尊重他人感受，能控制自己的情緒不影響溝通。
公民社會 (Socialization and citizenship)	瞭解關懷社會弱勢的方式及落實社會正義應有的價值，能發揮公民力量凝聚社群共識，並知道如何採取務實的公民行動，實際參與社會改良活動。	願意參與公共事務，願意探索社會中的民主價值，對於他人的需求會尊重並關心，對社會中存在的個別差異能給予尊重與包容。
美感素養 (Aesthetics literacy)	能注意及區辨生活中具美感的事物，並連結自身生活經驗和聯想力以體驗生活美感。懂得欣賞，知道如何提升自己的美感知能，並體認各種藝術價值及其文化脈絡。	能關心生活周遭細節，追求生活品質，品味生活的細微之處，並能主動且積極分配時間去瞭解、參與和規劃美感活動。
科學思辨 (Scientific thinking)	能認識自然科學的基本概念與應用價值，能運用所習得的科學方法探究科學知識，並且將科學知識應用於日常生活行為上。	願意瞭解及探求知識，常使用各種不同的角度來檢視問題，對於研究結果不作過度的推論，也不輕信權威或缺乏證據支持的敘述。

表 2

大學生基本素養測驗的認知層面與情意態度層面的內涵（續）

素養名稱	認知能力層面的內涵	情意態度層面的內涵
資訊素養 (Information literacy)	能區辨各種資訊媒體的功能與限制，瞭解資訊倫理與資訊安全，並能有效運用網路獲取資訊。	能以更敏銳且正面的態度面對資訊科技，對資訊科技有較高的接受度，也會注重資訊倫理的議題。
生涯發展 (Life and career)	能瞭解自己的能力性格與不同工作的符合程度，能考量不同因素以協調較好的生涯決策，並能規劃與執行具體可行的生涯目標。	能關注自己的未來，在生涯規劃方面會積極地探索和嘗試，並相信自己能決定自己的未來，有信心來面對和克服生涯挑戰與阻礙。

（二）訂定素養定義與向度

以溝通合作為例，由文獻可知，Verderber 與 Verderber (1995) 將人際溝通定義為有意義的互動歷程，雙方彼此對於溝通當時與溝通後所形成的意義均負有責任。對於人際溝通的成分，Scott (1998) 認為溝通三個要素為：知識、技巧和態度。Sypher、Applegate 及 Church (1981) 認為人際溝通包含表達、傾聽、同理心、行為多樣性、適應性、專注力、澄清、溝通控制，其中有偏向人格特質的同理心及適應性，也有像是技巧性的傾聽及澄清等構念。Gardner (1983) 的多元智能理論中的人際智慧是指能善用言語或適當策略來進行有效的人際互動，並善於用合作團隊的方式處理事件。綜合以上定義，可延伸出溝通合作的認知向度為：溝通能力、同理心、瞭解與尊重個別差異、互助合作；情意態度向度則包含：表達意願、傾聽同理、人際互動、理解與控制。本研究也依照相同程序，訂出各素養的定義和所包含的面向與指標，各素養詳細內容請參見大學生基本素養指導手冊（國立臺灣師範大學雲端測驗中心，2014）。

（三）訂定測驗題型與長度

本測驗所有試題皆為情境式題組，情境取自大學生日常生活中接觸到的事物，為使題目更接近其日常生活，有些情境或選項以圖片或影片呈現。每個題組包含三至五題認知能力試題與三至五題情意態度試題。認知能力試題有兩種題型，一種為四選一的選擇題，另一種為叢集式是非題（cluster true-false item）。情意態度試題則為李克特四點量表（Likert scale），選項為非常同意、同意、不同意、非常不同意。

學生作答的題本皆包含九個素養的題目，每個素養有二至三個題組，這二至三個題組會測量該素養的所有向度，學生一次測驗大約作答 20~25 個題組，完成作答約需 50~60 分鐘，所有題目均為電腦化線上施測，試題範例可參見附錄一。

(四) 命題和修審題程序

大學生基本素養測驗是依照標準化測驗編製流程發展測驗(郭生玉, 2012; 陳柏熹, 2011), 先根據蒐集而得的國內、外文獻, 確立各素養的重要向度後, 編製範例試題與命題手冊, 舉辦命題研習會議, 邀請高中職教師、大學教師、大學相關科系學生與研究生參與訓練並協助命題, 回收題目之後, 再依命題者的情境製作動畫或影片, 並進行題目初步修改。之後請各領域三至五位專家學者進行修審, 判斷題組情境與題目內容是否有專業知識上的錯誤、敘述表達適當與否、題目是否符合命題向度。試題修改完之後, 將試題組成六套題本, 每套題本均包含九個素養, 採用不等組變化式共同試題設計(non-equivalent group with variable anchor test design)(Chen, Kuo, & Sung, 2011), 每套測驗題本之間以共同試題連結(linking), 使所有試題參數能等化至同一量尺上。

(五) 測驗編製與資料蒐集

截至 2013 年年底已完成四次資料蒐集, 四次蒐集時間分別是 2012 年 11 月至 2013 年 1 月、2013 年 3 月至 4 月、2013 年 5 月至 6 月、2013 年 11 月至 12 月, 前三次邀請國立臺灣師範大學與臺北醫學大學的學生自由上網填答, 第四次則以便利取樣方式抽取全國 20 所大專院校參與施測。四次預試包含六所公立大專院校 4,590 名學生、14 所私立大專院校 6,368 名學生, 其中一般大學有 12 所共 6,746 名學生, 科技大學有八所共 4,212 名學生, 有效樣本共為 10,958 名學生。

學生背景資料如表 3 所示, 本研究根據 Holland (1973) 提出的六大類職業類型人格, 將全國學生根據其系所分為實用型(工程等類科)、研究型(數理類科)、事務型(財會類科)、藝術型(藝文類科)、商業型(商管等類科)、社會型(社會教育等類科), 各類科歸類方式如附錄二所示, 整體而言, 除事務類型僅 394 人, 其餘類型人格皆有 1,000 人以上, 年級和性別的分布上, 低年級的學生較多, 高年級的學生較少, 女生又比男生多一些。

貳、測驗效度評估

本研究將從建構效度、專家效度、效標關聯效度等三面向, 進行大學生基本素養測驗的效度評估, 以下將分述效度檢驗的方法與效度評估結果。

一、效度檢驗方法

(一) 建構效度

建構效度包含模式檢定、試題的適配度兩部分。在模式檢定上, 由於測驗皆為題組試題,

表 3
大學生基本素養測驗學生背景資料

職業類別	一年級		二年級		三年級		四年級		小計
	男	女	男	女	男	女	男	女	
藝術型	295	797	97	230	111	281	50	162	2,023
實用型	654	209	410	66	551	110	249	73	2,322
研究型	589	449	157	125	177	108	169	95	1,869
商業型	166	281	119	235	116	323	156	349	1,745
社會型	254	994	63	513	68	272	91	333	2,588
事務型	22	50	44	48	34	121	17	58	394
小計	1,980	2,780	890	1,217	1,057	1,215	732	1,070	

若題組效果量太大，將使測量精準度被過度高估（Wang & Wilson, 2005），單向度（unidimensionality）的假設也無法被滿足，需以題組反應模式進行分析，但題組反應模式會增加估計成本，故本研究先檢驗題組效應的大小，若題組效應不大，表示對測量精準度的影響不大，則可簡化以非題組模式分析。為檢視本測驗試題是否需考慮題組效果，研究團隊以第一次回收而得的資料（施測時間為 2012 年 11 月至 2013 年 1 月）為例，將各素養資料分別以 Rasch 模式（Rasch, 1960）與題組反應模式（testlet response model）進行分析，比較資料與兩個模式的適配程度，以確定題組效果的嚴重情形，若題組效果大到不容忽視，則以題組反應模式分析，若題組效果不大，將把叢集式是非題得分加總轉為多元計分，例如：第一題叢集式是非題包含四個小題，每小題答對給 1 分，答錯給 0 分，將四小題的分數加總，變成第一題叢集式是非題的分數，該題的分數範圍變成 0~4 分，轉換之後，再利用部分計分模式（partial credit model, PCM）進行各素養的資料分析。確定估計模式之後，四次蒐集而得的資料將同時行分析。

建構效度是指測驗是否能測出所欲測量的潛在特質架構，因此在試題反應理論（item response theory, IRT）中要驗證建構效度時，可以分析資料符合 IRT 模式的程度，作為建構效度之參考。本測驗參考 Wright（1994）之研究，若 infit MNSQ 介於 1 ± 0.3 之間， t 檢定值介於 ± 2.0 之間，則表示試題和模式適配度良好。除此之外，因本研究將叢集式是非題得分加總轉為多元計分，這類題目的閾難度值（step）應該不會有翻轉（reverse）的情形，較低分數的閾難度值必須低於較高分數的閾難度值，亦即得到高分比得到低分還要困難，本研究也會檢視是否有翻轉的情形。

（二）參數估計

本研究採用貝式期望後驗法（expected a posterior, EAP）進行能力值估算，EAP 估計法適合用在樣本數少、題數少的情況，且能估算出全部答對和全部答錯學生的能力值。Bock 與

Mislevy (1982) 也指出，在分分計較的實務情境下，考量先驗分配時，使用 EAP 估計法得到的結果，較最大概似估計 (maximum likelihood estimator, MLE) 的結果公平合理。大學生基本素養測驗每個素養一次只施測二至三個題組，在題數少的情況，用 EAP 估計可利用先驗分配 (prior distribution) 增加估計精準度，且本研究的樣本遍及全國大專院校，樣本數大、異質性也大，可以建立相對穩定的先驗分配，故選擇以 EAP 估計參數。

(三) 專家效度與效標關聯效度

為判斷各個素養是否測得該素養的特質，每道試題均邀請五至九位專家學者判斷：每個題目的內容取材是否符合所對應之向度描述 (向度符合度)、題目的敘述方式是否適切 (表達適切性)，每個判斷給予 1~5 分，分數愈高代表題目的向度符合程度愈佳，且題目敘述方式愈適當，表示愈具專家效度。

關於測驗效標的選取，多數研究會讓學生額外做相關的測驗或問卷以當成效標，但本研究考量坊間較無相關的基本素養測驗，以及本測驗的施測時間較長，學生多半不願再花時間做其他測驗，故本研究改用檢測性別差異、年級差異、各類型科系的優勢和弱勢素養等分析結果，以作為效度之輔佐。性別差異是以四次資料進行認知能力和情意態度的 t 檢定分析，高低年級部分，本研究將大學一、二年級的學生定義為低年級學生，三、四年級學生則定義為高年級學生，因許多學校只有高年級或低年級的學生作答，為怕資料有所偏頗，本研究以高、低年級參與人數皆有 500 人以上的某大學資料，進行低年級與高年級的學生在各素養認知能力和情意態度的 t 檢定。

二、效度評估結果

(一) 建構效度

1. 模式檢驗

表 4 呈現以 Rasch 模式與題組反應模式分析而得的偏差平方和 (deviance)、自由度 (degree of freedom) 以及兩種模式的卡方檢定 p 值，結果顯示，大部分的素養用這兩種模式分析並無顯著差異，只有資訊素養的資料在 Rasch 模式與題組反應模式的適配差異程度達顯著 ($p < .001$)，以題組反應模式進行分析的偏差平方和顯著比用 Rasch 模式分析的偏差平方和小，顯示資訊素養的資料和題組反應模式較適配。進一步檢視資訊素養的題目，發現該素養有很多叢集式是非題，猜測這可能是導致題組效果較明顯的原因，故素養發展團隊嘗試將資訊素養的叢集式是非題得分加總轉為多元計分，例如：第一題叢集式是非題包含四個小題，每小題答對給 1 分，答錯給 0 分，將四小題的分數加總，變成第一題叢集式是非題的分數，該題的分數範圍變成 0~4 分，再利用 PCM 進行資料分析。研究顯示，利用 PCM 分析時，其偏差平方和為 3,094.58 (自由度為 25)，小於用題組反應模式分析之偏差平方和值 3,094.63 (自由

表 4

大學生基本素養測驗 Rasch 模式與題組反應模式之適配指標比較

素養名稱	Rasch模式		題組反應模式		兩種模式差異檢定 <i>p</i> 值
	偏差平方和	<i>df</i>	偏差平方和	<i>df</i>	
溝通合作	2,198.65 [■]	10	2,211.35	16	
美感素養	3,422.06	16	3,415.52 [■]	22	.37
科學思辨	2,706.76 [■]	11	2,721.75	17	
資訊素養	4,986.59	32	4,950.21 [■]	40	< .01**
終身學習	3,753.91	15	3,745.20 [■]	23	.37
創新領導	2,488.02	11	2,486.73 [■]	17	.97
問題解決	5,059.90	27	5,048.14 [■]	33	.07
公民社會	2,731.17 [■]	14	2,741.00	20	
生涯發展	3,328.17 [■]	12	3,336.47	18	

註：(1)偏差平方和旁加註「[■]」代表該模式較佳；(2)溝通合作、科學思辨、公民社會、生涯發展因 Rasch 模式的偏差平方和已經小於題組反應模式的偏差平方和，故不再進一步進行差異檢定。

***p* < .01.

度為 34)，且兩者進行卡方檢定未達顯著差異，表示利用 PCM 模式可以解決題組效果的問題。根據以上結果，本研究將四次蒐集而得的學生資料以同時估計法等化，各素養並採用 PCM 進行分析。

2. 試題適配度

各素養的試題難度和加權的 MNSQ 如表 5 所示，所有題目的 MNSQ 皆落在 1 ± 0.3 之間，*t* 檢定值介於 ± 2.0 之間，符合 Wright (1994) 之研究標準。資料進行初次分析時，有 8% 的題目闕難度有翻轉的情形，這類題目往往有某些得分人數過少的狀況，例如：計分為 0~4 分的叢集式是非題，只有 10 個人得到 1 分。對於此種情形，本研究將人數少的得分合併到相鄰得分，例如：若得到 2 分的人數過少，則將 2 分和 1 分視為相同分數。經合併計分之後，只剩下 5% 的試題有闕難度翻轉的情形。

另外，素養認知試題的平均難度與難度範圍也可從表 5 得知，從平均難度來看，各素養為中等偏易，難度範圍大多落在 -3~2 之間。其中美感素養、溝通合作的平均難度雖偏簡單，但難度範圍分別介於 -4.15~3.27、-4.13~2.53，顯示仍然有難度較高的題目。資訊素養的平均難度則為 -1.55，難度範圍界在 -4.45~0.41，偏難的試題較少，需要增加較難的題目。而問題解決素養的平均試題難度接近 -0.14，平均試題難度適中。整體來說，各素養皆有簡單到難的試題，但未來仍需增加較困難之試題。

表 5

大學生基本素養認知能力試題難度平均值與範圍

單位：logit			
素養名稱	平均難度	難度範圍	適配度（MNSQ）範圍
溝通合作	-1.29	-4.15~3.27	0.95~1.19
美感素養	-0.95	-4.13~2.53	0.94~1.03
科學思辨	-0.52	-3.33~2.05	0.95~1.12
資訊素養	-1.55	-4.45~0.41	0.95~1.18
終身學習	-0.36	-4.21~3.72	0.93~1.03
創新領導	-0.59	-2.18~3.69	0.96~1.05
問題解決	-0.14	-3.46~3.94	0.93~1.14
公民社會	-0.85	-4.07~2.30	0.97~1.03
生涯發展	-0.64	-4.40~1.53	0.95~1.05

（二）專家效度

表 6 為各素養專家效度分析結果，結果顯示在認知能力測驗中，除了溝通合作及創新領導兩個素養的表達適切性的平均分數未達 3.5 分外，其餘各素養的向度符合度及表達適切性的平均分數皆在 3.5 分以上；在情意態度測驗中，除了溝通合作素養在表達適切性的平均分數未達 3.5 分之外，其餘各素養的向度符合度與表達適切性的平均分數皆在 3.5 分以上。顯示各素養在題目內容的設計上均能測量到該素養所要測量的向度，各素養均具有良好的專家效度，而平均數未達 3.5 分的素養也已針對專家建議加以修改。

（三）效標關聯效度

1. 性別差異分析結果

表 7 為各素養認知能力測驗性別差異分析結果， t 檢定值顯示只有科學思辨素養在性別上無顯著差異，學生在其他八個素養的認知能力測驗中，呈現明顯性別差異，女大學生的表現皆優於男大學生。但進一步檢視效果量，性別影響在各素養所造成的效果量，除了溝通合作、創新領導、公民社會三個素養上女生優於男生的效果量大於 .20 之外，其他素養皆小於 .20，根據 Cohen's d 的準則（Cohen, 1988, 1992），本結果的效果量偏小， t 值會顯著可能是樣本數太大造成的。整體來說，男生和女生在各素養的認知能力表現差距不大，此結果大致上與過去相關研究結果相符（林煥祥、劉聖忠、林素微、李暉，2008；陳儀蓉、黃芳銘，2006；Masters & Sanders, 1993；McManus & Furnham, 2006；Organization for Economic Co-operation and Development [OECD], 2007a, 2007b；Rottinghaus, Day, & Borgen, 2005；Rowland, 2004）。以科學思辨為例，根據 PISA 2000 的資料，各國在科學素養的表現上是否會因性別而有所不同？此點

表 6

大學生基本素養測驗專家效度評估

素養名稱	認知能力		情意態度	
	向度符合度	表達適切性	向度符合度	表達適切性
溝通合作	3.66 (2.33~4.50)	3.49 (2.33~4.50)	3.55 (2.50~4.30)	3.32 (2.57~3.78)
美感素養	4.44 (4.20~4.90)	4.16 (4.00~4.60)	4.52 (4.20~4.86)	4.34 (4.00~4.60)
科學思辨	4.30 (2.80~4.80)	3.99 (2.25~4.80)	4.29 (3.00~4.80)	4.29 (3.50~4.80)
資訊素養	3.79 (2.50~4.17)	3.81 (3.50~4.25)	3.93 (3.50~4.50)	3.93 (3.50~4.50)
終身學習	4.45 (3.80~5.00)	4.12 (2.50~5.00)	4.79 (4.25~5.00)	4.37 (3.75~5.00)
創新領導	3.49 (2.17~4.67)	3.24 (2.00~4.33)	4.26 (3.25~4.88)	4.29 (3.50~4.88)
問題解決	4.26 (2.63~5.00)	4.04 (3.00~4.75)	4.61 (4.00~5.00)	4.42 (3.63~4.88)
公民社會	4.88 (4.67~5.00)	4.38 (3.00~5.00)	4.82 (4.50~5.00)	4.58 (4.17~5.00)
生涯發展	4.57 (3.29~5.00)	4.16 (3.40~4.75)	4.79 (4.33~5.00)	4.68 (4.28~5.00)

註：括號內的值為專家所評的分數範圍。

表 7

大學生基本素養認知能力測驗性別差異分析結果

素養名稱	年級	人數	平均分數	標準差	t值	效果量
溝通合作	男	4,526	98.25	10.26	16.29**	.324
	女	6,094	101.32	8.61		
美感素養	男	4,514	99.63	4.53	7.34**	.146
	女	6,058	100.28	4.39		
科學思辨	男	4,089	99.99	7.42	0.07	.003
	女	5,697	100.01	7.18		
資訊素養	男	4,178	99.48	7.87	5.95**	.122
	女	5,754	100.39	6.98		

表 7

大學生基本素養認知能力測驗性別差異分析結果（續）

素養名稱	年級	人數	平均分數	標準差	<i>t</i> 值	效果量
終身學習	男	3,670	99.67	4.36	6.14**	.132
	女	5,264	100.23	4.11		
創新領導	男	4,040	99.54	3.37	11.34**	.233
	女	5,586	100.33	3.41		
問題解決	男	4,439	99.37	5.92	9.76**	.194
	女	5,885	100.49	5.63		
公民社會	男	4,489	99.52	3.76	11.55**	.231
	女	6,001	100.36	3.52		
生涯發展	男	4,044	99.80	5.52	3.17**	.065
	女	5,562	100.16	5.60		

***p* < .01.

尚未有一致的定論，從 PISA 2006 整體資料來看，全球男生和女生在科學素養上的表現並沒有顯著差異（OECD, 2007a, 2007b; Yip, Chiu, & Ho, 2004），但從 TIMSS 的資料則顯示，男生在科學的表現顯著比女生好。TIMSS 之所以和 PISA 有不同的結果，可能是因為 PISA 的問題比較偏生活能力和問題解決，TIMSS 的試題比較偏向科學知識和科學技能，由於試題取向不同，造成 PISA 和 TIMSS 的結果不同。而本研究科學思辨的題目，偏向因果推論和批判反思，比較類似 PISA 的試題，而本研究也發現，不同性別在科學思辨的表現上沒有顯著差異，此結果和 PISA 的結論相仿（Schmidt, Raizen, Britton, Bianchi, & Wolfe, 1997; Yip et al., 2004）。

表 8 為各素養情意態度測驗性別差異分析結果，*t* 檢定值顯示受測者在資訊素養的情意態度測驗中，沒有明顯性別差異，其他素養在情意態度表現上則有顯著的性別差異。其中在溝通合作、美感素養、科學思辨、問題解決、公民社會、生涯發展素養上，男生的態度較女生正向，在終身學習和創新領導素養則是女生的態度較男生正向。但進一步檢視效果量，性別影響在各素養所造成的效果量皆小於 .20，根據 Cohen's *d* 的準則（Cohen, 1988, 1992），本結果的效果量很小，*t* 值會顯著可能是樣本數太大造成的，故男生和女生在各素養的情意態度表現差距不大，此結果與過去相關研究結果相符（林煥祥等，2008；廖添富、黃景裕，1995；Liu & Sun, 2012; Ma, Ding, & Li, 2009; OECD, 2007a, 2007b; Pope-Davis & Twing, 1991）。以科學思辨為例，根據 PISA 2006 的資料，男、女生在不同面向的科學態度上，有不同的表現情形，以臺灣的資料來看，在科學學習興趣上，男生顯著優於女生，男生在科學自我概念上也表現得比女生好，但全球資料則顯示，性別差異雖存在於自我概念上，但兩者的差異並不大。而在環境議題態度上，臺灣資料顯示男生的覺知程度比女生好，女生的關心程度則優於男生，

表 8

大學生基本素養情意態度測驗性別差異分析結果

素養名稱	年級	人數	平均分數	標準差	t值	效果量
溝通合作	男	4,527	100.60	14.03	-3.75***	.074
	女	6,096	99.61	12.70		
美感素養	男	4,518	100.39	14.00	-2.52*	.050
	女	6,061	99.71	13.18		
科學思辨	男	4,086	100.64	15.13	-3.56***	.074
	女	5,686	99.57	13.96		
資訊素養	男	4,170	100.14	10.07	-1.15	.023
	女	5,762	99.92	8.80		
終身學習	男	3,667	99.24	15.59	4.28***	.093
	女	5,247	100.65	14.69		
創新領導	男	4,041	99.35	14.30	4.09***	.086
	女	5,587	100.51	12.68		
問題解決	男	4,446	100.59	16.83	-3.45*	.070
	女	5,896	99.50	14.36		
公民社會	男	4,494	101.02	16.10	-5.70***	.113
	女	6,019	99.28	14.71		
生涯發展	男	4,001	100.88	21.61	-3.53***	.074
	女	5,432	99.34	20.08		

* $p < .05$. *** $p < .001$.

在樂觀程度上，二者沒有顯著差別，在責任自覺程度上，女生顯著優於男生（OECD, 2007a, 2007b; Yip et al., 2004）。分各個子向度去檢視男生和女生在科學態度的表現時，男生和女生各有所長，但若整體來看，男生和女生在科學態度上則沒有非常顯著的差異。本研究科學思辨素養的態度題目，包含四個面向，整體來看，男、女生的表現並無顯著的不同，結果和過去文獻一致。

2. 年級分析結果

分析結果如表 9、10 所示， t 檢定值顯示在美感素養和科學思辨的認知能力上，高、低年級有顯著差異，低年級的美感素養略優於高年級。在終身學習、問題解決、生涯發展的情意態度上，高、低年級也有顯著差異，低年級的學生在終身學習、創新領導、生涯發展的情意態度略微高於高年級的學生。但進一步檢視效果量，年級在各素養所造成的效果量皆小於 .20，根據 Cohen's d 的準則（Cohen, 1988, 1992），本結果的效果量很小， t 檢定量會顯著可

表 9

大學生基本素養認知能力測驗年級差異分析結果

素養名稱	年級	人數	平均分數	標準差	t值	效果量
溝通合作	低年級	1,973	100.77	8.56	0.01	< .001
	高年級	685	100.77	8.31		
美感素養	低年級	1,941	101.22	4.43	2.04*	.093
	高年級	648	100.82	4.20		
科學思辨	低年級	1,967	101.99	6.92	2.58*	.116
	高年級	683	101.20	6.67		
資訊素養	低年級	1,904	100.89	6.80	-0.65	.030
	高年級	636	101.09	6.69		
終身學習	低年級	1,902	100.50	4.02	0.89	.041
	高年級	619	100.33	4.23		
創新領導	低年級	1,907	100.39	3.86	1.32	.059
	高年級	642	100.16	3.99		
問題解決	低年級	1,870	101.29	5.31	0.59	.028
	高年級	603	101.14	5.39		
公民社會	低年級	1,907	100.55	3.32	1.31	.058
	高年級	629	100.35	3.57		
生涯發展	低年級	1,896	100.75	5.27	-1.27	.058
	高年級	619	101.06	5.47		

* $p < .05$.

表 10

大學生基本素養情意態度測驗年級差異分析結果

素養名稱	性別	人數	平均數	標準差	t值	效果量
溝通合作	低年級	1,972	99.63	12.37	0.64	.028
	高年級	686	99.28	12.71		
美感素養	低年級	1,939	99.39	14.96	0.61	.027
	高年級	649	98.98	15.10		
科學思辨	低年級	1,971	100.52	14.09	0.54	.024
	高年級	681	100.19	13.49		
資訊素養	低年級	1,917	100.55	9.39	0.63	.030
	高年級	643	100.27	9.41		

表 10

大學生基本素養情意態度測驗年級差異分析結果（續）

素養名稱	性別	人數	平均數	標準差	t值	效果量
終身學習	低年級	1,892	101.87	15.44	4.01***	.188
	高年級	610	99.00	15.06		
創新領導	低年級	1,906	100.83	12.86	1.93	.088
	高年級	643	99.72	12.24		
問題解決	低年級	1,875	101.31	15.51	3.05**	.144
	高年級	603	99.13	14.68		
公民社會	低年級	1,921	100.21	15.83	1.43	.067
	高年級	638	101.22	14.43		
生涯發展	低年級	1,790	101.50	20.42	2.64**	.131
	高年級	544	98.88	19.54		

** $p < .01$. *** $p < .001$.

能是樣本數太大造成的。另外要特別說明的是，低年級到高年級的進步幅度應該如何才合理？目前尚無明確定論，本測驗中各個素養涵蓋的範圍較為廣泛，但學校課程往往會有目標性和特定性，會針對某一主題進行教學，例如學生在「電影賞析」課程會培養電影相關的美感素養，但無法有全面性的美感素養訓練，因此，學生的學習成效未必會反映在大學生基本素養測驗的表現上，未來本研究會檢視各校課程內容和本測驗的評量目標是否一致，也會繼續蒐集本測驗的效度證據，以對評估學校通識教育的教學成效提供較具有建設性的建議。

參、測驗信度評估

本研究從測量精準度、信度等面向，進行大學生基本素養測驗的信度評估，以下分述信度檢驗方法與評估結果。

一、信度檢驗方法

（一）測量精準度

IRT 的信度是採用訊息量（information），也就是測量精準度的概念來表現，它表示試題或測驗在不同能力點上有不同的測量精準度。本研究以 $I_i(\theta)$ 來表示試題 i 在能力值 θ 能提供的訊息量，計算方式如公式（1）（Birnbaum, 1968），其中， $P_i(\theta)$ 是指能力值 θ 的學生在 i 題上的答對率，本研究是以 PCM 模式計算而得之， $P_i'(\theta)$ 為答對率對能力值的一階微分。訊息量愈高代表試題對該能力點的測量精準度愈高，從另一方面來解釋，訊息量也反映出試題在

不同能力值的估計標準誤，訊息量和估計誤差 $SE(\theta)$ 兩者間的關係表示如公式 (2) (Birnbaum, 1968)，訊息量愈高表示估計標準誤愈小。本研究將提供各個素養抽取二個、三個、四個題組時的估計標準誤，作為評估題庫品質之參考。

$$I_i(\theta) = \frac{P_i'(\theta)^2}{P_i(\theta)[1 - P_i(\theta)]} \quad (1)$$

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (2)$$

(二) 信度

在 IRT 中，並不假設每個人的能力估計誤差相同，而是針對每個人的作答反應情形估計出每個人的能力估計誤差。所以每位受試者的能力測量標準誤 (standard error of measurement, SEM) 並不相同。本研究根據 PCM 模式估計出來的能力值和測量標準誤，利用公式 (3) 推算出條件化信度 (conditional reliability)， σ_{θ}^2 為所有受試者能力值的變異數， SEM_s^2 為該受試者能力參數之測量變異誤 (error variance) 的平均 (Raju, Price, Oshima, & Nering, 2007)，SEM 和 σ 為已知，故可推算出信度。

$$\text{信度} = \frac{\sigma_{\theta}^2 - SEM_s^2}{\sigma_{\theta}^2} \quad (3)$$

二、信度評估結果

(一) 測量精準度

圖 1~圖 9 是各素養在不同題組數目下的能力估計誤差，從圖中可以看出，只施測兩個題組的時候誤差較大，但施測三~四個題組時，誤差有明顯降低的趨勢，而當同一基本素養測驗的題組全部施測時，能力估計誤差約介於 .20~.60 之間。另外，分析結果顯示，中等能力受試者的能力估計誤差較小，這可能是因為本測驗大部分的試題屬於中等難度，因而能對中等能力的受試者提供較高的訊息量。

(二) 信度評估結果

表 11 呈現各素養認知能力測驗與情意態度測驗之信度評估結果，表內數據包含各素養在各個題本的平均試題數、平均信度、信度的標準差。在認知能力測驗部分，各素養的平均信度介於 .50~.81 之間，但大部分素養的信度皆在 .69 以上，惟溝通合作素養的信度較低，若從認知能力測驗信度的標準差來看，信度的標準差大多在 .05 以下，顯示變動幅度非常小，只有溝通合作素養的標準差較其他素養大。溝通合作素養在各題本的題數普遍較少，推測可

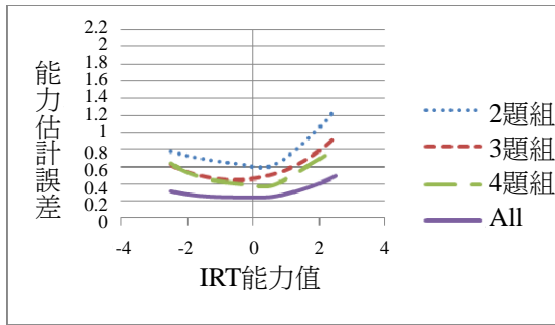


圖1. 美感素養認知能力測驗之能力估計誤差

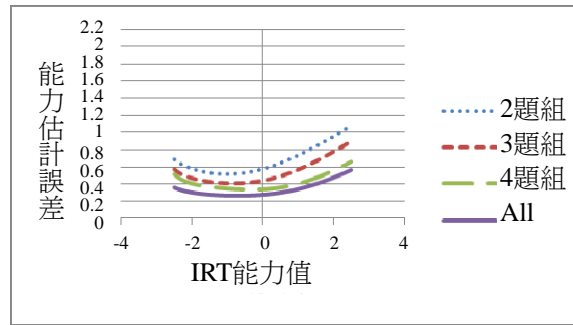


圖2. 溝通合作認知能力測驗之能力估計誤差

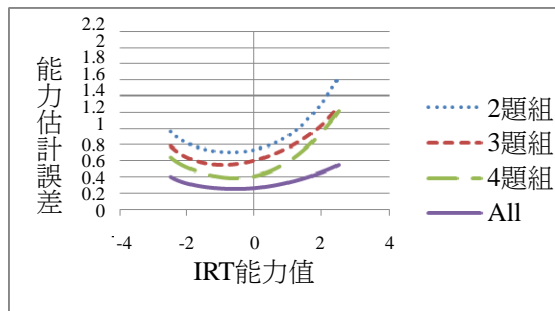


圖3. 創新領導認知能力測驗之能力估計誤差

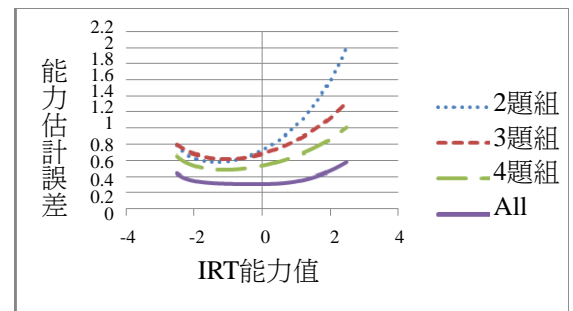


圖4. 公民社會認知能力測驗之能力估計誤差

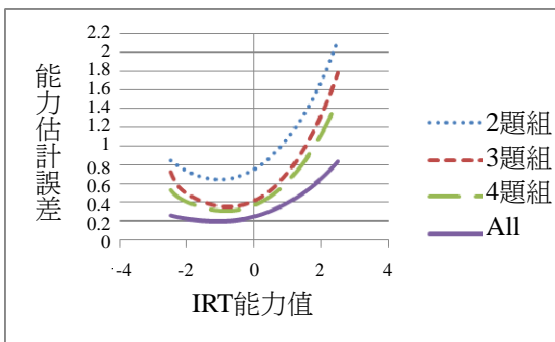


圖5. 資訊素養認知能力測驗之能力估計誤差

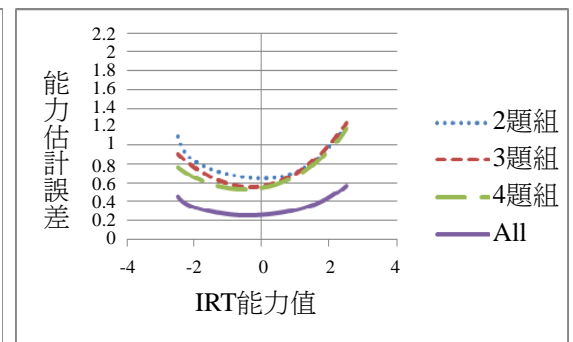


圖6. 生涯發展認知能力測驗之能力估計誤差

能是造成題本間信度較低且變動較大的原因。

在情意態度部分，各個素養的平均信度介於 .53~ .68 之間，惟問題解決素養的信度較低，從信度的標準差來看，各素養的信度標準差均在 .05 以下，顯示在各個素養中，題本間的信度變動幅度並不大。將情意態度和認知能力的信度相比，呈現出情意態度信度比認知能

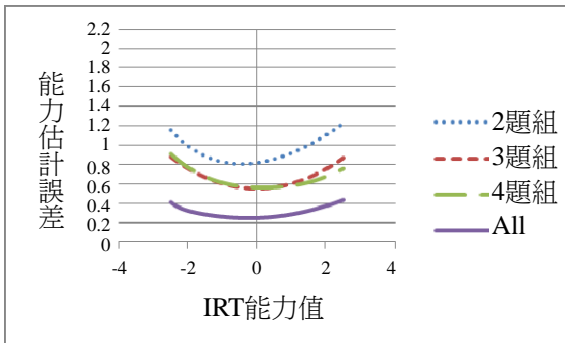


圖7. 終身學習認知能力測驗之能力估計誤差

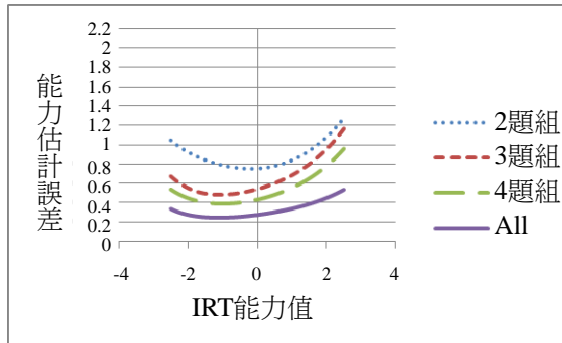


圖8. 科學思辨認知能力測驗之能力估計誤差

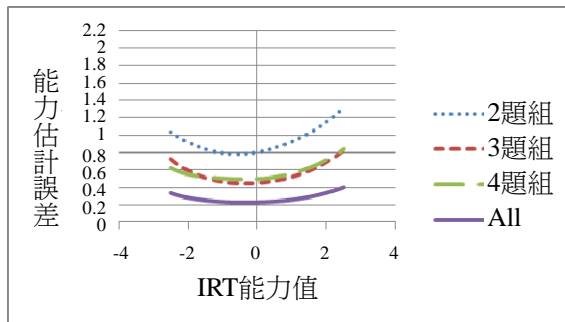


圖9. 問題解決認知能力測驗之能力估計誤差

表 11

大學生基本素養測驗各素養題本信度平均值和標準差

素養名稱	認知能力測驗		情意態度測驗	
	平均題數	信度	平均題數	信度
溝通合作	6	.50 (.11)	9	.65 (.04)
美感素養	10	.85 (.01)	7	.68 (.04)
科學思辨	6	.69 (.04)	8	.58 (.05)
資訊素養	8	.72 (.05)	6	.68 (.03)
終身學習	8	.81 (.01)	9	.61 (.03)
創新領導	6	.79 (.02)	6	.58 (.03)
問題解決	8	.79 (.02)	6	.53 (.06)
公民社會	7	.79 (.02)	9	.59 (.03)
生涯發展	8	.76 (.02)	9	.59 (.05)

力信度還低的趨勢。可能是因為情意態度沒有標準答案，是作答者對自身的認定情形，但對自身的認定往往受到外在環境的干擾，且容易有社會期許的問題，故學生在情意態度題目的作答反應變異較小，導致情意態度的信度略較認知能力的信度低。

肆、後續相關分析

學生在作答本測驗的同時也會填寫素養自評表，評估自己在各個素養的表現和同儕相較是很低、中等偏低、中等偏高或很高，並依序計為 1、2、3、4 分。表 12 為學生自評結果和實際認知能力與情意態度表現的相關性。整體而言，學生的自評表現和本測驗的認知能力分數的相關很低，但自評表現和本測驗的情意態度分數相較較高。由於大學生基本素養測驗的情意態度試題主要是測量學生的自信、動機與興趣，與學生自評表現一樣沒有正確答案，題型也類似，故兩者相關較高。但是本測驗的認知能力試題是學生在面對實際生活問題中所展現出來的能力高低，每題都有正確或較佳的答案，不是由學生自我認定，因此和學生自評表現的相關較低。由此結果可知，學生基本素養的認知能力試題與情意態度試題是評量不同面向的內涵，若要完整評量素養的概念，應將兩種題型都納入進行測量，不宜只用自陳式量表來進行測量。

表 12

學生自評表現與認知能力、情意態度表現的相關摘要

	美感 素養	溝通 合作	創新 領導	公民 社會	資訊 素養	生涯 發展	終身 學習	問題 解決	科學 思辨
認知能力	.08***	.06***	.01	.03**	.07***	.00	.02	.04***	.12***
情意態度	.27***	.26***	.29***	.29***	.20***	.38***	.36***	.31***	.19***

** $p < .01$. *** $p < .001$.

伍、結論與討論

大學生基本素養測驗是一份能協助學生瞭解自身基本能力並協助學校瞭解通識教育課程成效的測驗。本測驗為多媒體情境式測驗工具，評量學生在溝通合作、美感素養、科學思辨、資訊素養、終身學習、創新領導、問題解決、公民社會、生涯發展等九項基本素養的表現。每個題本皆包含九項素養，每個素養均有二至三個題組。本研究目的旨在探討大學生基本素養測驗的發展並分析其信度與效度。

在測驗目標部分，本研究先整理國內大專院校的通識教育目標與 ATC21S 提出的 21 世紀學生須具備的素養，歸納出大學生基本素養測驗應包含的向度，進而訂定命題向度和認知層

面及情意態度層面的各項指標。從命題和審題的流程來看，各項素養的題目均依照標準化測驗編製流程編製試題，出題者皆經過命題研習會議的訓練，題目不僅都經過素養發展團隊改良，並請各領域專家進行審題，接著再根據專家意見修改題目，確認題目敘述適當且設計良好，並測得欲測量的向度後，才進行組卷與預試，整個測驗編製流程控制良好。

本研究從建構效度、專家效度、效標關聯效度三面向進行大學生基本素養測驗的效度評估。由模式檢定來看，儘管大學生基本素養的題目全為題組題，但題組效果並未大到影響測量精準度，只要將叢集式是非題加總成多元計分題之後，就能大幅減少題組效果的影響，此時部分計分模式比題組反應模式更加適配本測驗的實際作答資料。在進一步檢視試題和模式的適配度後，發現所有試題皆符合部分計分模式。雖然初步分析資料時有部分題目的閾難度值出現翻轉的情形，但合併計分後可以改善閾難度值翻轉現象，正式題本的計分也將依此來進行。綜合以上，由模式檢定和試題適配度結果可知，大學生基本素養測驗有良好的建構效度。在專家效度上，各個素養均邀請五至九位專家學者判斷每個題目的向度符合度、表達適切性，並參酌專家意見修改，所有試題均具有良好的向度符合性，試題和測驗指標相對應，具良好的專家效度。

在效標關聯效度的探討上，本研究以性別差異、年級差異、自評成績作為效標關聯效度之參考指標。整體而言，不同性別與年級的學生在各基本素養的認知能力層面和情意態度層面的表現差異並不大。此結果合乎預期，因為大學生基本素養測驗評量到的並非學生的專業領域能力，而是學生通識性、綜合性、一般性的能力，較不會因為性別不同而不同，目前大專院校的養成教育著重在專業能力的培養，較沒有特別強調通識課程，故不同年級間的差異也不大。綜合以上所述，大學生基本素養測驗具備良好的效標關聯效度。

在信度方面，本研究是以能力估計精準度與條件化信度來評估大學生基本素養測驗的信度。估計精準度為 IRT 中表示測驗信度的方法，當各素養只施測兩個題組的時候能力估計誤差較大，但施測三~四個題組時，誤差明顯降低，而將該素養測驗的題組全部施測時（約 8~10 個題組），能力估計誤差介於 .20~.60 logit 之間。另外，進一步將估計精準度轉成古典測驗的信度時，單一題本的測驗信度可達到 .69 以上。整體而言，本測驗信度良好。

整體而言，大學生基本素養測驗有嚴謹的建置過程，試題的品質、測驗的信度、效度都具備良好的水準，且本研究提供不同題數的測量標準誤，讓測驗編製者可以根據測驗目標決定適當的測驗題數。本研究將持續維護題庫與新增試題，並建立大學生基本素養之常模，供高等教育機構擬訂通識教育政策與課程之參考。

誌謝

本研究感謝教育部「邁向頂尖大學計畫」與科技部「跨國頂尖研究中心計畫」(MOST104-2911-I-003-301) 的支持。

參考文獻

一、中文文獻

- 王保進 (2011)。以學生學習成效品質保證為核心之第二週期系所評鑑作業規範。評鑑雙月刊，33，7-11。
- 【Wang, B.-J. (2011). Student learning outcomes associating with quality assurance mechanism for department and graduate institute evaluation. *Evaluation Bimonthly*, 33, 7-11.】
- 林煥祥、劉聖忠、林素微、李暉 (2008)。台灣參加 PISA 2006 成果報告。科技部專題研究計畫成果報告 (NSC95-2522-S-026-002)。花蓮市：國立花蓮教育大學；高雄市：國立高雄師範大學。
- 【Lin, H.-X., Liu, S.-Z., Lin, S.-W., & Li, H. (2008). *PISA 2006 results in Taiwan*. Ministry of Science and Technology project report (NSC95-2522-S-026-002). Hualien, Taiwan: National Hualien University of Education; Kaohsiung, Taiwan: National Kaohsiung Normal University.】
- 郭生玉 (2012)。心理與教育研究法。新北市：精華。
- 【Guo, S.-Y. (2012). *Educational psychology and methodology*. New Taipei City, Taiwan: Jin-Hua.】
- 陳儀蓉、黃芳銘 (2006)。組織公民行為量表在男女員工群體上之測驗恆等性檢定。測驗學刊，53 (2)，297-326。
- 【Chen, Y.-J., & Hwang, F.-M. (2006). Assessing measurement invariance of organizational citizenship behavior citizenship behavior scale between male and female staff. *Psychological Testing*, 53(2), 297-326.】
- 陳柏熹 (2011)。心理與教育測驗：測驗編製理論與實務。新北市：精策教育。
- 【Chen, P.-H. (2011). *Psychology and educational testing: Theory and practice*. New Taipei City, Taiwan: Kinetics.】
- 廖添富、黃景裕 (1995)。我國大學生公民態度之研究—樣本取自北部地區。公民訓育學報，4，109-127。doi:10.6231/CME.1995(4)05
- 【Liao, T.-F., & Huang, C.-Y. (1995). A study of the civic attitude of college students on Taiwan, R.O.C.—Samples selected from northern Taiwan. *Bulletin of Civic and Moral Education*, 4, 109-127. doi:10.6231/CME.1995(4)05】
- 國立臺灣師範大學雲端測驗中心 (2014)。大學生基本素養指導手冊。取自 <http://140.122.69.221/ntnucit/index.php/services-mainnav/testing-of-students-literacy-mainnav>
- 【National Taiwan Normal University Research Center for Internet Testing. (2014). *Guilds for basic literacy test for university students*. Retrieved from <http://140.122.69.221/ntnucit/index.php/services-mainnav/testing-of-students-literacy-mainnav>】
- 國家教育研究院 (2014)。教育部提升國民素養專案辦公室。取自 <http://literacytw.naer.edu.tw/index.php?REFDOCID=0m8mptc3ioz0a9z8>
- 【National Acedemy of Educational Research. (2014). *Office of enhancing the basic literacy of people*. Retrieved from <http://literacytw.naer.edu.tw/index.php?REFDOCID=0m8mptc3ioz0a9z8>】

二、外文文獻

- Assessment & Teaching of 21st Century Skills. (2010). *21st century skills white paper*. Retrieved from <http://atc21s.org/index.php/resources/white-papers/>
- Aiken, L. R., & Groth-Marnat, G. (2005). *Psychological testing and assessment* (12th ed.). Boston, MA: Pearson Education.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-472). Reading, MA: Addison-Wesley.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6(4), 431-444. doi:10.1177/014662168200600405
- Chen, P.-H., Kuo, J.-W., & Sung, Y.-T. (2011, July). *Influence of pre-test design on the precision of the parameters estimation in the multidimensional items bank*. Paper presented at the IMPS 2011, the 76rd annual meeting of the Psychometric Society, Hong Kong, China.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey, NY: Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- European Association for Quality Assurance. (2009a). *ENQA position paper on quality assurance in the EHEA – In view of the leuven and louvain-la-neuve meeting of ministers responsible for higher education of 28-29 April 2009*. Retrieved from [http://www.enqa.eu/files/ENQA_position_paper%20\(3\).pdf](http://www.enqa.eu/files/ENQA_position_paper%20(3).pdf)
- European Association for Quality Assurance. (2009b). *The Bologna process 2020 – The European higher education area in the new decade*. Retrieved from http://www.enqa.eu/files/Leuven_Louvain-la-Neuve_Communique_April_2009.pdf
- Gardner, H. (1983). *Frames of mind*. New York, NY: Basic Books.
- Holland, J. L. (1973). *Making vocational choices: A theory of careers*. Upper Saddle River, NJ: Prentice Hall.
- Liu, T.-T., & Sun, H.-B. (2012). Gender differences on information literacy of science and engineering undergraduates. *International Journal of Modern Education and Computer Science*, 4(2), 23-30. doi:10.5815/ijmecs.2012.02.04
- Ma, F.-C., Ding, R., & Li, Z.-Z. (2009). Case study: An analysis of information literacy competence of university students in Wuhan area. *Document, Information & Knowledge*, 1, 24-29. (in Chinese)

- Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics, 23*(4), 337-341. doi:10.1007/BF01067434
- McManus, I. C., & Furnham, A. (2006). Aesthetic activities and aesthetic attitudes: Influences of education, background and personality on interest and involvement in the arts. *British Journal of Psychology, 97*(4), 555-587. doi:10.1348/000712606X101088
- Organization for Economic Co-operation and Development. (2007a). *PISA 2006: Science competencies for tomorrow's world, volume I analysis*. Paris, France: Author.
- Organization for Economic Co-operation and Development. (2007b). *PISA 2006: Science competencies for tomorrow's world, volume II analysis*. Paris, France: Author.
- Pope-Davis, D. B., & Twing, J. S. (1991). The effects of age, gender, and experience on measures of attitude regarding computers. *Computers in Human Behavior, 7*(4), 333-339. doi:10.1016/0747-5632(91)90020-2
- Raju, N. S., Price, L. R., Oshima, T. C., & Nering, M. L. (2007). Standardized conditional SEM: A case for conditional reliability. *Applied Psychological Measurement, 31*(3), 169-180. doi:10.1177/0146621606291569
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
- Rottinghaus, P. J., Day, S. X., & Borgen, F. H. (2005). The career futures inventory: A measure of career-related adaptability and optimism. *Journal of Career Assessment, 13*(1), 3-24. doi:10.1177/1069072704270271
- Rowland, K. D. (2004). Career decision-making skills of high school students in the Bahamas. *Journal of Career Development, 31*(1), 1-13. doi:10.1023/B:JOCD.0000036702.02293.1a
- Schmidt, W. H., Raizen, S. A., Britton, E. D., Bianchi, L. J., & Wolfe, R. G. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school science*. Dordrecht, the Netherlands: Kluwer Academic.
- Scott, B. P. (1998). Just what in a competency? *Training Minneapolis, 35*(6), 58-64.
- Sypher, H. E., Applegate, J. L., & Church, S. L. (1981, May). *Cognitive differentiation and verbal intelligence: Clarifying relationships*. Paper presented at the annual conference of the International Communication Association, Minneapolis, MN.
- Verderber, R. F., & Verderber, K. S. (1995). *Inter-act: Using interpersonal communication skills/voices: A selection of multicultural readings*. Belmont, CA: Wadsworth.
- Wang, W.-C., & Wilson, M. (2005). The Rasch testlet model. *Applied Psychological Measurement, 29*(2), 126-149. doi:10.1177/0146621604271053

Wright, B. D. (1994). Reasonable mean-square fit values. In B. D. Wright & J. M. Linacre (Eds.), *Rasch measurement transactions* (Part 2; p. 370). Chicago, IL: MESA.

Yip, D.-Y., Chiu, M.-M., & Ho, E. S.-C. (2004). Hong Kong student achievement in OECD-PISA study: Gender differences in science content, literacy skills, and test item formats. *International Journal of Science and Mathematics Education*, 2(1), 91-106. doi:10.1023/B:IJMA.0000026537.85199.36

附錄一 範例試題



你現在正在進行的是  美感素養分測驗

大學生基本素養測驗

南投縣仁愛鄉位居中央山脈之上，地勢高峻陡峭，觀光事業發達。而仁愛鄉的中正社區有條道路，配合了當地景象與民族生態，將整個社區營造出美麗且獨特的風格。如下圖所示。



根據圖片內容，請找出這張照片中最與眾不同的景象為何？

- (A) 坐擁山景，樹林綠意盎然，讓人心曠神怡。
- (B) 滿滿的檳榔樹，水土保持需要更加留意。
- (C) 路燈的造型非常特殊，引人注意。
- (D) 車輛與人煙稀少，空氣清新。

仁愛鄉中正社區這條道路的規劃，最有可能和以下哪一種設計理念有關呢？

- (A) 道路規劃整齊暢通，少彎道也能減少交通事故傷害。
- (B) 簡單的單行道設計，讓人能在綠色走道中盡情奔馳。
- (C) 空出旁邊視野，讓人能同時欣賞山景和周圍的作物。
- (D) 路燈造型經過特殊設計，彷彿蛇在跳舞般歡迎來賓。

請問如果想要更了解此道路的設計原理，來增加對於道路環境規劃的藝術涵養，以下哪些選項能讓你達到目的呢？

- 調查實際參觀人數，計算多少人對此產生共鳴。
- 請教居民當地的詳細傳統背景和民族精神。
- 將道路各種特色風貌拍攝下來，與有興趣的朋友分享。

根據自身狀況回答下列問題！	非常同意	同意	不同意	非常不同意
我認為大部分的公共藝術品都能提升我們的生活品質。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
在我的生活中，每個月都一定要規劃前往台灣各種特色景點旅遊。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
到一些特殊景點旅遊前，我一定會先詳細了解當地風俗民情之後才出發。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
即使一些保護文化資產的活動違反了公共行政的決策，我依然會支持並參與聲援活動來保留那些歷史。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

附錄二 根據 Holland 理論分類的科系範例表

類型	代表科系
藝術型	設計系、景觀系、美術系、音樂系、建築與室內設計等相關學系
實用型	土木工程學系、建築學系、資訊工程、電子工程、機械工程系、環境與安全衛生工程等相關科系
研究型	物理系、數學系、生科系、地球科學系、生命科學系、化學工程等相關科系
商業型	企業管理系所、行銷系、事業經營學系、國際企業系、經營管理學系等相關系所
社會型	心理輔導學系、社會工作系、幼兒教育學系、教育學系、護理系等相關科系
事務型	財務金融系、會計系、財經法律學系、國際貿易系等相關科系

Journal of Research in Education Sciences

2015, 60(3), 95-126

doi:10.6209/JORIES.2015.60(3).04

Evaluation of the Psychometric Properties of the General Literacy Test for University Students

Po-Hsi Chen

Department of Educational Psychology
and Counseling,
National Taiwan Normal University

Hsin-Ying Huang

Science Education Center,
National Taiwan Normal University

Yu-Hsin Chen

Department of Educational Psychology
and Counseling,
National Taiwan Normal University

Tai-Ting Yeh

Department of Educational Psychology and Counseling,
National Taiwan Normal University

Shao-Tsu Su

Department of Educational Psychology and Counseling,
National Taiwan Normal University

Abstract

This study evaluated the psychometric properties of the General Literacy Test for University Students. To develop the assessment framework, the educational objectives of general literacy courses of universities in Taiwan as well as the core competencies of Assessment and Teaching of 21st Century Skills were all reviewed and considered. The general literacy test is composed of nine literacy domains: communication and collaboration, esthetics, information, lifelong learning, career, leadership, problem solving, social concerns and citizenship, and scientific thinking. The items of the general literacy test were developed into a multimedia assessment. In addition, booklets including two to three testlets on each of the nine literacy domains were organized. Overall, 10,958 students from 20 universities participated. The results demonstrated that testlet effect could be ignored and the partial credit model could be used to analyze data. In addition, the items fit the partial credit model closely and the test demonstrated favorable construct validity. In addition, small differences existed between gender and grades. The results were consistent with those of previous studies and showed that the test's criterion-related validity was acceptable. Furthermore, the standard error of ability ranged from .20 to .60 logit and the reliabilities of the booklets were more than .69 in

Corresponding Author: Hsin-Ying Huang, E-mail: aessaess@gmail.com

Manuscript received: Oct. 6, 2014; Revised: Feb. 28, 2015; Accepted: Mar. 5, 2015.

◆ 126 ◆ 大學生基本素養測驗

陳柏熹、黃馨瑩、陳郁欣、葉泰廷、蘇少祖

the nine literacy domains. The evidence supports the high degree of reliability and validity of basic literacy assessment.

Keywords: reliability, validity, literacy assessment, scenario based test, online test