中學生現代公民核心素養之測驗編製

陳彥君、洪素蘋*

本研究旨在編製一信、效度良好的中學生現代公民核心素養測驗，以評估臺灣中學生現代公民核心素養狀況。測驗編製係參考現代公民核心能力指標以及國際公民教育與素養調查研究（International Civic and Citizenship Education Study [ICCS]）的評量架構，編擬符合中學生生活場域與經驗的試題。試題編製完成後經學科專家確認內容效度與試題預試，於正式施測階段，採分層便利取樣，蒐集來自全臺北、中、南、東（含離島）各區共 22 所學校的中學生共 1479 份有效樣本，以多向度部分給分試題反應理論模式（multidimensional partial credit model [MPCM]）進行資料分析。結果發現：中學生現代公民核心素養測驗具有良好試題內部一致性，認知題與情感行為題的試題區分信度係數（item separation reliability）皆高於.9，五向度的受試者區分信度係數（person separation reliability），介於.593 至.882 之間，測驗折半信度為.883。效度驗證程序採用試題反應理論效度驗證架構進行，結果顯示本測驗符合五向度測量建構，同時並提供測驗的內容、結構、本質、類推效度與效標關聯效度，顯示本研究所發展的測驗工具具有良好信度與效度。

關鍵詞：現代公民核心素養、試題反應理論模式、試題差異功能

陳彥君：國立成功大學政治系政治經濟學博士班研究生
* 洪素蘋：國立成功大學師資培育中心副教授
（通訊作者：sphung@mail.ncku.edu.tw）

Research Article

Journal of Educational Practice and Research
Vol. 32, No. 2, 39-80, Dec. 2019

Development of a Modern Citizen Core Literacy Test for Middle School Students

Yen-Chun Chen & Su-Pin Hung*

The purpose of this study was to design the Modern Citizen Core Literacy Test with well reliability and validity for middle school students. To develop the test, the core competency index as well as the assessment framework of the International Civic and Citizenship Education Survey were used as references. The Modern Citizen Core Literacy Test is developed to reflect the life experiences of middle school students. To establish content validity of the test, a panel of experts from middle schools were invited to review the first version of the test. After a pilot study, the test was administered to a total of 1,479 students from 22 middle schools in Taiwan; they were recruited with stratified purposive sampling. Multi-dimensional Partial Credit Model was used to assess the model-data fit. Five-core literacy were identified: moral literacy, democracy literacy, scientific literacy, media literacy and aesthetic literacy. Moreover, the test showed sufficient internal consistency with person separation reliability ranging from .593 to .882. The Split-half reliability is .883. In addition, the present study adopted item response theory based validation procedure. The results revealed that five dimensions of the Modern Citizen Core Literacy Test exhibited an acceptable model-data fit. Criterion-related validity, interpretability evidence, generalizability evidence, and substantive evidence were all supported, indicating that the test has good reliability and validity.

Keywords: ***differential item functioning, item response theory, modern citizen core literacy***

Yen-Chun Chen: Ph.D. Student, PhD Program of Political Economy, Department of Political Science, National Cheng Kung University

*Su-Pin Hung: Associate Professor, Center for Teacher Education, National Cheng Kung University (corresponding author: sphung@mail.ncku.edu.tw)

中學生現代公民核心素養之測驗編製

陳彥君、洪素蘋

壹、緒論

面臨國際間動態競爭與跨國性經濟體系的變動，思考如何培育未來公民使其能良好的適應 21 世紀的工作與生活世界，已然成為國際間對於人才培育的共識（張鈿富、吳慧子、吳舒靜，2010；Commission of the European Communities, 2005；Salganik & Stephens, 2003）。而培育公民素養的不二法門首當從教育著手，隨著時代的更迭，對於培育現代公民素養的內涵日益多元（李俊湖，2009）。董秀蘭（2016）曾剖析國民教育階段社會科學習領域的教學內涵與培育現代公民所需具備的核心素養，藉以連結兩者的關聯性；然則現代公民素養的養成，非僅限特定、單一學習領域，而是見諸於各學習領域，並且不限於正式與非正式學習途徑，自然也包含家庭教育。審視十二年國民基本教育課程綱要總綱（教育部，2014）也可發現，國民教育的課程發展是以素養為核心（呂秀蓮，2019）。素養的概念與國際間對於核心能力以及關鍵能力的倡議及其進展有關，各國對於核心能力或是關鍵能力的用字也有所不同，如：澳大利亞「key competencies」、紐西蘭「essential skills」、美國「workplace know-how」、英國「core skills」；另外也有「transferable skills」、「transition skills」、「enabling skills」、「basic skills」和「core competencies」（張鈿富等人）。在素養的英文用詞上，林永豐（2014）針對 literacy 與 competence 兩個詞進行字意與用法分析，literacy 一詞廣義地的解釋係指應達到足以因應社會參與需要的功能程度，是回應社會脈絡所需的能力；而 competence 雖然以往被認為屬於較狹隘的解釋，亦即職業訓練脈絡之下的能力，但時至今日，其用法上較常被用以表述整體的、具跨領域特性的概念，內涵較 literacy 所指的某一領域或範疇更為廣泛，如 12 年國教課綱中的 core competence，係指回應情境、因應未來社會複雜需要的一種通用性能力。基於上述考量，本研究以 literacy 作為英文主要用詞，與 competence 稍作區分。

審視過往國際組織以中學生為對象進行的素養相關測驗，包含了「國際學生評量

計畫」（Programme for International Student Assessment [PISA]），將中學生的數學、閱讀、資訊、科學等素養（literacy）列為重要評估項目，2018 年更增加了全球素養（global competence）（Organization for Economic Cooperation and Development [OECD]，2019），各素養輪流施測；「國際數學與科學教育成就趨勢調查」（Trends in International Mathematics and Science Study [TIMSS]）以四年級、八年級學童為對象，調查其數學與科學成就（Taiwan International Large-Scale Study Center [TILSSC]，2019）；「國際閱讀素養研究」（Progress in International Reading Literacy Study [PIRLS]）則是評估小學四到六年級學生的閱讀及數位閱讀素養（張貴琳，2013）。由上述國際素養調查研究顯見素養評估具有其重要性，而這些跨國性的素養測驗在編製過程也建立具有清楚的評量架構，如 PISA 的測量即包含對各類素養的知識、技能、態度與價值。國際大型教育調查研究具有作答時間較長、非每年施測之特性以及考題不完全釋放等，因此在一般學校情境中較難以取得做為自我檢測工具。目前國內有關現代公民核心素養培育的探討主要針對受高等教育者，試圖建構大學生核心能力架構（吳京玲，2011；吳京玲、林妙容、李信，2009）。為了審視大學階段學生的核心素養培育情況，陳柏熹、黃馨瑩、陳郁欣、葉泰廷與蘇少祖（2015）分析國內大專院校的通識教育目標與核心素養，並參考 21 世紀技能評量與教學計畫（Assessment & Teaching of 21st Century Skills [ATC21S]，2010）所提出的 21 世紀現代學生需具備的 10 項基本素養，發展大學生基本素養測驗，包含：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等 9 項基本素養。歐慧敏（2013）則以教育部人文社會科學教育計畫辦公室於 2012 年所列的倫理、民主、科學、媒體與美學等五大現代公民素養為架構（教育部，2012），發展大學生公民素養量表，採用自陳式問卷方式，評估大學生的現代公民素養能力。

儘管陳柏熹等人（2015）所發展的素養測驗具有良好的信度與效度，所使用的測量方式較具生態效度（趙子揚、黃嘉莉、宋曜廷、郭蕙寧、許明輝，2016），但其測驗的施測對象為大學生，測量架構主要以大學生所應具備的基本素養為基礎，而非教育部所列現代公民五大素養，就測量內涵上仍有所不同，若直接將這些素養分測驗使用於中學生，恐因試題難度不適切而降低對受試者能力評估的精準度。歐慧敏（2013）所使用的自陳量表編製架構係採用教育部所提出的現代公民五大核心素養為架構，此架構被廣泛接受，然則情境判斷判斷題組與自陳量表設計用途不盡相同，情境判斷測驗具有擬真情境的試題設計與生態效度（歐宗霖、邱皓政、孫國勛，2013），這也是

本研究選擇採用此較為複雜的測驗設計形式的主因。基於前述，本研究旨在發展一份具有良好信、效度的中學生現代公民核心素養測驗。

貳、中學生現代公民核心素養

現代公民核心素養的養成具有長時間性與切身性的特色（莊富源，2013），其養成需要同時顧及縱貫性的時間累積以及橫斷面的跨領域之間的串聯。教育部（2015A）所推動的十二年國民基本教育即以「提升國民基本知能，培養現代公民素養」為第一條總目標，顯見現代公民核心素養的培育在國民基本教育階段的重要性。過往我國的公民教育在國民教育階段主要仰賴學校中的生活與倫理、公民與道德等相關課程進行培育（鄧毓浩，2013），不過隨著時代更迭，為了因應時代的脈動，現代公民所應具有的核心素養的範疇也日益多元。教育部人文社會科學教育計畫辦公室即提出，現代公民核心素養的培育範疇，包含倫理、民主、科學、美學、媒體等五大素養（教育部，2012）。在此政策推動下，高等教育階段也透過通識課程展開對大學生現代公民素養的培育（鄧毓浩）。陳佳欣（2012）接受教育部計畫委託進一步針對達成此五大核心素養所需具備的能力進行調查，透過訪談、專家共識會議方式，臚列各素養下專家所提出培育該項核心素養不可或缺的能力項目，其中各專家成員為五大素養、公民核心能力、教育評量等相關專業領域的專家學者，其隸屬單位包含各素養相關專業系所和通識教育中心教師，共 46 位。

從個體社會化的過程來看，中等教育階段實屬公民核心素養發展的關鍵期。教育部（2014）於「十二年國民基本教育課程綱要總綱」亦揭示，第四學習階段（國中生）與第五階段（高中職）應持續提升所有現代公民應具備的核心素養。教育部（2015A）核心素養不宜以「學科知識」為學習的唯一範疇。顯見，核心素養的概念具有跨越學習領域既有疆界，強調統整性與生活實踐性的意味。從 Oana 與 Pop（2016）所進行的縱貫研究發現，青少年階段所培養不同的能力與素養，可幫助青少年因應未來世界的生活。基於此，本研究選擇以中等學習階段的受教者為對象，探討中學生現代公民核心素養的概況。

依據十二年國民基本教育課程綱要總綱（教育部，2014），國民教育階段所欲培育學習者的核心素養可從三面九項來進行，其目的在於養成國民核心素養，以提升國

民素質（洪裕宏，2011；陳伯璋，2010；蔡清田，2014；蔡清田、洪若烈、陳延興、徐彬瀚、葉雅卿，2012）。為了突顯國民教育階段對於培育現代公民核心素養的影響性，本研究嘗試將此 9 項核心素養與教育部（2015B）所提的現代公民所需具備的五大素養加以並列比較，並摘要整理如表 1。其中，現代公民五大素養核心能力項目係依據陳佳欣（2012）所進行三回合核心能力指標重要性的專家德菲法結果，並挑選其中達 3.6 分以上具代表性的核心能力指標。由表 1 可見，現代公民所需具備的五大素養之核心能力指標與十二年國民基本教育所欲培育的九項核心素養大抵能有所對應，顯示十二年國民基本教育階段所培育的核心素養能與現代公民所需具備的五大核心素養粗略相符。

表 1　現代公民核心素養與 12 年國民基本教育核心素養之構念與相對應關係

十二年國教核心素養概念	現代公民五大素養核心能力項目
A1 身心素質與自我精進	→ **I倫理素養：**
J-A1 國民中學教育 探索人性、自我價值與生命意義，積極實踐。	專家共識重要之指標：
U-A1 高級中等學校教育 透過自我精進與超越，追求至善與幸福人生。	I-1　道德與實踐能力（3.7）
A2 系統思考與解決問題	→ **III科學素養：**
J-A2 國民中學教育 運用適當的策略處理解決生活及生命議題。	專家共識重要之指標：
U-A2 高級中等學校教育 具備系統思考、分析與探索的素養，深化後設思考，解決人生的各種問題。	III-1　思考與批判能力（4.1） III-2　主動探索與研究精神（3.9） III-3　問題解決能力（4.0）
A3 規劃執行與創新應變	
J-A3 國民中學教育 主動學習與創新求變。	
U-A3 高級中等學校教育 具備規劃、實踐與檢討反省的素養，並以創新的態度與作為因應新的情境或問題。	
B1 符號運用與溝通表達	→ **IV媒體素養：**
J-B1 國民中學教育 能以同理心與人溝通互動。	專家共識重要之指標：
U-B1 高級中等學校教育 具備掌握各類符號表達的能力，與他人溝通並解決問題。	IV-1　使用資訊科技的能力（4.2） IV-2　理解媒體訊息內容（3.9）
B2 資訊科技與媒體素養	IV-3　反思媒體內涵（3.7）
J-B2 國民中學教育 具備善用科技、資訊與媒體以增進學習的素養，並察覺、思辨人與科技、資訊、媒體的互動關係。	IV-4　反思閱聽人的意義（3.6） IV-5　溝通表達能力（4.0）
U-B2 高級中等學校教育 具備適當運用科技、資訊與媒體之素養，並能反思科技、資訊與媒體倫理的議題。	
B3 藝術涵養與美感素養	**V美學素養：**
J-B3 國民中學教育 了解美感的特質、認知與表現方式，增進生活的豐富性與美感體驗。	→ 專家共識重要之指標：
U-B3 高級中等學校教育 具備藝術感知、欣賞、創作與鑑賞的能力，透過生活美學的涵養，對美善的人事物，進行賞析、建構與分享。	V-1　藝術與美感養成（3.7） V-2　美感體驗與品味生活（3.5） V-3　欣賞和美感表達（3.6）
C1 道德實踐與公民意識	→ **I倫理素養：**
J-C1 國民中學教育 培養道德思辨與實踐能力，關懷生命倫理議題與生態環境。	專家共識重要之指標：
U-C1 高級中等學校教育 具備對道德課題與公共議題的思考與對話素養，主動參與環境保育與社會公共事務。	I-1　道德與實踐能力（3.7） I-2　尊重與關懷（3.9）
C2 人際關係與團隊合作	
J-C2 國民中學教育 培育相互合作及與人和諧互動的素養。	→ **II民主素養：**
U-C2 高級中等學校教育 溝通協調及團隊合作的精神與行動。	專家共識重要之指標：
C3 多元文化與國際理解	II-1　社會參與能力（3.8） II-2　多元文化與國際視野（3.9）
J-C3 國民中學教育 接納多元文化的涵養。	
U-C3 高級中等學校教育 尊重欣賞多元文化，具備國際化視野。	

註：表中之底線為構念相對應之處，括弧內之數字為專家德菲法平均分數，滿分為 5 分。因篇幅受限，本表僅列所篩選出的核心能力指標作為五大核心素養內涵的呈現。本研究依據素養編製順序，分別編碼倫理Ⅰ、民主Ⅱ、科學Ⅲ、媒體Ⅳ、美學Ⅴ。

參、試題反應理論多向度部分給分模式

傳統古典測驗理論的試題分析，關注受試者在測驗上的總分並以此做為受試者能力的評估，缺乏試題與受試者反應的連結（Guion & Ironson, 1983；Wright, 1977），試題特性分析忽略受試者在各題的作答反應，因此試題難度（通過率）或是鑑別度等指標的建立完全受到樣本依賴的影響。反觀，試題反應理論使用受試者的作答反應資訊，並且依據受試者的反應分別進行試題特性與受試者能力的估算。試題與能力測量具有不變性，不受樣本依賴的影響外，使得不同版本測驗的等化更為可行（Embretson & Reise, 2000）。而當測驗滿足特定的試題反應理論模型，如：Rasch（1960）所建立的量尺更具有「等距」與「統計充份性」的特色。隨著試題反應理論模型的開展，適用於各種測驗情境的試題反應理論模型，已然被發展。例如：適用於二元計分的 Rasch 模型（Rasch）、適用於多元計分的等級評分模型（Rating Scale Model [RSM]）（Andrich, 1978）以及適用於混合計分的部分給分模型（Partial Credit Model [PCM]）（Masters, 1982）等。此外，因應許多構念所反映的特質並非單一潛在構念，因此前述的試題理論模型也逐一擴展為多向度試題反應理論模型（Reckase, 1997）。目前試題反應理論模式已廣泛運用於各類心理與教育測量工具的試題分析與信效度檢驗（王佳琪、何曉琪、鄭英耀，2014；王佳琪、鄭英耀、劉昆夏、何曉琪，2011；巫博瀚、賴英娟、施慶麟，2013；侯雅齡，2010；陳柏熹等人，2015；張貴琳，2013；黃宏宇、洪素蘋，2007；趙子揚等人，2016；Hung, 2018；Hung, Huang, & Chen, 2016）。

基於上述試題反應理論的優勢，以下簡述本研究所使用的多向度部分給分模式（Multidimensional Partial Credit Model [MPCM]）。由於 MPCM 同屬 Rasch 家族模式，其所衍生出來的量尺，同樣具有等距與統計充份性。以多向度 PCM 模式來說，每道試題均有各自的閾難度（threshold difficulty），公式如下：

$$P(X_{ijk}=1|\boldsymbol{\theta}_{jq},\boldsymbol{\delta}_{iy})=\frac{\exp\left[\sum_{y=1}^{k}\left(\sum_{q=1}^{D}w_{qiy}\boldsymbol{\theta}_{jq}-\boldsymbol{\delta}_{iy}\right)\right]}{1+\exp\left[\sum_{y=1}^{k}\left(\sum_{q=1}^{D}w_{qiy}\boldsymbol{\theta}_{jq}-\boldsymbol{\delta}_{iy}\right)\right]}$$

$P(X_{ijk})$為受試者 j 填答第 i 題的第 k 類別反應的機率。

$\boldsymbol{\theta}_{jq}$ 向量為受試者 j 在向度 1 到向度 D 的潛在能力特質參數

q 為本研究所欲測量之構念。

w_{qiy} 為加權數。

$\boldsymbol{\delta}_{iy}$ 為試題難度向量以及閾難度。

構念（q=1,…, D），以本研究而言所代表的是「倫理素養」、「民主素養」、「科學素養」、「媒體素養」與「美學素養」等五大核心素養；在此模式中，受試者做答某一題的能力即為多個潛在能力的加權總合，此加權數 w_{qiy} 為間斷非負數值（0, 1, 2, …）。

試題反應理論模式適配評估分別採用試題區分信度係數（Item separation reliability）與「受試者區分信度係數（Person separation reliability）做為評估指標。「受試者區分信度係數」愈大，表示受試者間的區分愈明顯，評分量尺也愈有效，測驗的信度也愈佳，以本素養測驗來說，愈能夠區別出不同素養能力水準的中學生。

試題區分信度係數（Item separation reliability）表示整份測驗試題難度層級的區分程度，其公式如下：

$$\text{ISR}=1-\frac{MSE_I}{SD_I^2}$$

ISR 為 Item separation reliability。

MSE_I 為試題測量誤差的均方。

SD_I^2 為試題觀察值的變異數。

當「試題區分信度係數」愈大，試題評分量尺分類階層愈佳，此時試題難度差異明顯，更可區分出受試者能力特質的差異；而受試者區分指標（Person separation index）表示測驗的信度，其概念等同於 Cronbach's α 的信度概念（Wright & Stone, 1999）。

其公式如下：

$$R_P = \frac{SA_P^2}{SD_P^2} = 1 - \frac{MSE_P}{SD_P^2}$$

R_P 為 Person separation reliability。

SA_P^2 為真實分數的變異數。

SD_P^2 為受試者觀察值的變異數。

MSE_P 為受試者測量誤差的均方。

量表的建構效度，則是透過均方誤指標（mean square error [MNSQ]）檢視各試題與部分給分模式（Partial Credit Model [PCM]）模式適配的狀況。均方誤指標，可區分為訊息加權統計量（information-weighted fit[Infit]）與極端值加權統計量（outlier-sensitive fit[Outfit]）」。兩者皆是自由度為 1 的卡方統計量，數值介於 0 到無限大之間，期望值為 1.0。Infit 均方誤主要係受反應組型影響，是指當試題難度對應於受試者能力時，試題作答反應的非預期敏感度。該指標所計算的是加權後的均方誤。當 Infit 指標有問題時，表示受試者的能力與試題難度相互匹配，受試者的能力與試題難度所在的觀察值與模式不相符，對於測量的威脅相對較大。Outfit MNSQ 是指受試者能力測量對於太簡單或太難的試題作答反應的非預期敏感度，容易受到能力極端者不合理的作答反應而影響。該指標所計算的為未加權均方誤。當 Outfit 指標有問題時，表示此題項受試者的能力與試題的難度所在的觀察值與模式不相符。Linacre（2002）建議，當 Infit MNSQ 及 Outfit MNSQ 數值介於 0.5—1.5，表示試題對於量尺建立具有貢獻性；當適配指標介於 1.5—2.0，表示試題對於測量量尺的建立較無重要性；當高 2.0，表示試題的存在將扭曲整體的測量構念，需要加以刪除。若本研究現代公民核心素養測驗當中的能力題與情感題的題目能吻合模式預期，就可將測量變項成功量尺化，這就是建構效度的重要依據（王文中，2004）。

肆、研究方法

一、評量架構與命題

本測驗在形成架構之初，審閱有關現代公民核心素養的內涵與定義（李俊湖，2009；教育部，2011；陳佳欣，2012；莊富源，2013；歐慧敏，2013），最後決定以教育部所提出的倫理、民主、科學、媒體與美學等現代公民五大素養（教育部，2015B），並參酌十二年國民基本教育課程綱要所揭示的三面九項素養為測量的基本架構。考量核心素養此構念，同時富涵知識、技能與情感層面。因此，本研究將現代公民核心素養定義為個體對於倫理、民主、科學、媒體與美學等方面的知識、技能、態度、價值與德行。

（一）確立測驗編製架構

參酌國內外評量各類素養的評量架構，發現各類素養的測量可大致區分為認知領域與情感行為領域兩大範疇（陳柏熹等人，2015）。為了顧及測量構念的完整性，本研究參考陳柏熹等人素養測驗試題設計方式以及教育成就評量國際協會（The International Association for the Evaluation of Educational Achievement [IEA]）所進行的國際公民教育與素養調查研究的評量架構（International Civic and Citizenship Education Study [ICCS]）（Schulz, Ainley, Fraillon, Losito, & Agrust, 2016 ; Schulz, Fraillon, Ainley, Losito, & Kerr, 2009），將五大核心素養的測量分成認知層面與情感行為層面。其中認知層面，考量分析與推論的認知概念層次較為接近，因此參酌 ICCS2016（Schulz et al. , 2016）將原 ICCS2009 年的架構改為：理解、推論與應用兩層次；本研究中有關情感行為層面的測量架構的產生，則同時參酌 ICCS2009 以及 ICCS2016 測量架構中影響情感行為領域的要素以及能預測實際行為出現的相關變項，如：行為意圖與行為承諾等（Ajzen, Czasch, & Flood, 2009）。基於此，本研究在情感行為層面則以受試者自陳認為各素養的價值性（重要性）、行為意圖與行為（想實踐以及會實踐）等二層面進行測量。

為了定義現代公民五大素養的測量內容領域，本研究除了從陳佳欣（2012）所進

行的專家共識調查結果為基礎，挑選出專家評估較具重要性的核心能力外，也綜合考量 12 年國民基本教育所培育的各項核心素養，最後臚列 15 項培育現代公民五大核心素養的重要能力項目，各核心素養所對應的核心素養指標暨能力項目的定義如表 2 所示。因此形成 5 大核心素養向度*15 項能力指標*5 認知與情感層面的測驗編製架構，試題題數規劃如表 3 所示。

表 2　現代公民核心素養暨能力指標概念型定義

核心素養	核心素養指標		核心素養指標概念型定義	
倫理素養	I-1	道德與實踐能力	I-1	日常的道德問題主動反省、在日常生活中能實踐倫理道德規範並能瞭解實踐道德行為的必要性。
	I-2	尊重與關懷	I-2	能理解人類對大自然可能的破壞、能愛護生活環境、能愛惜自己。
民主素養	II-1	社會參與能力	II-1	能了解政府的基本運作方式，並理解個人和社會之間相互依存關係，並具備經濟和消費技能。
	II-2	多元文化與國際視野	II-2	能具有基本世界歷史地理知識的能力，並關心全球化下衍生的重要議題，且掌握國際社會脈動。
科學素養	III-1	思考與批判能力	III-1	能收集證據，並判斷真偽，並區分事實與價值問題後指出問題核心。
	III-2	主動探索與研究精神	III-2	能運用所學的知能於生活中，主動關懷科技議題並發揮好奇精神及敏銳觀察的能力。
	III-3	問題解決能力	III-3	能提出解決問題的方案，針對問題收集資料並分析、組織與運用資料，並具備獨立思考及反省的習慣。
媒體素養	IV-1	使用資訊科技的能力	IV-1	能運用網路蒐集資料，且運用基本電腦及網路處理日常生活事物，並判斷資源來源的正確性。
	IV-2	理解媒體訊息內容	IV-2	了解並能應用媒體製作技巧與技術、媒體類型與敘事的產製方式，並了解不同媒體的呈現方式與限制。
	IV-3	反思媒體內涵	IV-3	能反思媒體對年齡、性別、種族、職業、階級等呈現的刻板印象，並比較媒體內涵與實際生活中的情境、人物、事件等媒介與社會真實的關係，並解讀媒體再現所潛藏的價值意涵與意識形態。
	IV-4	反思閱聽人的意義	IV-4	瞭解媒體的立場對閱聽人的影響，並反思媒體內容對閱聽人的影響、其所傳達的社會與文化意涵。
	IV-5	溝通表達能力	IV-5	能清晰地闡述自己的理念或意見，運用適當的表達與溝通技巧闡述個人意見並樂於與他人進行溝通。
美學素養	V-1	藝術與美感養成	V-1	能了解在地藝術特色，認識藝術的內涵和理念，並理解人類重要藝術文明。
	V-2	美感體驗與品味生活	V-2	養成參與藝文活動的生活習慣，將藝術融入在日常生活中並對周遭環境的美感加以詮釋或批判。
	V-3	欣賞和美感表達	V-3	能將藝術涵養運用在專業知識上，並對基本的藝術作品內涵有所了解，能聆賞不同創意表現方式的能力。

表 3　中學生現代公民核心素養雙向細目表

核心素養向度	核心素養指標	認知領域		情感行為領域	
		理解	推論與應用	價值性	行為意圖與行為
倫理素養	I-1道德與實踐能力	1	2	1	2
	I-2尊重與關懷	1	2	1	2
民主素養	II-1社會參與能力	1	2	1	2
	II-2多元文化與國際視野	1	2	1	2
科學素養	III-1思考與批判能力	1	2	1	2
	III-2主動探索與研究精神	1	2	1	2
	III-3問題解決能力	1	2	1	2
媒體素養	IV-1使用資訊科技的能力	1	2	1	2
	IV-2理解媒體訊息內容	1	2	1	2
	IV-3反思媒體內涵	1	2	1	2
	IV-4反思閱聽人的意義	1	2	1	2
	IV-5溝通表達能力	1	2	1	2
美學素養	V-1藝術與美感養成	1	2	1	2
	V-2美感體驗與品味生活	1	2	1	2
	V-3 欣賞和美感表達	1	2	1	2
	合計題數		45		45

（二）第二階段情境判斷題組題與情感題編製

1. 訂定素養定義

本部分分別依據認知領域與情感行為領域的內容定義說明。以倫理素養為例，共涵蓋道德與實踐能力以及尊重與關懷兩核心素養指標。而各核心素養指標可藉由具體的核心能力指標進行檢核。舉例而言，以I-1 道德與實踐能力為例，評估是否達到此核心素養，可以透過核心能力指標I-1-1 對日常的道德問題主動反省、I-1-2 在日常生活中能實踐倫理道德規範以及I-1-3 能瞭解實踐道德行為的必要性等三項核心能力指標進行檢核。依據核心能力指標的意涵，達成道德與實踐能力的理解層次；其概念包含：能知道日常生活中可能面臨到的道德問題有哪些，例如：公車讓座問題，能知道

自己應該要如何實踐道德，如：讓座。而推理與應用層次的概念則包含：了解日常生活中有的道德問題後，能去推演出哪些問題也是因為道德能力不足而衍生的。例如：從公車不讓座，舉一反三觀察日常生活中的相關道德議題：在公共場合抽菸；並知道自己有必要去阻止這樣的行為，詳如表 4。依此程序，本研究逐一針對各項核心素養項目指標所對應的認知層次進行定義。因此，評估受試者是否具備對此核心能力指標的理解以及推理與分析，則可以透過上述定義，編擬符合受試者生活經驗的情境判斷試題進行評估，如圖 1 所示。其中，認知領域試題採用二元計分，情感行為領域則使用李克特式四點量尺計分。

表 4　現代公民核心素養當中核心能力情境判斷題組題出題定義範本

素養面向	核心能力	核心能力指標		理解	推理與應用
倫理素養	I-1 道德與實踐能力	I-1-1	對日常的道德問題主動反省	能知道日常生活中可能面臨到的道德問題，例如：公車讓座問題。且能知道自己應該要如何實踐道德，如：讓座。	了解日常生活中有的道德問題後，能去推演出可能道德能力不足而衍生的問題。例如：從公車不讓座，舉一反三觀察日常生活中的相關道德議題：在公共場合抽菸；並知道自己有必要去阻止這樣的行為。
		I-1-2	在日常生活中能實踐倫理道德規範		
		I-1-3	能瞭解實踐道德行為的必要性		

一、君君是一名孕婦且無法久站，搭乘公車時沒有空位可入座，在公車上卻沒有任何人讓座。

(　　) 1. 有人需要座位卻沒有人起身讓座，這顯示生活中何種道德議題沒有受到重視？
(A) 基本人權保障
(B) 不同情況給予不同待遇
(C) 新住民家庭
(D) 性別平權

(　　) 2. 若你曾經遇到圖中類似的情況，下面選項何者是最能夠實踐倫理道德規範的行為？
(A) 若我是坐在位置上的學生，因為路途遙遠，我應該裝睡一下。
(B) 若我是坐在位置上的學生，我應該看著手機螢幕。
(C) 若我是站著的學生，座位上有行李，我會詢問主人是否可以移開。
(D) 若我是站著的學生，我會覺得孕婦與我同病相憐。

(　　) 3. 根據公車上幾位乘客的說法，哪一個人的立場最可能具有道德與實踐能力？
(A) 小明：孕婦站著一定很不舒服，但是我的目的地很遙遠，應該繼續坐著。
(B) 小寶：我媽媽以前也懷孕過，孕婦應該要多站著，坐在椅子上反而不舒服。
(C) 阿蘋：若孕婦有座位的需求，應該讓她有一個較舒適的搭車環境，並告知大家幫助的重要性。
(D) 阿君：若孕婦有座位的需求，坐在博愛座的人應該讓座。

圖 1　倫理素養──道德與實踐能力（認知題範本）

針對情感行為領域，本研究共分為價值性、行為意圖、行為承諾等三層面。依據前述的內容領域，進行每個核心能力情感題的試題內容定義，臚列如表 5；題目範例如圖 2。以I-1 道德與實踐能力為例，其價值性評估以受試者自陳該核心能力的重要性進行評估，例如：我認為倫理與道德很重要；行為意圖與行為則以想做出該行為的念頭進行評估以及會做出實際行動進行評估，例如：我想去實踐符應倫理與道德行為；我會去實踐符應倫理與道德行為。依此程序，本研究逐一針對 15 項核心素養指標所對應的情感層次進行定義。

表 5　現代公民核心素養當中核心能力情感行為題出題定義範本

核心能力	核心能力指標	價值性	行為意圖與行為
I-1-4 道德與實踐能力	I-1-4-1 對日常的道德問題主動反省 I-1-4-2 在日常生活中能實踐倫理道德規範 I-1-4-3 能瞭解實踐道德行為的必要性	對於道德與實踐能力的價值觀與態度，如：認為對日常的道德問題主動反省是重要的。	想做出符應倫理與道德的行為；會做出符應倫理與道德的行為。

4. 請依據您自身狀況，在數字上進行圈選。

關於倫理與道德	非常不同意	不同意	同意	非常同意
我認為倫理與道德很重要。	1	2	3	4
我想去實踐倫理與道德。	1	2	3	4
我會去實踐倫理與道德。	1	2	3	4

圖 2　倫理素養──道德與實踐能力（情感行為題範本）

2. 試題編製與試題修刪流程

情境判斷題組重視特定領域的情境建構，因此所設計的問題情境以符合中學生日常生活情境與經驗為主要考量。試題編製完成後，逐一確認個別試題是否符合雙向細目表、是否符合中學生認知發展階段，並確定測驗用詞符合社會科學強調的價值中立。經過與專家的逐一審題後，逐題做最終確認與細部修改，認定所挑選的試題符合中學生情境，並確認語意適當後，形成預試題本。

本研究共邀請 14 位具有中等學校教學經驗的教師協助進行審題，專家們主要任教領域包含：國中綜合活動領域、生物、數學、國文、理化、輔導、表演藝術、公民等學科專家教師，總計參與專家教師平均教學年資達 10.57 年，其中有 9 位專家教師具有導師與兼任校內行政工作經驗，嫻熟中等學校學生生活經驗。每位專家會拿到中學生現代公民核心素養概念表、試題題本與專家意見單，專家可依據概念表進行審題。本研究商請專家們審核 15 個核心能力情境判斷試題（共 45 題）、校正語句用詞

以及提供試題修改意見，包含質性意見與量化統計。量化部分為「不適合」、「修正後適合」與「適合」。本研究第一次專家審題均無「不適合」之題目，遂依據勾選「修正後適合」之題目進行重新設計，待修改語句不通順及語意含糊不清的試題後，進行第二次專家審題；第二次審題時，所有題目經過所有專家都認定為「適合」後，完成預試題本。預試題本完成後，先邀請幾位國高中生進行試答，用以評估中學生作答所需時間以及是否理解題意。預試對象與資料分析詳述如後，預試分析將適配度不佳的題目進行刪減後，形成正式題本。正式題本中的五大核心素養，皆包含兩個對應的核心能力情境判斷題組，各核心能力情境判斷題組共計 3 道試題，總計含 29 題情境判斷試題與 30 題情感行為領域的自陳題，合計共 59 題。

二、試題施測流程與資料分析

（一）研究對象

1. 預試樣本

預試施測對象究選取臺南市一所國中、高雄市一所綜合高中及嘉義市一所高中的 1 到 3 個年級（均為常態編班），用以檢驗預試階段測驗工具試題特性與試題品質。預試人數選擇依據古典測驗理論（Classical Test Theory [CTT]）預試人數的最低要求：施測人數須大於 300；而試題反應理論（Item Response Theory [IRT]），則以受試者人數需大於題目數量的十倍為最低預試人數要求（余民寧，2009）。基於此，本研究最後共施測 393 份樣本，共收得 373 份的有效樣本。其中，男女比為 153：220，平均年齡為 15.20 歲。

2. 正式樣本

正式施測部分，Karabatsos（2000）指出，當題數為 50 題時，樣本數至少要 1000 人。本研究將全臺中學生進行分層便利取樣，共發放了 1691 份問卷。總計共有來自 22 所中等學校學生參與，總計回收 1561 份，回收率為 92%，經剔除回答不完整或固定答案之無效問卷 82 份，總計有效問卷 1479 份，有效問卷比率為 87.5%。其中，北部共 448 人、中部共 460 人、南部共 421 人、東部共 126 人與離島共 24 人，北：中：南：東：離島的比例為 30：31：28：9：2；國高中職比例為 580：431：468，約各佔 39%、29%與 32%；男女比為 634：844，約為 43%與 57%；平均年齡為 15.47 歲。

（二）研究工具

本研究以陳柏熹等人（2015）大學生基本素養測驗作為效標。大學生基本素養測驗各個素養皆有二至三個題組。研究對象為全國大專院校一至四年級的學生，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，其以部分給分模式（Partial Credit Model [PCM]）（Masters, 1982）來估計，幾乎所有試題與模式都能適配，顯示建構效度良好。試題發展過程均經歷嚴謹修審題程序，具有良好專家效度證據。此外，不同性別和年級的學生在各素養的表現上差異不大，和過去文獻相符合，具有良好效標關聯效度。信度證據方面，各素養能力估計誤差約在.20—.60 評定單位（logit）之間，單一題本信度高於.69，顯示此測驗題數雖少，但信度大致良好。此測驗以線上施測的方式供大學生使用，本研究經由陳柏熹及其研究團隊同意，選取其中測量構念與本研究現代公民核心素養相近測量構念的試題進行效標關聯效度分析。其中以「公民社會」作為民主素養的效標；「問題解決」與「科學思辨」做為科學素養的效標；「美感素養」做為美學素養的效標。

（三）資料分析方法

本研究資料分析目的在檢驗中學生現代公民核心素養測驗的試題品質與正式測驗的信度與效度。因此，有關試題品質的評估採用各試題通過率、鑑別度、誘答力分析以及試題與模式適配程度指標等進行綜合判斷。而效度驗證程序則採用 Messick（1995）試題反應理論效度驗證架構進行中學生現代公民核心素養測驗的效度驗證，Messick 提出效度的六個面向，包含：內容（content）、本質（substantive）、結構（structural）、概化（generalizability）、外在（external）與後果（consequential），此驗證程序已廣泛被運用於國內外相關心理與教育測驗領域的量表編製程序之中（王佳琪等人，2011；王佳琪等人，2014；張貴琳，2013；Hung, 2018）。內容面向代表測驗內容是否符合測驗目標等與內容關聯的證據以及試題品質；本質面向代表受測者的反應如理論所預期；結構面向代表測驗實際測到所要測量的理論結構和特質的程度；概化面向代表分數的特性與詮釋能夠類推到其他群體或是作業的程度；外在面向包含聚斂證據與區辨證據以及效標關聯證據與應用價值；後果面向則是考量測驗使用與分數應用實際與潛在可能帶來的後果。

本研究受限於資料取得，因此乃針對上述所列內容（雙向細目表、專家德懷術分析）、結構、本質、概化（類推效度）與外在（效標關聯效度）效度證據進行說明。

信度估計分別使用 Cronbach's alpha 進行量表內部一致性信度分析與折半信度分析，並使用受試者區分信度係數（person separation index [PSR]）評估試題對於不同能力受試者的區分程度，該信度指標相當於 KR-20 以及 Cronbach's alpha（Linacre, 1997）；效標關聯效度分析則選取陳柏熹等人（2015）大學生基本素養測驗中的「公民社會」、「問題解決」與「科學思辨」以及「美感素養」等分向度測驗做為效標。本研究所發展測驗屬於混合二元計分與多點計分題型，因此選擇 Rasch 家族中的部分給分模式（Partial Credit Model [PCM]）（Masters, 1982）進行模式參數估計。以 ConQuest Version 2.0（Wu, Adams, & Wilson, 2007）進行資料分析，參數估計以多向度隨機係數多項洛基模式（multidimensional random coefficients multinomial logit model [MRCMLM]）（Adams, Wilson, & Wang, 1997），其可透過最大概似邊緣估計（marginal maximum likelihood [MML]）與 Bock 與 Aitkin（1981）所提的期望最大化算則（expectation-maximization [EM] algorithm）執行。由於貝氏期望後驗估計（expected a posteriori [EAP] estimate）（Bock & Mislevy, 1982）不僅估算較為快速且誤差較小（Embretson & Rise, 2000），因此，本研究採用 EAP 估計結果做為受試者的能力參數估計值，單位為 logits。試題與模式適配度指標採用 Infit MNSQ 及 Outfit MNSQ，其數值介於 0.5—1.5 之間，表示模式適配良好（Linacre, 2002）。除此之外，為了提供類推效度證據，也針對正式施測群體進行不同性別試題差異功能檢驗（differential item functioning [DIF]）（Holland & Wainer, 1993）。

伍、結果與討論

一、試題信度

為瞭解本研究所發展的測驗在正式樣本的施測狀況，本研究將正式題本施測結果進行內部一致性分析。根據 Jensen（1980）指出，信度在.80 以上便屬於中高信度。本研究之正式題本試題內部一致性信度為.882。將現代公民核心素養測驗依據測量內容折半，兩個短測驗皆包含五個面向核心素養情境題組（倫理、民主、科學、媒體、美學），包含認知題與情感行為題，兩個半測驗之間相關為.791，經 Spearman-Brown 校正公式校正後的折半信度為.883。除此之外，本研究以多向度部分給分模式進行資

料分析，結果發現現代公民核心素養測驗的認知題試題區分信度係數（Item separation reliability）為.995，情感行為題試題區分信度係數為.996，代表本研究受試群體已足夠區分出不同難度層級的試題。一般而言，試題區分信度係數大於.9 代表受試群體數量足以區分出不同難度層級的試題（Wright & Stone, 1999）。各個素養中，認知題的受試者區分信度係數（person separation reliability）：倫理認知為.593；民主認知為.729；科學認知為.675；媒體認知為.751；美學認知為.736，整體的內部一致性信度（Coefficient Alpha）為.82；情感行為題的受試者區分信度係數：倫理情感行為為.866；民主情感行為為.882；科學情感行為為.880；媒體情感行為為.873；美學情感行為為.880，整體的內部一致性信度（Coefficient Alpha）為.96，顯示試題對於不同能力受試者能有良好的區分性。本測驗認知題難度範圍介於-2.84 至 0.50 logits，情感行為題難度範圍介於-6.61 至-1.41 logits 之間，表示本研究所發展的測驗試題能夠測出不同素養程度的學生。整體而言，中學生現代公民核心素養測驗各分測驗的認知層面與情感行為層面試題區分信度係數皆達.8 以上，而測驗折半信度係數亦達.8 以上，顯示測驗內部試題同質性佳。

二、效度支持證據

（一）內容效度證據

為了確保試題編製與測驗目標契合，本研究首先針對測量內容領域進行定義（詳細定義如表 2），並據此發展測驗編製藍圖（如表 3）。為了檢視本測驗所編製的試題內容是否符合測驗目標，在完成測驗試題後乃依循測驗編製藍圖反覆檢視試題是否符合測驗目標以及試題內容是否涵蓋所有欲測量的內容領域，亦即現代公民所需具備的五大核心素養。此外，也商請 14 位專家們審核 15 個核心能力情境判斷試題（共 90 題）、校正語句用詞以及提供試題修改意見，包含質性意見與量化統計。量化部分為「不適合」、「修正後適合」與「適合」。本研究共進行兩回合專家效度問卷，以確保所有專家均認可所有試題切合所欲測量的內容領域。

除了透過質化程序檢視測驗內容是否吻合所欲測量屬性，也透過進行預試試題分析檢視試題品質，確保測量內涵的適切性。預試階段，試題通過率、鑑別度與選項誘答力分析結果發現，本研究所發展的現代公民核心素養測驗的測驗題通過率介於.13—.93 之間。根據 Ebel 與 Frisbie（1991）鑑別度（Discrimination，以下簡稱 D）判斷

標準，D 為.20—.29 表示試題尚可但須作局部修改；D 為.19 以下，表示試題低劣，需要刪除或修改。因此，本研究針對鑑別度數值低於.20 以下的試題進行修改或是刪除。由於預試題數較多，綜合考量未來量表正式施測的時間花費以及測量內涵的完整性，本研究規劃五大素養面向各保留兩個情境題組。依此目標並考量前述試題篩選標準進行試題刪除與修改，刪完後的試題形成正式題本，共計 59 題，並且在正式施測階段後進行試題信度與其他各項效度證據分析。

（二）結構效度證據

考量認知層面與情感行為層面可能反應不同的潛在能力，因此將兩部分的試題分開分析。如表 6 所示，將認知層面試題進行單向度 Rasch 模式與多向度 Rasch 模式進行比較後，發現卡方值達顯著（$\chi^2_{df=19}$= 245.75 > 6.844），表示五向度模式與單向度模型確實有所不同，而以五向度模式與資料適配程度較佳；將情感行為層面進行單向度 PCM 模式與多向度 PCM 模式進行比較後，發現卡方值達顯著（$\chi^2_{df=10}$= 7534.57 > 2.558），表示五向度模式與單向度模型確實有所不同，而以五向度模式與資料適配程度較佳；故本研究依據此分析結果與文獻探討所建立的假設，將現代公民核心素養測驗認知層面與情感行為層面分別以多向度模式進行呈現。

表 6　現代公民核心素養測驗向度分析模式比較摘要表

	認知層面		情感行為層面	
	偏差平方和	估計參數	偏差平方和	估計參數
多向度模式	46369.98786	49	46243.17727	101
單向度模式	46615.74212	30	53777.74779	91

N=1479

多向度模式分析結果整理如表 7，基於估計精準度的考量，部分情感行為領域試題由於試題在特定選項類別的人數過少，甚至沒有受試者填答，因此將部分試題的選項人次合併進行估計。從本研究分析結果來看，所有試題的 Infit MNSQ 指標均介於 0.5~1.5 之間且所有試題的閾難度皆符合預期，亦即受試者選答同意程度較低的選項其閾難度估計值也較小，反之同意程度愈強的選項其閾難度估計值也愈大，所有試題閾

難度的排序皆依序而未出現逆轉的情況。

表 7　中學生公民素養測驗試題難度平均值、難度與適配度範圍摘要表

		認知			情感行為		
		平均難度	難度範圍	適配度範圍	平均難度	難度範圍	適配度範圍
倫理素養	道德與實踐	-0.64	-1.70～0.50	0.94～1.10	-4.84	-5.44～-3.71	0.83～1.00
	尊重與關懷	-1.45	-2.84～-0.60	0.97～1.00	-5.08	-5.27～-4.87	0.83～1.00
民主素養	社會參與	-1.74	-2.20～-0.93	0.87～1.00	-2.04	-2.93～-1.41	0.96～1.10
	多元文化與國際視野	-0.84	-1.80～0.98	0.95～1.20	-3.06	-3.27～-2.76	0.83～0.90
科學素養	思考與判斷能力	-0.28	-1.30～0.96	0.93～1.00	-5.62	-6.61～-4.58	0.91～0.97
	問題解決能力	-1.04	-1.40～0.41	0.93～1.00	-5.06	-5.82～-4.11	0.82～1.00
媒體素養	理解媒體訊息內容	-0.89	-1.80～-0.05	0.09～1.10	-3.76	-4.35～-3.38	1.00～1.10
	反思媒體內涵	-0.04	-0.82～0.10	0.93～1.10	-3.54	-3.70～-3.39	0.94～0.99
美學素養	藝術與美感養成	-1.23	-1.50～-0.79	0.90～1.00	-4.05	-4.92～-3.40	1.00～1.20
	欣賞和美感表達	-0.86	-1.27～-0.62	0.95～1.10	-4.56	-5.23～-4.01	0.95～1.10

單位：logit

圖 3 為倫理、民主、科學、媒體與美學素養的受試者認知能力及題目難度對照圖。圖 4 為倫理、民主、科學、媒體與美學素養的受試者情感行為能力及題目難度對照圖。圖左邊的"X"標示受試者在倫理、民主、科學、媒體與美學素養的認知能力分佈情況，數值愈大表示受試者在該核心素養認知程度愈高；最右側則標示各試題的難度分佈，數值愈大表示愈困難；情感亦然。理想上，我們希冀試題難度範圍可以遍及所有受試者能力範圍，如此對於不同受試能力的受試者而言，均有相對應其能力的試題可供精準量測其能力。由圖 3 顯示，各素養的認知測驗難度大致與受試者能力分布範圍接近，

受試者在各素養向度的能力表現分布大致都在±3 範圍內，且各向度能力分布的標準差低於 0.69，惟對於高認知能力者來說，對應其認知能力的試題仍需持續增加。而由圖 4 顯示，在情感行為領域方面，受試者在各素養的能力分布範圍介於-7.00 到 9.00 logits 之間，而試題的難度範圍則主要落在-11.00 到 3.00 logits 之間，顯示對於高情感行為能力者來說，對應其能力的試題仍需持續增加，而在情感行為領域，受試者在各素養面向的能力分布落差也相對於認知層面來得大，除了在民主素養方面能力分布的標準差為 0.79，其他各向度的能力分布標準差均大於 1 以上。圖 5 為現代公民核心素養五個分測驗認知題訊息曲線與全認知測驗的試題訊息曲線與能力估計誤差圖，分測驗依序為倫理認知、民主認知、科學認知、媒體認知與美學認知；從圖中可以看出，現代公民核心素養認知對於受試能力介於-2.6 至 0.6 者，估計最為精準，估計誤差低於 0.5。圖 6 為現代公民核心素養五個分測驗情感行為題訊息曲線與全情感行為測驗的試題訊息曲線與能力估計誤差圖，分測驗依序為倫理情感行為、民主情感行為、科學情感行為、媒體情感行為與美學情感行為；從圖中可以看出，現代公民核心素養情感行為對於受試能力介於-6 至-2.4 者，估計最為精準，估計誤差低於 0.5。

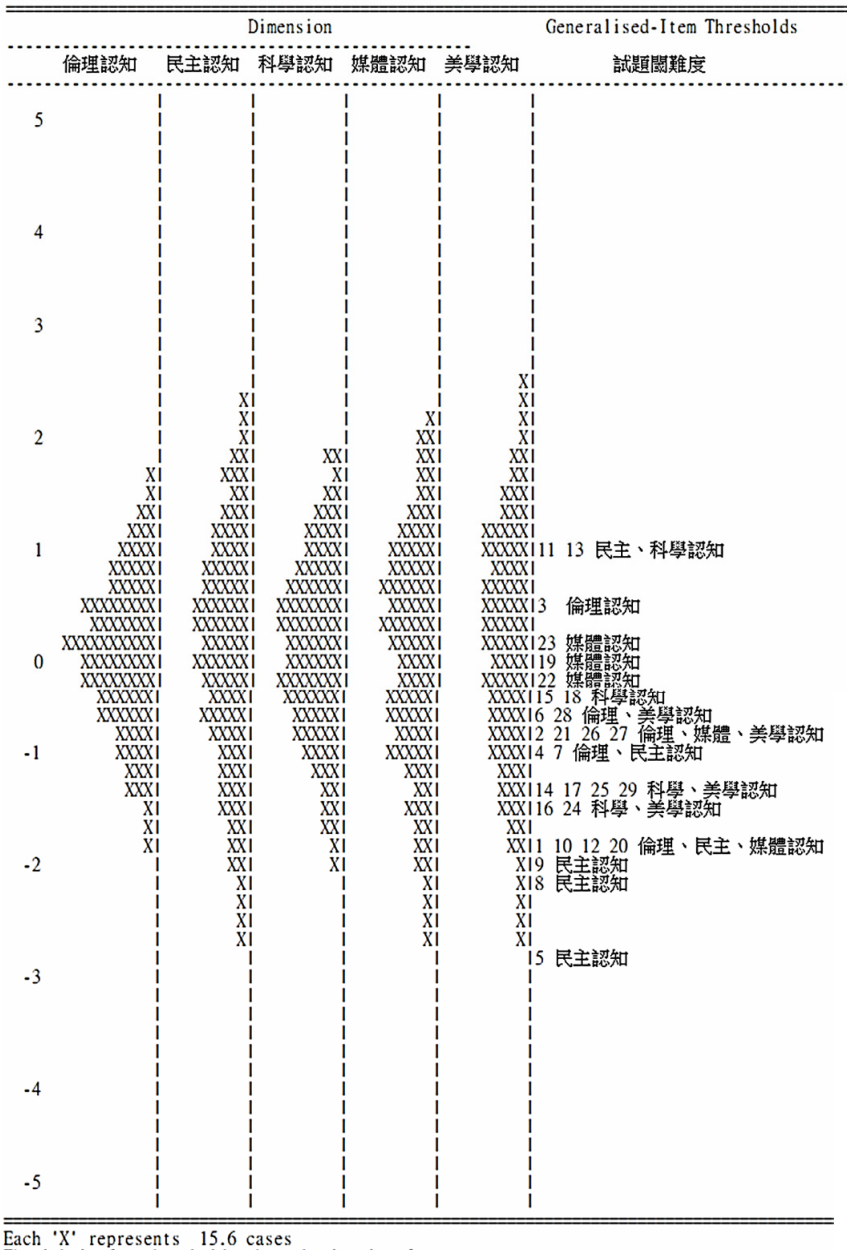

圖 3　各認知題向度的受試者能力及試題難度對照

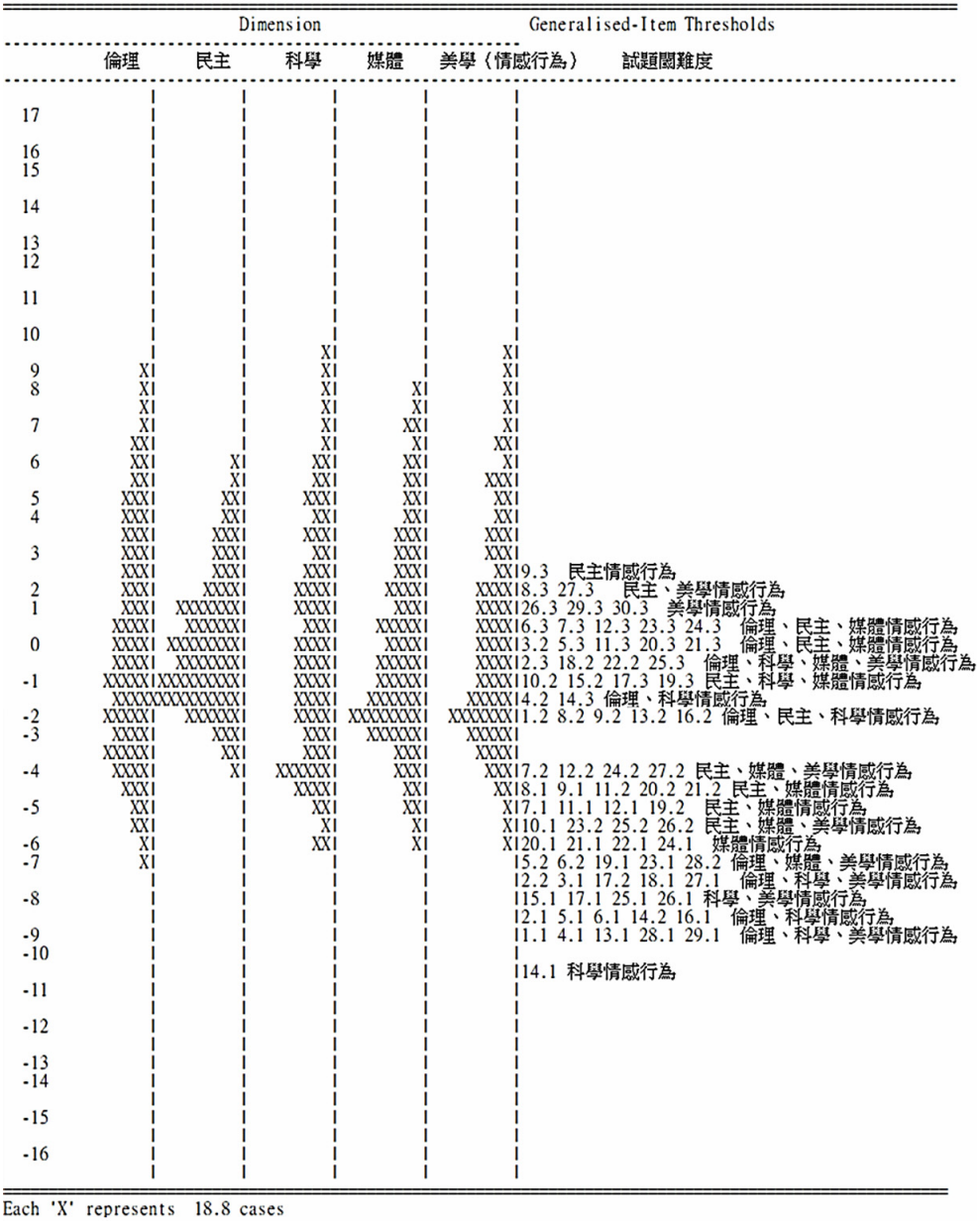

圖 4　各情感行為題向度的受試者能力及試題難度對照

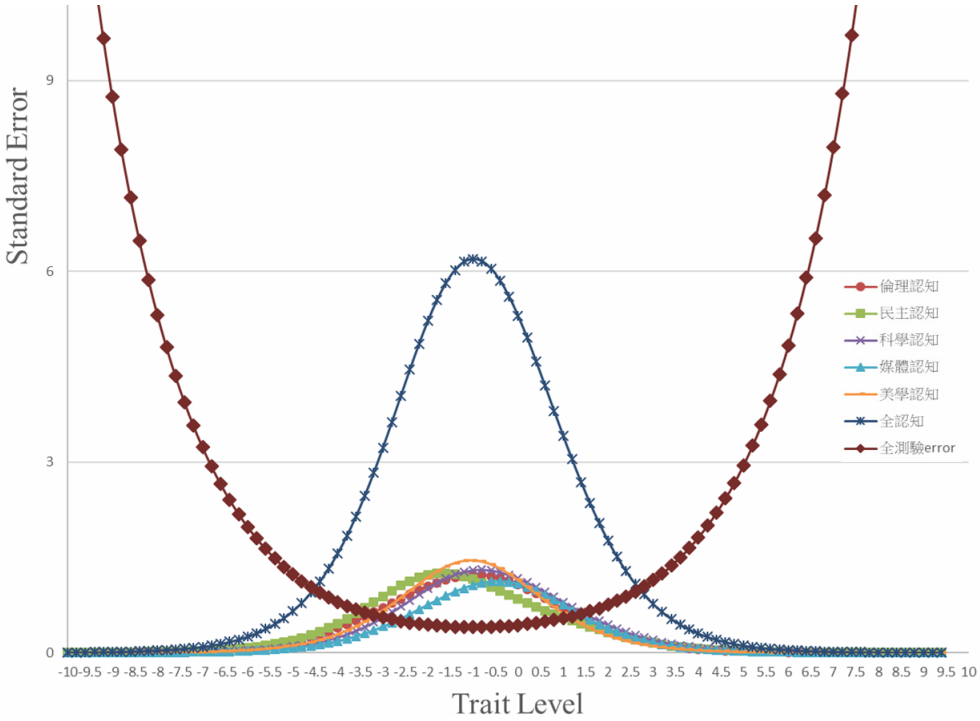

圖 5　認知測驗訊息曲線與全認知測驗的試題訊息曲線與能力估計誤差圖

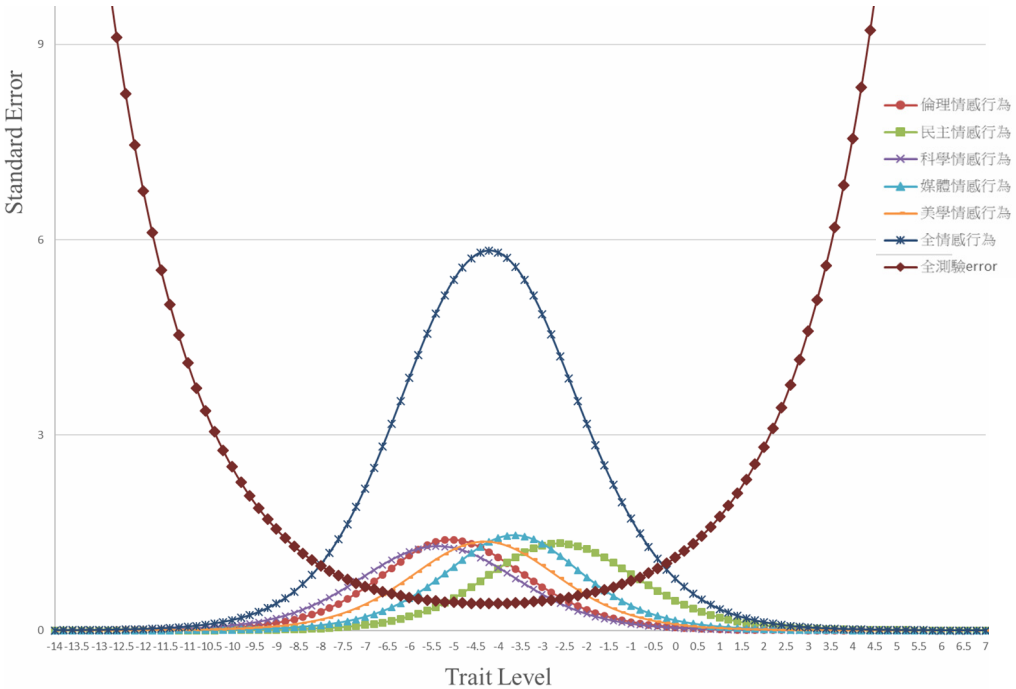

圖 6　情感行為測驗訊息曲線與全情感行為測驗的試題訊息曲線與能力估計誤差圖

（三）本質效度證據

本研究的素養試題設計，預期隨著學生的能力愈高，愈可能達到較高的程度。McFadden 與 Skaggs（2013）以本質效度作為量表的效度證據，本質效度代表構念特質高低能與作答者反應得分高低機率相連結；Messick（1995）亦提及，若受試者反應通過的程度與測驗尺度程度高低的假定符合原先架構的預期，亦即證據與推論之間的關係獲得驗證，便符合本質效度；從本研究資料符合 PCM 模式的狀況，可發現結果符合本質上的預期，隨著學生的能力增加，學生達到高 Level 的機率也會愈高，亦即所測量的試題內容能與欲測量的特質相符應。

單一試題部分，以民主素養—社會參與能力為例，透過多向度 PCM 分析結果，學生在認知題與情感行為題四個選項達到不同 Level 的機率如圖 7 與圖 8 所示。橫軸表示學生的社會參與能力，單位以 logit 表示，愈右邊表示學生能力愈高；縱軸表示學

生達到每一個 level 的機率。圖 7 代表學生社會參與能力認知答對的機率，由圖可以看出學生能力低於-0.93 個 logits 時，則答對機率將低於.5；反之，當學生能力高於-0.93logits 時，最答對機率則相對提高。圖 8 代表學生社會參與能力的情感行為題在四個 level 的機率，由圖可以看出學生能力低於-4.21 個 logits 時，最可能達到 Level0；當學生能力介於-4.21—-2.17 logits 之間，最可能達到 level 1；當學生能力介於-2.17—2.15 logits 時，最可能達到 level2；當學生能力高於 2.15 時，最可能達到 level3。以情感行為試題而言，隨著學生能力愈高，其選答較高階難度的機率也愈高。因此觀察每道試題的閾難度估計值是否有逆轉的情況也可以了解學生選答狀況是否吻合預期。以本測驗的情感行為試題來說，多數情感行為試題具有 2 到 3 個閾難度，顯示受試者在部分情感行為試題上對於不同意與非常不同意兩選項的選答情況較少。因此，將選答較少的選項合併進行分析，所有試題皆無出現閾難度反轉的情況，顯示本研究中所發展情感行為試題均符合本質效度預期。

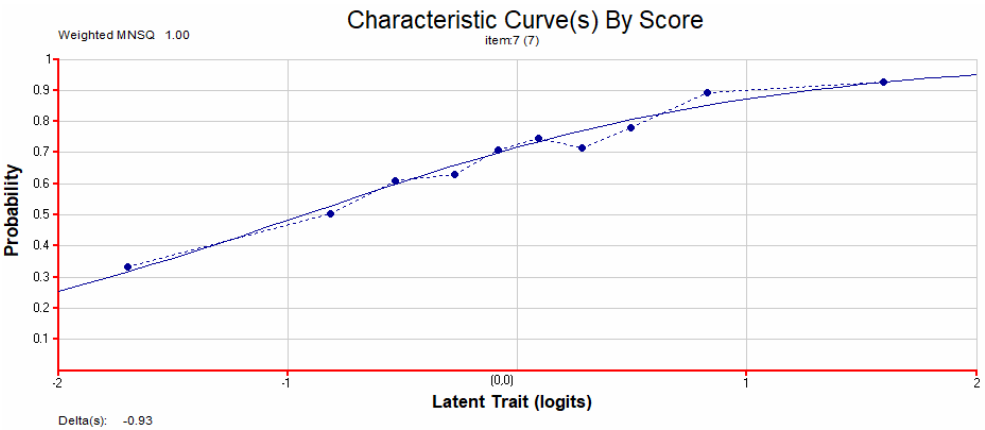

圖 7　社會參與能力「認知題：社會參與能力理解」II-1-1（認知第 7 題）

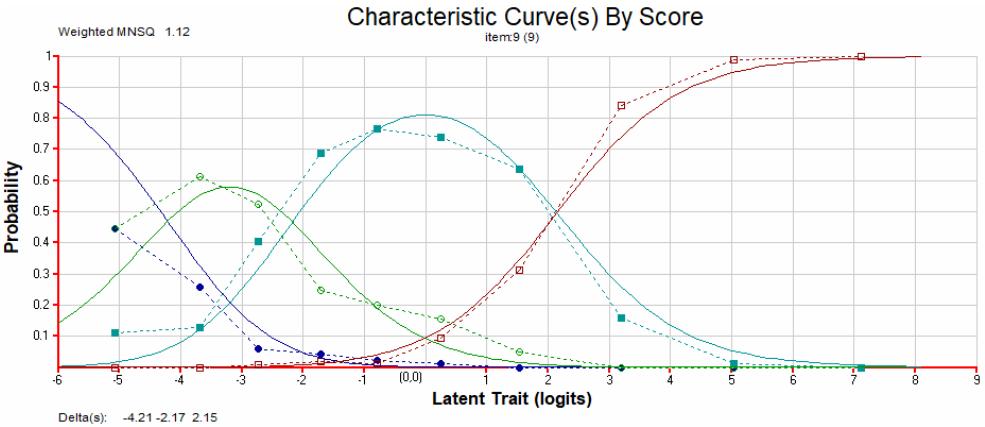

圖 8　社會參與能力「情感行為題：社會參與能力行為意圖」II-1-4-3（情感行為第 9 題）

本研究依據 46 位專家學者認定分屬五大素養之指標進行試題編製，從構念成分來看，受試者在五大核心素養的認知層面與情感行為層面潛在能力關聯情形表 8 所示，受試者在各素養的認知層面間的相關介於.671 至.818 之間；而在各素養的情感行為層面相關介於.741 至.864 之間。僅管各素養之間的關聯情形不低，不過若僅採用受試者在單一認知或單一情感行為表現用以解釋其另一認知或另一情感行為的表現仍有所不足，而透過五個素養向度間的關聯訊息以便能對於受試者現代核心素養的潛在認知能力估計與情感行為能力估計提供更為精準的估計。此外，受試者在各素養的認知能力與各素養的潛在情感行為相關較低，顯見學生認知能力高，情感行為能力不一定高。由於本研究中情感行為係採用自陳量表的形式，用以了解受試者的態度、行為意向，而無標準答案，然則認知層面則是評量受試者在各素養面向所展現的能力，屬於最大表現測驗。過去陳柏熹等人（2015）針對大學生素養測驗的研究結果也發現，受試者在自評的情感面向與認知面向間的做答反應未必有所關聯，因而素養的測量應同時涵蓋情意與認知的測量，方可完整評量素養的概念。

表 8　現代公民核心素養測驗相關矩陣

	倫理認知	民主認知	科學認知	媒體認知	美學認知	倫理情感行為	民主情感行為	科學情感行為	媒體情感行為	美學情感行為
倫理認知	1									
民主認知	.728**	1								
科學認知	.671**	.770**	1							
媒體認知	.686**	.818**	.814**	1						
美學認知	.682**	.766**	.729**	.777**	1					
倫理情感行為	.070**	.049	.060*	.039	.055*	1				
民主情感行為	.097**	.083**	.084**	.071**	.080**	.830**	1			
科學情感行為	.130**	.115**	.132**	.113**	.116**	.845**	.864**	1		
媒體情感行為	.148**	.139**	.149**	.133**	.130**	.741**	.860**	.854**	1	
美學情感行為	.061*	.055*	.068**	.054*	.055*	.747**	.816**	.796**	.760**	1

N=1479

*p<.05；**p<.01；***p<.001

（四）類推效度證據

差異試題功能（differential item functioning [DIF]）的檢驗係確保測驗品質與公平性的重要方法之一。本研究大部份題目的難度估計值在性別上的差距皆小於 0.43logits，表示男生與女生在作答反應上沒有差異功能存在；大於 0.638，屬於顯著差異功能（Paek, 2002；Wilson, 2005）。所有試題中大於 0.638 logits 的有倫理素養當中的「尊重與關懷」的情感行為領域題 10、11、12 等三題，差距為 0.930、0.708 與 0.654；民主素養「參與社會公共議題」的認知題 15，差距為 0.688。本研究男、女學生在多數試題上皆無差異功能，亦即不會因為試題本身品質而影響不同性別在該題的作答，但有 4 題有顯著差異功能，若要以本測驗進行男女差異之分析，應將此 4 題去除。整體而言，測驗題目在不同性別具相等的測量意涵。

（五）效標關聯效度證據

本研究以 135 位中學生為對象，同時施測大學生基本素養測驗與現代公民核心素養測驗，以進行效標關聯效度分析。由於大學生基本素養測驗與本研究之中學生現代公民核心素養測驗之架構並不相同，故僅擷取其中較相近的測量構念作為效標，分別

以大學生基本素養測驗中的公民社會、問題解決與科學思辨以及美感素養與本研究所發展的中學生民主素養、科學素養、美學素養等進行相關矩陣分析。結果如表 9 所示，本研究所發展的「中學生現代公民核心素養」當中的「民主素養」測驗，與大學生基本素養當中的公民社會有顯著相關，為.419，且較其他素養的相關高；本研究所發展的中學生現代公民核心素養當中的美學素養測驗，與大學生基本素養當中的美學素養有顯著相關，為.345，且較其他素養的相關高，顯示兩份測驗區辨效度良好。值得注意的是，本研究所發展的科學素養測驗，與大學生基本素養測驗的科學素養雖然達.269 顯著正相關，但相對於與大學生美學素養測驗間的關聯情況則稍低。而以受試中學生在大學生科學素養與美學素養的表現來看，達.558 顯著正相關。推敲可能的原因是研究者所編的科學素養測驗以及大學生素養測驗，當中科學議題的融入可能涉及與美學相關議題；此外，從受試者在科學與美學素養測驗上的關聯強度來看，也可發現這兩大素養測驗背後與測量內容領域也有一定程度的關聯情形。就整體看來，造成中學生素養測驗與大學生素養測驗間的中低度關聯強度的可能原因是中學生在作答大學生基本素養測驗時，因題目描述與使用電腦施測，以及科學素養試題的難度對中學生來說較難，因而產生相關程度不高的狀況。

表 9　中學生民主素養、科學素養、美學素養與大學生基本素養相關矩陣

Pearson 相關	1	2	3	4	5	6
1. 民主素養 （社會參與、多元文化與國際視野）	1					
2. 科學素養 （思考與判斷、問題解決能力）	.690**	1				
3. 美學素養 （藝術與美感養成、欣賞與美感表達）	.633**	.717**	1			
4. 大學生民主素養 （公民社會）	.419**	.316**	.319**	1		
5. 大學生科學素養 （問題解決、科學思辨）	.241**	.269**	.240**	.112	1	
6. 大學生美學素養 （美感素養）	.319**	.355**	.345**	.157	.558**	1

N=135

**p<.01

陸、結論與建議

本研究以中學生為評量對象編製信度與效度良好的中學生現代公民核心素養測驗。總結其特點如下：(1)本研究所發展的中學生現代公民核心素養測驗在測量內涵上能夠直接回應教育部顧問室所規劃的五大核心素養（教育部，2015B），測量構念相對於單一素養的測量工具而言較為完整；(2)在測驗命題上也採用符合受測對象的學習與生活經驗發展情境式測驗。採用情境式判斷測驗題型，不僅較能測量出受測者的真實能力，也因為命題情境較具生活化與更為擬真，更符合生態效度。目前這樣的題型也被運用來測量師資生教學潛能（趙子揚等人，2016）以及大學生核心素養等測驗（陳柏熹等人，2015）；(3)在試題編擬與審查方面，以嚴謹的程序進行測驗編製與試題修審，本研究以嚴謹程序彙集領域專家建議，針對試題進行逐一修改，以確保試題的品質；(4)提供良好的信度指標以及多元的效度支持證據。在信度方面，本研究所發展的測驗，各個素養的受試者區分信度係數（person separation reliability）均達.8 以上，顯示試題對於受試者的能力能有良好的區分性。測驗的折半信度為.883，顯示若將題本分為兩個半測驗的同質性甚佳；(5)本研究的效度驗證程序採用 Messick（1995）所提的試題反應理論效度驗證架構進行，結果顯示本研究所發展測驗符合五向度的測量模型，亦即中學生的現代公民核心素養，包含：倫理素養、民主素養、科學素養、媒體素養以及美學素養等五大素養，吻合測驗的建構。本現代公民核心素養對於受試能力介於±4 者，估計最為精準；類推效度證據部分，共有 4 道試題出現性別差異功能的狀況，若要以本測驗進行性別差異分析，應將此 4 題去除；本質效度證據部分，無論是認知測驗或是情意測驗，所有試題的閾難度（threshold difficulty）估計值皆符合測驗設計上的本質預期；(6)本研究提供測驗使用者測驗結果解釋，使得分數說明與解釋有所依循，詳見附錄一的中學生現代公民核心素養測驗等級描述範例。

本研究使用教育部顧問室所提的五大核心素養（教育部，2015B）作為編製的基本架構，此架構就測驗編製來說，具有可行性與可測量性，但不可諱言此舉也摒除了其他研究者對於現代公民素養的多元內涵的定義，因此就測量構念上就有所限縮，此乃本研究限制之一。其次，本研究在五大現代公民素養量測兼顧認知與情意，然而在情意測量方面仍未脫離採用自陳的方式進行評估，未來研究可針對受評者在這些素養

面向的內隱態度進行量測。此外，針對情意測量的試題評量內容，本研究希冀從價值性、行為意圖與行為等層面進行評量，因此預期此三試題的題目難度排序以行為實踐為最難，雖然幾乎全數試題都符合此預期，但是仍有少數試題與預期有些微落差。此結果突顯出測驗命題思維與作答者實際做答反應未必完全一致，未來研究亦可針對作答者答題反應歷程進行更深入的質性探討。再者，本研究在效度驗證方面以 Messick（1995）試題反應理論效度驗證架構進行各項效度證據的驗證，惟未來研究仍可透過長期的追蹤調查輔以行為觀察以持續針對量表的預測效度進行探討。

最後，本測驗編製之用意並非取代 PISA 科學素養等多種現行已有之素養測驗，而是提供教師在課室操作簡便且快速可診斷學生素養狀況的評測工具；本研究嘗試將課程綱要總綱中三面九項的核心素養與現代公民五大核心素養下的指標項目進行對應，惟如前所述，課程綱要總綱所強調的核心素養是一種通用性的能力，亦即課程綱要所強調的核心素養內涵已超越本研究所欲測量的架構。此外，針對本研究中各素養細部指標構念，本研究考量一般課室情境施測時間限制，最後僅保留各素養具代表性之兩個指標進行量測，惟在編製之初亦嚐試針對所有可測量的指標進行試題編製，未來仍可針對其他細部指標發展相關試題，俾使更貼近現代公民核心素養構念的全貌。

誌　謝

感謝匿名審查委員及編輯委員會提供之寶貴修正意見，以及參與本研究之專家學者的熱心協助，本研究為科技部計劃（MOST 106-2410-H-006-057-MY2）之部分成果，特此致謝。

參考文獻

王文中（2004）。Rasch 測量模式與其在教育與心理之應用。**教育與心理研究，27**（4），637-694。

[Wang, W. -C. (2004). Rasch measurement theory and application in education and psychology. *Journal of Education & Psychology, 27*(4), 637-694.]

王佳琪、何曉琪、鄭英耀（2014）。「科學創造性問題解決測驗」之發展。**測驗學刊，61**（3），337-360。

[Wang, C. -C., He, H. -C., & Cheng, Y. -Y. (2014). Development of the children scientific creative problem solving test. *Psychological Testing, 61*(3), 337-360.]

王佳琪、鄭英耀、劉昆夏、何曉琪（2011）。以 Rasch 分析檢驗「多向度幽默感量表」之信效度。**測驗學刊，58**（4），691-713。

[Wang, C. -C., Cheng, Y. -Y., Liu, K. -H., & He, H. -C. (2011). Validation of the multidimensional sense of humor scale using Rasch analyses. *Psychological Testing, 58*(4), 691-713.]

余民寧（2009）。**試題反應理論（IRT）導論與應用。**台北：心理。

[Yu, M. -N. (2009). Introduction and application of item response theory (IRT). Taipei: Psychological Publishing.]

呂秀蓮（2019）。課綱為本課程設計經驗之研究：以國中教師為對象。**教育實踐與研究，32**（1），1-32。

[Lu, H. -L. (2019). Standards-based unit design: The experiences of middle school teachers. *Journal of Educational Practice and Research, 32*(1), 1-32.]

吳京玲（2011）。建構大學生核心能力架構之研究：分析學術界與產業界的觀點。**通識教育學刊，7，**9-33。

[Wu, C. -L. (2011). The construction of college student core competency framework: Comparing academic and industry viewpoints. *Taiwan Journal of General Education, 7,* 9-33.]

吳京玲、林妙容、李信（2009）。大學生核心能力培育之研究－以「探索、體驗與自我成長」通識課程爲例。**通識教育學刊，4，**121-144。

[Wu, C. -L., Lin, M. -J., & Lee, -H. (2009). Cultivating college students' core competencies: An example of a general education course "Exploring, experiencing and self-growth". *Taiwan Journal of General Education, 4,* 121-144.]

李俊湖（2009）。公民素養與公民教育。**教育人力與專業發展，26**（1），1-2。

[Li, C. -H. (2009). Citizenship and civic education. *Educators and Professional Development, 26*(1), 1-2.]

巫博瀚、賴英娟、施慶麟（2013）。「Rosenberg 自尊量表」之試題衡鑑：評等量尺模型的應用。**測驗學刊，60**（2），263-289。

[Wu, P. -H., Lai, Y. -C., & Shih, C. -L. (2013). Evaluation of the Rosenberg self-esteem scale using the rating scale model. *Psychological Testing, 60*(2), 263-289.]

林永豐（2014）。素養的概念及其評量。**教育人力與專業發展，31**（6），35-47。

[Lin, Y. -F. (2014). The concept of literacy and its assessment. *Educators and Professional Development, 31*(6), 35-47.]

侯雅齡（2010）。「國民中學自然科學性向測驗」之編製。**測驗學刊，57**（1），29-58。

[Hou, Y. -L. (2010). The development of natural science academic aptitude tests for junior high school students. *Psychological Testing, 57*(1), 29-58.]

洪裕宏（2011）。定義與選擇國民核心素養的理論架構。**教育人力與專業發展，28**（4），15-24。

[Hung, Y. -H. (2011). Defining and selecting the theoretical framework of national core literacy. *Educators and Professional Development, 28*(4), 15-24.]

教育部（2011）。公民素養塑陶計畫。**教育部人文社會科學相關領域計畫入口網。**取自 http://hss.edu.tw/wSite/mp?mp=1

[Ministry of Education. (2011). *Citizen literacy plastic pottery plan.* Retrieved from Ministry of Education, Institute for Humanities and Social Sciences Web site: http://hss.edu.tw/wSite/mp?mp=1]

教育部（2012）。現代公民核心能力養成計畫。**教育部人文社會科學相關領域計畫入口網。**取自 http://hss.edu.tw/wSite/lp?ctNode=342&mp=4&idPath=&nowPage=1&pagesize=300

[Ministry of Education. (2012). *Modern citizen core ability development plan.* Retrieved from Ministry of Education, Institute for Humanities and Social Sciences Web site: http://hss.edu.tw/wSite/lp?ctNode=342&mp=4&idPath=&nowPage=1&pagesize=300]

教育部（2014）。十二年國民基本教育課程綱要總綱。**國家教育研究院。**取自 http://www.naer.edu.tw/files/15-1000-7944,c639-1.php?Lang=zh-tw

[Ministry of Education. (2014). *Outline of the Twelve-year National Basic Education Curriculum.* Retrieved from National Academy for Educational Research Web site: http://www.naer.edu.tw/files/15-1000-7944,c639-1.php?Lang=zh-tw]

教育部（2015A）。**十二年國民基本教育領域課程綱要核心素養發展手冊。**國家教育研究院課程及教學研究中心核心素養工作圈，未出版。

[Ministry of Education. (2015A). *Twelve-year National Basic Education Curriculum outline core literacy development manual.* National Institute of Education Curriculum and Teaching Research Center core literacy work circle.]

教育部（2015B）。**現代公民素養內涵。**取自 http://ms3.ntcu.edu.tw/pey/public_html/citizen.htm

[Ministry of Education. (2015B). *Modern citizen literacy.* Retrieved from Ministry of Education Web site: http://ms3.ntcu.edu.tw/pey/public_html/citizen.htm]

陳伯璋（2010）。臺灣國民核心素養與中小學課程發展之關係。**課程研究，5，**1-25。

[Chen, P. -C. (2010). The relation between the core competences and elementary and secondary school curriculum development in Taiwan. *Journal of Curriculum Studies, 5,* 1-25.]

陳佳欣（2012）。**現代公民核心能力指標報告。**教育部資訊及科技教育司現代公民核心能力養成計畫總計畫（編號：NSC99-2511-S-006-010），未出版。

[Chen, C. -H. (2012). *Modern citizen core competence indicator report.* Ministry of Education Information and Technology Education Council Report No. NSC99-2511-S-006-010.]

陳柏熹、黃馨瑩、陳郁欣、葉泰廷、蘇少祖（2015）。大學生基本素養測驗的發展及信度效度分析。**教育科學研究期刊，60**（3），95-126。

[Chen, P. -H., Huang, H. -Y., Chen, Y. -H., Yeh, T. -T., & Su, S. -T. (2015). Evaluation of the psychometric properties of the general literacy test for university. *Journal of Research in Education Sciences, 60*(3), 95-126.]

張貴琳（2013）。青少年線上閱讀素養評量工具之發展。**教育實踐與研究，26**（2），29-66。

[Chang, K. -L. (2013). The development of adolescent online reading literacy assessments. *Journal of Educational Practice and Research, 26*(2), 29-66.]

張鈿富、吳慧子、吳舒靜（2010）。歐盟、美、澳「公民關鍵能力」發展及其啟示。**教育資料集刊，48，**273-299。

[Chang, T. -F., Wu, H. -T., & Wu, S. -C. (2010). Development and their implications of civic key competencies of the EU, the USA, and Australia. *Bulletin of Educational Resources and Research, 48,* 273-299.]

莊富源（2013）。建構與釋義：大學生公民素養內涵探析。**教育研究月刊，227，**75-88。

[Chuang, F. -Y. (2013). Construction and interpretation: Citizenships of college students. *Journal of Education Research, 227,* 75-88.]

黃宏宇、洪素蘋（2007）。中文網路成癮量表之等級反應模式分析。**測驗學刊，54**（2），331-353。

[Huang, H. -Y., & Hung, S. -P. (2007). An IRT study of graded response model on the Chinese internet addiction scale. *Psychological Testing, 54*(2), 331-353.]

董秀蘭（2016）。社會領域：一個培養現代公民素養與核心能力的關鍵領域。**教育脈動，5，**1-12。

[Tung, H. -L. (2016). Social sphere: A key area for cultivating modern citizen literacy and core competencies. *Educational Pulse, 5,* 1-12.]

趙子揚、黃嘉莉、宋曜廷、郭蕙寧、許明輝（2016）。教師情境判斷測驗之編製。**教育科學研究期刊，61**（2），85-117。

[Chao, T. -Y., Huang, J. -L., Sung, Y. -T., Kuo, H. -N., & Hsu, M. -H. (2016). Construction of the teacher situational judgment test. *Journal of Research in Education Sciences, 61*(2), 85-117.]

歐宗霖、邱皓政、孫國勛（2013）。當測驗碰上情境：情境判斷測驗與自陳量表在量表發展上之比較。**測驗學刊，60**（2），239-262。

[Ou, T. -L., Chiu, H. -C., & Sun, K. -H. (2013). When test met scenario: The comparison between situational judgment test and self-report scale on scale development. *Psychological Testing, 60*(2), 239-262.]

歐慧敏（2013）。大學生公民素養量表簡介及信度、效度分析。**教育研究月刊，227，**89-102。

[Ou, H. -M. (2013). Introduction and analysis of reliability and validity: Citizens' competence inventory for college students. *Journal of Education Research, 227,* 89-102.]

蔡清田（2014）。**國民核心素養：十二年國教課程改革的 DNA。**台北：高等教育。

[Tsai, C. -T. (2014). *National core literacy: DNA of the Twelve-year National Curriculum reform.* Taipei: Higher Education.]

蔡清田、洪若烈、陳延興、徐彬瀚、葉雅卿（2012）。**K12 各教育階段核心素養與各領域課程統整研究。**國家教育研究院委託中正大學之總計畫期末報告（編號：NAER-100-16-A-1-02-00-2-03），未出版。

[Tsai, C. -T., Hung, R. -L., Chen, Y. -H., Hsu, P. -H., & Yeh, Y. -C. (2012). *Core literacy of K12 education stages and curriculum integration in various fields.* National Academy for Educational Research entrusted National Chung Cheng University Report No. NAER-100-16-A-1-02-00-2-03.]

鄧毓浩（2013）。臺灣公民教育的發展。**現代教育論叢，193，**37-50。

[Teng, Y. -H. (2013). The development of Taiwan citizen education. *Modern education, 193,* 37-50.]

Adams, R. J., Wilson. M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21,* 1-23.

Ajzen, I., Czasch, C., & Flood, M. G. (2009). From intentions to behavior: Implementation intention, commitment, and conscientiousness. *Journal of Applied Social Psychology, 39,* 1356-1372.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Assessment & Teaching of 21st Century Skills [ATC21S]. (2010). *21st century skills white paper*. Retrieved from ATCS Web site: http://atc21s.org/index.php/resources/white-papers/

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443-459.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6,* 431-444.

Commission of the European Communities. (2005). *Communication from the commission, COM (2005) final/2, modernizing education and training: A vital contribution to prosperity and social cohesion in Europe.* Brussel, Kingdom of Belgium: Author.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Embretson S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Guion, R. M., & Ironson, G. H. (1983). Latent trait theory for organizational research. *Organizational Behavior and Human Performance, 31,* 54-87.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Hung, S. P. (2018). Validating the creative self-efficacy student scale with a Taiwanese sample: An item response theory-based investigation. *Thinking Skills and Creativity, 28,* 190-203.

Hung, S. P., Huang, P. S., & Chen, H. C. (2016). Cognitive complexity in the remote association test-Chinese version. *Creativity Research Journal, 28*(4), 442-449.

Jensen, A. R. (1980). *Bias in mental testing.* New York, NY: The Free Press.

Karabatsos, G. (2000). A critique of Rasch residual fit statistics. *Journal of Applied Measurement, 1,* 152-176.

Linacre, J. M. (1997). KR-20 / Cronbach alpha or Rasch person reliability: Which tells the "Truth"? *Rasch Measurement Transactions, 11*(3), 580-581.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16,* 878-878.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149-174.

McFadden C., & Skaggs, G. (2013). Development and validation of the Sense of competence Scale-revised. *Journal of Applied Measurement, 14,* 318-331.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749. doi:10.1037/0003-066X.50.9.741.

Oana, N. S., & Pop, E. I. (2016). Longitudinal links between career adaptability and academic achievement in adolescence. *Journal of Vocational Behavior, 93,* 163-170.

Organization for Economic Cooperation and Development [OECD]. (2019). *PISA 2018 Global Competence.* Retrived from OECD Web site: https://www.oecd.org/pisa/pisa-2018-global-competence.htm

Paek, I. (2002). *Investigations of differential Item functioning: Comparisons among approaches, and extension to a multidimensional context* (Unpublished doctoral dissertation). University of California, Berkeley, California.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: The University of Chicago Press.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21,* 25-36.

Salganik, L. H., & Stephens, M. (2003). Key competencies for a successful life and a well-functioning society. In D. S. Rychen. & L. H. Salganik (Eds.), *Competence priorities in policy and practice* (pp.13-40). Cambridge, MA: Hogrefe & Huber Publishers.

Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. (2016). *IEA international civic and citizenship education study 2016: Assessment framework.* Paris, France: IEA Publications.

Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2009). *International civic and citizenship education study: Assessment framework.* Paris, France: IEA Publications.

Taiwan International Large-Scale Study Center [TILSSC]. (2019). *Trends in International Mathematics and Science Study (TIMSS).* Retrived from Taiwan International Large-Scale Study Center Web site: https://tilssc.naer.edu.tw/timss

Wilson, M. (2005). *Constructing measure: An item response modeling approach,* Mahwah, NJ: Lawrence Erlbaum.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14,* 97-116.

Wright, B. D., & Stone, M. H. (1999). *Measurement Essentials* (2nd ed.). Wilmington, DE: Wide Range.

Wu, M. L., Adams, R. J., & Wilson, M. R. (2007). *ConQuest (Computer software and manual)*. Victoria, Australia: Australian Council for Educational Research.

投稿收件日：2018 年 10 月 25 日

接受日：2019 年 11 月 29 日

附錄一　中學生現代公民核心素養測驗等級描述範例

倫理素養

認知題（理解、推理與應用）	
良好（5—6）	能夠將道德與實踐的能力、懂得尊重與關懷他人的能力完整融入生活，並透過這樣的能力實踐於生活中的倫理道德相關議題。
普通（3—4）	具備基本的道德與實踐的能力與態度，並懂得尊重與關懷他人，並透過這樣的能力實踐於生活中的倫理道德相關議題。
待加強（1—2）	在具備基本的道德與實踐的能力與態度、尊重與關懷他人上，需要透過這樣的能力實踐於生活中的倫理道德相關議題較需要更努力。
情感行為題（價值性、行為意圖與行為）	
正向積極（17—24）	願意主動瞭解道德與實踐能力，具備積極的道德實踐、尊重與關懷的價值觀與態度，並有想做出行為的念頭與確實做出對於道德實踐、尊重與關懷表現於外的行為。
普通（9—16）	願意瞭解道德與實踐能力，具備基本道德實踐、尊重與關懷的價值觀與態度，並曾經有想做出行為的念頭與曾經確實做出對於道德實踐、尊重與關懷表現於外的行為。
待加強（1—8）	對於瞭解道德與實踐能力，基本道德實踐、尊重與關懷他人沒有太大的興趣，較少具備道德實踐與尊重他人的意圖與較少能夠具備時間道德與尊重他人的行為。