

領域特定詞彙知識的測量： 三至八年級學生數學詞彙能力

吳昭容 國立臺灣師範大學教育心理與輔導學系／學習科學跨國頂尖研究中心教授

曾建銘 國家教育研究院測驗及評量研究中心副研究員

鄭鈴華 國立臺灣師範大學教育心理與輔導學系研究助理

陳柏熹 國立臺灣師範大學教育心理與輔導學系／學習科學跨國頂尖研究中心教授

吳宜玲 國立臺灣師範大學教育心理與輔導學系博士候選人

摘 要

學生的詞彙知識是理解教師講解、課堂討論、課本內容，以及題目題意的基礎。測量數學詞彙知識的詞目常是研究者主觀選定、只在特定單元或年級，工具未經信效度考驗，欠缺常模，更沒有複本。本研究建立了一至九年級數學課程中會面對的數學詞彙列表，包括術語、一般用語，以及符號，共 731 個詞，採比率抽樣從 A、B、C 年段（分別是三四、五六、七八年級）學過的詞目抽取詞彙來命題。經修審題，再經小型、第一次、第二次預試，及正式施測，共計 5,593 名受測學生，以年段為潛在迴歸變項進行 Rasch 模式的試題參數估計。三年段各三個複本在三類詞彙與四種數學主題的比率、試題難度分布，與測驗訊息量大致相當。本測驗之內部一致性信度達可接受範圍，並與數學成就、數學文字題、數學運算能力達中度以上相關。本研究以同時估計法將三個年段的題本垂直等化，而各年段的三個複本也經過水平等化，且以正式施測的 2,357 名學生建立三至八年級的常模，提供不同版本及不同年段的數學詞彙能力成長變化情形，也指出數學詞彙的難詞。

關鍵詞：詞彙知識、試題反應理論、數學符號、術語



The Measurement of Domain-Specific Vocabulary Knowledge: The Mathematical Vocabulary Ability of Third to Eighth Grade Students

Chao-Jung Wu

Professor, Department of Educational Psychology and Counseling/Institute for Research Excellence in Learning Sciences, National Taiwan Normal University

Chien-Ming Cheng

Associate Research Fellow, Research Center for Testing and Assessment, National Academy for Educational Research

Chien-Hua Cheng

Research Assistant, Department of Educational Psychology and Counseling, National Taiwan Normal University

Po-Hsi Chen

Professor, Department of Educational Psychology and Counseling/Institute for Research Excellence in Learning Sciences, National Taiwan Normal University

Yi-Ling Wu

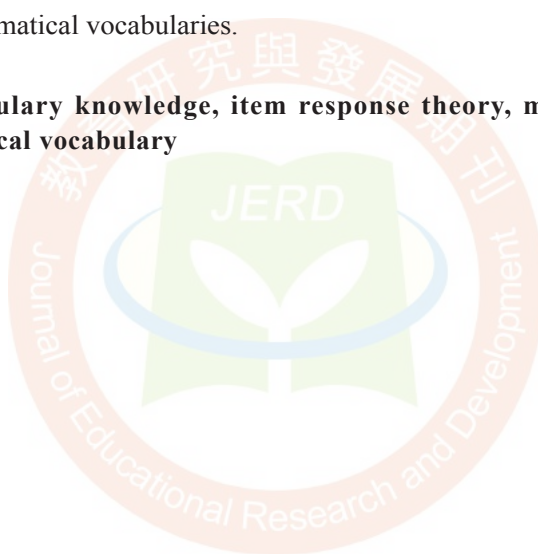
Doctoral Candidate, Department of Educational Psychology and Counseling, National Taiwan Normal University

Abstract

Students' vocabulary knowledge is a foundation for understanding lectures, classroom discussions, textbooks, and semantics of word problems. The lemmas that measure mathematical vocabulary knowledge usually are selected by researchers subjectively, limited only to specific units or grades, lacking in reliability/validity evidence, norm and parallel forms of measurement tools. This study aims to develop measurements of mathematical vocabulary that contains 731 words demonstrating in 1st through 9th graders' mathematic courses, including technical, general, and symbolic vocabularies. We selected vocabularies for the measurements with data of the acquired lemmas in A, B, and C levels (that is, 3rd–4th, 5th–6th, and 7th–8th grades) by proportional sampling method. After modification of items, small-scale trial, first pilot-test, second pilot-test, and final test were conducted, in which about 5,593 students participated. We adopted levels as latent regression variable and conducted parameter estimation of Rasch model. The three parallel forms of each level among the three kinds of vocabularies and four kinds of mathematical topics showed approximately equivalent

proportion, distribution of difficulty, and item information. The internal consistency reliability of this measurements was acceptable. Besides, the relationships between these measurements and the abilities of mathematical achievements, mathematical word problem solving, and mathematical calculation all indicated moderate positive correlations. Generally speaking, this study adopted concurrent calibration to build vertical scales for the three levels, and to develop horizontal scales for the three parallel measurements of each level. The norm of 3rd through 8th grades has been established by 2,357 students in the final test. This study showed not only the growth of students' mathematical vocabulary ability in different levels and parallel forms, but the varying difficulty of mathematical vocabularies.

Keywords: vocabulary knowledge, item response theory, mathematical symbol, technical vocabulary





壹、緒論

詞彙在語言學習的角色已被大量研究，但數學詞彙的學習研究似乎未受到足夠的重視。依據 Gough 與 Tunmer (1986) 所提「閱讀的簡單觀點模式」(the simple view of reading)，閱讀包含識字解碼和語言理解兩個成份。而學習者的詞彙能力已被確定是影響閱讀理解與溝通表達的重要成分(洪儷瑜、王瓊珠、張郁雯、陳秀芬，2006；洪儷瑜、陳心怡、陳柏熹、陳秀芬，2014；曾雅瑛、黃秀霜，2002；張莉萍，2012；蘇宜芬、張祐瑄、李孟峰、黃鈺茜，2016；Stæhr, 2008)。而特定領域的用語在日常生活多半是低頻詞，抑或其詞義或詞用不同於日常生活(Harmon & Hedrick, 2005; Shanahan & Shanahan, 2008)，因此人們並不容易在生活中自然地發展出這些領域特定詞彙(domain-specific vocabulary)的知識。學生從數學課室中累積與保留的數學詞彙知識的狀況如何？數學詞彙能力與其它各種數學能力成分(例如數學成就、文字題解題能力、運算能力)的關係如何？隨著年級的發展，數學詞彙知識的學習及其与其它數學能力成分的關係會有何變化？回答這些問題都需要能良好測量數學詞彙知識的工具。

近年來由於強調數學素養與數學溝通的重要性，有越來越多學者主張數學詞彙教學的必要性(Harmon & Hedrick, 2005; Livers & Elmore, 2018; Monroe & Orme, 2002; Riccomini, Smith, Hughes, & Fries, 2015)。然而，欠缺數學詞彙能力的評量工具，使得檢視這些研究的成效變得相當困難，例如必須由評分者逐筆分析學生撰寫的答案或口語使用詞彙的品質(Monroe & Pendergrass, 1997; Topping, Campbell, Douglas, & Smith, 2003)，不僅評分費時費力，有時甚至連評分者間信度都付之闕如，品質堪憂。

測量學生數學詞彙知識，必須解決一連串的問題。首先，何謂數學詞彙？數學詞彙是詞彙的子集，應該如何界定一個詞彙是或不是數學詞彙？其次，對正在學習數學的中小學生而言，哪些數學詞彙是應該理解的？課程綱要界定學生應該學的，還是實際上出現在數學課本的？第三，數學詞彙知識的內涵是什麼？從哪些表現得以界定學生的數學詞彙能力？最後，數學詞彙能力和數學成就或其它數學能力的關係如何？以下透過回顧文獻來回應前述問題。

一、數學詞彙的定義與類型

數學專有名詞不是唯一出現在數學學習脈絡並影響著學生學習的詞彙。Jacobson (1998) 指出學科文本中有四類會造成理解困難的詞彙：術語 (technical vocabulary) 是指在某一學科領域中獨有且重要的概念，如生物學裡的「基因」、化學領域的「碳氫化合物」、數學領域的「共軛複數」等；特殊詞 (special vocabulary) 是指在不同領域會有不同意義的詞彙，例如：幾何學中平面或立體形體的「頂點」和生活話語「這已經是企業發展的頂點」(秦麗花, 2006) 的意義不同，或者運動力學的「慣性」也不同于形容一個人「慣性犯罪」；生詞 (unfamiliar words) 意指生活中會應用到但學生不熟悉的詞彙，如「樣式」。回指詞 (anaphora) 是以代名詞來取代前文出現過的名詞，如「此」、「其」等，必須透過前後文脈絡方能理解哪個詞彙被取代。

Monroe 與 Panchyshyn (1995) 則主張，學生要發展數學概念必須學習以下四類的詞彙：術語、準術語 (subtechnical, 有些學者稱 semi-technical)、一般 (general)，和符號 (symbolic)。例如「函數」、「平行四邊形」是數學領域的術語；「volume」是準術語，這類詞彙在數學領域與日常生活中的意義不同，類似於 Jacobson 的特殊詞；一般詞彙是數學領域中會用到的且意義與日常生活運用一致的語詞，但 Monroe 與 Panchyshyn 指出數學課本中經常出現超越該年段閱讀課的詞彙，而造成學生的困擾，類似於 Jacobson 的生詞；符號意指非文字而傳遞數學意義的表徵，例如數字、等號、度量衡單位均屬之，例如以下各例的「2」的意義都不同：12、23、 4^2 、 $\frac{1}{2}$ 、 $\frac{2}{3}$ 、 m^2 ，甚至讀音也不同；運算符號和關係符號是另一類常見的數學符號： $+$ 、 $-$ 、 $\times$ 、 $\div$ 、 $\sqrt{\quad}$ 、 $=$ 、 $\geq$ 、 $\approx$ ；度量衡單位也被視為數學符號之一，這些單位常以縮寫方式呈現，例如 mm、cm、l、cc，縮寫也是造成困擾的因素之一。

Nation (2001) 則認為學科文本除了高頻詞之外，還有學術詞彙 (academic vocabulary)、術語、低頻詞三類。學術詞彙意指會出現在跨領域學科文本的學術語詞，例如「區分」(differentiate)、「下層」(subordinate) (Coxhead, 2000)；國內臺灣大學也建置 MyET (http://efreeway.avcenter.ntu.edu.tw/freeway/postgraduates/vocab/awl/awl_myet.html) 介紹英文學術詞彙。至於術語，Nation 認為術語的特徵是，鮮少出現在日常生活中，但在特定領域則出現頻率相對較高、

是特定領域知識結構的一部分、與特定領域之其它概念的關係明確。至於低頻詞，Nation 意指出現在學科文本中非高頻詞、非術語、非學術詞彙的語詞均屬之，類似於 Jacobson 的生詞。

綜合前述學者的分類與描述，數學詞彙可被分為術語、準術語（或稱為特殊詞）、一般（或特指低頻詞、生詞），以及數學符號四類。回指詞應從詞彙知識測驗中排除，乃因其與詞彙知識的關係較弱，而與句法或語句理解較有關連。此外，考量分類的可行性，本研究排除了準術語的類別。因為要確認一個語詞在不同領域脈絡（例如數學相對於日常生活、科學、文學、…）是否有不同意義，牽涉到意義是否相同的判定，且要核對的知識範疇太大，操作上有其困難。因此本測驗界定的數學詞彙，只包含了數學術語、一般詞彙，與數學符號三類。

二、數學詞彙的識別與範圍

在實務操作上，如何界定出一群數學詞彙作為測量的取樣範圍呢？一般的詞彙測驗多半從大型詞庫中取樣，而詞頻是重要的取樣依據（王瓊珠、洪儷瑜、張郁雯、陳秀芬，2008；洪儷瑜等人，2006；洪儷瑜等人，2014；黃秀霜，2001；曾雅瑛、黃秀霜，2002；Laufer & Nation, 1999），具有以客觀資料為基礎的優點；但領域特定的詞彙測驗的詞目則多由專家主觀選擇（秦麗花、邱上真，2004；Edwards, 1936; Kazima, 2007; Powell, Driver, Roberts, & Fall, 2017; Powell & Nelson, 2017）。

國內學齡階段學生適用的識字或詞彙測驗多以大型資料庫作為字詞取樣的來源。適用於一至九年級學生的「中文年級識字量表」（黃秀霜，2001）和「識字量評估測驗」（王瓊珠等人，2008；洪儷瑜等人，2006）分別以「新聞語料字頻統計表」（中央研究院，1993）和教育部編輯的「國小學童常用字詞調查報告書」（教育部，2000）的 5,021 字，依字頻分級取樣。而詞彙成長測驗（洪儷瑜等人，2014）則取樣自「國小學童常用字詞調查報告書」（教育部，2002）的 46,667 個詞，先以一些刪除原則刪了單字詞、數量詞、人名、專有名詞、縮詞、稱謂……等，接著將剩下的 36,641 個詞依據詞頻分成四級，最後再分級抽樣。

領域特定的詞彙測驗無法僅以一般性的詞庫作為取樣的來源，因為領域特定詞彙常常未被收錄於一般詞庫中。Chung 與 Nation（2004）指出了識別特定領域的術語的困難。他們以一本臨床解剖學教科書的 11 頁、共 876 條詞目作為分析材

料，評比四種術語識別方法的優缺點。專家評定法是由接受過分類訓練的領域專家逐一檢視詞目，評定各詞目是否屬於特定領域的術語。領域辭典法是以有無出現在特定領域的辭典來決定詞目是否為術語，Chung 與 Nation 選了兩本解剖學詞典重複檢驗，一本詞目較多，一本詞目較少。文章線索法意指文章作者所提供識別術語的方法，包括給定義、印刷線索（引號、粗體、斜體）、圖上標示（lable）。自動萃取法意指透過電腦技術抽取領域中與其它概念高關聯性的詞目。Chung 與 Nation 認為專家評定法雖然最費人力，但的確是最好的方法，理由是專家的評分者間信度高達 .95，且被專家辨識為術語詞彙被解剖學辭典（詞目較多的那本）收納的比率高達 .98。另外三個識別方法若以專家評定為效標，則辨識正確率在 .74 至 .87 之間。

雖然 Chung 與 Nation（2004）認為專家一一評定所有詞目是否為術語是最佳的方法，但三年級至八年級學生所接觸的數學文本包括各版本的國中小教科書，純以人工進行分析是不可行的。此外，Kwary（2011）也建議編製領域特定詞彙測驗可採用混合法，而 Brossard 與 Shanahan（2006）也實際採用混合多種方法的程序，並以客觀取樣的方式篩選科學詞彙。

Brossard 與 Shanahan（2006）以瞭解科學與科技術語的能力來界定科學素養，他們採用了以下程序來降低研究者主觀判斷對術語挑選的影響。首先，選擇一本科學辭典作為系統取樣的起點，作者選擇了有 9000 條詞目且領域廣泛的牛津科學辭典；第二，採用系統取樣的方式在每一頁抽一詞目，過程中會避開準術語（例如 Euclid 可能代表一位幾何學家，也可能 Ohio 州的一個城市），共選了 896 條詞目；最後，以一個新聞全文資料庫和隨機建構一週七天（由不同期間的週一到週日組成的，以避免特殊事件造成的影響），搜尋前述八百多條詞目的出現頻率，選擇七天平均頻率最高的 31 條（5%）詞目。

Brossard 與 Shanahan（2006）的方法值得本文參考，但因測驗目的的差異，仍有一些不同之處。Brossard 與 Shanahan 測量的是素養，所以採用新聞資料庫並選擇出現頻率最高的詞目是合宜的，但本文測量的是領域特定詞彙知識，不宜只採用如「新聞語料字頻統計表」（中央研究院，1993）或「國小學童常用字詞調查報告書」（教育部，2002）等領域一般的語詞資料庫，因為這類資料庫收納的學科詞彙相當有限。此外，Brossard 與 Shanahan 的評量工具旨在測量成人的科學素養，可以採用一般的新聞資料庫和科學辭典，但本文的評量對象是國中小學生，

必須納入中小學生適用的詞庫或數學辭典。

由於本測驗並非數學詞彙的成就測驗，所以除了納入數學課程綱要明訂該年級需學習的標準用語（教育部，2008，頁 196–207）和度量衡單位（教育部，2008，頁 185）之外，學生在數學學習情境中會面對的詞彙亦在取樣的範圍。例如課程綱要的標準用詞雖然遲至八年級才列入「線段」，但此一詞彙早在四、五年級就出現於數學課本，因此，透過參與研究的數學教師檢視課本，可納入各年級學生實際經驗到的詞彙。此外，數學課程中常出現的一般用語也需要透過數學教師推薦。建立了數學詞彙的候選詞目列表之後，本研究擬以數本數學辭典或資料庫來確認哪些詞彙是數學術語，其中會至少有一適用於中小學生的數學辭典或詞庫。剩下的候選詞目再經一般詞庫來確認哪些語詞是個詞彙，其中會至少有一適用於中小學生的詞彙資料庫。如此就可將文字所組成的語詞區分為數學術語或一般詞彙。非中文的符號，其外型明確，但仍須經專家會議確認是否適合納為數學符號。

三、數學詞彙知識的內涵與評量

發展數學詞彙測驗必須界定數學詞彙知識的內涵，以及可以如何測量這些內涵。參酌一般詞彙測驗的文獻，測量詞彙知識常以能否解釋該語詞、替換同義詞、選出反義詞、判斷圖像是否符合語詞意義、在文脈中選擇適當的語詞等方式，前四者與詞義有關，最後一個則與詞用有關，例如 Laufer 與 Goldstein（2004）就指出一般詞彙測驗的知識內涵通常包含詞義理解與詞用。而方金雅、陳新豐、王高帆（2003）也主張詞彙知識應該包括詞義能力、詞用能力，和構詞能力，例如：「比喻英雄豪傑的壯志 ① 鴻雁哀鳴 ② 鴻鵠之志 ③ 鴻飛冥冥 ④ 鴻鵠將至」屬詞義能力，又例如：「中國人認為[]是一種美德 ① 謙虛 ② 虛偽 ③ 虛弱」則是詞用能力，而「冰塊[]化 ① 溶 ② 容 ③ 融 ④ 熔」屬構詞能力。除了選擇題的題型，該文也舉了配合題型的詞彙測驗試題。洪儷瑜等人（2014）針對評量四至六年級學生的詞彙成長狀況，採用無文脈的相似詞選擇題型，例如「唐僧 ① 唐詩 ② 和尚 ③ 唐代」。整體而言，詞義和詞用比較適合數學詞彙的知識內涵，測量學生能否構造出一個合法的數學詞彙可能意義不大，因為能構出詞，未必理解其義，也未必能正確使用；替換近義詞或指認反義詞，也不適用於數學詞彙測驗，因為再相近的詞彙都存在微妙的差異，不符合強調精準的數學。

在領域特定的詞彙測驗似乎以測量術語的定義為主。Brossard 與 Shanahan (2006) 發展科學素養評量工具(見前文)時,評估科學概念的難度可能讓問答题難以回答,且不易有好的評分者信度,而非題的猜測率過高,所以採用填空題的題型。題幹上有關術語的敘述來自牛津科學辭典,例如:「以人為的方式讓個體產生免疫力的方式稱[]」,學生必須自行填答「預防接種(vaccination)」,結果效標關連效度不錯。Brossard 與 Shanahan 自評這個工具的缺點之一就是乏味,因為每一題都像字典上的定義。乏味的測驗容易讓學生只因沒有意願作答而得低分,這個缺點用測量定義來編製的詞彙測驗很容易產生的問題。

數學詞彙測驗也是以測量術語的定義為主,題型上常採配對題或選擇題。例如 Amen (2006) 針對五年級學生進行的行動研究,實驗教學前後採用的是寫定義作業,學生被要求以文字、畫圖,或舉例來定義 15 個數學詞彙;而在每一單元課程前後則採用 10 個題幹對 10 個選項的配對題,題幹為某個未知術語的定義或描述,選項則是數學術語。再例如 Edwards (1936) 針對 Georgia 大學合計八百多名大學生設計了一份數學詞彙測驗,測驗中包含的 100 個術語或符號,目的在於測量學生對數學術語與符號的認識。該測驗的題型為配對題,學生需在題幹後的空格中填入配對項的編號。題目包含兩類形式,一類的題幹是「an expression consisting of a single term」之類定義,配對項則是「rational number」、「transposition」等數學術語;另一類的題幹是「cube root」、「greater than」之類的數學術語,配對項則羅列各種數學符號,如「 $\sqrt[3]{}$ 」、「 $\cong$ 」。該測驗單數題與雙數題得分的折半相關達 .92。至於 Walt 等人 (Walt, Maree, & Ellis, 2008) 針對四至七年級學生設計的數學詞彙量表,僅說明題型為五選一選擇題,未介紹其詞彙知識的構念。

有鑑於數學概念研究顯示學生常能背誦定義,但卻未必能正確舉例 (Wilson, 1990),數學詞彙測驗除了測量學生對於定義的掌握之外,詞義理解的測量應該包含其它面向。Powell 等人 (Powell et al., 2017; Powell & Nelson, 2017) 發展的一、三、五年級的數學詞彙,其測量的層次包括憶取 (recall)、理解 (comprehension),和在複雜作業中的運用 (use in complex tasks),三層次的試題有不同的比率。為配合不同認知層次的題目, Powell 等人採用的作答方式也很多元,有配合題、選擇題、畫圖、簡答題。

歸納前述文獻,本文界定數學詞彙知識為兩大類,一是詞義,另一則含括判斷詞彙的正例、分辨性質的關聯性、詞用、符號辨識等。詞義和符號辨識的試題

可能在有文脈或無文脈下命題，而詞用則必然有文脈。題型則採用四選一的選擇題。

四、數學詞彙能力與其它能力的關係

有關數學詞彙測驗的表現與其它能力之關係的探討不常見，主要是因為欠缺客觀的數學詞彙測驗。前文介紹過的 Edwards (1936) 大學生數學詞彙測驗，以八百多名大學生為對象，結果測驗成績與兩群修習數學課程學生的期末成績之相關分別為 .64 和 .59，和數學普測、心理學測驗、英語分班測驗之相關也達 .47 至 .84，顯示數學詞彙能力與大學數學成績以及其它科目的成績都有中度以上的相關。但該研究並未排除智力或其它一般性認知能力的影響，所以這些相關可能來自其它更基礎的認知能力的貢獻。

Powell 與研究夥伴 (Forsyth & Powell, 2017; Powell & Nelson, 2017; Powell et al., 2017) 發現一、三、五年級學生的數學詞彙與一般詞彙、數學運算能力的相關都在 .59 到 .69 之間，且多層次線性回歸顯示以一般詞彙或運算能力可以預測 37% 至 45% 的數學詞彙分數的變異，她們據此主張一般詞彙能力與數學運算能力提供數學詞彙發展的基礎。然而，這些研究各年級的學生數僅來自同一學校的 65 至 128 人，且 Powell 等人承認她們的數學詞彙測驗並非嚴謹的測驗工具，既無常模也沒有信效度。秦麗花與邱上真 (2004) 在探討數學文本閱讀理解相關因素中，探討到數學詞彙與符號理解能力与其它數學能力間的關係。她們抽樣了 336 位四年級班級的學生，除了施測文本閱讀理解測驗之外，還包含其它六種測驗。發現數學閱讀特殊技能包含四個子技能，其標準化因素負荷的 R^2 分別為詞彙符號 .62、先備知識 .53、作圖程序 .50、數學圖示 .28，其中以詞彙符號最高。

考量國中小學階段評量數學詞彙能力的實用性，本研究以發展三至八年級的數學詞彙測驗為目的。基於詞彙知識的學習具有累積性，且測量的標的是能力而非特定年級的學習成就，故各年級的試題並非全都取樣自該年級所學的詞彙，而是以所學過的數學詞彙為取樣母群，因此題本的年級區隔並不需要非常明顯。本研究以 A 題本做為三、四年級學生的測驗卷，B 題本用於五、六年級，C 題本則使用於七、八年級，採試題反應理論 (Item Response Theory, IRT) 建立各年段題本的垂直等化；且為了提供檢視教學介入與延宕成效的平行題本，每個年段均有三個經 IRT 水平等化的複本。

綜合以上，本文的研究目的在於：

- 一、發展三至八年級跨年段垂直等化的數學詞彙測驗。
- 二、評估不同年段之數學詞彙測驗複本的信度與效度。
- 三、建立各年級的常模。
- 四、分析各年級數學詞彙的難詞。

貳、研究方法

一、受測者與有效資料

小型預試的受測者來自新北市與新竹市各一所國小與國中，三至八年級各一班，各年級學生在 46 至 56 人之間，共 298 人參與。其中新竹市的國中小離市中心較偏遠，弱勢族群的學生比率較高。

第一次預試參考 104 學年國小、國中學校別資料（教育部統計處，2016a、b）邀請北一區 4 個縣市、北二區 3 個縣市、中區 4 個縣市、南區 4 個縣市、東區及離島 3 個縣不同規模的學校參與。國小以每校三至六年級、國中每校七、八年級各 1 班的學生為對象，最後有 20 所國小及 19 所國中參與。三至八年級的參與人數依序為 399、426、414、422、507、521 人，總計 2,689 位學生。

第二次預試旨在補足 B 年段的題目。採便利取樣自臺北市、新北市、高雄市各一校的五、六年級各 1 班，五、六年級學生為 69 與 76 人，共 145 位學生參與。

常模樣本來自正式施測，同樣參考各縣市國中、小概況統計資料進行取樣，由於施測時間為九月，故國小以每校四至六年級、國中七至九年級各 1 班的學生為對象，作為三至八年級的常模樣本。最後參加測驗的有 17 所國小及 17 所國中，四至九年級的參與人數依序為 391、406、389、423、422、430 人，共計 2,461 位學生參與，如表 1。

表 1 參與數學詞彙正式測驗之各縣市學校規模、校數、與人數

序號	區域	校數 縣市	國小學校規模				國中學校規模			
			小	中	大	總計 (人數)	小	中	大	總計 (人數)
1	北一	臺北市	0	2	0	2 (121)	0	0	1	1 (76)
2		新北市	0	1	1	2 (142)	0	1	0	1 (79)
3		基隆市	0	0	1	1 (76)	0	0	1	1 (78)
4		宜蘭縣	—	—	—	—	—	—	—	—
5	北二	桃園市	0	0	1	1 (81)	1	0	1	2 (125)
6		新竹縣 (市)	0	0	1	1 (81)	—	—	—	—
7		苗栗縣	0	0	1	1 (73)	0	0	1	1 (82)
8	中	臺中市	0	0	1	1 (84)	1	0	0	1 (62)
9		彰化縣	0	0	1	1 (80)	0	1	1	2 (175)
10		南投縣	—	—	—	—	0	1	0	1 (74)
11		雲林縣	0	0	1	1 (82)	0	0	1	1 (82)
12	南	嘉義縣 (市)	0	0	1	1 (75)	0	0	2	2 (176)
13		臺南市	—	—	—	—	0	1	0	1 (81)
14		高雄市	1	0	1	2 (70)	1	0	0	1 (68)
15		屏東縣	0	0	1	1 (77)	—	—	—	—
16	東+離島	花蓮縣	0	1	0	1 (74)	1	0	0	1 (46)
17		臺東縣	—	—	—	—	—	—	—	—
18		澎湖縣	0	0	1	1 (70)	0	1	0	1 (71)
總計			1	4	12	17 (1,186)	4	5	8	17 (1,275)

註：學校規模以教育部公布 104 學年度國民小學校別資料及國民中學校別資料的班級數為主，國小部分以 6 班以下為小校、7～24 班為中校、25 班以上為大校；國中部分以 12 班以下為小校、13～36 班為中校、37 班以上為大校。

四次測驗的施測時間詳見後文「五、施測程序」。四次測驗的無效樣本比率分別為 0%、2%、0%、3%，故有效樣本分別為 298、2,638、145、2,399 人。無效樣本的判定原則，包括：（一）空白或作答不清楚的題數超過一半者。（二）答對率在 1/4 以下，且答題具某種樣式（如：12341234…等）者。（三）學校提供的特教生名單。

二、數學詞彙測驗編製

(一) 詞目蒐集與分類

本研究透過官方文件與中小學教師推薦兩種方式收集詞目，並就詞彙類型、數學領域，以及首度出現的年級三向度分類。教育部（2008）《國民中小學九年一貫課程綱要：數學領域》附錄四（p.196–207）之標準用詞，皆被列入詞目，各年級詞彙 19 至 74 個，共有 404 個。另經四位參與研究之國中小數學教師推薦，各年級詞彙 6 至 125 個，共有 352 個。兩個來源合計 756 個詞目。

詞彙類型的分類上，中文文字構成的詞目經以下步驟分成數學術語和一般詞彙，非中文文字的詞目則經專家會議確認是否為數學符號。中文文字構成的詞目若出現在以下三個數學辭典或詞庫之一，則列入「數學術語」。本測驗採用《簡明數學辭典》（凡異出版社，1987）、《數學名詞—高中（含）以下數學名詞》詞庫（國家教育研究院，2013）及《數學名詞》詞庫（國家教育研究院，2016）進行比對。未被列為數學術語的中文文字詞目，則進一步以教育部（2002）《國小學童常用字詞調查報告書》及中央研究院（2013）《現代漢語平衡語料庫》為參照，若出現在任一資料上，則該詞彙判定為「一般用語」。非中文文字構成的詞目，包括數字（0~9）、運算子（+、-、×、÷、...）、度量衡（cm、kg、l、m²、m³、°、...）、不等號（>、<、≥、≤、≠）、百分號與折扣（%、off%）、小數點（.）、幾何標記與性質（⊥、△、π、∴、∴、SSS、RHS、...）、及其它代號（ $f(x)$ 、AM、≈、|、...）等。

有 45 個中文文字構成的詞目既未出現在任一數學辭典或詞庫中，也未出現在常用字詞詞庫，理應視為非詞。但其中 13 個詞目來自課程綱要明列的標準用語，例如：小數點以下第幾位、差平方、公式解、...等，故仍保留於詞目列表，再經專家會議分類。另由詞彙的關聯性，例如個位至兆等位名、正號與負號，而補上遺漏的數個詞彙，最後共 731 個數學詞彙，完整列表參見附錄 1（<https://goo.gl/wC16gk>）。

詞彙所屬學習主題（數與量、幾何、代數、統計與機率）的分類，則依據課綱原本的分類，或者數學辭典、詞庫中的分類，又或者由專家會議中決定。詞彙所屬的年級，則依課綱出現的年級或者專家會議決定。分類統計表見表 2。

表 2 一至九年級各主題的術語、一般用語、符號個數

類型與 主題 年級	術語 (T)					一般用語 (W)					符號 (S)					合計				
	N	G	A	D	小計	N	G	A	D	小計	N	G	A	D	小計	N	G	A	D	總計
一	23	4	0	0	27	17	0	0	1	18	13	0	0	0	13	53	4	0	1	58
二	17	9	1	0	27	9	1	0	0	10	5	0	0	0	5	31	10	1	0	42
三	37	19	0	2	58	35	5	0	0	40	11	1	0	0	12	83	25	0	2	110
四	20	32	1	4	57	9	2	0	0	11	4	0	1	1	6	33	34	2	5	74
五	19	20	0	0	39	8	2	0	0	10	4	0	0	0	4	31	22	0	0	53
六	19	2	0	1	22	9	2	0	0	11	2	0	0	0	2	30	4	0	1	35
七	43	22	46	0	111	25	0	9	4	38	12	2	4	0	18	80	24	59	4	167
八	24	51	28	0	103	3	4	0	0	7	2	13	0	0	15	29	68	28	0	125
九	0	37	7	18	62	0	1	1	2	4	0	1	0	0	1	0	39	8	20	67
總計	202	196	83	25	506	115	17	10	7	149	53	17	5	1	76	370	230	98	33	731

註：N 為「數與量」、G 為「幾何」、A 為「代數」、D 為「統計與機率」。

（二）詞彙取樣

為編製足夠的試題，A、B、C 年段的詞彙取樣各 100 個詞，其中最高年段（分別為四、六、八年級）的詞彙數各僅占 8 個。詞彙取樣保留較大比率來自較早年級學習的內容，是基於詞彙學習的累積性，且詞彙知識保留對於新詞彙的學習非常重要。部分年段在代數或統計與機率主題的詞彙極少，會先將之納入後，再依類型、主題、來源的比例，決定各細格的詞彙比例與個數，以亂數表抽取。

（三）命題

四名數學教師參與命題，題型採四選一單選題。研究者先舉數個範例題，於會議中形成命題原則與共識。命題卡上包括詞目、年級、來源、詞彙類型、數學主題、作業類型、題幹、選項、正確答案、難度判定等。測量詞義的題數以 50% 為原則，而另外 50% 的題目則測量是否能判斷詞彙的正例、關係、使用、辨識等。

命題數達一半時，再次召開會議，形成題目修改之依據：**1. 題目內容不適**：測量標的有誤（測計算能力、量感、換算、數學知識）、圖形呈現不佳、不適合該年段或超過該年段的學習程度。**2. 題幹調整**：語意模糊、語句不通順、具多餘

訊息。**3. 選項修改**：有非唯一解的可能性、與詞彙無直接相關、缺少誘答力。命題教師據此完成命題與審查前的修改。表 3 為數學詞彙命題之示例題，以四個題目為例，呈現數學詞彙的命題樣式。

表 3 數學詞彙命題之示例題

詞目	年級	來源	類型	主題	作業 類型	示例題
分類	1	教師補充	一般用語	機率與統計	意義	小華將形狀相同的東西放在一起。小華在做什麼？① 數數 ② 分類 ③ 計算 ④ 統計
乘數	2	課綱	術語	數與量	正例	算式「 $20 \times 11 = 220$ 」中的「11」稱為什麼？① 商 ② 積 ③ 被乘數 ④ 乘數
相交	3	教師補充	術語	幾何	意義	兩條線「相交」的意思是什麼？ ① 兩條線平行 ② 兩條線交叉 ③ 兩條線重疊 ④ 兩條線相連
等式	7	教師補充	術語	代數	辨識	下列何者為「等式」？ ① x ② $x + 5$ ③ $x + 5 = 6$ ④ $x + 5 > 1$

（四）審題與修題

300 道題目分成四個審查題本，每個題本均包含 A、B、C 年段的題目，且都由一位數學教育專家及一位國語文專家審閱，合計數學教育專家和國語文專家各四位參與審查。評分包括：**1. 內容**須符合受測年段的學習程度，用詞應避免出現罕見字，避免超越詞彙而測到更複雜的知識。**2. 題幹**應該是學生熟悉的情境、語義要清楚明白、字數應適當，句子不宜過長等。**3. 正確答案**須無爭議性，並只有唯一解。**4. 錯誤選項**應具備誘答力。最後分三級：直接採用、需小幅修改、需大幅修改或不宜採用。

當數學教育與國語文專家意見一致時，依其意見納入、修改，或刪除題目；若兩人意見相左時，知識面的建議以數學教育專家為主，行文或語順方面則以國語文專家意見為主。修改後，再與命題老師會議討論。

（五）題本組題與留用

組題以 10 分鐘能答完為原則，且 A 年段題本在各次施測均加注音，以 A4 紙張印製題本。

1. 小型預試

各年段僅組一個題本，各 20 題，難度分配約為易 60%、中 30%、難 10%；詞彙類型分配約為術語 55%、一般用語 35%、符號 10%；學習主題分配約為數與量 50%、幾何 30%、代數 15%、統計與機率 5%。結果發現，每班最快答完的人在 5 至 8 分鐘，10 分鐘結束時未答完人數在 0 至 6 人之間、集中在 C 年段。顯示 10 分鐘對 A、B 年段較充裕，對 C 年段則適中。

刪除過於簡單的試題後，選出第一次預試的 8 個共同題，以對低年段較難、對高年段較易，並盡量涵蓋不同類型與主題的題目為原則，如表 4。「統計與機率」主題的詞彙，由於題數極少，經篩選後未有適合作為共同題的題目。

表 4 第一次預試各年段間所採用之共同題

共同題年段	數學詞彙	來源年段	類型	主題	通過率	難度	鑑別度
A、B	」（直角）	A	S	G	0.82	0.87	0.27
	驗算	A	T	A	0.55	0.42	0.54
	橫式	A	W	N	0.48	0.52	0.50
	積	B	T	N	0.84	0.81	0.38
B、C	四則運算	B	T	N	0.60	0.54	0.58
	下底	C	T	G	0.63	0.61	0.70
	依序	C	W	A	0.91	0.91	0.17
	≥（大於等於）	C	S	A	0.55	0.57	0.52

2. 第一次預試

A、B、C 年段題數分別為 22、21、20 題，各年段組四個複本，並盡量讓各複本的詞彙類型、主題，及難度相似。其中 B 年段和 A、C 年段各題本分別有 4 題共同題，如圖 1. 所示，故 A、B、C 年段的總題數各有 75、54、67 題。施測後刪除品質較差的題目，留用 177 個題目。

年 段	A 年段		B 年段		C 年段	
每本題數	22 題		21 題		20 題	
題數分配	18 題	4 題共同題 (3A、1B) ^{註 1}	13 題	4 題共同題 (1B、3C) ^{註 2}	16 題	
該 年 段	75 題		54 題		67 題	
總 題 數	(18×4 + 3 = 75)		(13×4 + 1 + 1 = 54)		(16×4 + 3 = 67)	

圖 1 各年段題本的共同題配置

註 1：共同題有 3 題是取自 A 年段、有 1 題是 B 年段。

註 2：共同題有 1 題是取自 B 年段、有 3 題是 C 年段。

3. 第二次預試

第一次預試因 B 年段採用較多 A、C 年段之共同題，使 B 年段總題數未達正式測驗所需之題數（如圖 1），因此第二次預試旨在補充 B 年段的題目，故僅有 B 年段 1 個版本 21 題。以第一次預試時鑑別度較好的 6 個題目做為定錨題。修改或重新命題的 15 個新題，施測後 13 道被留用。

4. 正式測驗

A、B、C 年段題數均 21 題，且各有三個複本，各複本間皆無共同題。題目由第一、二次預試的留用題中挑選 189 題，其試題鑑別度全都大於 0，A、B、C 年段試題鑑別度小於 0.20 的比率為 21%、24%、10%，在 A、B 年段多為過於容易的題目，在 C 年段則為過難的題目。組卷時盡量讓各複本的詞彙類型、主題，及難度相似。其中有 48 個題目經局部修改題幹敘述或調整選項次序。表 5 呈現各複本的詞彙類型與主題的題數。

表 5 正式測驗各複本於各詞彙類型與主題之題數

題本	類型與 主題	術語 (T)					一般用語 (W)					符號 (S)					合計				
		N	G	A	D	小計	N	G	A	D	小計	N	G	A	D	小計	N	G	A	D	總計
A1		6	4	1	0	11	6	1	0	0	7	3	0	0	0	3	15	5	1	0	21
A2		6	5	0	1	12	5	1	0	0	6	3	0	0	0	3	14	6	0	1	21
A3		6	5	0	1	12	5	1	0	0	6	2	1	0	0	3	13	7	0	1	21

(續下頁)

表 5 正式測驗各複本於各詞彙類型與主題之題數（續）

類型與 主題 題本	術語 (T)					一般用語 (W)					符號 (S)					合計				
	N	G	A	D	小計	N	G	A	D	小計	N	G	A	D	小計	N	G	A	D	總計
B1	6	7	0	1	14	3	1	0	0	4	2	0	1	0	3	11	8	1	1	21
B2	6	5	1	1	13	3	2	0	0	5	2	0	0	1	3	11	7	1	2	21
B3	6	5	0	2	13	3	2	0	0	5	2	0	1	0	3	11	7	1	2	21
C1	3	5	5	0	13	3	0	1	1	5	2	0	1	0	3	8	5	7	1	21
C2	4	2	5	0	11	2	2	1	0	5	2	1	1	1	5	8	5	7	1	21
C3	4	3	4	1	12	3	0	1	1	5	3	1	0	0	4	10	4	5	2	21
總計	47	41	16	7	111	33	10	3	2	48	21	3	4	2	30	101	54	23	11	189

最後，題庫留用 238 道試題，包括第一次預試的 177 題、第二次預試的 13 題，及正式測驗因微幅調整而被視為獨立的 48 題。

三、作答方式

考量三、四年級學生不熟悉畫卡的作答方式，故 A 年段為直接作答於題本上；B、C 年段則採答案卡。但第二次預試時，因僅有 B 年段 6 個班級參加，故採直接作答於題本上。

四、參照測驗

為提供建構效度之證據，本研究以數學文字題能力測驗（吳昭容，未出版）、數學運算能力測驗（楊心怡，未出版）、及數學成就測驗作為參照測驗。前二者與本測驗同時發展，同樣適用於三至八年級（分 A、B、C 年段的題本）。文字題測驗採多重是非題組型態，檢測學生閱讀應用題之後能否選出正確算式或方程式；運算能力測驗為三選一單選題，評量學生是否具備正確的數學運算知識、熟練的數學運算技巧，及靈活的數與符號的操作能力，具有可接受的信效度。各年級的成就測驗主要取自 2012 及 2016 年縣市學力檢測題目（國家教育研究院，2017），六年級題本除 6 題來自「臺灣學生學習成就評量資料庫」（Taiwan Assessment of Student Achievement, 簡稱 TASA）（國家教育研究院，2015）釋出

題之外，其餘自編。成就測驗旨在檢視三至八年級學生於一學年數學學習之後的概念理解、程序執行，與解題思考的綜合成效，採四選一單選題。各測驗之信效度見附錄 2 至 4 (<https://goo.gl/gw1fTY>)。

五、施測程序

小型預試安排在 105 年 12 月、第一次預試安排在 106 年 3 月、第二次預試安排在 106 年 5 月、正式施測安排在 106 年 9 月。

小型預試由計畫研究人員擔任主試；第一、二次預試、正式測驗則由各班導師或數學老師擔任主試者，施測程序依研究者提供的指導語及注意事項進行，施測地點在學生原班教室。施測時，由主試者先發下題本及 / 或答案卡，各年段題本於小型施測及第二次預試時只有單一版本，故同班學生的題本均相同；於第一次預試及正式測驗有複本，則均採班內隨機發放。

三個參照測驗依作答時間不同（數學文字題及數學運算題各為 20 分鐘、數學成就測驗為一節課），安排數學詞彙搭配數學文字題與成就測驗，或數學詞彙搭配數學運算題與成就測驗的方式進行施測，使每班的施測時間都控制為最多兩節課。因此每年級有 8 個班級的學生實施數學詞彙與數學文字題兩種測驗，其中有 4 個班級的學生另一節課施測成就測驗；有 8 個班級的學生實施數學詞彙與數學運算題兩種測驗，其中有 4 個班級的學生另一節課施測成就測驗。在此安排下，各年級兩兩測驗都有 8 個班級的學生參與。

六、試題參數估計與測驗等化

為了將不同題本的結果放在相同量尺上進行比較，本研究採用試題反應理論進行試題參數估計與測驗等化。常見的 IRT 測驗等化方法有（1）同時估計法（concurrent estimation）；（2）分離估計法（separate estimation）：又分為平均數\平均數法（mean\mean）、平均數\標準差法（mean\sigma）、特徵曲線法（characteristic curve methods）等（Kolen & Brennan, 2004）。由於同時估計法是利用共同題將所有題本的試題作答反應放在同一個檔案中進行估計，減少了測驗間連結時產生的誤差，且國內外許多文獻亦證實，採用同時估計法能獲得較佳的精準度（Hanson & Béguin, 2002；Kim & Cohen, 1998；郭伯臣、王暄博，2008；陳煥文，2011），因此本研究採用同時估計法。至於共同題的比例則是參考過

去學者的建議，以大於題本題數的五分之一為原則（Livingston, 2004; Holland & Dorans, 2006）。

當使用同時估計法進行測驗等化時，會將受測群體的能力分布視為單一群體並估計其平均數與標準差，這對於跨不同年級且明顯為不同能力群體的垂直等化設計（vertical equating）較不適合，會低估困難題本的試題參數，且高估簡單題本的試題參數（DeMars, 2003）。因此，在垂直等化情境中為了降低上述偏誤，Wu, Adams 與 Wilson（1998）建議可將不同能力群體視為不同能力分布參數來進行同時估計，並以潛在迴歸（latent regression）模式來進行試題參數與能力估計。

在本研究中，第一次及第二次預試皆是利用共同題將不同題本的作答反應合併成同一個作答反應矩陣，進行 IRT 試題參數估計，並以分析結果作為試題篩選與修改的參考。正式施測時，則是將前兩次預試中被選入正式題的作答反應資料與正式施測作答反應資料合併成一個作答反應矩陣，進行試題參數與能力估計，以便進行水平與垂直等化。

參、結果與討論

一、試題難度與受測者能力

本研究整併第一、二次預試與正式測驗共 5,182 名有效受測者的資料，將所有作答反應矩陣整併在一起進行同時估計，以 Rasch 模式進行分析，估計時將不同年段作為潛在迴歸變項並採用貝氏估計法進行試題參數與受測者能力估計（Wu et al., 1998），估計出 238 道試題參數及所有受試者能力值。表 6 為各年段的 IRT 試題難度與各年級受測者平均能力。A、B、C 年段的題目難度平均值為 -0.95、-0.17、1.21；六個年級的學生能力平均值介於 0.36 至 2.00，標準差介於 0.67 至 0.73，年級能力分布如圖 2，分布曲線逐年右移，表示學生數學詞彙能力隨著年級增長而增加。圖 3 為受測者能力與試題難度分布圖。由試題難度、學生能力，及分布情形，顯示試題對各年段學生而言都略顯簡單。

表 6 數學詞彙測驗各年段 IRT 試題難度與各年級受測者平均能力值

試題難度	A 年段 (81 題)	B 年段 (92 題)	C 年段 (79 題)
平均值	-0.95	-0.17	1.21
標準差	1.12	1.37	1.05
最大值	0.70	3.32	3.32
最小值	-4.48	-2.82	-1.53

受測者能力值	三年級	四年級	五年級	六年級	七年級	八年級
人數	782	825	863	906	886	920
平均值	0.36	0.75	1.20	1.23	1.81	2.00
標準差	0.67	0.68	0.67	0.70	0.73	0.72

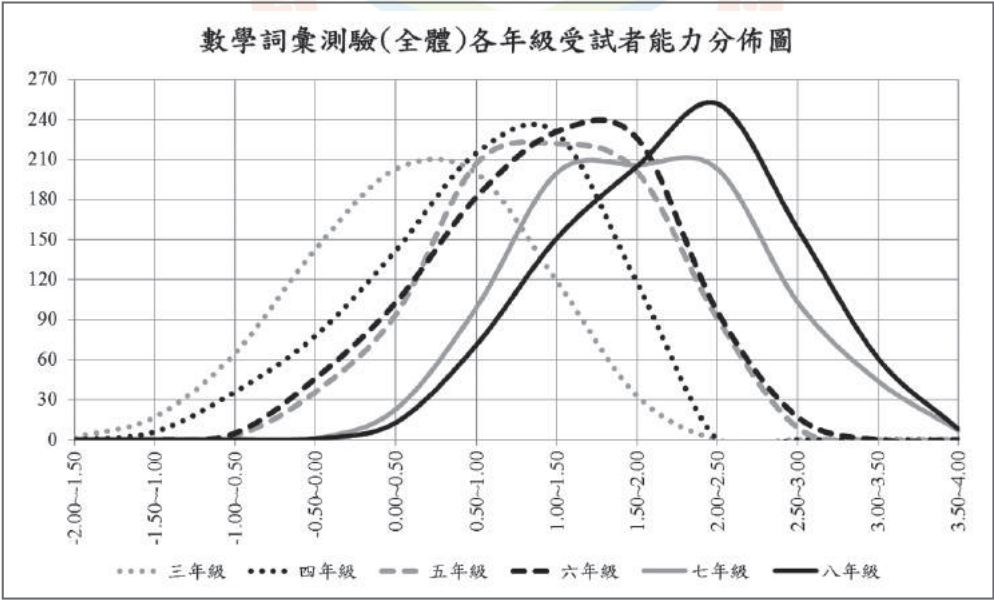


圖 2 數學詞彙測驗各年級之受測者能力分布圖

難度/能力值	受試者	試題		
		A 年段	B 年段	C 年段
4	I			
	XI			
	XXI			
	XI			
	XXI			
	XXXXI			
	XXXXX			
	XXXXX		120	120
	XXXXXX			233
	XXXXXXXX			
3	XXXXXXXX			145, 176
	XXXXXXXX			146
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
2	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
1	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
0	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
-1	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
-2	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
-3	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			
	XXXXXXXX			

每個 X 代表 6.0 位受試者

圖 3 數學詞彙測驗受測者能力與試題難度 Wright Map

註：方框標示的 8 個題號為預試時的共同題；灰底標示的 6 個題號則為預試與正式測驗時的共同題。

二、數學詞彙能力測驗的信度表現

本研究以 Cronbach's α 係數考驗正式測驗題本之內部一致性信度，並以 IRT 測驗訊息量檢視正式測驗的複本信度。表 7 呈現各版本之內部一致性信度，其值介在 .63 至 .80，而 CTT 試題難度則在 .64 至 .76。A 或 B 年段部分題本的信度略低，主因題目較為簡單，且受測者人數不多，未來需再收集更多樣本資料，以作為測驗優化的參考。

表 7 各年段的試題信度與難度

版本		人數 (n)	信度 (Cronbach's α)	CTT 試題難度 平均數 (全距)
A 年 段	A1	265	.69	.75 (.52 ~ .95)
	A2	260	.63	.76 (.56 ~ .94)
	A3	265	.65	.75 (.54 ~ .98)
B 年 段	B1	263	.72	.71 (.24 ~ .98)
	B2	264	.63	.71 (.33 ~ .97)
	B3	269	.69	.72 (.49 ~ .98)
C 年 段	C1	270	.80	.65 (.39 ~ .86)
	C2	272	.70	.64 (.33 ~ .85)
	C3	271	.74	.64 (.27 ~ .93)

試題反應理論以測驗訊息量表示試題對不同能力受試者所能提供的測量精準度，訊息量愈大，則表示試題所提供的測量精準度亦愈高，其計算公式為：

$$I_i(\theta) = \frac{P_i'^2}{P_i Q_i} ,$$

其中 $P_i'^2$ 為 P_i 對的一階微分的平方， $I_i(\theta)$ 為該題的訊息量， P_i 為該題答對率， Q_i 為該題答錯率。依據試題反應理論局部獨立性之假設，將測驗中所有題目的訊息量加總，可以得測驗訊息量 (test information)。以函數圖形表示，橫軸為學生能力值，縱軸為訊息量，三個年段共 9 個題本的訊息量函數圖如圖 4。依據表 7 各年段複本測驗的信度平均數 0.69 轉換所得的測量標準誤約為 0.55 (= $1 \times \sqrt{1 - 0.69}$)，而再轉換為所需的訊息量約為 3，故以訊息量 3 以上的區段對照

出最適合施測的能力值區間，結果 A、B、C 年段各複本最適施測能力值分別在 -2.4 至 0.8、-1.6 至 1.5、-0.4 至 2.7 間，且三個複本的函數圖大致相同，顯示具有良好的複本信度。

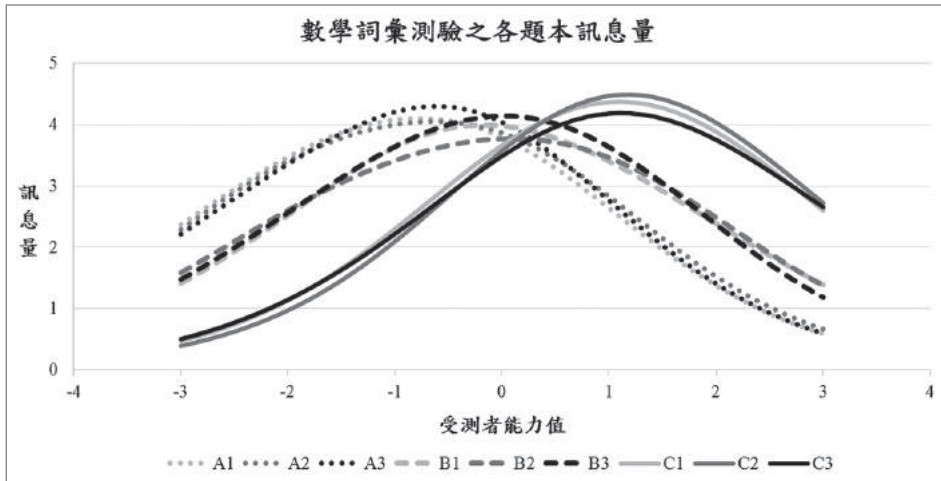


圖 4 各年段題本之測驗訊息量函數圖

三、數學詞彙能力測驗的效度表現

本研究利用 ConQuest 進行 Rasch 模型之參數估計與分析作為建構效度之一。參照 Wu 等人（1998）與陳柏熹、葉泰廷、黃馨瑩、陳郁欣、蘇少祖（2015）以適配值判斷試題良窳，適配標準為試題之訊息加權均方差（Weighted Fit MNSQ）介於 0.7 至 1.3 之間，大於 1.3（低適配）或小於 0.7（過度適配）皆為不適配（Bond & Fox, 2007）；本測驗全體 238 道題中僅有 8 道題目的適配值較差，試題與單參數 Rasch 模式的適配比例達 96.6%。檢視各題本與單參數 Rasch 模式的適配比例達 90.5% 至 100.0%，表示大多數試題都與 Rasch 模式符合，即題目內容與單向度模式相符，測驗試題測量到相同潛在特質，具有建構效度。

效度證據之二，為本測驗與參照測驗之相關，見表 8，各相關係數均達顯著（ $ps < .01$ ）。數學詞彙與成就測驗及文字題在六個年級均達中高度相關，相關係數在 .50 至 .75 之間，顯示越能掌握數學詞彙的學生，通常數學成就表現以及判斷文字題列式的能力也越高。相對地，數學詞彙與數學運算的關聯性則隨年級有所變

化，A 年段的三、四年級兩測驗之相關係數均未達 .40，僅中等偏低，B 年段的相關係數則在 .55 左右，但在 C 年段則升至 .65 與 .70，表示隨著年級的增長，數學詞彙能力與數學運算的關係越來越強。

表 8 各年段之數學詞彙與數學文字題、運算題及成就測驗之相關係數

相關係數	數學詞彙題					
	A 年段		B 年段		C 年段	
	三年級	四年級	五年級	六年級	七年級	八年級
數學成就測驗	.56** (n =183)	.62** (n =194)	.66** (n =180)	.58** (n =212)	.75** (n =208)	.65** (n =214)
數學文字題	.50** (n =199)	.65** (n =204)	.57** (n =198)	.62** (n =214)	.68** (n =202)	.65** (n =214)
數學運算題	.39** (n =183)	.37** (n =197)	.51** (n =184)	.57** (n =200)	.65** (n =196)	.70** (n =198)

註：** $p < .01$

四、常模建置及測驗分數的轉換與應用

常模樣本來源為 106 年 9 月正式測驗的有效受測者，學生能力均為等化後的能力值。依據教育部統計處（2017a、b）的北一、北二、中、南、東區及離島等五個區域之學生數，計算各年級有效樣本與母群分布之比例差，結果五年級的「東區及離島」與六至八年級的「北一」人數比例與實際分布情形的差距較大（> 0.10），因此五至八年級由人數較多的區域，採隨機方式刪除部分學生（合計 42 人），使各區抽樣比例差距在 ±0.10 以下，各年級人數、平均能力值及題本之 IRT 難度如表 9，並以 2,357 人建置本測驗的全國常模。

表 9 數學詞彙測驗常模之各年級受測者人數、平均能力值與標準差及題本 IRT 難度

受測者能力值	三年級	四年級	五年級	六年級	七年級	八年級
人數	387	403	381	409	368	409
平均值	0.48	0.87	1.34	1.23	1.99	2.00
標準差	0.62	0.65	0.66	0.70	0.76	0.74

（續下頁）

表 9 數學詞彙測驗常模之各年級受測者人數、平均能力值與標準差及題本 IRT 難度（續）

題本 IRT 難度	A1	A2	A3	B1	B2	B3	C1	C2	C3
平均值	-0.96	-0.82	-0.86	-0.03	-0.10	-0.15	1.11	1.24	1.14
標準差	1.15	1.12	1.05	1.31	1.35	1.15	0.96	0.90	1.08
最大值	0.66	0.55	0.70	2.74	2.27	1.44	2.65	3.18	3.32
最小值	-3.53	-2.53	-2.96	-2.82	-2.29	-2.59	-0.51	-0.17	-1.25

註：三至八年級的受測者，為剛升上四至九年級的學生。

本測驗常模取樣之六年級平均能力（1.23）較五年級（1.34）低，推測因施測時間在九月，六年級的受測者為剛升上七年級的學生，有較其他年級更長又無學習作業要求的暑假，易有更明顯的夏季失落（summer loss）現象（Entwisle, Alexander, & Olson, 1997）。我們以線性、二次、對數模式預測表 10 三至八年級的平均能力值，結果以線性模式具最佳解釋力（ R^2 達 99%），顯示數學詞彙能力應該隨著年級而線性增高。因此，我們以表 9 常模的三、四、五、七、八年級的平均能力值建立線性迴歸線（ R^2 達 96%），推估六年級的平均能力應為 1.53，與常模所得的能力值 1.23，兩者相差了 0.30，故常模的六年級學生之能力值應平移 + 0.30。我們有兩個數據支持上述對六年級能力值的調整。其一，表 10 為第一次預試（三月）合併第二次預試（五月）共 2,783 名學生的能力值，這群學生比常模樣本的學生（九月）少了兩、三個月的學習時間，除了六與八年級之外，其餘各年級能力值均較表 9 的常模能力值略低，而表 10 並未出現六年級能力值低於五年級的現象，支持常模取樣的六年級能力值為低估的解釋。其二，本研究之後在 107 年 6 月追蹤測驗，結果六年級學生在 6 月的平均能力值為 1.58，與表 9 的六年級能力值 1.23 落差相當大（經過暑假相差 0.35），其它年級在 6 月也比 9 月的能力值高一些，但落差最多為 0.16。

表 10 數學詞彙兩次預試的各年級受測者人數、平均能力值與標準差

受測者能力值	三年級	四年級	五年級	六年級	七年級	八年級
人數	395	422	481	492	485	508
平均值	0.25	0.64	1.09	1.23	1.67	2.00
標準差	0.69	0.69	0.66	0.69	0.67	0.70

表 11 為原始分數與能力值及 PR 值轉換表。原始分數對應的能力值若有缺漏，大部分發生在最高與最低分處，採對照同年段其它版本已知能力值間的差距來推算，其對照的 PR 值多數為 1 與 99。表中六年級的 PR 值乃依調整後的能力值來建立的。

五、各年級學生已學但表現不佳的數學詞彙

為了解各年級學過的數學詞彙有哪些難詞，本研究比對正式測驗 189 個試題難度值高過各年級學生能力平均值的數學詞彙，表示這些詞彙對於該年級一半以上的學生較為困難，結果見表 12。總計三至八年級有 8 至 15 個詞彙難度高於學生平均值，這些難詞約占各年級已學過且受測之詞彙的 6% 至 11%。多數難詞在下一個年級或年段就不再是難詞，例如「加數」對三年級學生而言是個難詞，但多數四年級以上的學生擁有超過該詞難度值的能力。少部分橫跨多個年級的難詞（如旋轉中心、終邊、對應邊），雖是四、五年級學習過的詞彙，其難度值甚至高於七、八年級學生的平均能力值。觀察這少數幾個特別困難的難詞，可發現其使用的情境過於狹隘，除了在特定單元會用到之外，其它情況很少使用。顯示數學詞彙的難易度和一般詞彙類似，詞頻是一個重要的影響因素。

表 11 數學詞彙測驗原始分數與能力值及 PR 值轉換表

原始 分數	A1 題本			A2 題本			A3 題本		
	能力值	三年級	四年級	能力值	三年級	四年級	能力值	三年級	四年級
		PR	PR		PR	PR		PR	PR
1									
2									
3									
4									
5									
6									
7	-1.17	1	1	-1.07	1	1	-1.08	1	1
8	-0.97	1	1	-0.88	1	1	-0.89	1	1
9	-0.79	1	1	-0.71	3	1	-0.72	2	1
10	-0.62	5	2	-0.54	5	4	-0.56	5	3

(續下頁)

表 11 數學詞彙測驗原始分數與能力值及 PR 值轉換表（續）

原始 分數	A1 題本			A2 題本			A3 題本		
	能力值	三年級 PR	四年級 PR	能力值	三年級 PR	四年級 PR	能力值	三年級 PR	四年級 PR
11	-0.45	5	4	-0.36	9	6	-0.39	8	5
12	-0.27	11	6	-0.17	15	7	-0.21	12	7
13	-0.08	17	8	0.02	23	11	-0.02	19	10
14	0.12	25	12	0.22	33	14	0.17	29	13
15	0.33	38	16	0.43	47	22	0.38	42	19
16	0.55	52	25	0.65	61	34	0.59	56	29
17	0.78	65	38	0.87	73	45	0.81	69	41
18	1.03	76	50	1.12	82	63	1.05	81	57
19	1.30	87	67	1.39	94	76	1.32	91	72
20	1.59	95	82	1.67	97	89	1.61	96	85
21	1.89	99	94	1.95	99	99	1.90	99	97
原始 分數	B1 題本			B2 題本			B3 題本		
	能力值	五年級 PR	六年級 PR	能力值	五年級 PR	六年級 PR	能力值	五年級 PR	六年級 PR
1									
2									
3									
4									
5									
6	-0.56	1	1	-0.61	1	1	-0.61	1	1
7	-0.35	1	1	-0.39	1	1	-0.39	1	1
8	-0.17	2	1	-0.19	1	1	-0.20	1	1
9	0.01	3	1	0.00	2	1	-0.03	2	1
10	0.19	6	3	0.18	6	3	0.14	4	1
11	0.37	9	6	0.37	8	6	0.31	7	4
12	0.57	14	11	0.56	11	9	0.50	11	7
13	0.77	22	16	0.77	18	14	0.69	16	12
14	0.98	30	24	0.98	27	22	0.89	25	18
15	1.20	41	33	1.20	35	30	1.10	32	26

（續下頁）

表 11 數學詞彙測驗原始分數與能力值及 PR 值轉換表 (續)

原始 分數	B1 題本			B2 題本			B3 題本		
	能力值	五年級 PR	六年級 PR	能力值	五年級 PR	六年級 PR	能力值	五年級 PR	六年級 PR
16	1.43	53	44	1.42	49	39	1.31	44	35
17	1.67	64	57	1.65	61	52	1.54	57	48
18	1.93	78	69	1.91	74	64	1.79	69	61
19	2.20	91	81	2.18	89	78	2.05	83	73
20	2.48	97	91	2.45	97	88	2.33	95	85
21	2.74	99	96	2.71	99	95	2.61	99	93

原始 分數	C1 題本			C2 題本			C3 題本		
	能力值	七年級 PR	八年級 PR	能力值	七年級 PR	八年級 PR	能力值	七年級 PR	八年級 PR
1									
2									
3									
4									
5	0.34	1	1	0.44	1	1	0.36	1	1
6	0.52	1	1	0.62	3	2	0.54	2	1
7	0.69	4	2	0.78	7	4	0.71	5	3
8	0.86	8	4	0.95	10	7	0.88	8	6
9	1.03	11	9	1.12	14	13	1.05	13	11
10	1.20	16	15	1.29	19	19	1.23	17	17
11	1.38	21	23	1.47	27	26	1.42	25	24
12	1.57	30	29	1.65	34	36	1.60	31	31
13	1.75	35	37	1.84	41	41	1.79	38	39
14	1.94	44	45	2.02	49	51	1.98	47	47
15	2.14	53	53	2.22	60	60	2.18	56	56
16	2.35	63	62	2.42	68	68	2.39	65	65
17	2.57	72	73	2.64	78	79	2.62	75	77
18	2.80	81	83	2.87	87	89	2.85	84	87
19	3.04	90	90	3.11	94	93	3.10	93	91
20	3.29	95	95	3.34	97	98	3.34	97	97
21	3.52	98	98	3.57	99	99	3.57	99	99

表 12 各年級的數學詞彙難詞

IRT 難度值	三年級	四年級	五年級	六年級	七年級	八年級
0.48	(三年級平均能力值)					
0.50	邊長					
0.55	被加數					
0.66	加數					
0.70	和					
0.81	除數					
0.87	(四年級平均能力值)					
0.98	減數	減數				
1.01	四則運算	四則運算				
1.03		底邊				
1.05	商	商				
1.06		腰				
1.13		鈍角				
1.20		周角				
1.25						
1.30						
1.30		無條件進入法				
1.32						
1.34		(五年級平均能力值)				
1.36			公倍數			
1.44		底角	底角			
1.47			% off (折扣表示法)			
1.53			(六年級平均能力值)			
1.55		同分母分數	同分母分數	同分母分數		
1.64		≈ (省略)	≈ (省略)	≈ (省略)		
1.71			因數	因數		
1.89			球心	球心		

(續下頁)

表 12 各年級的數學詞彙難詞（續）

IRT 難度值	三年級	四年級	五年級	六年級	七年級	八年級
1.99					(七年級平均能力值)	
2.00						(八年級平均能力值)
2.20				大約等於	大約等於	大約等於
2.23					打擊率	打擊率
2.26					α (成正比)	α (成正比)
2.27		旋轉中心	旋轉中心	旋轉中心	旋轉中心	旋轉中心
2.30			十萬分位	十萬分位	十萬分位	十萬分位
2.41					至多	至多
2.52					重合	重合
2.53						性質
2.60					倒數	倒數
2.64					遞移律	遞移律
2.65					不等號	不等號
2.72					等量公理	等量公理
2.74		終邊	終邊	終邊	終邊	終邊
3.18					異號數	異號數
3.32			對應邊	對應邊	對應邊	對應邊
總計(個)	8	13	11	9	14	15
難詞比例 *	10%	11%	8%	6%	8%	8%

* 註：各年級以 189 個試題中已學過的詞彙數作為分母。三～八年級分母依序是 84、116、138、147、185、189。

肆、結論與建議

一、研究成果與發現應用

（一）本測驗基於客觀評估中小學生數學詞彙的知識之目的，透過納入課程

綱要之標準用語與教師推薦收集詞目，並以數學辭典 / 資料庫與常用字詞資料庫將之分類後，建立了符合一至九年級學生數學學習經驗的數學詞彙列表，再以比率取樣來編製試題。該詞彙列表保存與公開於網路，便於教師與學者後續的運用。

（二）本測驗旨在測量學生之數學詞彙知識的「能力」，而非「學習成就」，因此詞彙取樣方式以學習過的詞彙為取樣來源，新近學習（最高年級）的詞彙比率不超過 10%，反映的是受測者數學詞彙的學習與保留的整體狀況。雖然文獻上常採下定義作為領域特定詞彙的測量方法，但本研究除了測量受測者對數學詞彙意義的了解之外，另外也有一半的題目測量是否能判斷詞彙的正例、關係、使用、辨識等，此種設計符應中小學生運用詞彙知識的情境，也避免通篇測驗都測定義而產生排斥感。試題編修過程也發現，領域特定詞彙知識的試題設計必須拿捏分寸，以防題目超越詞彙知識的範圍，例如測到了更深入的數學知識，或者測到與該詞彙相關的量感、單位換算，或計算能力。

（三）本研究發展適用於三至八年級的數學詞彙測驗，採 Rasch 模式垂直等化與水平等化了三個年段各三個複本，信效度資料顯示具有可接受的品質，並建立各年級的全國常模。由於有三個複本，有利於教師或研究者比較教學介入前後的數學詞彙能力的變化；且由於三個年段的測驗已經垂直等化，可追蹤學生在不同年級的數學詞彙能力之成長變化。

（四）對三至八年級學生而言，各對應的 A、B、C 年段測驗卷的難度都偏簡易，研究者認為此結果符合已學過的領域特定詞彙知識應有的水準。數學詞彙知識是具備與應用數學知識的基礎，也是數學閱讀理解或聽覺理解的前提，所以多數學生具備大多數已經學過的數學詞彙知識應是合理的表現。從各題本訊息量函數圖所呈現的最適施測能力值也顯示，本測驗運用於各年級能力較高的學生時，測量誤差會較大。因此，本測驗適合用來篩選低能力的學生。從難詞分析也可發現，這些已學過但較學生平均能力來得高的詞彙，約僅佔各年級受測詞彙的 10%，而且多數的難詞都在下一年級或年段就不再是難詞了。而少部分橫跨多個年級的難詞，其使用情境相當狹隘，例如「旋轉中心」僅在「旋轉角」單元出現，由於接觸機會少，即使學生年級增長，對此種出現頻率非常低的數學詞彙，學生仍然難以掌握。這一方面顯示數學詞彙與一般詞彙類似，使用頻率是形成數學詞彙知識的重要因素，另一方面也提醒教師，經常在脈絡中運用數學詞彙是讓學生發展數學詞彙知識的重要方法。

(五) 本測驗常模取樣的六年級受測者實際是九月份的七年級學生，在常模資料中，六年級受測者的平均能力值略低於五年級學生。然而，此一不合預期的現象，並未出現於僅用三月第一次預試合併五月第二次預試的結果。因此研究者以這群剛升上七年級的學生有較長的夏季失落來解釋。本研究採解釋力最佳的線性模式，以常模樣本的其它五個年級之平均能力值推估六年級學生應有的能力值，並依能力值平移 +0.30 的資料建立六年級常模的原始分數與 PR 值的對照表。研究團隊後續進行的追蹤，也支持此一調整的合理性。雖然本測驗以較廣泛的詞彙取樣來源來降低近期學習經驗的影響，但在一個學年的不同時間點施測，仍影響學生的表現，因此施測時間的選擇以及資料解釋宜留意，例如六年級升七年級的學生可能有特別明顯的夏季失落，因此建議剛進國中的七年級新生，應避免在剛開學的 9 月初施測。

(六) 本研究指出數學詞彙知識與數學成就以及理解文字題並選擇正確算式的能力在跨三至八年級均有中高相關 (.50 至 .75)，但與運算能力的相關則隨年級而漸增，小學中年級的相關係數不到 .40、小學高年級則在 .50 至 .60 之間，到了中學七八年級則相關係數達 .65 至 .70。此一結果顯示掌握數學詞彙與熟練小學較低年級的算術能力較無關係，然而越高年級或國中階段的數學運算涉及概念理解的成分越高，數學詞彙能力就會與數學運算能力有中高相關。

二、未來研究建議

(一) 依據閱讀理解理論，我們預期數學詞彙能力是高數學成就、文字題表現，以及國中階段數學運算能力的前提。雖然本研究已探討了數學詞彙能力與數學成就、數學文字題能力，以及數學運算能力的相關，但未來需要有多個時間點的追蹤資料，以建立這些數學能力之間的因果模式。

(二) 本測驗每個數學詞彙只命一道試題，主要反映的是詞彙知識的廣度 (breadth)，而無法呈現其深度 (depth) (Wesche & Paribakht, 1996; Laufer & Goldstein, 2004)，例如我們無法知道，理解在部分整體模式下的「分數」意義的學生，是否了解商模式或比例模式下的分數意義？是否掌握分數與整數、小數的關係？能否判斷此一數學詞彙詞用的適當性？此外，本測驗施測時不設時限，測驗表現不能反應受測者的解碼流暢性。然而，一名能理解「平方」就是二次方、「立方」就是三次方，但每每需要一兩秒才能成功轉譯的學生，在充斥著「平方」、「立

方」的課堂上，就會因為這一兩秒的延宕而無法理解老師的講解。數學詞彙解碼流暢性的測量，是下一步值得發展的測驗工具。

（三）研發數學詞彙知識的教學方案。運用數學詞彙能力測驗篩選出數學詞彙知識低落的學生，下一步就該如同語文科詞彙教學一般，發展能改善學生知識狀況的教學方案（陳昱君、連育吟、李俊仁，2016），而本測驗能協助教學成效的考核，同時教學成效也能回饋本測驗的效度。

誌謝

本研究承蒙科技部專題研究計畫（MOST 105-2511-S-003-039-MY3）與教育部補助國立臺灣師範大學高等教育深耕計畫「學習科學跨國頂尖研究中心」經費支持，並感謝國家教育研究院測評中心楊元傑先生的協助，特此致謝。

參考文獻

- 凡異出版社（1987）。**簡明數學辭典**。新竹市：作者。
- 中央研究院（1993）。**新聞語料詞頻統計表**。臺北市：作者。
- 中央研究院（2013）。**現代漢語平衡語料庫**。臺北市：作者。
- 方金雅、陳新豐、王高帆（2003，6月）。**電腦化詞彙測驗題型之研究**。論文發表於僑務委員會舉辦之「第三屆全球華文網路教育研討會」，臺北市。
- 王瓊珠、洪儷瑜、張郁雯、陳秀芬（2008）。一到九年級學生國字識字量發展。**教育心理學報**，39（4），555-568。
- 吳昭容（未出版）。**數學文字題能力測驗**。
- 洪儷瑜、王瓊珠、張郁雯、陳秀芬（2006）。**識字量評估測驗**。臺北市：教育部。
- 洪儷瑜、陳心怡、陳柏熹、陳秀芬（2014）。**詞彙成長測驗**。臺北市：中國行為科學社。
- 秦麗花（2006）。從數學閱讀特殊技能看兒童數學閱讀的困難與突破。**特殊教育季刊**，99，1-12。
- 秦麗花、邱上真（2004）。數學文本閱讀理解相關因素探討及其模式建立之研究－以角度單元為例。**特殊教育與復健學報**，12，99-121。
- 國家教育研究院（2013）。**數學名詞－高中（含）以下數學名詞**。取自 <https://goo.gl/sHXpz1>
- 國家教育研究院（2015）。**臺灣學生學習成就評量資料庫**。取自 <https://goo.gl/Ncs5T3>
- 國家教育研究院（2016）。**數學名詞**。取自 <https://goo.gl/QmuJ1f>
- 國家教育研究院（2017）。**學力檢測公布系統**。取自 <http://rap.naer.edu.tw/>
- 張莉萍（2012）。對應於歐洲共同架構的華語詞彙量。**華語文教學研究**，9（2），77-96。
- 教育部（2000）。**國小學童常用字詞調查報告書**。臺北市：作者。
- 教育部（2002）。**國小學童常用字詞調查報告書**。臺北市：作者。
- 教育部（2008）。**國民中小學九年一貫課程綱要：數學學習領域**。臺北市：作者。
- 教育部統計處（2016a）。**104 學年國民小學校別資料**。取自 https://quality.data.gov.tw/dq_download_xlsx.php?nid=6240&md5_url=8b23f290682c20dc56ce69c7f0239

df8

教育部統計處（2016b）。**104 學年國民中學校別資料**。取自 https://quality.data.gov.tw/dq_download_xlsx.php?nid=6239&md5_url=8fc3a80422ac4a707d86faa6764d21f2

教育部統計處（2017a）。**105 學年國民小學校別資料**。取自 https://quality.data.gov.tw/dq_download_xlsx.php?nid=6240&md5_url=8c547f6bdef258dccc5ded420d8a7c3d

教育部統計處（2017b）。**105 學年國民中學校別資料**。取自 https://quality.data.gov.tw/dq_download_xlsx.php?nid=6239&md5_url=418e6d1e051f631f72a5ba0c4d107aa5

郭伯臣、王暄博（2008）。大型測驗中同時進行垂直與水平等化效果之探討。**教育研究與發展期刊**，4（4），87-120。

陳昱君、連育吟、李俊仁（2016）。透過詞彙教學方案增加低成就學童閱讀能力。**臺東大學教育學報**，27（2），51-76。

陳柏熹、葉泰廷、黃馨瑩、陳郁欣、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。

陳煥文（2011）。垂直等化連結特性之研究：六種連結方法的比較。**測驗學刊**，58（4），559-583。

曾雅瑛、黃秀霜（2002）。國民小學中文詞彙閱讀測驗之編製。**測驗年刊**，49（2），199-216。

黃秀霜（2001）。**中文年級認字量表**。臺北市：心理。

楊心怡（未出版）。**數學運算能力測驗**。

蘇宜芬、張祐瑄、李孟峰、黃鈺茜（2016）。國小二至六年級朗讀流暢度篩檢準確度及切截點分析。**教育科學研究期刊**，61（4），33-57。

Amen, J. (2006). *Using math vocabulary building to increase problem solving abilities in a 5th grade classroom*. Retrieved from <http://scimath.unl.edu/MIM/files/research/AmenJ.pdf>

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Brossard, D., & Shanahan, J. (2006). Do they know what they read? Building a scientific

- literacy measurement instrument based on science media coverage. *Science Communication*, 28(1), 47-63.
- Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System*, 32(2), 251-263.
- Coxhead, A. (2000). *Headwords of the academic word list*. Retrieval from <http://www.victoria.ac.nz/lals/resources/academicwordlist/awl-headwords>
- DeMars, C. E. (2003, April). *Recovery of graded response and partial credit parameters in MULTILOG and PARSCALE*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Edwards, A. S. (1936). A mathematics vocabulary test and some results of an examination of university freshmen. *Journal of Educational Psychology*, 27(9), 694.
- Entwisle, D.R., Alexander, K.L., & Olson, L.S. (1997). *Children, schools and inequality*. Boulder, CO: Westview Press.
- Forsyth, S. R., & Powell, S. R. (2017). Differences in the mathematics-vocabulary knowledge of fifth-grade students with and without learning difficulties. *Learning Disabilities Research & Practice*, 32(4), 231-245.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education*, 7(1), 6-10.
- Hanson, B. A., & Béguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26(1), 3-24.
- Harmon, J. M., & Hedrick, W. B. (2005). Research on vocabulary instruction in the content areas: Implications for struggling readers. *Reading & Writing Quarterly*, 21, 261-280.
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 187-220). Westport, CT: Praeger.
- Jacobson, J. M. (1998). *Content area reading: Integration with the language arts*. Albany, NY: Delmar.
- Kazima, M. (2007). Malawian students' meanings for probability vocabulary. *Educational studies in mathematics*, 64(2), 169-189.

- Kim, S. H., & Cohen, A. S. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurement*, 22(2), 131-143.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer.
- Kwary, D. A. (2011). A hybrid method for determining technical vocabulary. *System*, 39(2), 175- 185.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51.
- Livers, S. D., & Elmore, P. (2018). Attending to precision: Vocabulary support in middle school mathematics classrooms. *Reading & Writing Quarterly*, 34(2), 160-173.
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ: Educational Testing Service.
- Monroe, E. E., & Orme, M. P. (2002). Developing mathematical vocabulary. *Preventing School Failure: Alternative Education for Children and Youth*, 46(3), 139-142.
- Monroe, E. E., & Panchyshyn, R. (1995). Vocabulary considerations for teaching mathematics. *Childhood Education*, 72(2), 80-83.
- Monroe, E. E., & Pendergrass, M. R. (1997). Effects of mathematical vocabulary instruction on fourth grade students. *Reading Improvement*, 34(3), 120-132.
- Nation, I. S. (2001). *Learning vocabulary in another language*. New York, NY: Cambridge University Press.
- Powell, S. R., & Nelson, G. (2017). An investigation of the mathematics-vocabulary knowledge of first-grade students. *The Elementary School Journal*, 117(4), 664-686.
- Powell, S. R., Driver, M. K., Roberts, G., & Fall, A. M. (2017). An analysis of the mathematics vocabulary knowledge of third-and fifth-grade students: Connections to general vocabulary and mathematics computation. *Learning and Individual Differences*, 57, 22-32.
- Riccomini, P. J., Smith, G. W., Hughes, E. M., & Fries, K. M. (2015). The language of mathematics: The importance of teaching and learning mathematical vocabulary.

Reading & Writing Quarterly, 31(3), 235-252.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40-59.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139-152.

Topping, K., Campbell, J., Douglas, W., & Smith, A. (2003). Cross-age peer tutoring in mathematics with seven-and 11-year-olds: Influence on mathematical vocabulary, strategic dialogue and self-concept. *Educational Research*, 45(3), 287-308.

Walt, M. V. D., Maree, K., & Ellis, S. (2008). A mathematics vocabulary questionnaire for use in the intermediate phase. *South African Journal of Education*, 28(4), 489-504.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40.

Wilson, P. S. (1990). Inconsistent ideas related to definitions and examples. *Focus on Learning Problems in Mathematics*, 12, 31-47.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1998). *Acer ConQuest*. Melbourne, Australia: ACER Press.

附錄 1 一至九年級數學詞彙列表（2016 年整理）

因篇幅限制，附錄 1 將長期置於雲端空間，網址為 <https://goo.gl/wC16gk>

附錄 2 至 4

因篇幅限制，附錄 2 至 4 將長期置於雲端空間，網址為 <https://goo.gl/gw1fTY>