

## 三至九年級學生數學運算能力等化測量 與多向度分析

吳宜玲\* 楊心怡\*\* 吳昭容\*\*\* 陳柏熹\*\*\*\*

### 摘要

數學運算能力是學習必備基礎與生活所需技能，在小學階段主要是數字與運算符號的連結，國中階段包含內容知識、代數與函數，內涵廣泛且跨幅大，因此，數學運算能力發展是值得探討的議題。本研究使用數學運算能力測驗，以測驗等化技術進行小學三年級至國中九年級的橫斷資料4,991筆及追蹤測驗511筆縱貫資料分析，使用多向度隨機係數多項洛基模式比較不同階段之數學運算能力發展。研究結果為橫斷資料逐年成長，縱貫追蹤資料僅在使用相同題本之重複測驗中有成長，藉由數學詞彙測驗為參照測驗，發現隨著學生的年級增長，其數學運算能力與數學詞彙能力的關係越強。

**關鍵詞：**多向度隨機係數多項洛基模式、測驗等化、試題反應理論、數學運算能力



DOI: 10.6869/THJER.202112\_38(2).0004

投稿日期：2021年12月23日，2022年2月10日修改完畢，2022年2月17日通過採用

\* 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，E-mail: irene54510@gmail.com

\*\* 楊心怡，國立臺北科技大學技術及職業教育研究所副教授，E-mail: hcy103@ntut.edu.tw

\*\*\* 吳昭容，國立臺灣師範大學教育心理與輔導學系教授，E-mail: cjwu@ntnu.edu.tw

\*\*\*\* 陳柏熹（通訊作者），國立臺灣師範大學教育心理與輔導學系教授，E-mail: lorenz1020@gmail.com

## 壹、緒論

數學運算能力對於數學成就而言雖然不是最關鍵的能力，但掌握數學算則並能正確執行是全面地發展數學能力的重要因素（Cowan et al., 2011; Fuchs et al., 2014; Rittle-Johnson, Schneider, & Star, 2015; Sisco-Taylor, Fung, & Swanson, 2015）。數學運算能力的相關議題相當多，例如：探討學習基本數學事實（basic math facts）、算術能力（arithmetic ability）、運算流暢性（computational fluency），以及代數操作（algebraic operation）等（Dowker, 2019; Geary, Saults, Liu, & Hoard, 2000; Wong, 2021）。Lin與Powell（2021）整合250篇數學能力相關之研究論文進行後設分析，研究後發現，無論是年紀的增長與時間的推移，初始數學運算能力對於往後數學綜合能力表現有穩定的預測力。鑑於數學內容涉及多個領域，數學能力的發展是由許多不同單一技能之數學能力累積而成，因此，數學能力不著重於任何單一數學技能，許多縱向研究表明早期的數學知識與能力，對於往後的數學成就有重要的影響（Geary, Hoard, Nugent, & Bailey, 2013; Lin & Powell, 2021; Watts, Duncan, Siegler, & Davis-Kean, 2014; Watts et al., 2015）。國際教育成就評估協會（IEA）所推行的國際數學與科學教育成就趨勢調查（Trends in International Mathematics and Science Study，簡稱TIMSS），其所關注的數學成就即為數學的學習表現，由TIMSS所編製的測驗內容來評定學生的數學表現；而TIMSS的數學評量架構分為內容向度（content domains）與認知向度（cognitive domains），評量的學生為小學四年級與國中八年級，測驗結果表現即為數學成就，屬於一個廣泛的數學成就表現，而非單一的數學能力。因此，數學運算能力僅為數學能力中，著重於運算的單一能力，無法以數學運算能力來表示整體的數學能力，也難以單一數學運算能力呈現數學成就。

在許多數學能力測驗研究中，多以單一向度來探討數學能力，而各研究所指之數學能力定義分屬該研究評量內容目的（王暄博、郭伯臣、呂玉如，2013；曾芬蘭、林奕宏、邱佳民，2017；曾建銘、吳慧珉、趙珮晴，2019），僅少數採用多向度能力觀點探究的研究。例如：余明杰、李茂能與楊德清（2016）將學童數常識能力分為四個向度，探究各

向度能力間不同的發展；又如：洪儷瑜與連文宏（2019）將數學運算能力分為「分解與重組」能力、「數學事實提取」能力、「複雜計算」能力，而在這些能力之間進行彼此影響與預測的研究。然而在數學運算能力的發展過程中，各個階段關於數學運算的學習內容也會隨著年紀增長而改變，因此，本研究將數學運算能力依各個學習階段不同，分為「初階」數學運算能力、「中階」數學運算能力與「進階」數學運算能力三個向度，探究各向度能力發展，採用多向度隨機係數多項洛基模式（multidimensional random coefficients multinomial logit model, MRCMLM）（Adams, Wilson, & Wang, 1997），進行分析並探討橫跨多個學習階段之數學運算能力。MRCMLM為Rasch模式的延伸，從測驗觀點來看，Rasch模式是等距的量尺，它可以將估出來的學生能力值與試題參數線性運算進行比較，也具客觀的測量，學生能力估計值及試題難度互相獨立，也就是說，學生能力估計值不受試題難度估計值影響，試題難度估計值也不受學生能力估計值影響（王文中，2004）。

這類探究橫跨多個學習階段的測驗，需建立涵蓋不同認知層次與評量內容的測驗題庫，因此，無法讓學生於短時間內全數完成所有題庫試題。在此限制下，依測驗目標，測驗所使用的題本亦分為「初階」數學運算題本、「中階」數學運算題本與「進階」數學運算題本。為了能確保不同測驗間分數能相互比較，必須建立共同量尺進行測驗等化，考量同年級不同測驗間之等化（equating）、不同年級不同測驗間之等化、同年級不同年度測驗間之等化、不同年級不同年度測驗間之等化，才能使不同題本（booklet）的測驗分數能建立在同一個量尺上（郭伯臣、王暄博，2008）。本文旨在達成上述分析，以了解學生經過不同向度之數學運算課程學習，在數學運算能力變化的關係。

縱貫性研究可以長期追蹤學生的學習變化，Jordan、Kaplan、Raminen與Locuniak（2009）進行幼兒園至小學一年級數學運算能力的縱貫研究，僅探究數字、計數與簡易的數學運算能力，並在小學三年級時追蹤，發現早期的數學運算能力對於三年級時的數學能力表現有影響。Geary（2011）進行小學一年級至五年級數學運算能力的縱貫研究，探究數字、計數與數學運算能力的成長模型，研究發現基本的數學運算能力是學習更複雜數學的基礎，但未擴展至中學領域。Geary、

Nicholas、Li與Sun（2017）進行小學二年級至中學八年級數學運算能力的縱貫研究，但僅探究數字、分數與數學運算能力的增長情形，研究發現早期的數學運算能力會影響後續數學運算能力的增長，整體呈現穩定增長。Korpiäkä等人（2017）進行小學一年級到中學七年級閱讀與數學運算能力之間的發展關係，Gilbert與Fuchs（2017）進行小學二年級至三年級三次的單詞與數學運算能力之間的發展關係追蹤研究，可以發現探討數學運算能力的縱貫性研究是有其必要性，但上述之縱貫研究，均使用固定測驗進行能力變化的探究，存在些許測驗誤差。本研究聚焦在小學三年級到國中九年級學生的算術與代數運算能力，探討處理跨多年級數學運算能力的等化技術，連結測驗以降低測驗誤差（Kim & Cohen, 1998），以及分析學生運算能力的發展現象。綜上所述，本研究欲探討跨越不同學習階段的數學運算能力成長變化情形，採用數學運算能力測驗（楊心怡，未出版），此測驗架構編製了橫跨小學三年級到國中九年級之數學運算試題。

## 一、數學運算能力的定義與類型

美國數學教育成就評量（National Assessment of Educational Progress，簡稱NAEP）提出數學架構包含五類內容成分（content strands）、三種數學能力（mathematical abilities）和三類數學力（mathematical power）（National Assessment Governing Board, 2003）。其中程序性知識（procedural knowledge）是和概念理解（conceptual understanding）、問題解決（problem solving）分庭抗禮的三種數學能力之一。學生的程序性知識顯現在適當選擇與正確執行程序、正確判斷模型與符號，以及將程序應用與調整在問題解決上。就NAEP的架構，五類數學內容成分上都有程序性知識的展現，其中算術能力與代數操作應該是最被關注的程序性知識（Dowker, 2019; Nelson, Parker, & Zaslofsky, 2016; Siegler et al., 2012）。例如Siegler等人（2012）以英國和美國的資料庫分析發現，小學高年級時的分數知識和除法運算，對高中的代數成績和整體數學成績的預測力，顯著高於加法運算、減法運算、乘法運算、語文智商、非語文智商、工作記憶、父母教育程度，以及家庭收入

的預測力，而且兩國的模式類似。

數學運算能力是典型的程序性知識，依據Anderson的ACT-R理論（Adaptive Control of Thought-Rational）（Anderson et al., 2004），其發展過程需要經歷理解意義（認知的，cognitive）、掌握算則（聯結的，associative），最後達到流暢地執行（自動化，autonomous）。例如：要算出四個星期有多少天？對一個還不會乘法表的兒童而言，透過累加的認知就能完成任務；對學過 $7 \times 4 = 28$ 算則的兒童，則從陳述性知識中搜尋出此一乘法事實；對練習九九乘法足夠多次的兒童，則答案可以很自動化地被提取出來。因此，數概念能預測數學運算能力（Wong, 2017），而數學運算能力的測量除了以正確率來反映是否掌握算則之外，也有以速度測驗來評量流暢性（Geary et al., 2000; Russell, 2000）。

十二年國民基本教育課程綱要指出數學教學首應注重數、量、形，讓學生在實作與實測熟悉數、量、形及相互關係，再逐步抽象化與程序化，才能順利發展數學能力。而在教材安排上，小學一、二年級第一學習階段以數的概念、認識算則、透過各種情境的應用熟練兩步驟解題的加、減、乘簡易運算；小學三、四年級第二學習階段認識基本四則運算的算則，在兩步驟解題以加減與除、連乘、連除、乘除為主的基礎運算，培養流暢的數字感，本研究稱此階段數學運算能力為初階數學運算能力，使用數學運算能力測驗之初階數學運算題本；小學五、六年級第三學習階段在整數部分則為三步驟以上的運算與併式，四則混合運算含分配律與結合律的算則，簡易的比例關係，並以常用的數量關係，解決生活問題，本研究稱此階段數學運算能力為中階數學運算能力，使用數學運算能力測驗之中階數學運算題本，由此可知，小學所學習的數學運算能力是所有計算能力的基礎；在國中第四學習階段，在數方面，開始學習負數、根式與指數，在代數方面，則學習簡單的函數、不等式、方程式、比例式與多項式運算等更高階的數學運算概念，本研究稱此階段數學運算能力為進階數學運算能力，使用數學運算能力測驗之進階數學運算題本。數學運算能力評量內容架構圖，如圖1所示，各國關於數學運算能力的課程教學安排大致遵循類似順序（教育部，2008，2018）。

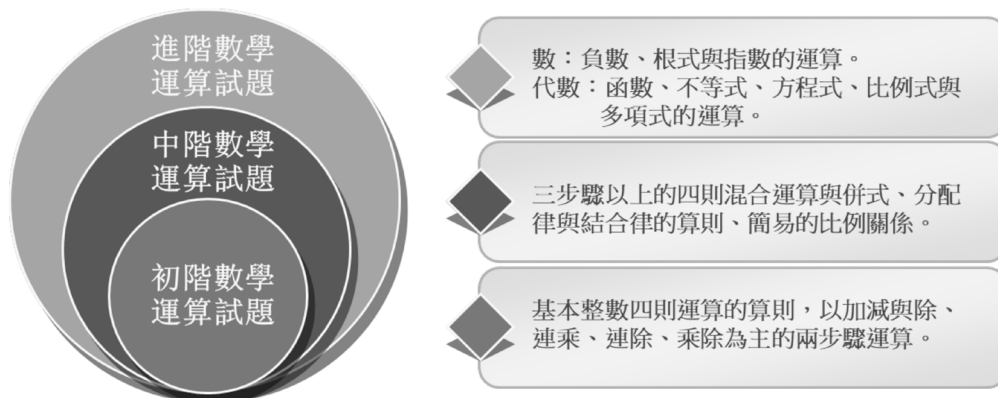


圖1 數學運算能力評量內容架構圖

在TIMSS的評量架構中，「數」(number)，分為「全數」(whole numbers)、「算式、簡單方程式與關係」(expressions, simple equations, and relationships)及「分數與小數」(fractions and decimals)；而在NAEP的評量架構中，「數的性質與運算」(number properties and operations)之內容，則包括「數感」、「估計」(estimation)、「數的運算」(number operations)、「比和比例的推理」(ratios and proportional reasoning)及「數與運算的性質」(properties of number and operations)。研究發現，在學生計算能力的發展歷程中，會先使用猜測與估計的策略，經過後天學習使用手指數數或口語數數進行個位數加減；隨著學習經驗的累積開始建立算式的連結，將原有的數數策略轉化為「分解與重組」策略(decomposition and regrouping strategy)，甚至可以從大腦中直接提取答案利用「數學事實提取策略」(math fact retrieval strategy)來解決問題，進而發展為複雜計算能力(洪麗瑜、連文宏，2019；Baroody, 2006; Beishuizen, 1993; Beishuizen, van Putten, & van Mulken, 1997; Jordan, Kaplan, Ramineni, & Locuniak, 2008; Lucangeli, Tressoldi, Bendotti, Bonanomi, & Siegel, 2003)。而在Imbo與Vandierendonck(2007)的研究指出，四年級到六年級的運算策略發展，分解與重組策略並未完全被數學事實提取策略所取代，在計算過程中彈性運用各個策略是相當重要的。本研究所使用之數學運算能力測驗，僅針對整數、比例，和代數等內容命題，故分數與小數的運算內容，並未納入本研究討論；再者，此測驗為難度測驗而非

速度測驗，因此，研究目的在探討學童的數學運算能力，於多種運算類型選擇題的能力表現。

## 二、能力估計模式

在單向度試題反應理論（unidimensional item response theory, UIRT）中，每道試題的試題參數與每個學生的能力會影響每道試題的答對機率，而所有題目都是測量同一向度與局部獨立性（local independency），是重要的基本假設。對於相同能力的學生，每一道試題答對的機率不會受其他題目是否答對而有所影響，因此，在這個基本假設下，接受不同測驗試題的學生，只要所測量的題目均是同一能力向度，所估出的能力值就可以互相比較。然而，實際上，有些測驗試題並非僅測量單一能力或潛在特質，這違反UIRT的理論假設，也無法藉著各向度能力的相關性來提升對各向度能力估計的精確性，需要增加試題數量，才能到達某個信度水準。因此，藉由多向度試題反應理論模式（multidimensional item response theory, MIRT）可以解決每一道試題只能測量一種能力的限制。利用MIRT模式估計能力，除了進行能力估計也同時納入各向度間的相關，因此，提升了各向度能力估計的精確性（陳柏熹，2006；Kelderman, 1996）。

Adams等人（1997）提出MRCMLM，是由Rasch模式延伸的MIRT模式，其反應模式如下：

$$f(X_{in} = 1; \mathbf{A}, \mathbf{B}, \boldsymbol{\xi} \mid \boldsymbol{\theta}_j) = \frac{\exp(\mathbf{b}'_{in} \boldsymbol{\theta}_j + \mathbf{a}'_{in} \boldsymbol{\xi})}{\sum_{k=1}^{K_i} \exp(\mathbf{b}'_{ik} \boldsymbol{\theta}_j + \mathbf{a}'_{ik} \boldsymbol{\xi})}, \quad (1)$$

其中  $X_{in} = \begin{cases} 1, & \text{第 } i \text{ 題作答第 } n \text{ 個反應類別} \\ 0, & \text{其他} \end{cases}$ ； $\boldsymbol{\xi}' = (\xi_1, \xi_2, \dots$

$\xi_p)$ ： $p$  個試題參數； $\boldsymbol{\theta}_j' = (\theta_{j1}, \theta_{j2}, \dots, \theta_{jD})$  第  $j$  個學生第 1 到  $D$  個能力向度； $\mathbf{a}_{ik}$ ：（ $i = 1, \dots, I$  和  $k = 1, \dots, K_i$ ）第  $i$  題中第  $k$  個反應類別所形成的向量，每個向量長度為  $p$ ，所有  $\mathbf{a}_{ik}$  形成  $\mathbf{A}'$  矩陣； $\mathbf{b}_{in}$ ：第  $i$  題中第  $n$  個反應類別得分； $\mathbf{b}_{ikd}$ ：第  $i$  題中第  $k$  個反應類別在第  $d$  個能力向度計分

形成  $b_{ik} = (b_{ik1}, b_{ik2}, \dots, b_{ikD})'$ ，再形成第  $i$  題的計分子矩陣  $B_i = (b_{i1}, b_{i2}, \dots, b_{iD})'$ ，最後成為整份測驗的計分矩陣  $B = (B'_1, B'_2, \dots, B'_i)'$ 。

由於MRCMLM是一個混合係數模型（mixed coefficients model），試題難度階參數可對應至本研究所定義之多向度數學運算能力，適用於本研究所定義之初階、中階與進階數學運算能力三個向度的能力估計。目前有學者已經開發完成的估計軟體可供使用，故本研究使用ACER ConQuest 2.0的估計軟體，採用MRCMLM模式，進行試題參數與多向度能力值估算（Adams, Wu, & Wilson, 2012）。

測驗架構可依據試題所測量的能力數量分為：(1) 題內多向度測驗（within-item multidimensional test）：測驗中每道試題不只測量一種能力，如圖2左方，中階數學運算試題裡的三步驟以上的運算涵蓋了初階數學運算試題裡的兩步驟基礎運算；(2) 題間多向度測驗（between-item multidimensional test）：測驗中每道試題只測量一種能力，如圖2右方（Adams et al., 1997）。

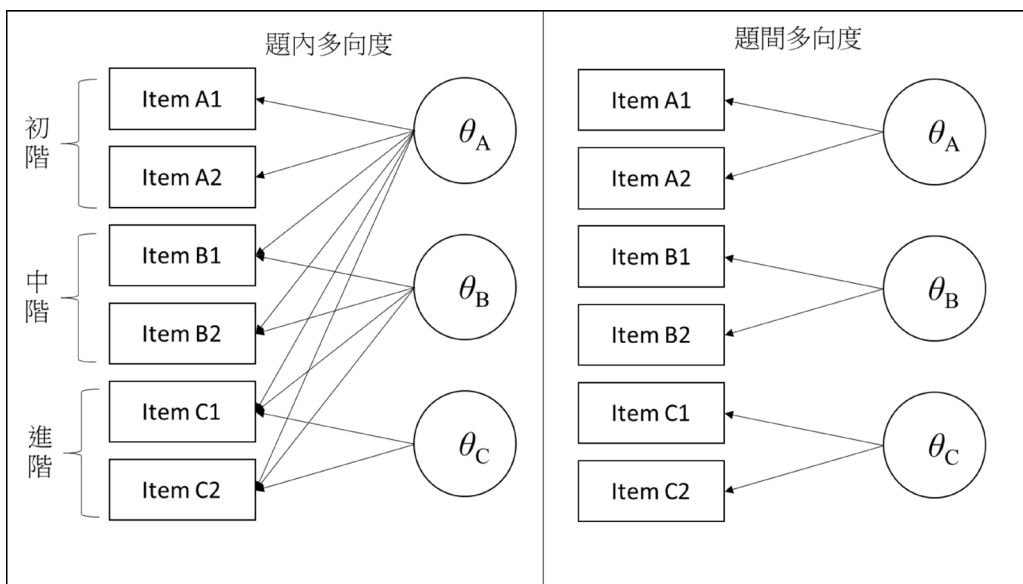


圖2 多向度評量架構圖

跨越各學習階段的數學運算能力通常被視為累進性的，也就是能完成較高階段的數學運算試題需要同時具備較低階段的能力。這在運用



MIRT進行能力估計時，可視為題內多向度的測驗架構，定義後面階段的數學運算能力具備前面階段的數學運算能力；也就是說，每一個階段的數學運算能力認知處理歷程不同，在每一個學習的過程當中，較高階之數學運算能力的形成，須具備較低階之數學運算能力。例如Siegler等人（2012）的橫斷資料研究指出，加、減、乘、除以整數之數學運算能力，會涉及往後代數運算能力。本研究的初階數學運算能力定義為小學三、四年級學生所應具備之數學運算能力，中階數學運算能力定義為小學五、六年級學生所應具備之數學運算能力，進階數學運算能力定義為國中學生所應具備之數學運算與代數能力。因此，在具備中階數學運算能力的情況下，也應有初階數學運算能力的基礎；同樣的，在具備進階數學運算能力的情況下，也應有初階數學運算能力與中階數學運算能力的基礎，這在運用MIRT進行能力估計時，可視為題內多向度的測驗架構，如圖2左方，每道試題會測量到一至三種能力，其數學運算能力評量內容架構圖，如圖1所示。也就是說，在中階數學運算題本之試題，測量中階數學運算能力時，測量內容亦涵蓋了初階數學運算能力，經由學習過程能力累加而形成中階數學運算能力；在進階數學運算題本之試題，測量進階數學運算能力時，測量內容亦涵蓋了初階數學運算能力與中階數學運算能力，經由學習過程能力累加而形成進階數學運算能力。另一種觀點是，將各學習階段的數學運算能力視為不同的認知能力。例如，國中代數運算中的數值，以整數而言通常為個位數或二位數，少見如五、六年級進行三、四位數的運算。因此，五、六年級時著力發展的複雜數字運算，並非完成國中階段代數運算絕對必要的能力；國中所學習的代數能力也不完全等於數學運算能力，而是著重於代數的單一數學代數能力，這在運用MIRT進行能力估計時，可視為題間多向度的測驗架構，也就是說，每道試題僅測量一種能力，而將這三個階段的數學運算能力分別視為不同的認知能力。但若將此三個階段的數學運算能力視為不同的認知能力，即無法進行各能力間的比較與推估。因此，本研究將以實徵資料採用MRCMLM題內多向度模式進行分析研究。

### 三、測驗等化

如何檢視學生是否隨著年級不同而能力有所不同是一個值得關注的議題，透過垂直等化能使不同年級的學生分別接受符合其能力之試題，將測驗結果建置在同一量尺進行比較，可以為學生的縱貫追蹤資料進行學習成長與年級間的差異。在測驗等化過程當中，必須在不同題本間加入一些相同的試題，稱為共同題。根據這些共同題來比較不同的試題參數。測驗等化方式有(1) 同時估計法（concurrent estimation）；(2) 分離估計法（separate estimate）：又分為平均數／平均數法（mean/mean）、平均數／標準差法（mean/sigma）；(3) 特徵曲線法（characteristic curve methods）等（Kolen & Brennan, 2004; Meyers, Miller, & Way, 2008）。Hanson與Béguin（2002）以模擬研究指出，在大樣本（例如  $n > 20$ ）的情況下同時估計法會比分離估計法有較小的誤差，小樣本（例如  $n = 20$ ）的情況則是分離估計法比同時估計法佳，而特徵曲線法則在大小樣本下都居中。Lei與Zhao（2012）使用垂直等化進行線性成長估計，其模擬研究指出，在  $n < 100$  的狀態下，使用同時估計法所得到的誤差較分離估計法低。

實徵上，許多研究也使用同時估計法降低測驗誤差，例如：陳煥文（2011）將四年級與六年級的數學科測驗利用同時估計法，進行兩個不同年級的測驗垂直等化連結至同一量尺進行測驗結果之比較。吳昭容、曾建銘、鄭鈴華、陳柏熹與吳宜玲（2018）將三年級至八年級學生數學詞彙能力測驗利用同時估計法，進行測驗水平與垂直等化，該研究亦指出，若是群體間有明顯不同能力，則可以使用潛在迴歸模式，將不同能力群體視為不同能力分布參數來進行同時估計法。曾建銘（2020）在三至八年級資料與可能性能力測驗中，也利用同時估計法來進行試題參數估計與測驗的垂直以及水平等化，透過縱貫性施測調查了解學生在三至八年級資料與可能性能力的變化。而採用同時估計法進行模擬研究有吳慧珉、郭伯臣、許天維與陳婉寧（2015）參考TASA2007數學科測驗，模擬四、五、六年級學生資料，利用BIB等化設計，採同時估計法進行測驗等化，發現以PV法在題數較少的情況下，會得到較佳的能力估計，而在題數較多的參數估計則是使用EAP法較好。因此，本研究採用同時

估計法進行水平與垂直等化，將所有學生的作答反應放入一個大矩陣內同時估計，藉以減少測驗連結時所產生的測驗試題參數估計誤差。

本研究使用數學運算能力測驗測得小學三年級至國中九年級學生的數學運算能力表現，每一道試題分別測量一至三個數學運算能力，利用不同階段數學運算題本及同階段數學運算題本之複本的共同題，連接等化三個不同階段能力的題本。由於每一種難度的數學運算題本間均有三個平行複本，有利於教師比較教學前後數學運算能力的變化，亦可使用不同難度的題本，追蹤學生在不同年級間的數學運算能力成長變化的情形。

綜合以上，本文的研究目的在於：

1. 進行跨多年級數學運算能力測驗之等化與試題分析。
2. 分析小學三年級至國中九年級各向度之數學運算能力。
3. 以橫斷資料與縱貫追蹤資料探討不同向度之數學運算能力的發展。

## 貳、研究方法

### 一、測驗工具

本研究使用數學運算能力測驗，適用於小學三年級至國中九年級學生。該測驗分為三種題本，初階數學運算題本適合小學三、四年級學生使用，其測驗評量內容為基本整數四則運算的算則，以加減與除、連乘、連除、乘除為主的兩步驟運算，所測量出來的能力為初階數學運算能力；中階數學運算題本適合小學五、六年級學生使用，其測驗評量內容為三步驟以上的四則混合運算與併式、分配律與結合律的算則、簡易的比例關係，所測量出來的能力為中階數學運算能力；進階數學運算題本適合國中七、八、九年級學生使用，其測驗評量內容在數方面，包含負數、根式與指數的運算，在代數方面，包含簡單的函數、不等式、方程式、比例式與多項式的運算，所測量出來的能力為進階數學運算能力，圖1為數學運算能力測驗各題本之評量內容架構圖，其命題之示例請見表1。每一種題本有三套平行複本，試題均為三選一之單選題，用

以測驗學生是否具備正確的數學運算知識、熟練的數學運算技巧、數與符號的運算能力，以及簡易的代數運算能力，測驗時間為20分鐘，測驗題長為30題。各測驗題本之內部一致性信度（Cronbach's  $\alpha$ 係數）介於.73至.89，而CTT試題難度則在.61至.83，如表2所示，結果顯示各測驗題本均具有可接受的信效度。

表1

數學運算能力測驗命題之示例題

類型	題本	主題	示例題
運算	初階數學運算	乘以一位數	$23 \times 6 = ?$
		連除	$48 \div 6 \div 2 = ?$
	中階數學運算	乘法對加法的分配律	$23 \times (6 + 14) = ?$
		比與比值	$3 : 2 = ( ) : 4$
代數	進階數學運算	一元一次方程式的解法	$a + 32 = 58$ ，求 $a = ?$
		一次函數	$f(x) = 4x - 8$ ，求 $f(2) = ?$

表2

數學運算各題本的試題信度與CCT難度

正式測驗 版本		人數 ( $n$ )	信度 (Cronbach's $\alpha$ )	CTT試題難度 平均數 (全距)
三、四 年級	初階1	258	.73	.71 (.34~.92)
	初階2	258	.86	.71 (.31~.93)
	初階3	263	.87	.72 (.35~.96)
五、六 年級	中階1	269	.86	.83 (.54~.98)
	中階2	272	.81	.82 (.52~.97)
	中階3	267	.89	.77 (.55~.95)
國中	進階1	276	.86	.61 (.22~.81)
	進階2	275	.86	.62 (.17~.85)
	進階3	272	.87	.61 (.18~.79)

初階數學運算能力題本的各平行複本間有6題共同題，藉以進行測驗水平等化。這6題共同題中的3題會同步出現在中階數學運算能力題本，在試題參數估計的同時，除了水平等化亦同步進行垂直等化。中階數學運算能力題本除了有3題初階數學運算能力試題之外，另有4題中階數學運算試題會同步出現在進階數學運算題本，藉以進行測驗垂直等化。進階數學運算能力題本的各平行複本間有6題共同題，藉以進行測驗水平等化，另有4題中階數學運算能力題本試題為共同題，藉以進行測驗垂直等化，圖3為各題本之共同題配置圖。

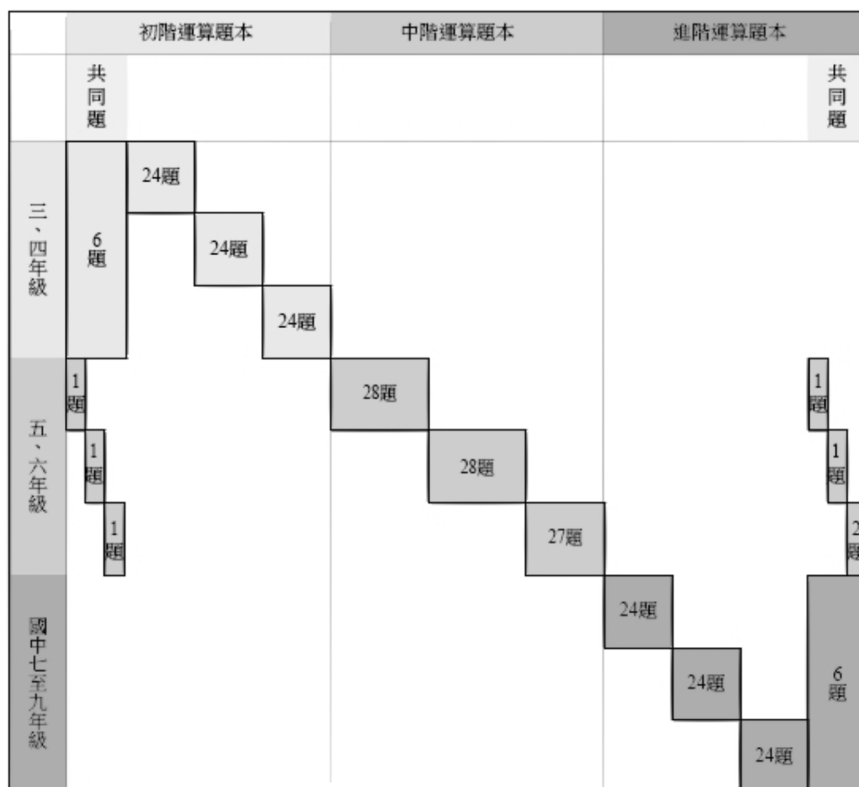


圖3 各題本共同題配置圖

## 二、資料來源

本研究所使用之數學運算能力測驗與參照測驗數學詞彙測驗資料來源分為預試、正式測驗與追蹤測驗，於2017年3月預試，同年9月正式測

驗，這兩次資料合併為本研究的橫斷資料；2019年5月部分受試者進行追蹤測驗，形成縱貫資料。數學運算能力測驗預試使用18個題本，322道試題，經過預試試題測驗結果分析，刪除不良試題後，保留9個正式題本，239道試題為正式測驗試題，追蹤測驗所使用之試題亦為正式測驗題本，橫斷資料的分析只採用品質良好試題之作答反應資料。

預試參考104學年國小、國中學校別資料（教育部統計處，2016a，2016b）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19所國小及19所國中參與。正式測驗，同樣參考各縣市國中、小概況統計資料進行取樣，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，最後有18所國小及17所國中參與。追蹤測驗以參加過正式測驗的部分學校為主要對象，總計有8所國小及25所國中參與。

本研究之數學運算能力測驗於小學三年級至國中九年級橫斷資料的有效樣本數依序為665、704、709、921、709、912、371筆，共計4,991筆，同時間也進行參照測驗數學詞彙測驗之學生於小學三年級至小學六年級的有效樣本數依序為317、109、99、70筆，共計595筆。數學運算能力追蹤測驗，分別為六年級、八年級與九年級的學生，縱貫追蹤資料的有效樣本數依序為168、172、171筆，共計511筆，同時間也進行參照測驗數學詞彙測驗之學生為國中九年級學生，有效樣本數為351筆。

### 三、試題參數估計與測驗等化

試題反應理論是可以描述受測者能力、試題參數（試題難度、試題鑑別度與試題猜測度）及作答反應結果的模式，依據所包含的試題參數個數分為單參數的Rasch模式、二參數模式和三參數模式。本研究採MRCMLM模式並以ConQuest進行試題分析（Adams et al., 2012），是Rasch模式的延伸模式，將不同題本的結果放在相同量尺上進行比較，進行試題參數估計與測驗等化，不同測驗卻測量相同的潛在特質或能力，可以透過連結（linking）和等化（equating）這兩種方法來比較測驗結果。水平等化目的在於校正測驗艱難度差異而非內容差異，垂直等化

目的則在於連結不同難度等級但測驗內容相似的測驗。

本研究採同時估計法，利用共同題將所有題本的試題作答反應合併成同一個作答反應矩陣，同時進行試題參數估計，以減少測驗連結間所產生的誤差，能獲得較佳的測驗精準度，以便進行水平與垂直等化（郭伯臣、王暄博，2008；陳煥文，2011；Kim & Cohen, 1998）。每一個數學運算能力題本題數均為30題，初階數學運算試題總題數為78題，中階數學運算試題總題數為87題，進階數學運算試題題數共74題，總題數共239題，表3為各題本間共同題所評量之主題與試題參數。

表3

各題本間之共同題主題與參數

來源	共同題題本	等化	主題	Rasch 難度	MNSQ
初階數學運算	初階數學運算	水平	減法	-2.71	0.99
			除	-1.07	0.88
			連除	-0.90	1.14
	初階、中階 數學運算	水平與 垂直	加法	-3.40	1.05
			連乘	-1.74	0.97
			減與除	0.90	1.11
中階數學運算	中階、進階 數學運算	垂直（中階）	連除	-0.62	1.01
			比例	-0.73	0.90
		水平（進階）	四則運算	-0.19	1.05
			四則運算	-0.34	0.94
進階數學運算	進階數學運算	水平	四則運算	-0.89	0.95
			四則運算	-1.02	0.95

由試題訊息量可以得知試題對於不同能力受測者所能提供之測量精準度，訊息量越大表示該題所提供之測量精準度越好，其計算公式為：

$$I_i(\theta) = P_i(\theta)(1 - P_i(\theta)) \quad (2)$$

其中 $I_i(\theta)$ 為第 $i$ 題的訊息量， $P_i$ 為第 $i$ 題答對率。

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），其計算公式為：

$$TI(\theta) = \sum_{i=1}^I I_i(\theta) \quad (3)$$

圖4為三個不同能力，共9個題本之測驗訊息量圖，可以發現三個能力各複本間的訊息量函數圖形大致相同，具良好複本信度。

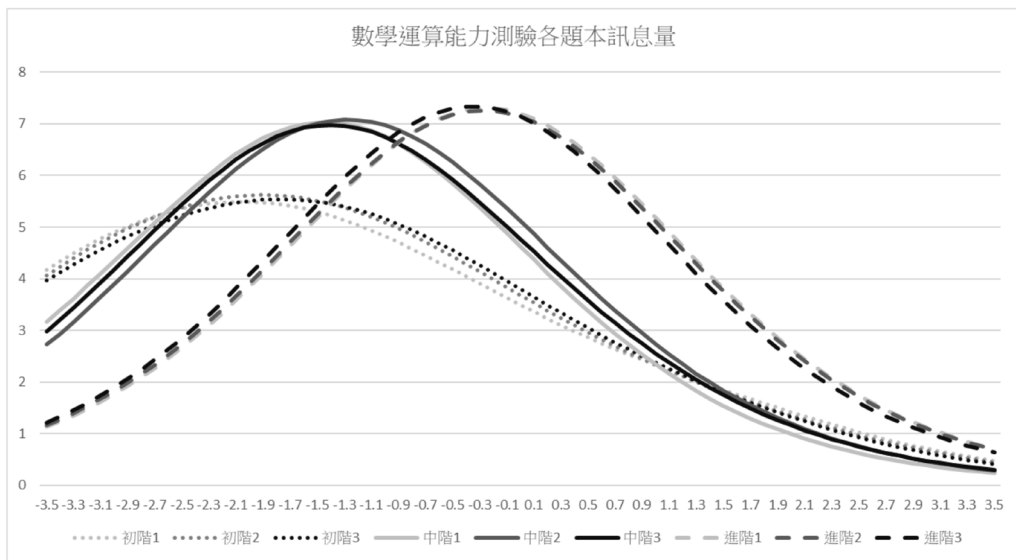


圖4 各題本之測驗訊息量函數圖



## 四、試題難度與受測者能力

本研究整併預試、正式測驗共4,991名有效受測者的資料，縱貫追蹤測驗共511名有效受測者的資料，以MRCMLM題內多向度模式進行試題參數估計，採用貝氏期望後驗法（expected a posterior, EAP）進行受測者能力估計，因為EAP估計法是利用先驗分配（prior distribution）來增進估計精準度，本研究之樣本遍及全國樣本數大、異質性也大，可以建立相對穩定之先驗分配，又能估算出全部答對與全部答錯之受測者能力值，且其所得之能力值結果較最大概似估計法（maximum likelihood estimator, MLE）合理（Bock & Mislevy, 1982）。

## 參、結果與討論

### 一、試題難度

本研究之數學運算能力測驗適用於小學三年級至國中九年級的學生，所使用的資料是從數學運算能力測驗資料庫取得，採MRCMLM模式利用題內多向度模式進行分析，將所有題本的試題作答反應合併成同一個作答反應矩陣以共同題進行串連，進行試題參數與能力值估計，為同時估計法，藉以減少測驗連結間所產生的試題參數誤差，獲得較佳的測驗精準度，同時進行水平與垂直等化三個不同向度之數學運算能力。因每一個題本的題數均為30題，答對題數未達11題視為無效樣本予以刪除，試題分析結果如表4所示。Wu、Adams與Wilson（1998）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, Yan, & Heene, 2021）；本測驗全部239道題中僅有2道題目的適配值略差（編號16、41），皆在進階數學運算題本，其Weighted Fit MNSQ值為1.35與1.37，考量進階數學運算題本的涵蓋測驗內容範圍較廣，試題內容包含國中七年級至九年級所學習的數學運算與基本代數試題，故不刪除任何試題，均予以保留。

表4  
數學運算能力各試題之分析結果

初階數學運算試題			中階數學運算試題			進階數學運算試題		
編號	Rasch 難度	MNSQ (CI)	編號	Rasch 難度	MNSQ (CI)	編號	Rasch 難度	MNSQ (CI)
1	-3.435	0.98 ( 0.70, 1.30)	1	-1.922	1.04 ( 0.65, 1.35)	1	-0.671	1.02 ( 0.83, 1.17)
2	-3.255	0.96 ( 0.88, 1.12)	2	-2.549	0.99 ( 0.31, 1.69)	2	-0.554	0.96 ( 0.87, 1.13)
3	-2.208	0.99 ( 0.54, 1.46)	3	-1.459	0.96 ( 0.78, 1.22)	3	-0.510	0.97 ( 0.88, 1.12)
4	-2.585	1.05 ( 0.85, 1.15)	4	-1.760	0.93 ( 0.70, 1.30)	4	-0.399	0.90 ( 0.90, 1.10)
5	-3.007	0.88 ( 0.95, 1.05)	5	-1.368	1.03 ( 0.80, 1.20)	5	-0.889	0.95 ( 0.88, 1.12)
6	-2.924	0.97 ( 0.94, 1.06)	6	-1.848	0.88 ( 0.67, 1.33)	6	-1.018	0.95 ( 0.85, 1.15)
7	-2.924	0.99 ( 0.89, 1.11)	7	-1.848	1.05 ( 0.67, 1.33)	7	-0.544	0.95 ( 0.87, 1.13)
8	-1.149	1.11 ( 0.95, 1.05)	8	-1.083	1.06 ( 0.84, 1.16)	8	-0.213	0.95 ( 0.93, 1.07)
9	-2.845	0.98 ( 0.73, 1.27)	9	-1.664	0.97 ( 0.73, 1.27)	9	-0.399	0.96 ( 0.90, 1.10)
10	-2.378	1.10 ( 0.85, 1.15)	10	-1.406	1.11 ( 0.79, 1.21)	10	-0.188	0.99 ( 0.93, 1.07)
11	1.442	1.10 ( 0.82, 1.18)	11	-1.563	0.97 ( 0.75, 1.25)	11	0.059	0.99 ( 0.93, 1.07)
12	-1.796	1.01 ( 0.77, 1.23)	12	-1.308	0.98 ( 0.81, 1.19)	12	0.151	0.94 ( 0.92, 1.08)
13	-2.263	1.01 ( 0.78, 1.22)	13	-1.149	1.09 ( 0.83, 1.17)	13	-0.209	0.92 ( 0.93, 1.07)
14	-2.897	1.03 ( 0.78, 1.22)	14	-1.120	1.02 ( 0.84, 1.16)	14	-0.428	0.92 ( 0.89, 1.11)
15	-2.018	1.12 ( 0.90, 1.10)	15	-1.682	0.83 ( 0.72, 1.28)	15	-0.274	0.95 ( 0.92, 1.08)
16	1.377	1.08 ( 0.79, 1.21)	16	-1.628	0.93 ( 0.74, 1.26)	16	-0.245	0.92 ( 0.92, 1.08)
17	-1.086	1.05 ( 0.84, 1.16)	17	-1.701	0.86 ( 0.72, 1.28)	17	-0.028	1.03 ( 0.93, 1.07)
18	1.173	0.88 ( 0.86, 1.14)	18	-1.021	1.04 ( 0.85, 1.15)	18	-0.238	1.00 ( 0.92, 1.08)
19	-1.344	1.04 ( 0.85, 1.15)	19	-1.393	0.84 ( 0.79, 1.21)	19	0.240	1.37 ( 0.91, 1.09)
20	-1.175	1.01 ( 0.78, 1.22)	20	-1.611	0.91 ( 0.74, 1.26)	20	0.158	1.28 ( 0.92, 1.08)
21	-1.149	0.90 ( 0.86, 1.14)	21	-1.419	0.90 ( 0.79, 1.21)	21	-0.025	0.98 ( 0.93, 1.07)
22	-0.659	0.95 ( 0.86, 1.14)	22	-1.368	0.82 ( 0.80, 1.20)	22	0.127	0.96 ( 0.92, 1.08)
23	-4.206	1.04 ( 0.91, 1.09)	23	-1.343	0.91 ( 0.80, 1.20)	23	0.179	0.95 ( 0.92, 1.08)
24	-1.061	0.95 ( 0.87, 1.13)	24	-1.074	1.05 ( 0.85, 1.15)	24	-0.116	1.04 ( 0.93, 1.07)
25	-3.395	0.95 ( 0.90, 1.10)	25	-0.064	1.13 ( 0.92, 1.08)	25	-0.031	0.93 ( 0.93, 1.07)
26	-1.065	0.88 ( 0.90, 1.10)	26	-0.786	1.06 ( 0.88, 1.12)	26	0.295	1.18 ( 0.90, 1.10)
27	-1.740	0.90 ( 0.90, 1.10)	27	-1.092	1.02 ( 0.84, 1.16)	27	-1.204	0.99 ( 0.56, 1.44)
28	-2.708	0.89 ( 0.91, 1.09)	28	-2.281	1.05 ( 0.42, 1.58)	28	-0.576	0.95 ( 0.86, 1.14)
29	0.903	0.89 ( 0.91, 1.09)	29	-1.518	1.03 ( 0.74, 1.26)	29	-0.619	0.99 ( 0.84, 1.16)
30	-3.638	1.02 ( 0.67, 1.33)	30	-1.601	0.98 ( 0.72, 1.28)	30	-0.366	0.92 ( 0.90, 1.10)
31	-3.550	0.97 ( 0.90, 1.10)	31	-1.550	1.14 ( 0.73, 1.27)	31	-0.672	0.98 ( 0.83, 1.17)

(續下頁)

32	-3.391	1.00 (0.88, 1.12)	32	-1.852	0.99 (0.63, 1.37)	32	-0.164	0.97 (0.93, 1.07)
33	-2.577	1.07 (0.68, 1.32)	33	-1.534	1.15 (0.74, 1.26)	33	-0.390	0.97 (0.90, 1.10)
34	-2.787	0.99 (0.71, 1.29)	34	-1.153	1.05 (0.82, 1.18)	34	-0.164	0.99 (0.93, 1.07)
35	-2.864	1.11 (0.82, 1.18)	35	-1.403	1.03 (0.77, 1.23)	35	0.069	1.00 (0.93, 1.07)
36	-2.890	1.05 (0.79, 1.21)	36	-1.444	1.08 (0.76, 1.24)	36	0.196	0.92 (0.92, 1.08)
37	-1.200	0.97 (0.78, 1.22)	37	-1.674	0.96 (0.70, 1.30)	37	-0.240	1.01 (0.92, 1.08)
38	-2.409	1.21 (0.78, 1.22)	38	-0.849	1.01 (0.87, 1.13)	38	-0.247	0.98 (0.92, 1.08)
39	-1.894	0.87 (0.90, 1.10)	39	-1.503	0.99 (0.74, 1.26)	39	-0.378	0.91 (0.90, 1.10)
40	1.448	1.08 (0.83, 1.17)	40	-1.503	0.97 (0.74, 1.26)	40	-0.317	0.91 (0.91, 1.09)
41	-1.239	1.08 (0.87, 1.13)	41	-1.225	0.96 (0.81, 1.19)	41	-0.084	1.08 (0.93, 1.07)
42	-2.787	1.00 (0.86, 1.14)	42	-1.403	0.91 (0.77, 1.23)	42	-0.141	1.05 (0.93, 1.07)
43	-2.555	0.97 (0.79, 1.21)	43	-0.763	1.16 (0.88, 1.12)	43	0.277	1.35 (0.91, 1.09)
44	-1.847	1.00 (0.82, 1.18)	44	-1.163	0.93 (0.82, 1.18)	44	0.159	1.19 (0.92, 1.08)
45	-1.666	1.02 (0.87, 1.13)	45	-1.534	0.94 (0.74, 1.26)	45	-0.114	0.98 (0.93, 1.07)
46	1.637	0.84 (0.88, 1.12)	46	-1.827	1.02 (0.64, 1.36)	46	0.119	0.99 (0.93, 1.07)
47	-1.372	1.00 (0.85, 1.15)	47	-0.842	1.04 (0.87, 1.13)	47	0.095	0.99 (0.93, 1.07)
48	-0.560	1.05 (0.90, 1.10)	48	-0.842	0.94 (0.87, 1.13)	48	-0.131	1.03 (0.93, 1.07)
49	-1.149	0.89 (0.91, 1.09)	49	-1.183	0.97 (0.81, 1.19)	49	-0.075	0.99 (0.93, 1.07)
50	-0.608	0.90 (0.90, 1.10)	50	-0.297	1.17 (0.92, 1.08)	50	0.274	1.07 (0.91, 1.09)
51	-1.162	0.80 (0.91, 1.09)	51	-0.043	1.13 (0.93, 1.07)	51	-0.712	1.04 (0.88, 1.12)
52	-1.770	0.89 (0.90, 1.10)	52	-2.115	0.97 (0.64, 1.36)	52	-0.609	0.91 (0.90, 1.10)
53	-0.299	1.02 (0.91, 1.09)	53	-2.090	1.10 (0.65, 1.35)	53	-0.357	1.00 (0.93, 1.07)
54	-3.523	1.02 (0.69, 1.31)	54	-1.630	1.04 (0.78, 1.22)	54	-0.443	0.91 (0.92, 1.08)
55	-3.256	0.95 (0.89, 1.11)	55	-1.900	0.95 (0.71, 1.29)	55	-0.684	0.90 (0.88, 1.12)
56	-2.551	0.98 (0.53, 1.47)	56	-1.753	1.06 (0.75, 1.25)	56	-0.210	1.10 (0.94, 1.06)
57	-3.124	1.01 (0.73, 1.27)	57	-1.117	1.10 (0.86, 1.14)	57	-0.524	0.95 (0.91, 1.09)
58	-3.156	1.13 (0.82, 1.18)	58	-1.461	0.99 (0.81, 1.19)	58	-0.360	0.94 (0.93, 1.07)
59	-3.093	1.01 (0.75, 1.25)	59	-1.978	1.07 (0.69, 1.31)	59	-0.049	0.94 (0.95, 1.05)
60	-1.201	0.98 (0.75, 1.25)	60	-1.814	0.97 (0.74, 1.26)	60	-0.214	0.97 (0.94, 1.06)
61	-1.188	1.03 (0.76, 1.24)	61	-1.214	1.01 (0.85, 1.15)	61	-0.603	0.99 (0.93, 1.07)
62	-2.210	1.04 (0.90, 1.10)	62	-1.865	0.97 (0.72, 1.28)	62	-0.468	0.97 (0.93, 1.07)
63	-1.641	0.88 (0.90, 1.10)	63	-1.029	0.94 (0.87, 1.13)	63	-0.563	0.92 (0.93, 1.07)
64	1.289	1.16 (0.85, 1.15)	64	-1.503	0.97 (0.81, 1.19)	64	-0.360	0.85 (0.93, 1.07)
65	-1.456	1.13 (0.89, 1.11)	65	-1.643	0.85 (0.78, 1.22)	65	-0.088	1.05 (0.95, 1.05)
66	-2.618	1.00 (0.87, 1.13)	66	-1.753	0.92 (0.75, 1.25)	66	-0.098	1.09 (0.95, 1.05)
67	-2.618	0.96 (0.82, 1.18)	67	-0.928	1.08 (0.88, 1.12)	67	0.012	1.14 (0.95, 1.05)

(續下頁)

68	-1.568	0.95 ( 0.82, 1.18)	68	-1.253	0.83 ( 0.84, 1.16)	68	0.129	1.26 ( 0.94, 1.06)
69	-1.869	0.98 ( 0.89, 1.11)	69	-1.191	1.02 ( 0.85, 1.15)	69	-0.202	0.98 ( 0.94, 1.06)
70	1.320	0.92 ( 0.88, 1.12)	70	-0.726	1.00 ( 0.90, 1.10)	70	0.038	0.97 ( 0.95, 1.05)
71	-0.849	0.95 ( 0.87, 1.13)	71	-1.229	0.98 ( 0.85, 1.15)	71	0.004	0.97 ( 0.95, 1.05)
72	-0.973	0.99 ( 0.91, 1.09)	72	-1.312	0.87 ( 0.84, 1.16)	72	-0.219	0.95 ( 0.94, 1.06)
73	-0.716	0.84 ( 0.91, 1.09)	73	-1.338	0.88 ( 0.83, 1.17)	73	-0.126	0.97 ( 0.95, 1.05)
74	-0.825	0.85 ( 0.91, 1.09)	74	-0.858	1.04 ( 0.89, 1.11)	74	0.111	1.07 ( 0.95, 1.05)
75	-1.023	0.77 ( 0.91, 1.09)	75	-1.696	1.04 ( 0.65, 1.35)			
76	-4.248	0.88 ( 0.91, 1.09)	76	-1.318	0.96 ( 0.64, 1.36)			
77	-0.644	0.88 ( 0.91, 1.09)	77	-1.295	1.03 ( 0.69, 1.31)			
78	-0.900	1.14 ( 0.93, 1.07)	78	-1.318	1.20 ( 0.70, 1.30)			
			79	-0.562	1.15 ( 0.69, 1.31)			
			80	-1.589	1.02 ( 0.86, 1.14)			
			81	-1.915	0.98 ( 0.74, 1.26)			
			82	-1.915	1.04 ( 0.64, 1.36)			
			83	-0.865	1.02 ( 0.64, 1.36)			
			84	-0.621	1.01 ( 0.93, 1.07)			
			85	-0.188	1.05 ( 0.97, 1.03)			
			86	-0.335	0.94 ( 0.96, 1.04)			
			87	-0.729	0.90 ( 0.92, 1.08)			

## 二、受測者能力

由橫斷資料顯示，藉由水平等化之結果，初階數學運算能力題本平均Rasch難度值與研究者繪製之各年級學生初階數學運算能力分布圖如圖5，初階數學運算題本的試題Rasch難度平均值為-1.72，初階數學運算能力的分析結果，三年級的學生初階數學運算能力平均值為-0.91，四年級的學生初階數學運算能力平均值為0.55，標準差分別為0.86與1.08，可以發現學生的初階數學運算能力隨著年級增長而增加，各年級之初階數學運算能力分布曲線亦逐年右移；中階數學運算能力題本平均Rasch難度值與研究者繪製之各年級學生中階數學運算能力分布圖如圖6，中階數學運算題本的試題Rasch難度平均值為-1.35，中階數學運算能力的分析結果，五年級的學生中階數學運算能力平均值為-0.34，六年級的學

生中階數學運算能力平均值為0.14，標準差分別為1.07與1.10，可以發現學生的中階數學運算能力隨著年級增長而增加，各年級之中階數學運算能力分布曲線亦逐年右移；進階數學運算能力題本平均Rasch難度值與研究者繪製之各年級學生進階數學運算能力分布圖如圖7，進階數學運算題本的試題Rasch難度平均值為-0.23，進階數學運算能力的分析結果，七年級的學生進階數學運算能力平均值為-0.26，八年級的學生進階數學運算能力平均值為0.21，九年級的學生進階數學運算能力平均值為0.44，標準差分別為0.67、0.93與0.96，可以發現學生的進階運算能力隨著年級增長而增加，各年級之進階數學運算能力分布曲線亦逐年右移。

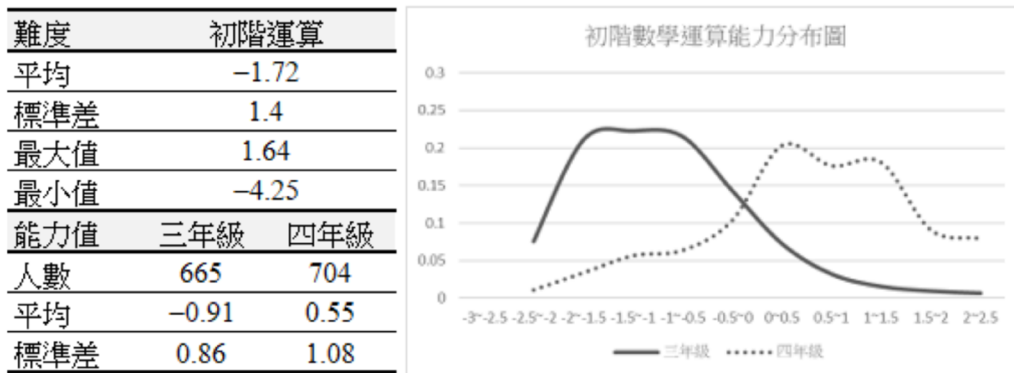


圖5 初階數學運算題本平均Rasch難度與各年級學生初階數學運算能力分布圖

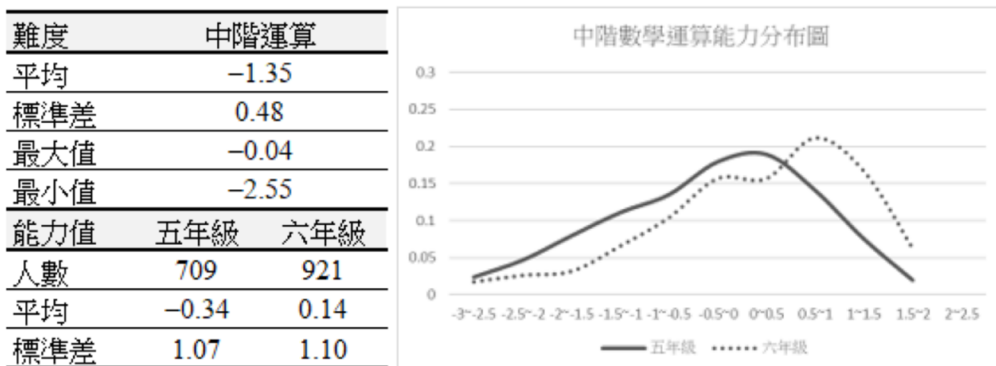


圖6 中階數學運算題本平均Rasch難度與各年級學生中階數學運算能力分布圖

難度	進階運算		
平均	-0.23		
標準差	0.32		
最大值	0.3		
最小值	-1.2		
能力值	七年級	八年級	九年級
人數	709	912	371
平均	-0.26	0.21	0.44
標準差	0.67	0.93	0.96

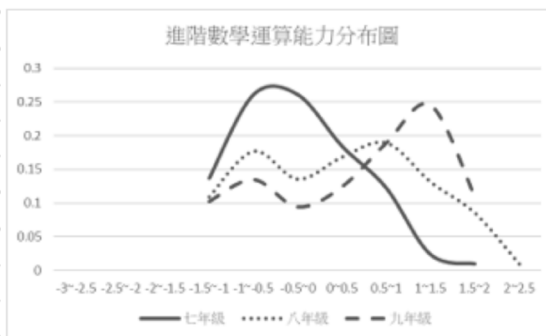


圖7 進階數學運算題本平均Rasch難度與各年級學生進階數學運算能力分布圖

### 三、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，利用共同題將所有題本的試題作答反應放在同一個矩陣中進行估計，以減少測驗連結時所產生的誤差，為同時估計法。學者建議共同題的比例以大於題本題數的五分之一為原則（Holland & Dorans, 2006; Livingston, 2004）。本研究所採用的數學運算能力測驗中，初階數學運算題本與進階數學運算題本的平行複本，均有6題共同題，比例占題本總題數30題的五分之一，而在中階數學運算題本，沒有共同題可以進行平行複本的測驗等化。

跨階段數學運算能力的垂直測驗等化，在初階數學運算題本與中階數學運算題本的測驗連結，三本中階數學運算題本各僅以1題初階數學運算題本的試題進行垂直等化；在中階數學運算題本與進階數學運算題本的測驗連結，三本中階運算題本各僅以1至2題中階數學運算與進階數學運算題本的試題進行垂直等化，因為垂直等化的共同題數過少，低於學者建議之共同題比例，導致垂直等化效果不盡理想，建議提升垂直等化的共同題題數，採用同一階段的水平等化設計標準，在每個跨階段題本中放入至少6題共同題，才能讓該題本與前後階段題本有較佳的等化效果。

然而，小學三年級至國中九年級的學生數學運算能力題本，在各個

階段的試題無法進行有效的共同題連結，原因為初階數學運算試題僅著墨於兩步驟解題，尤以加減與除、連乘、連除、乘除為主的基礎運算，在進行反覆的大量運算下，只能以數值的變化，加以增加試題的難度；但在中階數學運算試題相較於初階運算試題數值變化較少，改以著墨為三步驟以上的運算與併式，四則混合運算含分配律與結合律的算則，亦增加簡易的比例關係。因此，在中階數學運算題本若與初階數學運算題本共同題題數比例達五分之一，測驗題長受限於30題，測驗時間受限於20分鐘，導致如此安排試題會使得測驗試題內容無法涵蓋小學五、六年級之數學運算課程內容。同樣的狀況亦發生在國中階段的進階數學運算題本，試題若僅以測驗計算的部分，整數而言通常為個位數或二位數，少見如五、六年級進行的三、四位數的運算，若代數涉及整數運算也相對容易，在進階數學運算試題相較於中階數學運算試題數值變化更少，改以著墨為負數、根式與指數的運算，在代數方面，則是簡單的函數、不等式、方程式、比例式與多項式運算等更高階的數學運算概念，因此，在進階數學運算題本若與中階數學運算題本共同題題數比例達五分之一，測驗題長受限於30題，測驗時間受限於20分鐘，導致如此安排試題會使得測驗試題內容無法涵蓋國中之所有運算與基礎的代數課程內容。若是在不同的階段採用不同的題長與測驗時間，就能增加共同題的數量。

綜上所述，本研究所使用之數學運算能力測驗，使用在跨階段的能力比較與推估需要更加謹慎，因此，在追蹤測驗僅比較同個階段的數學運算能力變化情形，例如：僅使用進階數學運算題本之七年級學生與九年級學生之進階數學運算能力的變化。

#### 四、追蹤測驗結果

本研究亦進行追蹤測驗，希望藉由共同學生進行前後測驗的縱貫追蹤資料，了解學生經過學習之後，數學運算能力的變化情形，分別有三種不同的樣本群體，四年級與六年級、六年級與八年級、七年級與九年級，學生分別在這兩個時間點進行前後測驗，其中在七年級與九年級這個群體中有55人，採用固定的相同題本。研究者繪製之四年級與六年

級追蹤測驗學生數學運算能力分布圖如圖8，四年級學生之初階數學運算平均能力值為0.50，六年級學生之初階數學運算平均能力值為0.44，標準差分別為1.10與1.08，表現差異不大；研究者繪製之六年級與八年級追蹤測驗學生數學運算能力分布圖如圖9，六年級學生之中階數學運算平均能力值為0.09，八年級學生之中階數學運算平均能力值為0.08，標準差分別為1.14與1.11，表現差異不大，此兩群學生均是跨不同題本進行追蹤測驗，分別為初階數學運算題本與中階數學運算題本，以及中階數學運算題本與進階數學運算題本，其能力沒有明顯提升，但可以發現，在較高的年級有較多的高能力者。在跨題本間的能力會因為各題本的難易度變化與測驗主題改變而有所影響，而用以連接此兩階段之垂直等化共同題過少，僅以1題共同題作為題本間的連結，導致測驗垂直等化效果不彰，例如，前述之追蹤四年級學生與六年級學生初階數學運算能力結果，在六年級學生所使用之中階數學運算題本僅使用1題初階數學運算試題進行測驗垂直等化，因此，在此共同題的作答反應結果會影響能力值的估計，而導致初階運算能力在四年級與六年級變化差異不大。

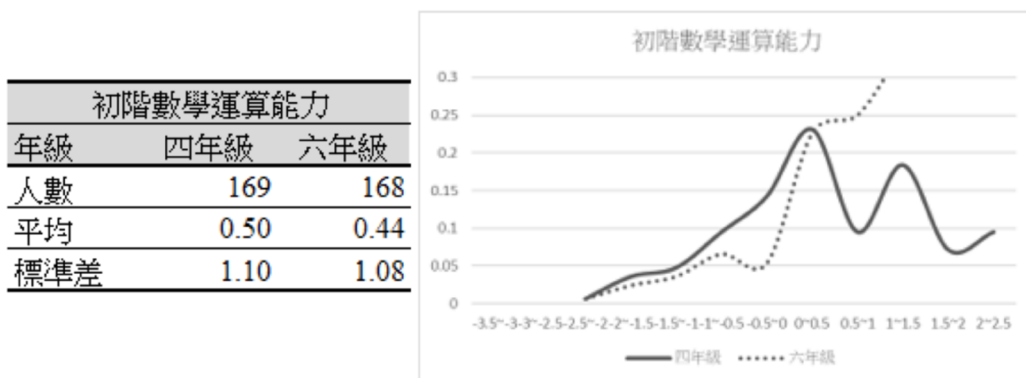


圖8 四年級與六年級追蹤測驗學生初階數學運算能力分布圖



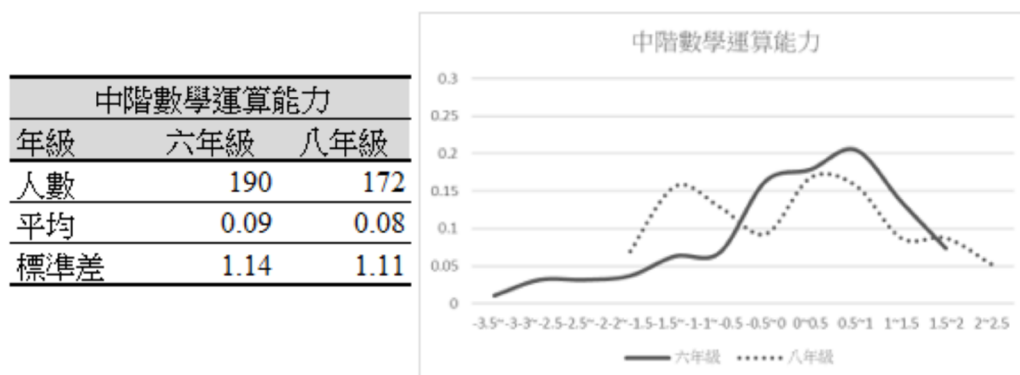


圖9 六年級與八年級追蹤測驗學生中階數學運算能力分布圖

研究者繪製之七年級與九年級追蹤測驗學生數學運算能力分布圖如圖10，七年級學生之進階運算平均能力值為-0.26，九年級學生之進階運算平均能力值為0.16，標準差分別為0.73與0.95，此群學生之追蹤測驗以水平等化為測驗連結，共同題比例為五分之一，可以發現學生在沒有跨階段題本間的能力表現，即有成長的趨勢，能力分布均逐年向右移，人數也有增加的趨勢，表示學生數學運算能力隨著年級增長而增加。

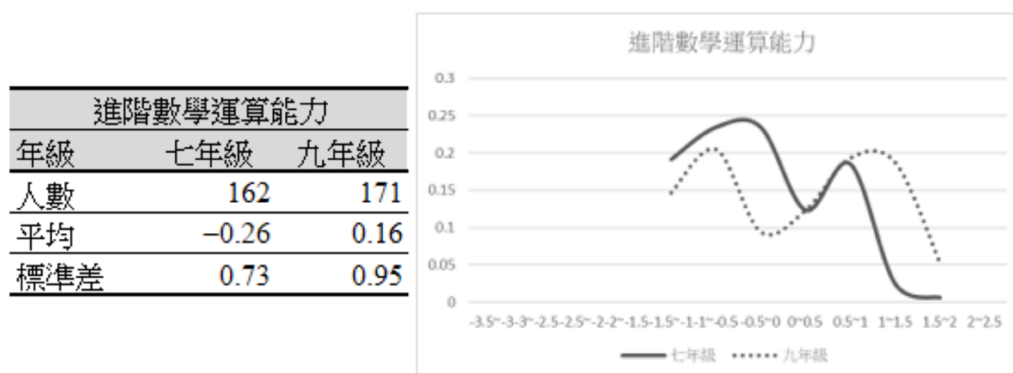


圖10 七年級與九年級追蹤測驗學生進階數學運算能力分布圖

為探討平行複本間可能造成的測驗偏誤，在七年級與九年級的追蹤測驗，有55人採用固定的相同題本，研究者繪製之追蹤測驗受測者能力分布圖如圖11，七年級學生之進階運算平均能力值為-0.49，九年級學生之進階運算平均能力值為0.00，標準差分別為0.80與0.71，與平行複本間的追蹤測驗結果相同，亦有隨著年級增長而能力值增長，而能力分布亦均逐年向右移，人數也有增加的趨勢，表示學生數學運算能力隨著年級增長而增加。由於每一種難度的數學運算題本間均有三個平行複本，因此，有利於教師使用同一階段難度的題本，比較教學前後數學運算能力的變化，追蹤學生在不同年級間的數學運算能力成長變化的情形。

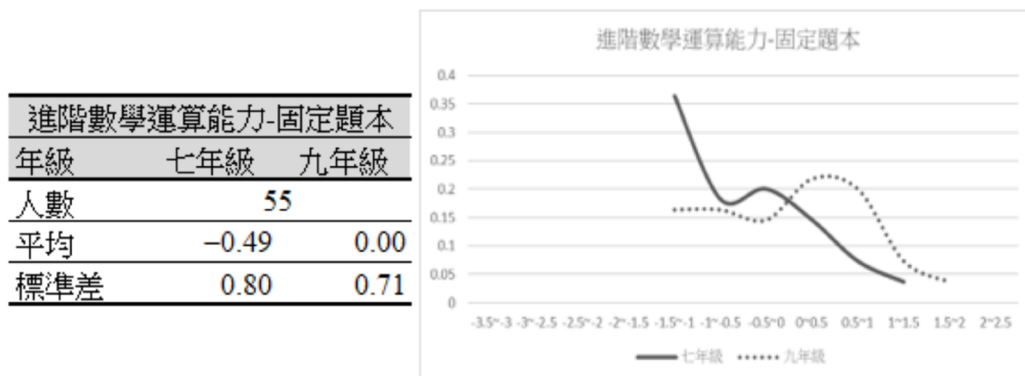


圖11 七年級與九年級追蹤測驗採固定題本學生進階數學運算能力分布圖

## 五、數學運算能力測驗的效度表現

為提供建構效度之證據，本研究以數學詞彙測驗為參照測驗，資料來源來自適用於小學三年級至國中八年級的學生使用之數學詞彙測驗（吳昭容、曾建銘、陳柏熹，2020），此兩份測驗為同時發展之測驗，取之樣本為同時實施數學運算能力測驗與數學詞彙測驗的學生表現。表5為數學運算能力測驗與數學詞彙測驗的相關係數，可以發現數學運算能力測驗與數學詞彙測驗的關聯性隨年級有所變化，小學三、四年級學生之初階數學運算能力與其數學詞彙能力的相關係數為 .35，小學五、六年級學生之中階數學運算能力與其數學詞彙能力的相關係數為

.42，而國中學生之進階數學運算能力與其數學詞彙能力的相關係數則為.62，表示隨著年級的增長，數學運算能力與數學詞彙能力表現的關係越來越強。

表5

數學運算與數學詞彙測驗之相關係數

相關係數	初階數學運算能力	中階數學運算能力	進階數學運算能力
數學詞彙能力	0.35**	0.42**	0.62**
	( <i>n</i> = 426)	( <i>n</i> = 169)	( <i>n</i> = 351)

註：\*\*  $p < .01$

## 肆、結論與建議

### 一、研究成果與發現應用

本研究使用小學三年級至國中九年級的數學運算能力測驗，應用MRCMLM模式，採同時估計法，利用共同題垂直與水平等化三個難度階段各三個平行複本的數學運算能力測驗題本，進行試題參數與能力估計，藉以減少測驗連結間所產生的試題參數估計誤差，獲得較佳的測驗精準度，由試題信效度資料顯示此測驗具有可接受的品質。由於每一個難度階段的數學運算能力測驗均有三個平行複本，初階數學運算題本與進階數學運算題本共同題所連接之題數達該題本總題數五分之一，水平等化結果良好，有利於教師比較教學介入前後的數學運算能力變化。此三個難度階段的數學運算能力測驗雖可以藉由垂直等化，進行測驗連結，但垂直等化之共同題題數較少，若使用此測驗追蹤學生在不同階段的數學運算能力之成長變化，如初階數學運算能力至中階數學運算能力的變化，需要更為謹慎。

在不同階段之數學運算能力，以MRCMLM題內多向度模式進行能力累加分析是適配的，每一個階段的數學運算能力認知處理方式歷程皆不同，數學運算能力都經過了質的改變，而產生新的數學運算能力，而這些數學運算能力，均建構在先前所積累的數學運算能力之上，使得下

一個階段的數學運算能力有累加前面階段的數學運算能力。舉例來說，在計算 $38 + 62 = ?$ 這道試題時，小學三、四年級的學生僅學習兩步驟基礎數學運算，此時學生的作答方式，會採用進位的兩步驟方式做計算，但到了小學五、六年級時，除了在學校的數學學科學習，生活中的生活技能，購物消費也常用100以內的數學運算，此時，學生的作答方式可能直接提取大量反覆練習的運算記憶答案，或是生活中購物消費所需的基本運算答案。由此可知，這兩種能力是截然不同的。本研究發現中階數學運算能力的發展為內化初階數學運算能力所擴張的一種新能力；同樣的，進階數學運算能力的發展也為內化初階數學運算能力與中階數學運算能力所擴張形成的一種新能力。

本研究所使用之初階數學運算題本適用於小學三、四年級，內容涵蓋：基本四則運算的算則、以加減與除、連乘、連除、乘除為主的兩步驟基礎運算；中階數學運算題本適用於小學五、六年級，內容涵蓋：四則混合運算含分配律與結合律的三步驟以上運算與併式算則、簡易的比例關係；進階數學運算題本適用於國中七至九年級，內容涵蓋：負數、根式、指數、簡單的函數、不等式、方程式、比例式與多項式的運算，試題內容皆符合各年級學習之內容。

由相同難度階段的數學運算題本橫斷資料分析數學運算能力的發展，可以發現學生在每一個階段的數學運算能力均隨著年級增長而增加，各年級之數學運算能力分布曲線亦逐年右移，學生經過學習後均達能力提升之結果。而由相同難度階段的數學運算題本縱貫追蹤資料中，可以發現數學運算能力的發展，無論是否採用固定題本，其成長結果為隨著年級而增長，而能力分布亦均逐年向右移，人數也有增加的趨勢，表示學生數學運算能力隨著年級增長而增加。唯在跨不同階段的比較與推估，因測驗垂直等化效果不彰，須審慎推估比較。

此外，採用MRCMLM模式估計的多向度數學運算能力與數學詞彙能力的關聯性隨年級越長而增加，此一結果與吳昭容等人（2018）的研究結果類似，雖然該研究對數學詞彙能力的估計方式與本文不同，但兩者都顯現學生在小學三、四年級時，數學運算能力與數學詞彙能力的相關不到 .40，學生在小學五、六年級時，數學運算能力與數學詞彙能力相關約 .40 ~ .60之間，而學生在國中時，數學運算能力與數學詞彙能力

相關高於 .60，這顯示學生掌握國小三、四年級的數學運算知識與技能和數學詞彙能力有一定的關聯性，而且此一關聯性隨著數學運算知識與技能的複雜度增加，與數學詞彙能力有更為密切的關係，數學詞彙能力越高者數學運算能力越佳，反之亦可以說，數學運算能力越好者掌握數學詞彙能力的狀態也越好。

## 二、未來研究建議

小學五、六年級的數學運算相關課程內容相較於小學三、四年級的課程內容較為廣泛，而國中階段的數學運算相關課程內容涵蓋範圍又增加了代數的課程內容；若採取20分鐘的測驗時間、總題數30題的題本作施測，將難以全面測量到完整的數學運算能力。建議在未來可擴增測驗題庫，並針對不同階段的數學運算題本增加測驗時間與測驗題長。

本研究所使用之數學運算能力測驗，在不同難度階段題本間的垂直等化共同題題數較少，僅有一本題本在中階數學運算題本與進階數學運算題本以2題共同題進行連結，其餘題本在跨階段能力的題本均只使用1題共同題做連結。若是可以提升垂直等化的共同題題數，則可以有效降低測驗連結時的試題參數估計誤差，在能力估計的精準上更為提升，建議採用初階數學運算題本與進階數學運算題本的水平等化設計標準，在每個跨階段題本中放入至少6題共同題，才能讓該題本與前後階段題本有較佳的等化效果。

在追蹤測驗的樣本分布方面，本研究僅針對三群學生，分別在四年級與六年級、六年級與八年級以及七年級與九年級進行追蹤測驗，建議未來可以增加樣本多樣性，得以更精確的了解數學運算能力在不同年級間變化情形。

本研究僅探討數學運算能力與數學詞彙能力間的相關程度，但在國中階段僅有九年級學生同時進行上述2個能力測驗，因此，在國中階段之數學運算能力與數學詞彙能力間的相關，不宜擴展至七、八年級學生，未來需要有七、八年級學生同時進行此2個能力測驗，得以更完整地建立國中學生數學運算能力測驗與數學詞彙能力之間的因果模式。

## 誌謝

本研究承蒙科技部專題研究計畫MOST 105-2511-S-027-002-與MOST 106-2511-S-027-002-經費支持，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 and psychology*, 27(4), 637-694.】
- 王暄博、郭伯臣、呂玉如（2013）。大型測驗等化群體不變性之探究：以2007年臺灣學生學習成就評量資料庫國中二年級數學科為例。**測驗學刊**，60(3)，489-518。
- 【Wang, H.-P., Kuo, B.-C., & Lu, Y.-J. (2013). Exploring the population invariance of equating in the large-scale assessments: Using the Taiwan assessment of student achievement as an example. *Psychological Testing*, 60(3), 489-518.】
- 余明杰、李茂能、楊德清（2016）。國小四年級學童數常識三階段診斷測驗工具的發展與應用之研究。**科學教育學刊**，24(1)，89-114。doi:10.6173/CJSE.2016.2401.04
- 【Yu, M.-C., Li, M.-N., & Yang, D.-C. (2016). The development and application of number sense three-tier test for fourth graders. *Chinese Journal of Science Education*, 24(1), 89-114. doi:10.6173/CJSE.2016.2401.04】
- 吳昭容、曾建銘、陳柏熹（2020）。國民中小學數學詞彙知識測驗。臺北市：心理。
- 【Wu, C.-J., Cheng, C.-M., & Chen, P.-H. (2020). *Mathematical vocabulary ability test for elementary and junior high school education*. Taipei, Taiwan: Psychological Publishing.】

吳昭容、曾建銘、鄭鈴華、陳柏熹、吳宜玲（2018）。領域特定詞彙知識的測量：三至八年級學生數學詞彙能力。**教育研究與發展期刊**，**14**(4)，1-40。doi:10.3966/181665042018121404001

【Wu, C.-J., Cheng, C.-M., Cheng, C.-H., Chen, P.-H., & Wu, Y.-L. (2018). The measurement of domain-specific vocabulary knowledge: The mathematical vocabulary ability of third to eighth grade students. *Journal of Educational Research and Development*, 14(4), 1-40. doi:10.3966/181665042018121404001】

吳慧珉、郭伯臣、許天維、陳婉寧（2015）。以可能值方法為基礎之多向度能力值垂直等化探究。**測驗學刊**，**62**(2)，95-126。

【Wu, H.-M., Kuo, B.-C., Sheu, T.-W., & Chen, W.-N. (2015). The research in estimating multidimensional traits under vertical equating based on plausible value method. *Psychological Testing*, 62(2), 95-126.】

洪儷瑜、連文宏。（2019）。國小學童計算能力之測量、發展軌跡與預測路徑。**測驗學刊**，**66**(1)，51-82。

【Hung, L.-Y., & Lien, W.-H. (2019). The measurement and developmental trajectories of arithmetic competences and prediction routes of arithmetic competences in elementary students. *Psychological Testing*, 66(1), 51-82.】

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。

【Ministry of Education. (2008). *Mathematics learning areas of General Guidelines of Grades 1-9 Curriculum for Elementary and Junior High School Education*. Taipei, Taiwan: Author.】

教育部（2018）。十二年國民基本教育課程綱要國民中小學暨普通型高級中等學校：數學領域。取自<https://cirn.moe.edu.tw/Upload/file/27405/61868.pdf>

【Ministry of Education. (2018). *Curriculum Guidelines of 12-Year Basic Education for Elementary, Junior and Senior High School Education: Mathematics areas*. Retrieved from <https://cirn.moe.edu.tw/Upload/file/27405/61868.pdf>】

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=8b23f290682c20dc56ce69c7f0239df8](https://quality.data.gov.tw/dq_download_xlsx.php?nid=6240&md5_url=8b23f290682c20dc56ce69c7f0239df8)

【Department of Statistics, MOE. (2016a). *2015 elementary school information data*. Retrieved from [https://quality.data.gov.tw/dq\\_download\\_xlsx.php?nid=6240&md5\\_url=8b23f290682c20dc56ce69c7f0239df8](https://quality.data.gov.tw/dq_download_xlsx.php?nid=6240&md5_url=8b23f290682c20dc56ce69c7f0239df8)】

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](https://quality.data.gov.tw/dq_download_xlsx.php?nid=6239&md5_url=8fc3a80422ac4a707d86faa6764d21f2)

【Department of Statistics, MOE. (2016b). *2015 junior high school information data*. Retrieved from [https://quality.data.gov.tw/dq\\_download\\_xlsx.php?nid=6239&md5\\_url=8fc3a80422ac4a707d86faa6764d21f2](https://quality.data.gov.tw/dq_download_xlsx.php?nid=6239&md5_url=8fc3a80422ac4a707d86faa6764d21f2)】

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)，89-90。

【Kuo, B.-C., & Wang, H.-P. (2008). A simultaneous vertical and horizontal equating of large-scale assessments. *Journal of Educational Research and Development*, 4(4), 89-90.】

陳柏熹（2006）。能力估計方法對多向度電腦化適性測驗測量精準度的影響。**教育心理學報**，38(2)，195-211。

【Chen, P.-H. (2006). The influences of the ability estimation methods on the measurement accuracy in multidimensional computerized adaptive testing. *Bulletin of Educational Psychology*, 38(2), 195-211.】

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。

【Chen, H.-W. (2011). A comparison of the differences in linked score properties for vertical scaling obtained using six linking methods. *Journal of Educational Research and Development*, 58(4), 559-583.】

曾芬蘭、林奕宏、邱佳民。（2017）。監控評分者效果的Yes/No Angoff標準設定法之效度檢核：以國中教育會考數學科為例。**測驗學刊**，



64(4), 403-432。

【Tseng, F.-L., Lin, Y.-H., & Chiou, J.-M. (2017). Validation of the rater-effects-monitored Yes/No Angoff standard-setting method: Using the Taiwan Comprehensive Assessment Program for Junior High School Students math exam as an example. *Psychological Testing*, 64(4), 403-432.】

曾建銘（2020）。三～八年級資料與可能性能力測驗的發展及信效度分析。測驗學刊，67(4)，301-331。

【Cheng, C.-M. (2020). The development and evaluation of data and chance ability test for third to eighth grade students. *Psychological Testing*, 67(4), 301-331.】

曾建銘、吳慧珉、趙珮晴（2019）。臺灣學生數學學習內容表現之探討：來自TASA和TIMSS的跨資料庫比較。測驗學刊，66(1)，27-50。

【Cheng, C.-M., Wu, H.-M., & Chao, P.-C. (2019). Taiwan students' mathematics performance: A comparative analysis of TASA and TIMSS database. *Psychological Testing*, 66(1), 27-50.】

楊心怡（未出版）。數學運算能力測驗。

【Yang, H.-I. (Unpublished). *Mathematical arithmetic ability test*.】

Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1-23. doi:10.1177/0146621697211001

Adams, R. J., Wu, M. L., & Wilson, M. R. (2012). ACER ConQuest: Generalised item response modelling software [Computer software]. Victoria, Australian: Australian Council for Educational Research.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 1036-1060. doi:10.1037/0033-295X.111.4.1036

Baroody, A. (2006). Why children have difficulties mastering the basic number combinations and how to help them. *Teaching Children Mathematics*, 13, 22-31. doi:10.5951/TCM.13.1.0022

- Beishuizen, M. (1993). Mental strategies and materials or models for addition and subtraction up to 100 in Dutch second graders. *Journal for Research in Mathematics Education*, 24, 295-323. doi:10.2307/749464
- Beishuizen, M., van Putten, C. M., & van Mulken, F. (1997). Mental arithmetic and strategy use with indirect number problems up to one hundred. *Learning and Instruction*, 7, 87-106. doi:10.1016/S0959-4752(96)00012-6
- 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
- Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). New York, NY: Routledge. doi:10.4324/9780429030499
- Cowan, R., Donlan, C., Shepherd, D. L., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology*, 103(4), 786-803. doi:10.1037/a0024556
- Dowker, A. (2019). *Individual differences in arithmetic: Implications for psychology, neuroscience and education* (2nd ed.). New York, NY: Routledge.
- Fuchs, L. S., Powell, S. R., Cirino, P. T., Schumacher, R. F., Marrin, S., Hamlett, C. L., ...Changas, P. C. (2014). Does calculation or word-problem instruction provide a stronger route to pre-algebraic knowledge? *Journal of Educational Psychology*, 106, 990-1006. doi:10.1037/a0036793
- Geary, D. C. (2011). Cognitive predictors of individual differences in achievement growth in mathematics: A five year longitudinal study. *Developmental Psychology*, 47, 1539-1552. doi:10.1037/a0025510
- Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS One*, 8(1), e54651. doi:10.1371/journal.pone.0054651

- Geary, D. C., Nicholas, A., Li, Y., & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *Journal of Educational Psychology*, 109, 680-693. doi:10.1037/edu0000159
- Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology*, 77(4), 337-353.
- Gilbert, J. K., & Fuchs, L. S. (2017). Bivariate developmental relations between calculations and wordproblems: A latent change approach. *Contemporary Educational Psychology*, 51, 83-98. doi:10.1016/j.cedpsych.2017.06.008
- 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. doi:10.1177/0146621602026001001
- 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. doi:10.1006/jecp.2000.2594
- Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology*, 96, 284-309. doi:10.1016/j.jecp.2006.09.001
- Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2008). Development of number combination skill in the early school years: When do fingers help? *Developmental Science*, 11, 662-668. doi:10.1111/j.1467-7687.2008.00715.x
- Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, 45, 850-867. doi:10.1037/a0014939

- Kelderman, H. (1996). Multidimensional Rasch models for partial-credit scoring. *Applied Psychological Measurement*, 20, 155-168. doi:10.1177/014662169602000205
- 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. doi:10.1177/01466216980222003
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking* (2nd ed.). New York, NY: Springer. doi:10.1007/978-1-4757-4310-4
- Korpipää, H., Koponen, T., Aro, M., Tolvanen, A., Aunola, K., Poikkeus, A.-M., ...Nurm, J.-E. (2017). Covariation between reading and arithmetic skills from grade 1 to grade 7. *Contemporary Educational Psychology*, 51, 131-140. doi:10.1016/j.cedpsych.2017.06.005
- Lei, P. W., & Zhao, Y. (2012). Effects of vertical scaling methods on linear growth estimation. *Applied Psychological Measurement*, 36(1), 21-39. doi:10.1177/0146621611425171
- Lin, X., & Powell, S. R. (2021). The roles of initial mathematics, reading, and cognitive skills in subsequent mathematics performance: A meta-analytic structural equation modeling approach. *Review of Educational Research*, 92(2), 288-325. doi:10.3102/00346543211054576
- Livingston, S. A. (2004). *Equating test scores* (without IRT). Princeton, NJ: Educational Testing Service.
- Lucangeli, D., Tressoldi, P. E., Bendotti, M., Bonanomi, M., & Siegel, L. (2003). Effective strategies for mental and written arithmetic calculation from the third to the fifth grade. *Educational Psychology*, 23, 507-521. doi:10.1080/0144341032000123769
- Meyers, J. L., Miller, G. E., & Way, W. D., (2008). Item position and item difficulty change in an IRT-based common item equating design. *Applied Measurement in Education*, 22(1), 38-60. doi:10.1080/08957340802558342
- National Assessment Governing Board. (2003). The 1990-2003 mathematics framework. *National Assessment of Educational Progress*.

- Retrieved from <https://nces.ed.gov/nationsreportcard/mathematics/previousframework.aspx>
- Nelson, P. M., Parker, D. C., & Zaslofsky, A. F. (2016). The relative value of growth in math fact skills across late elementary and middle school. *Assessment for Effective Intervention, 41*(3), 184-192. doi:10.1177/1534508416634613
- Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*(4), 587-597. doi:10.1007/s10648-015-9302-x
- Russell, S. J. (2000). Developing computational fluency with whole numbers in the elementary grades. *The New England Math Journal, 32*(2), 40-54. doi:10.5951/TCM.7.3.0154
- Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., ...Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science, 23*(7), 691-697. doi:10.1177/0956797612440101
- Sisco-Taylor, D., Fung, W., & Swanson, H. L. (2015). Do curriculum-based measures predict performance on word-problemsolving measures? *Assessment for Effective Intervention, 40*, 131-142. doi:10.1177/1534508414556504
- Watts, T. W., Duncan, G. J., Chen, M., Claessens, A., Davis-Kean, P. E., Duckworth, ...Susperreguy, M. I. (2015). The role of mediators in the development of longitudinal mathematics achievement associations. *Child Development, 86*(6), 1892-1907. doi:10.1111/cdev.12416
- Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352-360. doi:10.3102/0013189X14553660
- Wong, T. T. Y. (2017). The unique and shared contributions of arithmetic

- operation understanding and numerical magnitude representation to children's mathematics achievement. *Journal of Experimental Child Psychology*, 164, 68-86. doi:10.1016/j.jecp.2017.07.007
- Wong, T. Y. T. (2021). Components of mathematical competence in middle childhood. *Child Development Perspectives*, 15(1), 18-23. doi:10.1111/cdep.12394
- Wu, M., Adams, R. J., & Wilson, M. R. (1998). *ACER ConQuest*. Victoria, Australian: Australian Council for Educational Research.

# **Multidimensional Test Equating the Arithmetical Abilities of Third to Ninth Grade Students**

Yi-Ling Wu\* Hsin-I Yung\*\* Chao-Jung Wu\*\*\*  
Po-Hsi Chen\*\*\*\*

## **Abstract**

Arithmetical ability is a fundamental part of learning and a necessary skill for life. Arithmetical abilities in elementary school are mainly about the connection between numbers and operation symbols, while those in middle school include mathematical computation content knowledge, algebra, and functions with a wide range of content and scope. Therefore, the development of arithmetical abilities is a subject worth exploring. In this study, using a Mathematical Arithmetic Ability Test, 4,991 cross-sectional and 511 longitudinal data were analyzed. These tests equated estimate item parameters and student's arithmetical abilities using multidimensional analysis. The subjects ranged from third graders of elementary school to the ninth graders of junior high school. Multidimensional Random-Coefficients Multinomial Logit Model (MRCMLM) was used to compare the development of arithmetical abilities at different stages. The results revealed that the cross-sectional data grew year by year, while the longitudinal data grew only in the repeated measurement using the same booklet. The mathematical vocabulary test was also used as a reference test to show that the relationship between arithmetical abilities and mathematical vocabulary abilities increases as the grade level advances.

**Keywords:** multidimensional random coefficients multinomial logit model, test equating, item response theory, arithmetical abilities



---

DOI : 10.6869/THJER.202112\_38(2).0004

Received: December 23, 2021; Modified: February 10, 2022; Accepted: February 17, 2022

\* Yi-Ling Wu, Ph. D. Candidate, Department of Educational Psychology and Counseling, National Taiwan Normal University, E-mail: irene54510@gmail.com

\*\* Hsin-I Yung, Associate Professor, Graduate Institute of Technological and Vocation Education, National Taipei University of Technology, E-mail: hcy103@ntut.edu.tw

\*\*\* Chao-Jung Wu, Professor, Department of Educational Psychology and Counseling, National Taiwan Normal University, E-mail: cjwu@ntnu.edu.tw

\*\*\*\* Po-Hsi Chen (corresponding author), Professor, Department of Educational Psychology and Counseling, National Taiwan Normal University, E-mail: lorenz1020@gmail.com