

# 數學素養調查的實施現況與國際借鏡

吳昭容

國立臺灣師範大學教育心理與輔導學系教授

鄭鈐華

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副教授

許志農

國立臺灣師範大學數學系教授

## 一、前言

隨著全球就業市場的快速變遷和資訊科技的蓬勃發展，基本的知識與技術、良好的學習態度、批判性的思考能力，以及與他人溝通的技能等，皆為因應環境變遷、面臨未來工作的重要素養。我國行政院在2011年9月核定「十二年國民基本教育實施計畫」，為了建立學生能力的監測機制，期能透過實證佐證學生能力的變化，教育部將「提升國民素養實施方案」列為「十二年國民基本教育實施計畫」的29項配套方案之一，自2012年至2017年執行六年計畫，以建構素養的核心概念與指標，並進行18歲國民素養的調查。此一方案旨在對我國即將完成國民教育的人力素質進行長期的系統調查（教育部綜合規劃司，2013），關照與檢視十二年國民基本教育下的國民素養。

在「提升國民素養實施方案」下，「國民素養」指的是重要的、每位學生都有權利獲得且必須獲得的能力。目前該方案訂有五項素養向度，包含語文素養、數學素養、科學素養、數位素養，

以及教養/美感素養。本文作者為2014年度數學素養計畫案的執行成員，目的在承接2013年數學素養團隊所建構的指標架構，規劃與執行實際的調查方式與資料分析。本文聚焦在數學素養，就數學素養的定義與內涵、國際上相關調查的執行方式、臺灣推動的概況與問題，並從國外執行經驗借鏡，做為未來數學素養調查執行方式的參考。

## 二、數學素養的定義與內涵

國際經濟合作暨發展組織（Organisation for Economic Cooperation and Development, OECD）所推動的國際學生評量計畫（Programme for International Student Assessment, PISA），其數學素養是指，學生能意識到數學在生活世界的意義，在情境中做出有效的數學判斷，以一個積極、關懷、反省的公民態度，致力於解決生活上出現的數學問題。因此PISA 2012將數學素養定義為：個體在不同情境脈絡中，形成、應用以及詮釋數學的能力，包含以數學概念、程序、事實、推理，以及運用工具來描述、解釋並預測數學現象。此一

定義涵納四個數學內容領域（變化與關係、空間與形狀、數量、不確定性與數據）、涉及三種認知歷程（形成問題情境；應用數學概念、事實、程序以及推理；詮釋與評鑑結果），以及兼顧四類情境脈絡（個人、社會、職業、科學）。簡單來說，數學素養使個體能辨識數學在日常生活中的意義，並且以一個公民該有的態度，做出具有建設性、投入性和反思能力的判斷和決策（OECD, 2010）。

第三次國際數學與科學教育成就研究（Third International Mathematics and Science Study, TIMSS）在1999年採用「數學與科學素養」這個概念時，關注的是能在日常生活中流暢運用數學與科學的能力（Wilkins, Zembylas, & Travers, 2002）。該調查的三個重要基礎是內容（學校教了什麼）、脈絡（數學內容知識如何出現在真實生活情境）、歷程（學生是否能邏輯且合理地解決真實生活中的問題）。該次調查的數學素養共38題，分成數感、代數感/資料、測量與估計三個領域。

數學素養的文獻通常強調其功能性及認知分析。從功能性來說，Jablonka（2003）主張，在日常生活周遭已形成一種與數學息息相關的文化，而現代公民必須具備能對此社會文化做出批判與評價的能力，此即為數學素養。若以認知分析觀點來看，Yore、Pimm和Tuan（2007）提出數學素養即為「認知與後設認知能力」、「數學思考和量化推理」、「情意傾向與心智習慣」、「數學語言」、和「資訊溝通科技」等五種基礎識感（fundamental senses）。

整合前述對於數學素養的觀點，教育部提升國民素養專案辦公室（2013）於「教育部提升國民素養專案計畫報告書」中指出，數學素養的定義與內涵為：個人的數學能力與態度，使其在學習、生活、與職業生涯的情境脈絡中面臨問題時，能辨識問題與數學的關聯，從而根據數學知識、運用數學技能、並藉由適當工具與資訊，去描述、模擬、解釋與預測各種現象，發揮數學思維方式的特長，做出理性反思與判斷，並在解決問題的歷程中，能有效地與他人溝通觀點。

### 三、國際現行數學素養調查方式

目前國際上有一些團體執行數學素養相關的大型調查，底下針對國際學生能力評量計畫（Programme for International Student Assessment, PISA）、國際數學與科學教育成就趨勢調查研究（Trends in International Mathematics and Science Study, TIMSS）、美國全國教育進展評估（The National Assessment of Educational Progress, NAEP）及全港性系統評估（Territory-wide System Assessment, TSA）等摘要重要資訊，作為了解國民素養計畫中數學素養實施方式之參照。雖然後三個調查系統未必直指素養為其評量目標，然其部分試題特性與規模仍可供參考。

#### （一）國際學生能力評量計畫（PISA）

OECD為回應會員國，能定期掌握到學生知識和技能的可靠資料，及檢視教育制度品質的需求，在90年代中期開始推動PISA（OECD, 2013a）。PISA於

2000年實施第一次調查，之後每三年舉辦一次。調查的核心領域為閱讀、數學（含問題解決）和科學素養，每次以一個核心領域為主，另二領域為副，對象為15歲的學生。以PISA2012而言，65個經濟體、約51萬名學生參與（OECD, 2014）。

PISA強調素養題目必須能引起學生的興趣與符應生活經驗，才能真正評量出學生是否具積極性、投入性及反思能力的現代公民（OECD, 2013a）。因此數學素養题目的脈絡情境分成個人、職業、社會、科學四類，盡可能包含各種學生感興趣的情境（臺灣PISA國家研究中心，2012）。個人情境如準備食物、購物、遊戲、運動、個人健康、理財、個人交通等與自我、家庭或同儕相關的議題；職業情境包括建築物的測量、成本和材料的訂購、工資/會計、

調度/庫存、設計/架構、與工作相關的決策等職場相關的議題；社會情境像是投票制度、公共交通、公眾政策、人口統計、廣告等與公眾事務相關的議題；科學情境則有如氣候、生態、醫學、太空科學、遺傳學、測量等數學可應用於自然界、科學與科技的議題（林素微，2013）。

由於PISA旨在提供系統（或國家）層級的評量回饋，而非評比學生個人的表現，因此，每個學生沒有必要接受完全相同的測驗題組。題本共有13式，學生僅作其中一式，時間約2小時。試題使用三種題型，分別為選擇式反應題（例如選擇題、多重是非題）、封閉式建構反應題（可由電腦判讀、編碼），與開放式建構反應題（需有解題步驟或說明理由，答案必須人工計分），這三種題型在一式題本上的題數約略相等

## 調味醬

PM924Q02 - 0 1 9

**問題 2：調味醬**

假設你正在調製沙拉醬。

這裡有一個調製 100 毫升 (mL) 沙拉醬的食譜。

沙拉油：	60 mL
醋：	30 mL
醬油：	10 mL

要調製 150 mL 的沙拉醬需要多少毫升 (mL) 的沙拉油？

答案：..... mL

圖1 PISA2012公開數學「調味醬」樣本題

資料來源：國立臺南大學（2012）。PISA 2012數學樣本試題【中文版】（頁41）。

臺南市：臺灣 PISA國家研究中心。取自[http://pisa.nutn.edu.tw/download/sample\\_papers/PISA\\_2012\\_items\\_for\\_release\\_Chinese\\_final.pdf](http://pisa.nutn.edu.tw/download/sample_papers/PISA_2012_items_for_release_Chinese_final.pdf)

(OECD, 2013a)。

PISA的學生問卷旨在瞭解學生在PISA的表現程度和各種因素（如移民、性別和學生的社經背景，以及學生對學校的態度和他們的學習方法）之間的關聯。問題多採李克特氏量表（Likert scale），內容和學生背景相關，例如家庭背景、學生的數學學習態度、解題經驗、和電腦使用情形等資訊。學生問卷共3式，每位學生僅作其中一式，約30分鐘（OECD, 2013b）。

以PISA 2012數學樣本公開題「調味醬」（見圖1）為例，此題的背景脈絡屬個人的情境，內容領域則為數量的範疇，題旨在於能在日常生活情境中應用比的概念，計算食譜中調味醬的需求量。題目中沒有提及「比」或「比例」的語詞，學生必須以所學習過的數學概念來解決調製沙拉醬的生活性問題。

直至2006年為止，PISA都使用紙筆測驗。但自2009開始，部分國家陸續實施電腦化評量，預計2015年將全面電腦化。電腦化評量需要學生具備基本的資訊科技（Information and Communication Technology, ICT）能力，如鍵盤和滑鼠的使用。對於現今的數位時代而言，具備高度的ICT素養是重要的，而軟體工具在問題解決上的使用亦是普遍的，並且提供了有力的幫助（OECD, 2013a）。

電腦化評量的優點在於豐富了問題表徵，如可移動物件、3D物件的呈現與旋轉、可與問題情境更為彈性的互動等，而且關心「資訊豐富（information-rich）」的問題，包括需要找出或評估網際網路或社群網站上的訊息、瀏覽不熟悉的網頁，和決定哪些資訊和題目有

關。這使得學生能回答更多非口語的題目，並且評量學生解決生活中不熟悉的問題，或是跨課程領域的認知技巧（江芳盛、李懿芳，2009；OECD, 2013a）。

以PISA 2012釋出的數學電腦化題目「圍欄」（見圖2）為例，學生可以拖曳電腦螢幕中的白點，則畫面中長方形的大小及下方顯示長、寬、使用的圍欄長度、面積的數據，也會同步隨之改變。因此學生解題時不必使用筆算或心算，而是使用電腦移動物件，並觀察數字的變化關係來回答問題。

## （二）國際數學與科學教育成就趨勢調查（TIMSS）

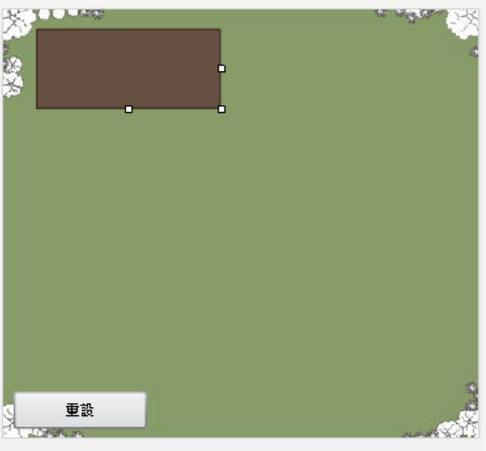
國際教育學習成就調查委員會（International Association for the Evaluation of Educational Achievement, IEA）自1990年開始推動國際性的大型成就評量，1999年舉辦的「第三次國際數學與科學教育成就研究後續調查（Third International Mathematics and Science Study, TIMSS-Repeat）」，對象為13歲學生（相當於我國的八年級學生），共有包括臺灣在內的38個國家參加（Gonzalez & Miles, 2001）。IEA自2003年起，固定每四年辦理一次調查，並正式定名為「國際數學與科學教育成就趨勢調查（TIMSS）」，對象鎖定為四年級和八年級的學生。

TIMSS是課程導向的研究，評比的科目只聚焦於數學及科學兩個學科，目的在於評量各國學生在數學與科學領域上學習成就的發展趨勢，並分析比較各國的文化背景、教育制度的差異，其與成就間的相關性，內容涵蓋：數學課程、

**圍欄**

小麗是一位庭院設計師，她正在設計一個有圍欄的長方形花園。她可用的圍欄長度是100單位。拖曳圖中的白點(白色小正方形)，你可以改變花園的大小。下表顯示每次花園形狀變更後，各項的測量值。

長方形	
35	長
15	寬
100	使用的圍欄長度
525	面積



**問題一：圍籬**

下列有關圍籬長度和花園面積的陳述是否正確？請為每個陳述點選「對」或「錯」。

如果小麗想增加花園的面積，使新的花園面積比原來設計的525個平方單位大，那麼她必須購買更多的圍籬。  對  錯

如果小麗知道她花園的面積，則不論如何建造此花園，所使用的籬笆長度永遠不會變。  對  錯

圖2 PISA2012所釋出的數學電腦「圍籬」樣本測試題

資料來源：臺灣 PISA國家研究中心（2012）。電腦化樣本試題範例：數學樣本試題範例－中文版。取自<http://pisa.nutn.edu.tw/cm/CM012/www/index.html>

25

小珮和小勇是學生會主席的候選人，選舉的結果如下：

小珮 80%

小勇 20%

隨機找一個學生來問，這位學生把票投給小珮的可能性為何？

(A) 這位學生一定把票投給小珮。

(B) 這位學生很可能把票投給小珮。

(C) 這位學生不太可能把票投給小珮。

(D) 這位學生一定沒有把票投給小珮。

圖3 TIMSS 2011數學八年級公開樣本題

資料來源：國立臺灣師範大學科學教育中心（2011）。TIMSS 2011數學八年級\_題本\_5（頁19）。取自[http://www.sec.ntnu.edu.tw/timss2011/downloads/T11\\_G8\\_M06.pdf](http://www.sec.ntnu.edu.tw/timss2011/downloads/T11_G8_M06.pdf)

家庭支持對學生的影響、學校資源和學習氣氛、教師備課和學生參與教室課程的程度，並評估學生是否掌握了社會所需的知識與技能（Mullis, Martin, Foy, & Arora, 2012）。

TIMSS採用紙筆測驗，數學和科學的測驗時間各為45分鐘。因TIMSS的題目眾多，因此將題目分配到區塊中，然後再將區塊組合成數個題本。以TIMSS 2011為例，總共有28個區塊，數學與科學分別各有14個區塊，每個題本由4個區塊組成（數學與科學各2個區塊，兩兩題本有共同題），共組合成14個題本（Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009）。題型包括選擇題與建構反應題兩類，評量架構以2011年八年級的數學科為例，分為內容領域（數、代數、幾何、資料分析與機率）與認知領域（知道事實、應用、推理）兩個向度。

由於TIMSS的調查目的，在於評量學生於數學與科學的學習成就，因此題目內容會控制在大部分國家課程的範圍之內。雖然TIMSS是成就測驗，但其題目與PISA一樣強調與生活的關連性。圖3以TIMSS 2011所釋出的數學八年級公開題為例，內容領域為資料分析與機率，認知領域則側重在知道機率的定義。

### （三）美國全國教育進展評估（NAEP）

NAEP是美國最大型的國家教育評量。自1973年起，NAEP定期蒐集學生數學成就的訊息（National Assessment Governing Board, 2012），實施的對象及架構等均由「國家評量指導委員會」（National Assessment Governing Board）決定。

NAEP數學測驗依據目的不同有兩個組成。一個測驗是「長期趨勢評估（Long-Term Trend Assessment）」；一個是「主要NAEP評估（Main NAEP Assessment）」。前者目的在於測量9、13、17歲學生在閱讀與數學成就的長期趨勢，每次都使用相同的基本設計。主要是在對學生的基本技能和定義的記憶進行測驗。採紙筆測驗，試題本包含三個15分鐘的問題區塊（U.S. Department of Education, 2013）。數學長期趨勢評估自2004年起，每四年做一次評量。最新一次報告在2012年。後者目的為透過評量4、8或12年級學生在閱讀、數學、科學、寫作、美國歷史、公民、地理及其它科目的教育成就，評量結果提供給聯邦政府、各州及學區作為擬訂及改革教育政策的參考。數學主要NAEP評估始於1990年，每兩年評估一次。

NAEP的數學題目測量五個內容領域，包括數字屬性和操作、測量、幾何、代數，以及資料分析、統計和機率。題型和PISA類似，有三類：選擇題、簡答式建構反應題，和申論式建構反應題。選擇題的施測時間約佔一半，另兩類建構反應題合計佔另一半的施測時間（National Assessment Governing Board, 2012）。

正如數學可以分為純數學和應用數學，NAEP也有這兩類的題目，有些題目涉及純粹的數學思想和概念，有些則以真實世界問題為背景來出題。例如圖4兩道12年級的NAEP題目，題旨都是「能運用三角函數求直角三角形的一股長」，但第一個題目是純數的，第二個題目則是真實世界問題的情境。

<p>Example Pair 2: Pure Mathematical Setting Grade 12 Measurement: Measuring physical attributes</p>	<p>Source: Modified NAEP Item Calculator Available</p>
<p>In the triangle below <math>\overline{AC} = 50</math> feet and <math>\angle C = 40^\circ</math>.</p>  <p>What is the length of <math>\overline{AB}</math> to the nearest foot?</p> <p>Answer: _____</p> <div style="border: 1px solid black; padding: 5px; display: inline-block;">Correct answer: 42 feet</div>	
<p>Example Pair 2: Contextual Mathematical Problem Grade 12 Measurement: Measuring physical attributes</p>	<p>Source: 2005 NAEP B3M12#15 Percent correct: 41% Calculator available</p>
<p>A cat lies crouched on level ground 50 feet away from the base of a tree. The cat can see a bird's nest directly above the base of the tree. The angle of elevation from the cat to the bird's nest is <math>40^\circ</math>. To the nearest foot, how far above the base of the tree is the bird's nest?</p> <p>A. 32 B. 38 C. 42 D. 60 E. 65</p> <div style="border: 1px solid black; padding: 5px; display: inline-block;">Correct answer: C</div>	

圖4 NAEP中純數背景的（上）和真實世界情境的（下）範例題

資料來源：Mathematics Framework for the 2013 National Assessment of Educational Progress (p.62), by National Assessment Governing Board, 2012, Washington, DC: U.S. Government Printing Office.

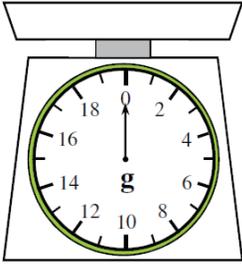
#### （四）全港性系統評估（TSA）

TSA是由香港教育局在2001年委託考試及評核局（以下簡稱「考評局」）發展與施行中、英、數三科的基本能力評估，目的在於讓香港政府及學校了解全香港整體學生於這三科的基本能力水準，藉此改善教與學，並讓政府為有需要的學校提供支援。TSA首先於2004年在小學三年級施行；2005年，小學三年級及六年級學生均須參與評估；至2006年迄今，已推展至中學三年級階段（香港

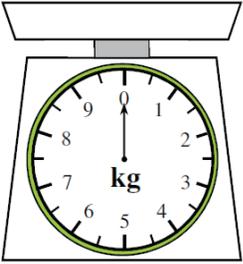
考試及評核局，2014a）。

TSA為了能提供充分資料回饋給政府及學校，因此評估內容涵蓋三科所臚列的基本能力，題數眾多；加上TSA旨在評估學校學生的整體表現，而非個別學生的能力，因此TSA與前述大型調查類似，採用多式題本，但有共同題的設計，讓每位學生只需作答其中一式，再經統計技術推估學生的表現（香港考試及評核局，2014a、2014b）。每式數學科的卷子均涵蓋「數與代數」、「度量、圖形與

Q10/M2



磅 A



磅 B

上圖顯示磅 A 和磅 B，美盈想找出一枝牙籤的重量。下列的方法中，哪個是最合適的？

- 美盈用磅 A 量度 100 枝牙籤的重量，然後把該重量除以 100。
- 美盈用磅 B 量度 100 枝牙籤的重量，然後把該重量除以 100。
- 美盈用磅 A 量度 2 枝牙籤的重量，然後把該重量除以 2。
- 美盈用磅 B 量度 2 枝牙籤的重量，然後把該重量除以 2。

圖5 2013年全港性系統評估數學科中學三年級公開題

資料來源：香港考試及評核局（2014b）。2013年全港性系統評估第一至第三學習階段中國語文科、英國語文科、數學科學生基本能力報告－數學科中學三年級（頁379）。取自[http://www.bca.hkeaa.edu.hk/web/TSA/zh/2013tsaReport/chi/Ch8c\\_S3\\_Math\\_TSA2013C.pdf](http://www.bca.hkeaa.edu.hk/web/TSA/zh/2013tsaReport/chi/Ch8c_S3_Math_TSA2013C.pdf)

空間」及「數據處理」三個範疇，題數約50題，作答時間限時65分鐘，學生無須填寫任何問卷（香港考試及評核局，2014c）。中學數學科的評估以紙筆形式進行，但可使用香港考評局核准之計算機（香港考試及評核局，2014d）。

這些題目是由各有關科目的專業人士（如考評局科目經理、教育局課程主任和資深教師）參照各科的課程和「基本能力」文件擬定的（香港考試及評核局，2014b，2014c）。題型則包含多種常見類型的題目，如選擇題、简答题（含填充、圈選、繪圖、排序等）、長答题（含繪圖、填表、文字解說、列式計算題、證明題等）。圖5為香港考試評核局公布的2013年TSA中學三年級的試題

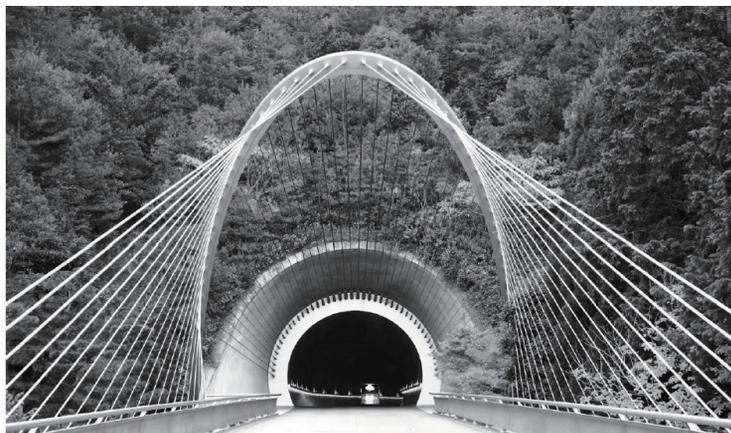
之一。該題的學習範疇為「度量、圖形與空間」，題旨在於評估學生就現實生活的量度，能選用適當的量度工具和策略。評核結果顯示，部分學生選擇了選項「C」，反映出學生大多能選用適當的工具，但卻忽視運用合適的策略（香港考試及評核局，2014b）。

#### 四、臺灣實施概況與問題

2014年數學素養計畫於五月在北、中、南、東14所高中職進行「試調查」，調查對象為31班的高二、高三班級，共1150位學生；並於十月配合國教院統一的取樣方式，在16所高中職進行「預調查」，調查的高三學生從各班抽樣，有效樣本共1172位。

### 圓弧形美術館隧道

下圖是某個美術館的入口。  
該館入口是圓弧造形的隧道，地面寬度估計為 8 公尺。工作人員站在路面的中點，測得圓弧形入口最高處也是 8 公尺。



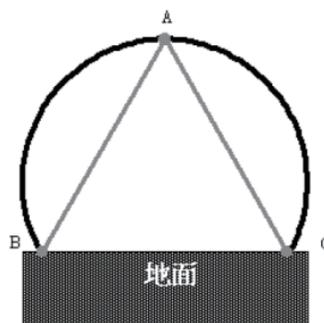
圖片來源：<http://bee1955.blogspot.tw/2012/05/miho.html>

圖6 「圓弧形美術館隧道」(公開題)的題幹文字敘述

#### Q1

該館週年慶，工作人員想在圓弧形入口的最高點 A 點，至地面兩端點 B 與 C，懸掛彩帶佈置(如下圖所示)，試問彩帶 ( $\overline{AB} + \overline{AC}$ ) 的長度至少需要多少公尺？ ( $\sqrt{2} \approx 1.414$ ,  $\sqrt{3} \approx 1.732$ ,  $\sqrt{5} \approx 2.236$ )

- (A) 8
- (B) 12
- (C) 15
- (D) 18



#### Q2

請問圓弧形入口最寬的部分是幾公尺？

答：

圖7 「圓弧形美術館隧道」题目的題型內容

數學素養預調查包含了「數學素養情意態度」與「數學素養認知能力」。情意態度問卷包含行為、態度、自我效能、自我概念四個向度，共31題，採四點量表。認知能力題包含數量、空間與形狀、變化與關係、不確定性與數據四個內容領域，分成3式題本，每式題本6個題組（各13個小題），兩兩題本有共同題組，以求等化。

認知能力題的題型參考PISA、TIMSS、NAEP及TSA，採用四選一選擇題與建構反應題（包含列式計算題、文字說明題）兩類。由於建構反應題的作答較費時，爲了在有限的時間內有較多的試題取樣，故安排這類試題的數量略少於總題數的一半。題目設計成題組的型式，包含了貼近真實生活情境的題幹敘述，再搭配2至3小題的問題。以圖6和圖7「圓弧形美術館隧道」爲例，該題題旨爲「應用圓方程式、畢氏定理、圓冪性質、解析法，或正弦定理，測量弦長與圓半（直）徑」，但題目中完全不出現數學專有名詞，學生需能活用課堂所學數學知識方能解決問題。

本次調查與PISA、TIMSS、NAEP、TSA等大型調查類似，目的不在個別學生能力的評估，而重在系統層級的整體評量，且受限於每位學生施測時間僅50分鐘，爲擴大試題廣度，因此採用多式題本的設計。每位學生只回答其中一式，再透過試題的等化以及統計技術的估計，推估我國整體高三學生的數學素養。

電腦化測驗是國民素養調查未來的目標，然因在數位介面輸入數學符號與算式有其技術上待克服的問題，因此

2014年預調查仍採用紙筆測驗。電腦具有可做爲測驗時的計算、表徵、視覺、修改、探索和實驗的工具，也有呈現各種數學物件、現象和過程等優勢，並且提供命題者在編寫題目時能設計出更多互動、真實和引人入勝題目的機會。例如電腦可採用新的題目格式（如圖形的拖曳和縮放）、展示真實世界的資料（如大量的、可排序的數據群集）、或使用顏色和圖表讓評量作業在視覺上更吸引人，是傳統紙筆測驗所缺乏的（OECD, 2013a）。然而Noyes與Garland（2008）提醒電腦化測驗存在一些待克服的問題，例如：電腦化評量和紙筆測驗能否等值（equivalence），尚須比較不同測驗方式的統計特徵方能確定。一般而言，數學的紙筆測驗能保留較完整的解題過程，提供評分者更全面的判斷依據。此外，大型調查若採電腦測驗，在設備無法提供所有受測者同時上線受測的情況下，如何確保試題內容的保密性；就長時間的測驗而言，盯著電腦螢幕比紙筆作答更累；數位化測驗的前置工作與環境控制，比紙筆測驗更複雜等，都是有待解決的問題。

## 五、借鏡國際調查的經驗

由於我國日後將實施高三學生的素養調查，觀察國際現行數學素養或大型的數學成就測驗所引發的討論，我們可以琢磨出較佳的執行方式。

### （一）調查設計能否測出學生在真實生活中實踐知能的疑慮

Dohn（2007）指出，雖然PISA宣稱旨在評量學生能否「活用知識與技能於

生活中」，但受限於評量方法，僅能測到「評量情境下的知識與技能」。類似的質疑，Sjøberg（2007）也懷疑由文字化的紙筆題目能否評量出學生因應「真實生活挑戰（real-life challenges）」的能力。

De Corte、Verschaffel和Greer（2000）指出，許多完成國小教育的學生已經發展出一套有關解決數學應用問題的信念與假定，藉此能在給定數字的問題中將解題簡化成一堆運算的組合和執行。但未對問題真實情境的可能限制做認真考量下，可能危及他們解題時應用的標準模型和解決方案的適當性，例如NAEP的巴士題型「328位年長者要去旅行。一輛巴士可坐40人，總共需要多少輛巴士才能讓所有年長者去旅行呢？」是傳統問法，但如果嵌入更貼近真實情境的設定（如可以使用電話叫小型計程車），學生在回答後者時，應有更多元且真實的答案。因此，題目和真實現象結合的程度會影響學生的思考和解題，也和能否真正評量出學生在真實世界的知能有關。

## （二）調查題目內容對課程的衝擊

這些大型調查的多數題目是保密且重覆使用的，但每年都會釋出一些不再使用的題目讓大眾知悉。美國國家教育統計中心（The National Center for Educational Statistics）主張這些公開的題目可以用來做為課堂教學的補充材料，並且讓師生瞭解評量的內容。然而，NAEP出現了部分題目和學校課程無關的爭論。例如Klein（2011）指出，要求學生完成一個樣式（pattern），或將其他

幾何形狀填入一個幾何形狀的內容，如同拼圖的作業，似乎更接近智商測驗的題目，而非數學課程的學習目標。Stake（2007）也提到，NAEP對於教學和課程設計也產生了負面影響。當NAEP標榜其題目內容或型態時，許多測驗命題者及教師也被相繼吸引照著出題或教學，反而限制和阻礙了學生學習的多樣性。

TSA對香港教師在課程設計、教學和評量方面的負面影響，也出現在一些評論中（黃毅英、林智中、陳美恩，2009；廖佩莉，2013）：在課程設計方面，教師會注重TSA的考試內容，例如課程會著重讀寫或觀察等訓練，但卻忽視學生其它認知和品德情意的發展。在教學方面，為了應付TSA，教師會設計相關練習幫助學生應試，但容易將課題「擬題過深」或「愈教愈深」，或是教師花了太多時間為學生做TSA的操練及補課，反而阻礙了自身教學的發揮空間。在評量方面，由於過往TSA評卷有過於挑剔或標準不一的情形（如2007年小學在鐘面上畫出時針和分針以表示「5時40分」的時鐘題目，評分標準要求學生必須將時針畫在5時半與6時之間才算正確，但隔年類似題「10時40分」則改為時針畫在10與11時之間即可），造成教師平時就反覆以TSA的模擬試卷訓練學生的應試技巧，造成學生成績雖有進步，但由於題目僵化，教師教學失去啟發與引導的意義，學生也因長期面對類似的題目，反而導致對學習提不起興趣的結果。

## （三）過於強調評比，將導致評量績效領導學習

PISA是當前比較教育中最知名也最

成功的大型測驗，每當辦理完新的一次調查、產生新的結果，報紙就會充斥著各國的排名表，而社會大眾也會要求政府對國家教育體系上的缺陷作出回答。如此巨大的影響力，當然引發各方的爭議（Hopmann & Brinek, 2007），包括批評PISA未能全面涵蓋整個學校教育，或把PISA描述成尋求測驗市場一塊大餅的企業。Sjøberg（2007）也指出PISA和TIMSS型塑了學校品質的印象，已經超越了它們實際上有加以探究的面向，卻少有學者提醒這些調查的侷限性。例如：2001年PISA2000結果公布時，挪威的報紙標題以「Norway is a school loser」刊登出來。但少有針對這些數據的公開辯論，甚至研究者也很少會去修正和提醒大眾有關調查推論的侷限性。

由於這些國際型調查儼然成為學校和教育成效的代理人，儘管囿於紙筆測驗方式以及與各國教學目標未必完全吻合，但因其國際上的權威地位，因此所提出的報告就會對教育政策產生深遠的影響，而使得某些國家會為了表現教育上的績效而另行設計出專門測驗的結果，例如加拿大的泛加拿大評量計畫（the Pan-Canadian Assessment Program）就是其中一個為表現績效而設計的例子（Aikenhead, Orpwood, & Fensham, 2011）。

Stoneberg（2005）則提出使用NAEP分數呈現美國各州排名的疑慮，Loveless（2013）也對NAEP成績排名提出警告。Stoneberg 和Loveless 認為NAEP雖然小心翼翼地（scrupulously）計算出各州評量的成績和標準差，但也是僅從某一年度

的某個年級在某個科目，使用這些資料呈現排名順序和交替跨州的比較。因此在使用NAEP成績進行排名時應該需要將一些誤差考量在內，例如：各州NAEP測驗的學生是隨機樣本，不是每位學生都參加，因此所報導的成績只能視為是各州平均分數的估計值；應該用跨年級、科目、年度、或它們的組合的各種方法，來整合NAEP的估計分數和標準差，再去比較各州的表現，這對於排名才可能是有助益的。

香港教育專業人員協會（Hong Kong Professional Teachers' Union）於2011年調查發現，TSA已嚴重遠離初始辦理的目的，成為高風險的「評核怪獸」。學校為了準備TSA使得補課情形相當普遍，考試操練已下移至小學一年級開始，中小學各年級還會使用額外的TSA補充教材作為練習，TSA牽制了各校的教學規劃及測驗與考試模式（香港教育專業人員協會，2014），不但讓學生被迫成為考試機器，亦使得不少教師要分析每年的考卷趨勢，造成學習內容由TSA主導，而其他非TSA的考核內容（如與創意思維、學習態度、同儕相處技巧等課程）都被排除在外，不利於學生的全面發展（梁亦華，2013）。因此，香港教育局期望各校，不要將TSA數據視為評估學生表現的唯一指標，或在校內或與其他學校進行片面的排名或比較，應該要善用TSA的題目分析報告，了解學生在各學習範疇的強項及弱項，從而回歸優化學與教的初衷（教育部香港事務局派駐人員，2014）。

## 六、結語

教育變革過程需要有評估改革成效的指標，僅從硬體的改善或表面數字的變化（例如多少人就學、多少人畢業），不足以反映實際改革的績效。學生能力與人品的發展是教育的標的，國民素養調查強調對學生能力及教養、美感的長期觀察與整體性的評估，此種回歸教育基本目的的評量方向，值得肯定。本文除了回顧數學素養相關的大型調查方式，說明2014年數學素養預調查的作法之外，透過借鏡國際經驗，我們指出執行國民素養調查宜努力與謹慎之處。

調查方式必須盡快突破紙筆測驗的限制，如此才能透過數位介面展示更接近真實世界的問題情境，以及容許更多元的答題反應方式，使得調查結果可以更趨近學生在真實生活中知能的實踐。

此外，從國際大型調查的經驗可以發現，當調查結果的呈現可以進行各

國、各州、各校的比較時，就容易產生扭曲的效果。人們傾向將有限的調查資料當作整體教育成效的代言人，過度解讀調查結果，而學校行政單位與教師難以抗拒追求好成績的壓力，使得學校教育產生向調查重點或方式傾倒的現象。雖然，我們有時候也需要透過診斷各校學生學習狀況來提供補救方向的回饋，但基於目的與功能的差異，國民素養實施方案強調重在整體高三學生素養的評估，不作校間的比較，此一作法對維持其良善的立意相當關鍵，未來調查結果的公布方式是非常值得留意的議題。

## 致謝

感謝103年度「提升國民素養實施方案—數學素養計畫」的阮錫琦、馬雅筠、陳玫樺、陳家豪、陳敏皓、黃淑華、董涵冬、廖森游、鄭博元、謝宗益等十位數學老師的命題與批閱，及吳宜玲、楊元傑助理的協助，讓數學素養預調查得以順利完成。

## 參考文獻

- 江芳盛、李懿芳（2009）。國際學生評量計量（PISA）試題特色分析及其對我國教育之啓示。教育資料與研究雙月刊，87，27-50。
- 林素微（2013）。PISA2012數學素養評量架構與意涵。新北市教育，6，52。
- 香港考試及評核局（2014a）。2014年全港性系統評估（中學）便覽—第一部分（全港性評估）。取自[http://www.bca.hkeaa.edu.hk/web/TSA/en/2014QuickGuideSec/QG\\_S\\_Part\\_1\\_Chi.pdf](http://www.bca.hkeaa.edu.hk/web/TSA/en/2014QuickGuideSec/QG_S_Part_1_Chi.pdf)
- 香港考試及評核局（2014b）。2013年全港性系統評估第一至第三學習階段中國語文科、英國語文科、數學科學生基本能力報告—數學科中學三年級。取自[http://www.bca.hkeaa.edu.hk/web/TSA/zh/2013tsaReport/chi/Ch8c\\_S3\\_Math\\_TSA2013C.pdf](http://www.bca.hkeaa.edu.hk/web/TSA/zh/2013tsaReport/chi/Ch8c_S3_Math_TSA2013C.pdf)
- 香港考試及評核局（2014c）。2014年全港性系統評估（中學）便覽—第四部分（數學科）。取自[http://www.bca.hkeaa.edu.hk/web/TSA/en/2014QuickGuideSec/QG\\_S\\_Part4\\_Chi.pdf](http://www.bca.hkeaa.edu.hk/web/TSA/en/2014QuickGuideSec/QG_S_Part4_Chi.pdf)
- 香港考試及評核局（2014d）。2013年全港性系統評估第一至第三學習階段中國語文科、英國語文科、

數學科學生基本能力報告—全港性系統評估施行模式。取自[http://www.bca.hkeaa.edu.hk/web/TSA/zh/2013tsaReport/chi/Ch3\\_TSA2013C.pdf](http://www.bca.hkeaa.edu.hk/web/TSA/zh/2013tsaReport/chi/Ch3_TSA2013C.pdf)

香港教育專業人員協會（2014）。全面消除操練誘因，取消小三TSA—全港性系統評估（TSA）問卷調查結果公布。香港：作者。取自[https://www.hkptu.org/education/wp-content/uploads/20140107\\_TSA.pdf](https://www.hkptu.org/education/wp-content/uploads/20140107_TSA.pdf)

國立臺南大學（2012）。PISA 2012數學樣本試題【中文版】。臺南市：臺灣 PISA國家研究中心。取自[http://pisa.nutn.edu.tw/download/sample\\_papers/PISA\\_2012\\_items\\_for\\_release\\_Chinese\\_final.pdf](http://pisa.nutn.edu.tw/download/sample_papers/PISA_2012_items_for_release_Chinese_final.pdf)

國立臺灣師範大學科學教育中心（2011）。TIMSS 2011數學八年級\_題本\_5。取自[http://www.sec.ntnu.edu.tw/timss2011/downloads/T11\\_G8\\_M06.pdf](http://www.sec.ntnu.edu.tw/timss2011/downloads/T11_G8_M06.pdf)

教育部香港事務局派駐人員（2014年4月）。減少全港系統性測驗 教育局冀紓緩學生壓力。教育部電子報，610。取自[http://epaper.edu.tw/windows.aspx?windows\\_sn=15073](http://epaper.edu.tw/windows.aspx?windows_sn=15073)

教育部提升國民素養辦公室（2013）。教育部提升國民素養專案計畫報告書。臺北：作者。取自<http://literacytw.naer.edu.tw/data/cht/20140801/20140801u4v194.pdf>

教育部綜合規劃司（2013年2月26日）。提升國民素養E起來—歡迎瀏覽教育部提升國民素養專案【即時新聞】。取自<http://www.edu.tw/news1/detail.aspx?Node=1088&Page=18250&Index=1&WID=45a6f039-fcaf-44fe-830e-50882aab1121>

梁亦華（2013）。「全港性系統評估」的風險。載於葉建源、黃家樂（主編），全港系統性評估（TSA）評論文集（190-191頁）。香港：香港教育專業人員協會。

黃毅英、林智中、陳美恩（2009）。「基本能力」還是「基本」嗎？數學教育期刊，28，2-9。取自<http://www.hkame.org.hk/html/modules/tiny2/content/v28/02%20WongNY%20BCA&TSA.pdf>

廖佩莉（2013）。倒流效應：香港全港性系統評估（TSA）對小學中國語文教師的影響。教育研究月刊，228，86-102。

臺灣 PISA國家研究中心（2012）。電腦化樣本試題範例：數學樣本試題範例—中文版。取自<http://pisa.nutn.edu.tw/cm/CM012/www/index.html>

臺灣PISA國家研究中心（2012）。PISA應試指南—數學應試指南。臺南市：作者。取自[http://pisa.nutn.edu.tw/download/sample\\_papers/2009/2013\\_0311\\_guide\\_mathematics.pdf](http://pisa.nutn.edu.tw/download/sample_papers/2009/2013_0311_guide_mathematics.pdf)

Aikenhead, G. S., Orpwood, G., & Fensham, P. (2011). Scientific literacy for a knowledge society. In C. Linder, L. Ostman, D. A. Roberts, P-O. Wickman, G. Erickson, & A. MacKinnon (Eds.), *Exploring the landscape of scientific literacy* (pp. 28-44). New York, NY: Routledge, Taylor and Francis Group.

De Corte, E., Greer, B., & Verschaffel, L. (2000). Connecting mathematics problem solving to the real world. In Rogerson, A. (Eds.), *Proceedings of the International Conference on Mathematics Education into the 21st Century: Mathematics for living* (pp. 66-73). Amman, Jordan: The National Center for Human Resource Development.

Dohn, N. B. (2007). Knowledge and skills for PISA-Assessing the assessment. *Journal of Philosophy of Education*, 41(1), 1-16.

Gonzalez, E. J. & Miles J. A. (2001). *TIMSS 1999 user guide for the international database*. Chestnut Hill,

- MA: International Study Center, Boston College, Lynch School of Education. Retrieved from [http://timssandpirls.bc.edu/timss1999i/data/bm2\\_userguide.pdf](http://timssandpirls.bc.edu/timss1999i/data/bm2_userguide.pdf)
- Hopmann, S. T., & Brinek, G. (2007). Introduction: PISA according to PISA-Does PISA keep what it promises? In S. T. Hopmann, G. Brinek, & M. Retzel (Eds.), *PISA according to PISA-Does PISA keep what it promises* (pp. 9-19). Vienna, Austria: LIT-Verlag.
- Jablonka, E. (2003). Mathematical literacy. In A.J. Bishop, M. A. Clements, C. Keitel & F. K. S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 75-102). Dordrecht, The Netherlands: Kluwer.
- Klein, D. (2011). What do the NAEP math tests really measure? *American Mathematical Society*, 58(1), 54-55. Retrieved from <http://www.ams.org/notices/201101/rtx110100053p.pdf>
- Loveless, T. (2013, November 13). *Be Wary of Ranking NAEP Gains* [Web blog message]. Retrieved from <http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2013/11/13-interpreting-naep-gains-loveless>
- Mullis, I. V.S., Martin M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- National Assessment Governing Board. (2012). *Mathematics Framework for the 2013 National Assessment of Educational Progress*. Retrieved from <http://files.eric.ed.gov/fulltext/ED542064.pdf>
- Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51(9), 1352-1375.
- OECD (2010). *PISA 2012 Mathematics Framework*. OECD, Paris. Retrieved from <http://www.oecd.org/pisa/pisaproducts/46961598.pdf>
- OECD. (2013a). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Retrieved from [http://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book\\_final.pdf](http://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book_final.pdf)
- OECD. (2013b). *PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs (Volume III)*. Retrieved from <http://www.oecd.org/pisa/keyfindings/PISA-2012-results-volume-III.pdf>
- OECD. (2014). *PISA 2012 Results: What Students Know and Can Do-Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014)*. Retrieved from <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-I.pdf>
- Sjøberg, S. (2007). PISA and “real life challenges”: Mission impossible? In S. T. Hopman, G. Brinek, & M. Retzel (Eds.), *PISA according to PISA – Does PISA keep what it promises* (pp. 203-225). Vienna, Austria: LIT-Verlag.
- Stake, R. E. (2007). NAEP, report cards and education: A review essay. *Education Review* 10(1), 1-22.
- Stoneberg, B. D. (2005). Please don't use NAEP scores to rank order the 50 states. *Practical Assessment*,

*Research, & Evaluation*, 10(9). Retrieved from <http://pareonline.net/pdf/v10n9.pdf>

U.S. Department of Education. (2013). *NAEP 2012 Trends in Academic Progress*. Retrieved from <http://nces.ed.gov/nationsreportcard/subject/publications/main2012/pdf/2013456.pdf>

Wilkins, J. L. M., Zembylas, M., & Travers, K. J. (2002). Investigating correlates of mathematics and science literacy in the final year of secondary school. In D. F. Robataille & A. E. Beaton (Eds.), *Secondary analysis of the TIMSS results: A synthesis of current research* (pp. 291-316). Boston, MA: Kluwer Academic.

Yore, L. D., Pimm, D., & Tuan, H. L. (2007). The literacy component of mathematical and scientific literacy. *International Journal of Science and Mathematics Education*, 5(4), 559-589.