

林素微（2022）。

「他山之石，可以攻錯」：從 PISA 視角談數學教育研究取徑。

臺灣數學教育期刊，9（2），113–120。

doi: 10.6278/tjme.202210\_9(2).005

## 「他山之石，可以攻錯」：從 PISA 視角談數學教育研究取徑

林素微

國立臺南大學教育學系

從臺灣參與 PISA 2006 開始，PISA 測量構念以及臺灣學生表現一直是教育關注的重點。隨著其國際能見度逐漸廣泛及對各國教育的影響逐漸深化，臺灣教育如何因應是重要的議題；例如：如何教？如何學？如何評量？甚至如何藉由 PISA 經驗或者資源進行研究探討？因應本期專題，筆者在此推薦 Stacey 與 Turner 所主編的《Assessing Mathematical Literacy: The PISA Experience》一書，書中詳細介紹 PISA 數學評量發展歷程，包含評量架構及研發流程，以及 PISA 對於許多國家所造成的衝擊。本文以此書的內涵簡介為出發點，進一步討論研究取徑的可能方向，期能透過本文的拋磚引玉，啟動學術的腦力激盪，共同為臺灣數學教育的進步而努力。

**關鍵詞：**PISA、數學評量、研究取徑

通訊作者：林素微，e-mail：swlin0214@gm2.nutn.edu.tw

收稿：2022 年 9 月 19 日；

接受刊登：2022 年 9 月 30 日。

## 壹、緒論

隨著十二年國教課綱的推動，核心素養已然是臺灣教育改革的重點；在這波教改中，素養的強調並非是臺灣的獨特議題，而是一個全球趨勢。在教育實踐中，課程、教學、評量是息息相關的環節，因此，在整個實踐過程中，評量雖居末節，但卻扮演著不可或缺的角色。在重視核心素養的時代潮流下，素養導向的評量是趨勢所在，而最受矚目的素養導向評量應屬 OECD（Organisation for Economic Cooperation and Development）所推動的國際學生能力評量計畫（the Programme for International Student Assessment, 簡稱 PISA）。PISA 從 2000 年開始啟動第一波的調查，至今（PISA 2022）已經進行到第八次評比，隨著調查次數的增加，參與國家愈來愈多，已然是全球最大的國際評比調查之一，可見 PISA 評量理念的被認同度與能見度愈來愈高。

自 PISA 2000 年推動以來，各次執行前後的各種文件，包括評量架構、技術報告、國際報告等都可在其官方網站（<https://www.oecd.org/pisa/>）取得，其調查結果的全球數據，亦可在官方網站下載進行分析與探討。筆者於 2022 年 9 月 18 日於 ProQuest 資料庫以「PISA」為關鍵詞進行檢索，得到 438,307 筆結果；進一步「PISA & Math」兩個關鍵詞進行檢索，結果有 66,432 筆，其中，書籍有 44,201 筆、學術期刊有 10,545 筆，換句話說，從 PISA 執行以來，與數學相關的主題，平均每年約有 2000 本書籍（包含學位論文）及 500 篇期刊論文產出，可見 PISA 的高能見度及對數學教育學術的貢獻力，就此而言，PISA 評量可稱之為全球性的指標測驗。

自教育部在 2014 年《十二年國民基本教育課程綱要總綱》中明確指出基本教育應實踐素養導向之課程與教學開始，至今已經進入第八個年度，但隨著時間推演，教育現場對於素養一詞的界定似乎仍未達明確共識，在教育實踐與學術研究上也仍有爭議的空間。筆者在此推薦 Stacey 與 Turner 所主編的《Assessing Mathematical Literacy: The PISA Experience》一書，希望可以提供臺灣數學教育研究在探討素養導向的理念、評量設計及研究取徑的相關參考。以下，筆者將先對此書的內涵進行簡介；最後，則分享筆者對於 PISA 研究的可能取徑及 PISA 評量設計的反思。

## 貳、Assessing Mathematical Literacy: The PISA Experience 簡介

書名：Assessing Mathematical Literacy: The PISA Experience

主編：Kaye Stacey & Ross Turner

出版年：2015

出版社：Springer Cham

頁數：321

ISBN : 978-3-319-10120-0 (Hardcover)

978-3-319-34606-9 (Softcover)

978-3-319-10121-7 (eBook)

DOI : <https://doi.org/10.1007/978-3-319-10121-7>

本書的主要作者群來自於負責前五次 PISA 評比（也就是 PISA 2000、PISA 2003、PISA 2006、PISA 2009 和 PISA 2012）的澳洲教育研究委員會（Australian Council for Educational Research, 簡稱 ACER）及數學專家小組（Mathematics Expert Group, 簡稱 MEG）的成員。由於 PISA 旨在探討參與國家的表現、排名、跨時間趨勢、教育均等、以及學校和教師等特徵和學生表現的關聯。但 MEG 成員認為除了這些評量結果的面向之外，也應適切地讓讀者瞭解 PISA 的理論和實踐發展，他們期待透過 PISA 數學評量發展的「inside story」，能讓讀者對於 PISA 數學素養（mathematical literacy）有更全面且深入的理解，後續方能更加善用 PISA 的相關結果。全書分為三個主要部分，以下將逐一介紹：

## 第一部分：PISA 數學的基礎

第一部分主要針對 PISA 數學素養的概念及理論背景進行回顧，這部分囊括了數學素養的起源與及數學教育領域中與數學素養其密切相關但可能部分不同的概念及相關術語（例如 quantitative literacy 和 numeracy）；探討數學素養所包含的數學能力及如何利用這些數學能力來估計 PISA 題目的認知需求；由於 PISA 旨在評量 15 歲學生面對現實生活問題的能力，因此，第三章聚焦於現實世界和數學世界之間評量的聯繫。第五章，則以一位數學家參與 PISA 設計後的個人反思作為此主題的結束。各章內容簡介如下：

第一章是 Kaye Stacey 和 Ross Turner 的《The Evolution and Key Concepts of the PISA Mathematics Frameworks》，本章描述了 PISA 評量架構的關鍵概念，以及這些概念在 PISA 及數學教育思維中的歷史和起源。

第二章是 Mogens Niss 的《Mathematical Competencies and PISA》，本章針對在 PISA 數學評量架構佔有一席之地的一組數學能力，包含這些能力的起源及它們在“做數學”所扮演的意涵。

第三章則是 Kaye Stacey 主筆的《The Real World and the Mathematical World》，在此描述 PISA 所重視的數學建模及數學建模如何將數學與日常生活實務進行橋接；以及如何透過真實情境的設計下，可以進行跨群體和跨文化之間的數學評量。

第四章則是由 Ross Turner、Werner Blum 和 Mogens Niss 共同撰寫之《Using Competencies to Explain Mathematical Item Demand: A Work in Progress》本篇研究呈現了如何透過 PISA 數學能力的解構來理解 PISA 數學作業的認知需求。

第五章則是 Zbigniew Marciniak 的《A Research Mathematician's View on Mathematical Literacy》，本文旨在呈現作者（一位數學家）在參與 PISA 後的個人反思，主要聚焦於 PISA 參與如何影響他對數學教育重要因素的觀點與看法；此反思彰顯了重要的理論考量的確會對個人觀點有深遠的影響。數學教育與數學家之間的「數學戰爭（math wars）」，作者建議透過 PISA 的共同投入或許會是一個解套的方式。

## 第二部分：PISA 調查的實施：協作、質量及複雜度

第二部分從內部角度描述了 PISA 調查的研發和實施方面的重要細節，包含了從紙筆測驗轉換到電腦化評量研發的過程、作答反應的編碼與問卷研發；如何借鏡全球知名機構的測驗研發經驗和成果來進行開發；藉由 PISA 公布試題描述數學素養試題的特徵及研發流程；以及數學素養不同水準如何訂定等議題。此部分還描述了 PISA 2012 調查如何收集數據來衡量各國學生的數學學習機會（opportunity to learn）；當國家的培育目標是數學素養時，如何透過 PISA 的證據找到符合 PISA 理念及傳統數學教學的平衡點。第二部分共包含五章，各章內容簡介如下：

第六章為 Ross Turner 所撰寫的《From Framework to Survey Data: Inside the PISA Assessment Process》，作者從身為 PISA 主要執行單位的角色，透過 PISA 評量過程的內部故事介紹了每次 PISA 調查的開發和實施所涉及的要素。

第七章是由 Dave Tout 和 Jim Spithill 合著的《The Challenges and Complexities of Writing Items to Test Mathematical Literacy》，他們以身為試題研發者的角度，描述數學素養試題開發的過程。文中透過 PISA 公布試題來舉例說明這些過程。

第八章則是 Caroline Bardini 所撰寫的《Computer-Based Assessment of Mathematics in PISA 2012》，作者描述了與 PISA 試題電腦化相關的理論和實踐問題，並搭配 PISA 2012 的公布試題進行說明。

第九章是 Agnieszka Sułowska 的《Coding of Mathematics Items in the PISA Assessment》，文中探討 PISA 評量中有關數學題目的編碼，從 PISA 國家評量中心的角度，針對學生對 PISA 試題的反應編碼進行了非常實用的解釋。

第十章是 Leland Cogan 和 William Schmidt 的《The Concept of Opportunity to Learn（OTL）in International Comparisons of Education》討論了在 PISA 2012 的學生問卷中，有關學習數學機會相關的創新問題。

### 第三部分：PISA 對全球的影響：啟發與適應

本書的第三部分是關於 PISA 影響的議題。此部分同樣包含了五個篇章。此部分介紹了 14 個國家的數學教育工作者從不同背景與觀點，描述 PISA 的理念和評量方法如何影響教學實務、課程安排、不同層面的評量實務，以及國家中引起的教育爭論。其中有些作者探討了國家的 PISA 分數隨著時間推移而有所進步，並希望使用 PISA 結果來進一步激勵教育革新，同時作為政策制定者的參考資訊。

第十一章是 Toshikazu Ikeda 的《Applying PISA Ideas to Classroom Teaching of Mathematical Modelling》，文中討論了如何運用 PISA 架構中所提倡的數學建模概念，並將其應用在日本課堂實踐之中。

第十二章是 Manfred Prenzel、Werner Blum 與 Eckhard Klieme 共同合著的《The Impact of PISA on Mathematics Teaching and Learning in Germany》，文中描述了德國一開始的 PISA 成績出乎意料的低落，此項衝擊導致後續因應 PISA 時，學校和教育系統的變化。

第十三章則由 Ferdinando Arzarello、Rossella Garuti 和 Roberto Ricci 撰寫的《The Impact of PISA Studies on the Italian National Assessment System》，文中描述了義大利國家評量系統受到 PISA 構念的影響，致使全國努力於數學課室實踐的改革，期待讓學生能充分掌握 PISA 的評量重點。

第十四章是由 Kai-Lin Yang 和 Fou-Lai Lin 所著之《The Effects of PISA in Taiwan: Contemporary Assessment Reform》，文中針對 PISA 對臺灣的影響有深入的描述，本文介紹 PISA 高表現國家之一的臺灣，其數學教育工作者為因應 PISA 衝擊的相關積極作為，並介紹當初引起爭議的北北基特色招生評量計畫。

第十五章則是由 Kaye Stacey 編纂的《PISA's Influence on Thought and Action in Mathematics Education》，此章透過一系列的短文描述 PISA 在十個國家的影響與反思，包含了智利、丹麥、法國、印尼、伊朗、以色列、韓國、新加坡、西班牙、美國。透過上述各國作者的摘要呈現，說明了 PISA 對於世界各地的影響力，以及其與許多教育工作者關注的一致性。

### 參、他山之石可以攻錯，從 PISA 談數學教育可探討的研究取徑

PISA 已是全球最大型的教育調查之一，《Assessing Mathematical Literacy: The PISA Experience》的主編和多數作者都直接參與了 PISA 數學評量架構和數學素養試題的創建與設計，如同主編在前言所指出，本書旨在介紹全球最大的數學評量的“inside story”。雖然 ACER 在 PISA 2012 後畫下 PISA 任務的句點<sup>1</sup>。但後續的數學素養評量架構仍然以其為基礎持續修

---

<sup>1</sup> PISA 2015 起由 ETS (Educational Testing Service) 所負責。

訂，因此，此書提供的 PISA 評量發展經驗，仍應值得數學教育工作者參考。PISA 此類的大型國際評比，參與的國家和學生數量眾多，這樣巨大的教育投入，應有相應的回饋功能方能持續運作。換句話說，當 PISA 結果能夠用來改善參與國家的教育成果，並對教育投入提供生產性、建設性的支撐時，才值得各國在 PISA 投入如此多的能量與努力。以下，針對臺灣數學教育如何藉由 PISA 的研發和研究經驗從中進行研究取徑，可以從以下幾個方向著手：

### 一、數學素養構念的再概念化

一份數學評量從開發、實施、到後果影響的精緻度取決於工具研發與實施過程的嚴謹程度，過程愈嚴謹，評量工具所具備的信度和效度就愈高，我們也就愈有信心相信評量結果能夠深入瞭解學生的數學表現。

如同本書的許多作者所持的觀點，數學素養的概念因應教育研究的脈動發展而來，但此概念在數學教育也是一個獨特的新貢獻；此概念因應 PISA 的能見度而讓全球各地的學者有這個契機將相關概念和術語進行討論與統整。國內、外的數學教育學者，陸陸續續針對數學素養的定義有所發表，雖然數量在增加中，但素養概念的明確性與各定義之間的釐清，仍有許多對話的空間。

### 二、PISA 試題實作與評量再研發

如緒論所述，因應這一波教育改革，素養的培育已是作為大多數教育系統培育的主要目標。在本書第三部分中，一些案例國家也嘗試運用 PISA 數學素養架構並進而形成了國家課程與評量發展的框架。除了可以處理比較教育的議題外，數學素養評量的發展也是重要的議題。在嘗試進行數學素養試題的研發時，筆者認為初始階段可先嘗試以 PISA 的數學評量試題進行探索，再從實作經驗中汲取研究設計想法。截至目前為止，PISA 的數學公布試題數量已經相當可觀，許多教育研究也開始用來作為教師專業發展和評量模式的發展。也就是說，PISA 從數學教育的實踐中發展而來，反過來又影響了數學教育的發展方向。值得注意的是，由於 PISA 涉及跨國的評量，在試題設計上難免會有些因地制宜的限制，所以光複製 PISA 型態的試題在培育數學素養上可能會略有不足，但這些公布試題卻能為改進教學提供了想法和方向；它們還可以刺激類 PISA 任務或試題的製作和課堂使用，透過這些任務和試題設計，提升學生數學建模和數學素養將更可期待。

### 三、試題認知成分解構與學生認知診斷

PISA 提供的評量架構和公布試題除了提供研究者可在研究場域中進行再測之外，有關數學素養能力的解構、認知負荷的分析其實也是另一個可行的方向。透過理論有據的認知成分

(請參考第二、四章)，針對試題描述逐一進行認知成分的編碼，結合實測所得的試題難度，將可掌握測驗難度的可能來源，這對測驗編製者設計試題及教師課室布題都有實質的幫助。近十幾年來，認知診斷分析是測驗學界蓬勃發展的重要議題，主要分析概念就是透過試題所需的屬性(認知成分)所形成的試題屬性矩陣，結合學生實際的作答表現，可以明確診斷學生的認知狀態。認知診斷模式可以從兩個角度著手，一個是根據理論的認知成分來設計試題，再進行資料蒐集與分析；另外一個則是以既有的測驗資料庫來處理，先針對測驗資料庫中的試題進行成分界定，再結合作答反應據以分析。就 PISA 此類的資料庫而言，可以採第二種角度切入，目前已有一些國內外學者嘗試針對 PISA 進行認知診斷探討。

#### 四、善用 PISA 資料庫探討教育議題

PISA 的資料庫是公開的，這個資料庫包含豐富的變項，涉及國家、學校、學生各個層面；藉由這些豐富的資訊進行深入探討是最多學者採用的方式。每次 PISA 調查的國際報告中，OECD 原則上多僅檢視這些單一變項和素養的關聯，報告中通常會先考慮社經地位和性別的可能影響，在控制了這兩個重要因素之後再檢視各變項與素養的關聯；但由於通常只著眼於單一變項，對於研究者而言，依據理論納入多重變項進行關聯探討的空間相當寬廣。

PISA 的取樣是採取先抽學校、再抽學生的兩步驟抽樣方法，此抽樣方式通常建議分析時應考慮資料的階層性，而 PISA 除了蒐集學生層級的訊息之外，也進行了學校與教師問卷的填答，因此，針對 PISA 豐富多元的資料，進行多層次的次級資料分析應有相當廣大的空間。值得注意的是，許多新進研究者會採用資料探勘的方式，但這種沒有理論的探索性分析通常對於教育的實質意涵較為薄弱，建議應採理論依據的資料分析才能有較高的學術生產性。

國際評比歷經嚴格的抽樣程序，通常較具母體代表性。除了全體學生資料的分析，筆者建議也可以著眼於特定學生，例如：不同精熟水準、不同社經地位、不同變項交織下的分類(如：綜整社經地位與成就分類：高社經低成就、低社經高成就；興趣與成就分類：高興趣低成就、高成就低興趣...)。此類特定學生的分析，往往可以關注教育中的特定議題，對於教育實務的實質意義頗高。

須謹記的是，雖然 PISA 已是世界上最大的教育研究調查之一，但它仍然無法回答教育中所有的問題。例如，它不能直接回答有關教育改革最佳方向的問題。也有許多數學教育者常批判 PISA 數學任務或試題的測量目標，由於 PISA 試題多是為了檢核學生在真實情況下的素養展現而設計，而且為了因應題庫的評量架構而必須有所取捨，在這樣的情況下，有些題目可能會被排除，但這些題目可能可以測量到學生重要的數學思維。此外，因涉及跨國比較，大型測驗通常會考量閱卷或作答編碼的合理性和經濟性，因此可能會忽略許多關於學生表現的細緻訊息。其他對於 PISA 的批判則包含試題設計和施測中的品質監控措施及跨文化間的測

驗偏誤。整體而言，PISA 並非是一個足夠完美的教育研究資料庫，因此，筆者建議應對於 PISA 結果有充分的理解，從理解與反思中出發，如此將更能善用這些測量結果所提供的相關資源，作為後續數學教育實務、研究及政策制定的參考基礎。

### 參考書目

Stacy, K., & Turner, R. (Eds.). (2015). *Assessing mathematical literacy: The PISA experience*. Cham, Switzerland: Springer. doi: 10.1007/978-3-319-10121-7