

PISA 2012 數學評量的啟示

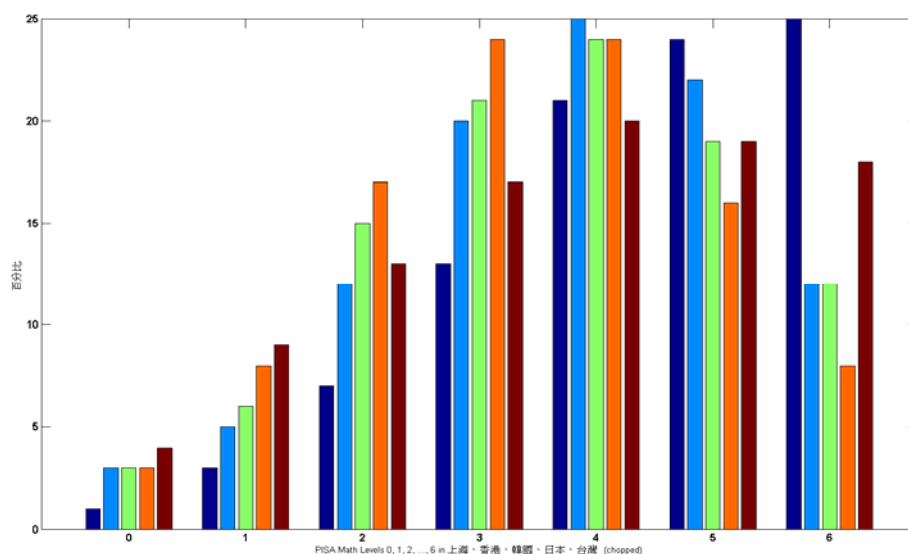
單維彰·102 年 12 月 9 日

去年 12 月 3 日下午，全球「同步」公布了 PISA 2012 的評量結果，針對某些國家的區域性報告也一併出爐了（不含台灣）。至本篇截稿為止，各大報紙都報導了台灣在數學、閱讀、和科學三方面的排名，國科會和國民教育署也發佈了「官方」報告。但整體而言，網路上能搜尋到的各方詮釋與反應還很少；相信在大家讀到本篇專欄之前，已經有很多新的消息散播於媒體之上了。

本欄在民國 97 年 1 月和 2 月，分別介紹過 PISA 國際評量的背景，以及它對數學能力或數學素養的定義。當時，根據 2006 年的數學測驗結果，就曾提出一個值得留意的現象，我國的評量分數「標準差」過高：事實上是世界第三高。2006 和 2009 年的「主要」測驗是科學和閱讀，所得的數學評量資訊還可以說不太精確。但 2012 年的「主要」測驗就是數學了，我們不能再迴避這個事實：我國 15 歲國民在 2012 年的國際數學評量中，**所得分數的標準差是世界第一高**；國科會的報告說『標準差 116，相較第二高的國家（105）有明顯的差距。』

PISA 產生的浩瀚資料，可以讓研究者各取所須地工作個一兩年。在這兩頁的篇幅裡，我只想對《科學月刊》的讀者們說一件事：關於標準差過大的這一件。我國受測學生的樣本，乃是按照主辦單位 OECD 的抽樣規定辦理，且根據國科會報告：『我國應試樣本包含 163 所學校（包括國中、五專及高中職），實際參與評量學生為 6,037 名，淨出席率達到 96.1%，評量參與情況良好。』讓我們假設其統計數據有足夠的信心水準。

標準差代表數據的分散程度，大的標準差表示我國學生的數學成績懸殊。



請參考圖一，裡面有 7 叢長條圖，其橫坐標 1, 2, 3, 4, 5 依序表示評量表現落在第一、二、三、四、五級數學能力的範圍內，而 0 表示未達第一級，6 表示達到第六級或者更高。每一叢裡面的五條長方形，最左邊代表上海（中國僅以上海市參加評量），最右邊代表台灣，中間是香港、韓國、日本。縱軸表示各地區受測學生在各等級的百分比；但是我要關注 0, 1, 2 的相對比例，所以裁切了百分之 25 以上的部分。

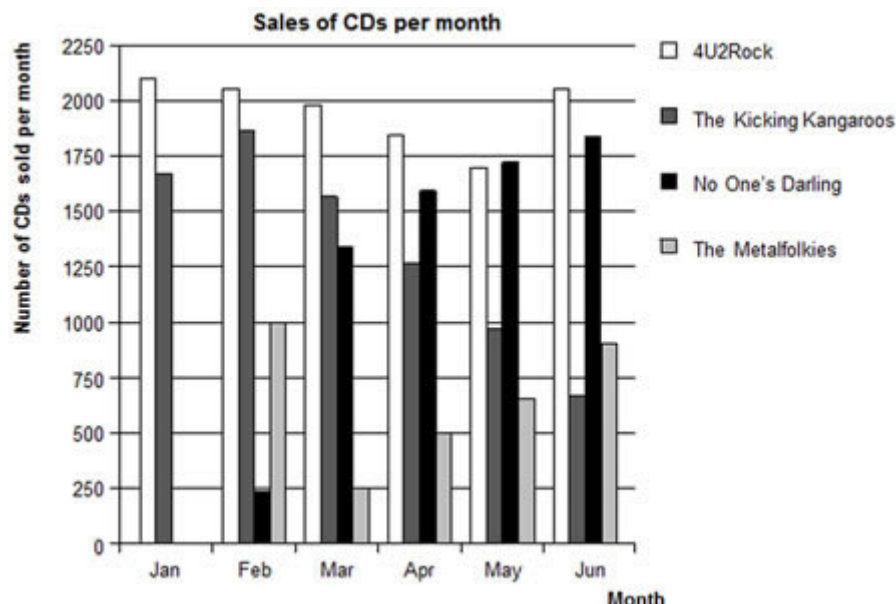
從圖一的 6 那一叢，可以推論上海和臺灣的「超標」學生特別多。這群學生是讓臺灣站上國際評比第四名的「功臣」（但 PISA 的官方報告指出，名列三、四、五的香港、臺灣和韓國的成績，在統計上並無差異）。不論他們自認為數學能力如何？看待數學的態度如何？PISA 測驗顯示我們有一群優秀的少年，她們不但掌握了基礎的數學知識，也能夠有效地運用它來解決實際問題。

但是，看看 0 和 1 那兩叢，就是我們那「世界第一高標準差」的根源了。所謂 0 是連第一級的問題都無法解決的意思；那麼，第一級是什麼？不要讀定義，讓我們看看 PISA 提供的標準範例題，如下。

第一級標準例題

4U2Rock 和 The Kicking Kangaroos 樂團在一月發行新 CD 唱片，No One's Darling 和 The Metalfolkie 樂團在二月也發行了新唱片。下圖顯示這四張 CD 從一月到六月的銷售量，試問在哪一月 No One's Darling 的唱片銷售量首度超過 The Kicking Kangaroos 樂團？（注：我沒有查詢臺灣的官方翻譯版，隨手自己翻譯了，但為了配合英文版的插圖而沒有翻譯樂團的名稱）

- A) 不曾發生 B) 三月 C) 四月 D) 五月



相信很多讀者同意，按照臺灣國中數學的標準，這根本不能算是一道數學題。但是我國有 4% 的學生無法回答這個等級的問題。我們有 9% 的學生能夠回答這類問題，因此達到 PISA 的第一級數學能力，但是無法達到第二級。判定

第二級數學能力的官方範例問題如下。

第二級標準例題

大美剛買一輛有時速錶的新腳踏車。那個儀錶可以報讀她騎車的距離和平均速率。在某一趟騎車的旅程中，大美在前 10 分鐘騎了 4 公里，接下來的 5 分鐘騎了 2 公里，試問以下敘述何者正確？

- A) 前 10 分鐘的平均速率大於接下來的 5 分鐘
- B) 前 10 分鐘的平均速率與接下來的 5 分鐘相等
- C) 前 10 分鐘的平均速率小於接下來的 5 分鐘
- D) 給定的條件不能判斷大美的平均速率

這就像個數學問題了，但是，至少在都會型的學校裡，它可能屬於小學五年級。的確，PISA 暗示，第一級的數學能力（若不考慮語文上的需求），應該在小學二、三年級即可具備。

讀到這裡，我們知道臺灣有 18% 的少年，具備最高級（或者更高）的數學素養，這個百分比高居世界第二（僅次於上海）。在頻譜的另一端，我們有 13% 的少年未達第二級的數學素養；若將這個數據從小排到大，我們在世界第十位，是參與評量的東亞國家（或經濟體）之中最高的。而且我們透過典型範例問題，大約知道了「未達第二級」是什麼意思。

所謂「山不厭高，海不厭深」，我們應該不至於嫌第六級的學生太多了而想要自廢武功，就像我們不曾聽說有人嫌自己的孩子長得太高了而為他做脛骨截短手術。更何況 PISA 測驗的題目，應該讓大家相信這些少年並不是被填鴨的書呆子，而是有能力活用數學解決實際問題的人。對於數學教育已經做得很好的這一部分，我們應該讓它維持著。

我們須要設定的任務，是降低未達二級的學生百分比。在這個問題上，切忌「急病亂投醫」。有一種議論是說數學太多或者太深，要把課程改得更淺；但是這樣做就好像要長得高的人截斷脛骨以降低身高，實在不可思議。比較可行而且已經在做的，是大規模的補救教學。但是補救教學的資源又以「公平」之名而不能集中效能，台北市中心的學校與台灣任何學校，都分到同樣比例的補救教學經費。而且設計課程的專家們，並沒有將數學內容做個輕重緩急的分疏，使得補救教學經常還是將所有內容再講一遍，或者把設計補救教材的責任放到教師身上。

我提出兩項建議。第一，課程綱要在設計的時候，就可以提出「層次核心」的建議，讓補救教學的效率提高。舉國中一年級面臨的第一個課題為例：「負數」。負數是相當基礎的數學語言，沒有負數則不可能在坐標平面上畫出完整的函數圖形，可見負數真的很重要。但是，大家不妨打開 PISA 的範例題庫，看個仔細，究竟有哪幾題須要用到負數的計算？幾乎沒有。在絕大多數的實際情境問題裡，只要知道小（正）數減大數得到負數的演算法和解讀方式就夠了。由此可見，就連「負數」這麼基礎的課題，都可以有層次之分。在補救教學的時候，教師應該有個課綱提供的取舍標準可供參考。

第二個建議已經是老生常談，但也是我們長期的痛：為什麼數學課堂始終不用資訊工具？！人人都知道，機械工具是體力的輔具，它讓我們跳得更高、跑得更快、舉得更重、擲得更遠更準。而資訊工具則是腦力的輔具。用機械工具對體力的輔助來類比資訊工具將能提供的腦力輔助，就可以戰慄地想像，我們因為拒絕這個工具而侷限了多少的腦力資源！對於已經能夠勝任現行課程的學生，倒是無所謂；但是對於須要補救的學生而言，從課程到教材到評量，我們對資訊工具有明顯而迫切的需求。

縮小標準差，或降低未達二級的學生比例，並不是為了臺灣爭取更高的國際排名，而是為了全體國民的福祉，更是為了我們共處的這個社會裡的每位公民，都有基本的共同知識與計算推論能力，得以有效地參與社會公共議題的討論和決策。這是我們一定要達成的任務，但是絕對不能用降低全體學生的學習內容，來達成這項任務。