

1992國際數理教育評鑑IAEP

—我們能夠學到什麼？—

楊榮祥

國立臺灣師範大學科學教育中心

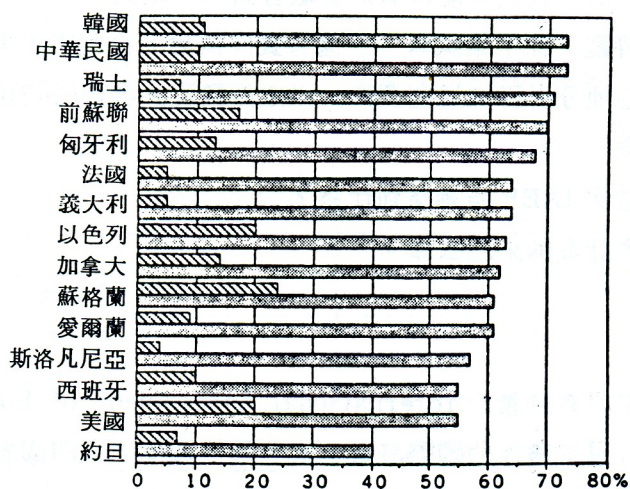
先後進行四年，包括美、英、蘇、加等，共有二十個國家所參與的國際數理教育評鑑（原名為 International Assessment of Educational Progress，簡稱 IAEP）終於有了最後的結果。主辦國美國在今（八十一）年二月五日，於華府的記者俱樂部（National Press Club）召開記者招待會，邀請世界主要的傳播媒體，正式發表 IAEP 的結果與其重要的發現。主辦單位教育測驗社（Educational Testing Service，簡稱 ETS）還邀來美國教育部長（Education Secretary）Lamar Alexander、國家科學基金會的副主席 Luther S. Williams、國家教育統計資料中心執行長 Emerson J. Elliott 等科學教育有關主要官員來會說明，並回答記者們的質詢。

筆者有幸，蒙教育部 IAEP 指導委員會指派，代表我國本計畫研究人員列席備詢。在會上聽到各傳播媒體記者熱烈討論並提出詢問，感慨良多。當天晚上的電視新聞、晚報、及次日各家早報，都在相當顯著的版面（有許多大報都做為頭版新聞），以很大的篇幅報導此會。

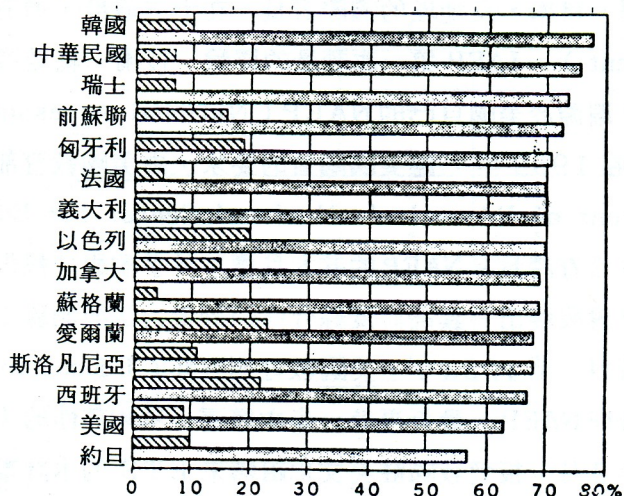
報導的重點，不外乎美國的教育問題。各家媒體討論重點似乎都在：為什麼美國在 IAEP 這項國際性教育評鑑中，不單未能名列前茅，反而幾乎要敬陪末座？根據這項國際性的研究計畫所得資料，美國的教育投資最大，學校環境似乎也是最好，却無論九歲或十三歲級的數學或科學，學童們所表現的成績都不如別國。倒是「在家看電視每日超過五小時」的學童數則比誰都多（圖表一）。這就難怪新聞週刊（News Week）將布希總統在不久之前，所宣稱“Education 2000”（美國一項教育改革計畫。即，在第廿一世紀時，美國教育將成為世界第一）譏為“Mission Impossible”，意為“如此下去美國將永遠無法達成此項任務”。

但，記者會上美國教育部、國家科學基金會，都明確地表示，將根據 IAEP 所得資料，重新檢討其中小學的數學與科學教育，並宣布已經編入龐大的預算，以改進其科學師資和課程教材。

(數學科)



(自然科學)



 每日看電視五小時以上學生百分率
  IAEP 平均答對率

〔圖表一〕 IAEP 13 歲群數學科與自然科學各國 平均答對率與每天看電視五小時以上學生百分率之比較 (1992 年 2 月 6 日 美國華盛頓郵報刊載)

我國在這次 IAEP 中，無論十三歲級或九歲級學生的數學或科學的成績，就其總平均答對率而言，都名列前茅，和韓國不相上下，都被美國記者引以質詢：中、韓二國的

教育環境不如美國（例如，每班平均學生數都很大，教育經費又少……），却都有如此優秀的表現等等。但，我們不能只看這些總答對率，就判定我們的科學教育最好。

IAEP 這次評鑑，除了筆試以外，還有實作測驗，也收集不少其他背景資料。如果我們更仔細而虛心地分析研究這些資料，必定能發現許多新的問題和觀念，值得我科教界來深入探討研究。

究竟我們從這次 IAEP 能夠學到什麼？

首先，介紹本計畫的來龍去脈。

IAEP 的緣起

國際數理科學教育評量（International Assessment of Educational Progress）曾在 1988 年舉行過試驗性的國際評量，有美國、加拿大（四個省分別參加）、英國、愛爾蘭、韓國，及西班牙等六國九群參加。評量的是十三歲學生數學及科學的成就。

這一次（1991 年）IAEP 第二次國際評量則有二十國共卅五群，分別參加十三歲及九歲學生的數學、科學，及地理的國際評量。在十三歲群，還有實作測驗（Performance Assessment）。IAEP 第一次評量的經驗、技術，就是這一次 IAEP II 的基礎。

其實，IAEP 淵源於美國自己的 NAEP（National Assessment of Educational Progress）。早在 1969 年，應美國國會的要求，由美國教育部的教育統計中心（U. S. Department of Education's National Center for Education Statistics）主辦過多次的國家教育評量。NAEP 的基本目標不外乎監視並報告美國的教育發展情形與其趨勢。NAEP 曾發展出一套高信度綜合性的評量方法與測驗工具，以收集國家教育成就有關的數據資料。至 1983 年，美國教育測驗社（Educational Testing Services，簡稱 ETS）接手辦理 NAEP。最近再進一步主導這一項國際性的大型研究計畫，IAEP。我國於民國 78 年 2 月，接到我國駐美文化組轉來 ETS 的本計畫主持人 Archie E. Lapointe 邀請函，由教育部委託國立台灣師範大學科學教育中心主辦，並由筆者擔任會議代表，負責國際間連繫以及主導國內有關活動及測試工作。

IAEP 的評量目標

此次國際性教育評量，由美國國家科學基金會（National Science Foundation，NSF）及美國教育部（U. S. Department of Education）所支助，希望能達成下列三項基本目標：

〔圖表二〕 IEAP 參與國一覽表（* 為參加學科別）

參 與 國〔語言分群〕	13 歲 數學・科學	9 歲 數學・科學	13 歲 實作測驗	13 歲 地 理
巴西 (Brazil)	*			
加拿大 (Canada)				
亞伯達省	*		*	*
英屬哥倫比亞省	*	*	*	*
曼尼托巴省〔英、法語群〕	*			*
新伯倫瑞克省〔英、法語群〕	*	*	*	*
紐芬蘭省	*		*	*
新斯科夏省	*		*	
安大略省〔英、法語群〕	*	*	*	*
魁北克省〔英、法語群〕	*	*		*
薩克其萬省	*		*	*
中國大陸 (Mainland China)	*			
英格蘭 (England)	*	*	*	*
法國 (France)	*			
希臘 (Greece)	*	*		
匈牙利 (Hungary)	*	*	*	*
愛爾蘭 (Ireland)	*	*	*	*
以色列 (Israel)〔希伯來文語群〕	*	*		
義大利 (Italy)	*	*		
約旦 (Jordan)	*	*		
韓國 (Korea)	*	*		*
莫三鼻克 (Mozambique, 後來退出)	*			
葡萄牙 (Portugal)	*	*		
蘇格蘭 (Scotland)	*	*	*	*
南斯拉夫斯洛凡尼亞 (Slovenia, Yugoslavia)	*	*		*
蘇聯 (Soviet Union)	*	*	*	*
西班牙 (Spain)	*	*	*	*
瑞士 (Switzerland)〔德、法、義語群〕	*			
中華民國台灣 (Taiwan, ROC)	*	*	*	
美國 (United States)	*	*		*

- 經由國際比較，以指認影響教育成就的文化背景及教育制度各項因素。
- 利用NAEP的工具、步驟與技術做國際性比較研究的可行性之探討。
- 將NAEP的經驗轉移給任何有興趣的國家，幫助其評量研究自己國家的教育進展或其他相關研究。

主辦國美國似乎希望以這次國際比較來徹底檢討其國內科學教育。因為他們已經舉辦多年的NAEP，雖然找出許多問題，却苦於無法定出一基準來比較，所以ETS希望邀請更多國家來做國際性比較，以訂定其科學教育發展的基準線（Baseline）。其他參加IAEP各國也都有類似的目標。當然也難免有些國家的某些人士，也會將此國際性合作研究計畫，當做一種國際競賽。

無論如何，這一項國際性研究，對於各國的科學教育發展，應有遠大的貢獻。因為IAEP能提供許許多多具體的寶貴資料給各國，做為其改革（進）科學教育的參考或依據。

各國參與情形

IAEP的評量分四個部分。主要評量對象為十三歲學生的數學與科學；以及九歲學童的數學與科學。另外還有試辦性質的十三歲實作測驗，主要評量目的為運用或操作器材以解決數學或科學問題的能力。此外，尚有地理一科，我國因其課程教材較特殊，經研討後決定放棄，但參加其他全部各年齡群各科各項評量。

圖表二為各國參加各科各項評量情形。加拿大各省有其獨立的教育系統。加拿大十省之中有九省個別參加IAEP。其中四省，還分成法語群（French-speaking students）和英語群（English-speaking students），分別用不同語文工具參加。有些國家則因其地理上，或文化背景上特殊原因，以城市，或部分地區參加評量。例如，巴西以二城市分別參加測驗；莫三鼻克也只有二城市分別參加，而且只參加十三歲群的數學科（後來不幸因內戰未能完成測驗），又如中國大陸，只在東部20省市參加評量。另外還有些國家因技術上的考慮，只能限在某種語文群中進行測驗，例如，以色列只評量用希伯來文的族群，蘇聯只評量講俄語的族群等。

IAEP評量架構

在學校裏，學生們學什麼，教師們教什麼，以及用什麼教材，都會反映其國家社會的教育文化和歷史的背景、道德價值觀，以及其社會經濟目標。每一國家的課程架構

都有其特殊性，都能反映其課程目標歷史的演變與進展。

但在這種高度歧異的教育國際中，仍然有其相同性。由於人類科技之發展，尤其全球傳播媒體之普及，國際間合作種種活動的增多，甚至國際上經濟的競爭，都提高了國際間許多教育文化的共通性。

許多國家以龐大的經費來支持教育，提高學生的就學率，改進師資與學校設備，以及增設校舍與校區。但，在教育品質或素質的評量上，則仍然留著很大的問題。尤其經濟快速發展的國家，其教育投資不斷地增加，却在另一方面，又擔心或懷疑此項龐大投資的功效。有些人認為教育的投資報酬率太小，也有些人却認為不小。究竟如何？應有一套適當有效的工具和方法來評量。

(一) 基本觀念

在教育統計上，有一個基本觀念，統計就是要比較。只靠一測量值，不論其精確度多高，沒有比較就完全無意義。為評量教育的成就進行國際比較，應可以學到更多，但，由於國際間教育環境的歧異很大，究竟能否做到「公平比較」？當然，如果你要比較各國的公民教育，那幾乎是不可能而且無意義的事，因為各國的政治哲學相差太遠。不過，如果我們比較各國的數學教學成就，則比較有意義。因為國際間數學課程的內容大致相同，而且各國師資培育機構以及課程專家也都將其基本架構放在國際常規(International scale)上。所以各國數學教材、教學方法，以及教科書內容都相差甚少。自然科學的情形也如此。

(二) 尋求共識設計評量架構的過程

在國際性評量研究計畫中，最重要的一件事，就是在評量目標及評量資料上達成共識(Consensus)。IAEP在正式測驗之前，召開過許多次國際會議，也利用快捷郵件，交換評鑑過各國教材及評量資料。目的就是要在評量目標上，獲得一個相當程度的共識，以便發展出一套IAEP的評量目標，做為此項國際性評量的基礎。

另外還有一件重要事必須注意的是：即使各國都達成共識的目標，不一定就能完全“公平”。因為，這些達成共識的目標，可能涵蓋某一國數學課程的60%，但，在另一國則僅代表其40%。各國教育系統中，對某一教學目標的相對重視程度也不一定完全相同。有些國家的教育，可能重視計算技巧，而另外有國家則可能特別重視測量和幾何的教材。甚至在更基本的層次上，各國文化背景上，都可能會有更重要的歧異，例如，在某些國家，某些文化下，視數學為教育上最關鍵性優先的課程；但另一方面，也有些國家認為語文、文學才是最基本而重要的課程。

以上所提國際間教育文化上的歧異，應在解釋評量結果時，適切地一併考量。所以IAEP懇切希望各國了解這種情形，IAEP所計畫評量的部分，只在各國間所能達成共識的部分，也就是說，我們只能運用各國都能同意的評量工具與方法，來測量各國課程中，某些特定而有限的部分課程。

參加本計畫各國代表所關心的另一件事，就是我們的評量內容。究竟評量的主題有那些？評量什麼能力？經由大家慎重討論，最後達成共識。數學與科學的評量架構，都應含有二大領域，其一為教材內容（Content），其二為技能（Skill），這些架構都可以用雙向矩陣（two-dimensional matrices）來表示。

首先，各國代表都仔細檢討過過去三次研究與調查結果，包括1988年第一次IAEP，1986年和1990年的NAEP。這三次的評量架構中，都含有主要教材內容與技能的描述。在1990年的NAEP的評量目標中，則有更詳細的評量教材內容與技能的說明。經由一番熱烈討論之後，IAEP II的參加國代表都達成共識，修訂接受了這一套評量架構。

各國分別根據這一套架構，各自召集課程專家及有關學者與教師，共同研討其本國國小到國中數學與科學的課程。就教材內容與技能兩類，評估單元教材內容與技能，在整個課程中所佔的比率。

1990年（民國79年）5月，在加拿大蒙特婁召開的IAEP各國代表會議中，經各國互相比較與討論，求得各科各項評量教材與技能的類別與佔分比率，做為IAEP II的評量架構（IAEP Frameworks）。

（三）數學科評量架構

參加IAEP的各國代表，都同意採用1990年美國NAEP的評量架構，只需稍做修訂，既省事省力又省錢。數學科的代表們從NAEP 1990年的試題與各國所提供的試題中，根據這套架構選出正式測驗所需要的試題。

這套試題要測驗下列三種數學技能（Mathematics skills）：

- 概念理解（Conceptual understanding）
- 過程知識（Procedural knowledge）
- 解題能力（Problem solving）

以及五個教材內容範圍（Content area）

- 數與演算（Numbers and operations）
- 測量（Measurement）

- 幾何 (Geometry)
- 數據解釋、統計和機率 (Data analysis, Statistics, and Probability)
- 代數與函數 (Algebra and Function)

圖表三圖示「教材內容」與「技能」在評量架構中的關係。圖表四、五則分別表示

〔圖表三〕 IAEP 9歲群及13歲群數學科評量架構

	數與演算	測 量	幾 何	數 據 解 釋 統計、機率	代 函 數 數
概念理解					
過程知識					
解 題					

〔圖表四〕 9歲群及13歲群數學技能評量百分比

數學技能	測驗題佔分百分率	
	9 歲群	13 歲群
概念理解	35	33
過程知識	35	33
解 題	30	33

〔圖表五〕 9歲群及13歲群數學教材內容評量百分率

教材內容	測驗題佔分百分率	
	9 歲群	13 歲群
數與演算	50	30
測 量	15	15
幾 何	15	20
數據解釋 統計、機率	10	15
代數、函數	10	20

各項教材與技能在本次 IAEP 測驗中代表所佔比率。請注意 9 歲及 13 歲群中不同的評量比率。這代表各國中小學數學科教育上不同的重點。還有一點要特別聲明，這些比率或重點並不代表任何國家的數學課程架構，也不代表任何一位專家的想法。這不過是代表參加 IAEP 各國專家們所認定重點比率的平均近似值而已。

關於數學技能 (Mathematical skills)，經過慎密討論，IAEP 各國接受 1990 年 NAEP 數學技能的評量架構。這項架構是美國數學教育家們，經過嚴慎的建立共識過程後所獲得的成果。如上述數學技能可分為三類：即，概念理解、過程知識和解題能力。但，這種分類並不意味著這三類技能之間有任何階段性 (hierarchical)。在各類技能之內，都含有不同複雜性的問題。在任何年齡階段，解題能力包括知識與技能之間的交互作用 (interaction between knowledge and skills)。某一年齡階段時，可能就是複雜的「解題」問題，在另一年齡階段中，則可能就被認為是「概念理解」或「過程知識」的問題。此外，在 9 歲和 13 歲群的學生，由於其不同的認知發展，其個別的解題方法都可能有所不同。運用圖表、數表、圖示和文字敘述，都可以評量相同的概念或技能。問題 (試題) 含義 (Context) 應能幫助我們來辨識它在評量那一項 (類) 技能或概念。

1. 概念理解 (Conceptual Understanding)

當學生能表現下列各項行為或事實時，我們可以說他理解某一概念：

- 能辨識、指認，舉出實例或反例。
- 能運用有關概念的模型、圖表及各種表徵。
- 能指認並運用有關原理。
- 能運用事實與定義以說明概念。
- 能比較、比對，並統整有關概念與原理。
- 能指認、說明，及運用有關概念與符號、代號或術語。
- 能解釋數學概念有關的假設與關係。

這些理解能力有助於解決數學問題，以及數學原理的運用。

2. 過程知識 (Procedural Knowledge)

當學生有如下所列各項行為表現時，我們可以說他具有數學的過程知識：

- 能判別或判斷具體模型或符號運用方法 (symbolic methods) 過程的正確性或適切性。
- 能修改或充實解題過程。

- 能運用各種不同的數學邏輯 (numerical algorithms) 以有效率地解決數學問題。
 - 能讀、能設計圖表以表現過程。
 - 能執行幾何構圖。
 - 能操作非計算技能，如：四捨五入、排序等。
3. 解題能力 (Problem Solving)

此項能力包括下列各項能力：

- 推理 (reasoning) 與分析 (analytic) 能力。
- 指認並形成問題的能力。
- 判辨數據的充分性 (sufficiency) 與均質性 (consistency) 。
- 運用策略、數據模型。
- 產生、修訂、充實過程。
- 空間推理、歸納推理、演繹推理、統計推理、比例推理等能力。
- 判斷問題答案或方法的正確性。

(四) 自然科學科評量架構

這次 IAEP 自然科學的評量架構，也和數學的一樣，修訂採用最近二次 NAEP，以及上次 1988 年 IAEP 所用的評量架構，包含三項認知技能 (Cognitive skills)：

- 知識 (Knows)
- 應用 (Uses)
- 統整 (Integrates)

以及四個教材內容領域 (Content areas)：

- 生命科學 (Life sciences)
- 物質科學 (Physical science)
- 地球與太空科學 (Earth and Space science)
- 科學的本質 (Nature of science)

圖表六的雙向表說明這三項技能和四個教材領域的評量架構。圖表七、八則表示其不同的佔分比率。但這些佔分比率並非代表，或主張其課程標準，也不是代表任何國家的課程結構。這些佔分的分配，只是用以保持評量試題範圍的不偏頗而已。

〔圖表六〕 I AEP 9 歲及 13 歲群自然科學評量架構

	生命科學	物質科學	太 空 與 地 球 科 學	科學的本質
知 識				
應 用				
統 整				

〔圖表七〕 9 歲及 13 歲群科學技能評量百分率

科 學 技 能	測驗題佔分百分率	
	9 歲群	13 歲群
知道科學事實，概念與原理	45	40
能應用科學知識以解決單純問題	35	35
能統整科學知識以解決複雜問題	20	25

〔圖表八〕 9 歲及 13 歲群自然科學各科內容評量百分率

教材內容	測驗題佔分百分率	
	9 歲群	13 歲群
生命科學	35	35
物質科學	30	35
地球與太空科學	20	15
科學的本質	15	15

科學技能可分三層次，係取自 1896 年 NAEP，測試三層不同的解題能力，分述如下：

1. 科學知識 (Knows Science)

在這一個層次的試題，是在測驗學生對事實 (fact) 現象的單純記憶或認識，包括：

- 能記憶特定事實、現象、概念、原理與科學方法。
- 能正確運用科學術語 (terminology)。
- 能在不同的情境內指認所學之科學知識。

- 能轉移所學知識為另一型態 (format)。

這一層次的認知活動是所謂的 one-step cognitive approach，包括單純的記憶和轉移，相當於 Bloom 氏 Taxonomy 中的 1.00 知識 Knowledge 和 2.00 Comprehension 的一部分 (2.0 轉移)。

2. 知識的運用 (Uses Science) 在這一個更上一層的層次，所評量的是“能將所學的知識連結有關原理、法則及公式”，用於新的情境。包括下列技能：

- 能將所學的基本知識或技能用於新的、具體的情境。
- 能運用科學基本知識以解釋數據。
- 能指認所觀察之概念、事實、原理之間的相互關係。
- 能指認所收集資料與有關概念、法則、原理之間的相互關係。

這一層次的認知活動為所謂的 two-step process，相當於 Bloom's Taxonomy 的 2.00 Comprehension 的一部分 (2.30 推理) 和 3.00 應用的層次。

3. 知識的統整 (Integrates Science)

在這一層次所要評量的是，解決問題過程中各項因子的組織能力，包括綜合與分析能力：

- 能運用科學概念、原理，以分析問題。
- 能運用邏輯根據數據下適當的結論。
- 在特定情境解決問題過程中，能指出 (評鑑) 解決問題最佳途徑。
- 能運用其他高層次的認知技能，以解決問題等。

總之，這一層次的認知技能，包括通則化 (Generalizing)、假設 (hypothesizing)、類比推論 (reasoning by analogy)、歸納演繹 (induction and deduction)，以及綜合與模型化 (synthesizing and modeling) 等能力，相當於 Bloom 氏分類的分析、綜合及評鑑等三層最高成就。

(五) 評量試題之編製與調適

根據各科評量架構，再收集或自行開發適當的評量試題。在評量試題的發展方面，首先都由各國根據評量架構，分別提供其數學與科學的評量試題。這些五花八門各色各樣的測驗題，都由 ETS 整理過後，用 DHL (國際快捷郵件) 分送給各國代表審查，提供修訂意見，並註明所評量的教材內容或技能是否包含在其本國課程範圍內。

我們先後審查過二千題以上由各國所提供的數學與科學測驗題。每一題都做過 Rating (評分)，以表示「願不願接受該試題為此項國際評量的試題」。ETS 根據各國

的評分，經篩選而產生「候選試題」約150條，在蒙特婁會議，由各國代表逐題討論，最後選出約120條試題，於1990年元月，在各國分別舉行試測（Pilot test）。根據各國試測結果所得之數據，再經過討論後做成正式測驗的試題。

雖然IAEP採用美國NAEP的架構，但其試題則由各國審查或提供，也就是說：這一套評量工具都是每一個參與國所同意的試題。其中有三分之二（在13歲群）與二分之一（在9歲群）的試題，都由美國以外的國家所提供。所以我們可以這麼說：IAEP的評量試題，雖然不能代表每一個參加國家各該科的課程，但，至少都可以代表各國的共識。

在非英語國家，還有翻譯的問題。各國分別各自翻譯成其本國語文之後，還得送至ETS，由語文專家再翻譯回英文（Back translation），以比對其翻譯內容的正確性。但，由於各國文化背景不同，某些程度的修改是被允許的，例如，數學的小數點的表示方法（用“點”式“逗點”），測量的單位（公尺、英呎、……）、人物姓名、地名、物名、動植物名稱等，都可做適度的調整。除此之外，必須“忠實地”翻譯原題，以維持評量內容的一致性。

（六）背景資料問卷

為研究可能影響數理教育品質的諸因素，特地發展學生問卷、學校問卷，及國家問卷等三種工具，以收集資料，做為分析研判評量結果的依據或參考資料。

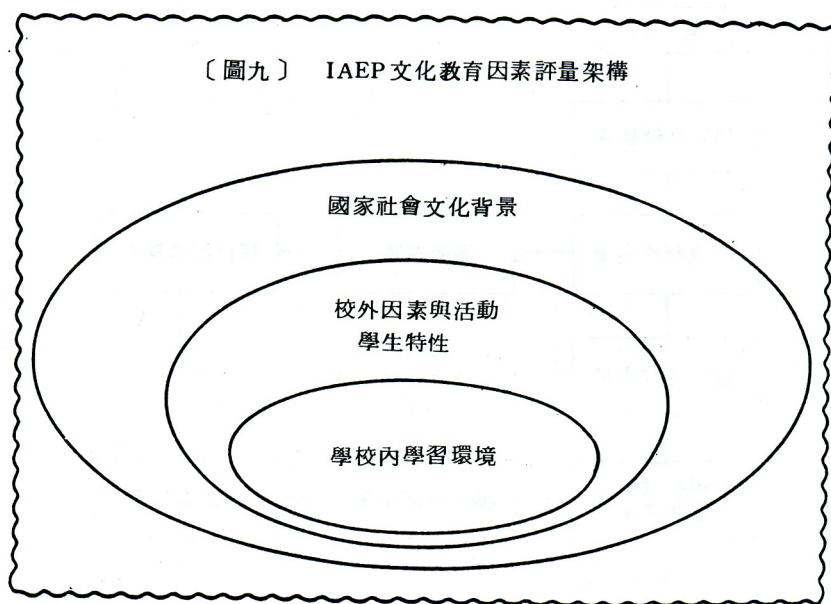
不同的國家（族群）對同樣問題的回答，都可能有其不同的背景。例如，有關家庭作業的量，1988年第一次IAEP分析結果顯示韓國、西班牙的學生都表示「做很多家庭作業」，但在美國，只有高成就的學生和低成就的學生，才做較多的家庭作業，而普通一般學生則很少有家庭作業。這些問題，必須在解釋取樣時候要一併考慮。雖然如此，這些背景資料有關問卷，應能幫助我們分析研判評量結果，使我們更了解教育的現況與問題。

三種問卷內容簡介如下：

- 學生問卷（The Student Questionnaire）：為收集受試學生的家庭及學校背景有關資料，包括其在學校的經驗，家庭環境，其求學的態度及抱負，特別是對於數學及自然科學的學習態度等。這些問卷就是各群各科題本的第五部分，在筆試時由全部受試學生填寫。
- 學校問卷（The School Questionnaire）：有關學校特性、教職員、教育資源、數學及自然科學課程與教材等資料。本問卷請各受試學校的校長負責填答。

- 國家問卷 (The Country Questionnaire) : 由於此次 IAEP 參加國達 20 國, 分佈至歐美、亞洲、非洲、中東, 正可收集很多不同的資料, 以便研究各國不同教育制度與文化下, 數理教育的成就等。當然單靠這些問卷資料, 僅能做粗略的解釋, 例如, 某些國家教育依賴或受制於嚴密的考試系統, 另一些國家則可能在較開放的評量制度下發展等。這些不同, 當可反映其基本的教育價值觀和教育方法論的不同。

為明確表示本計劃的研究架構, IAEP 特地發展 IAEP 文化與教育因素評量架構 (如圖九)。這是由本計劃中一個背景問卷委員會中, 由各國代表共同擬定, 做為各種問卷設計的依據。這個模型為三個同心圓的架構, 表示國家、學校、家庭, 對於數學與科學學習成就的可能影響因素。

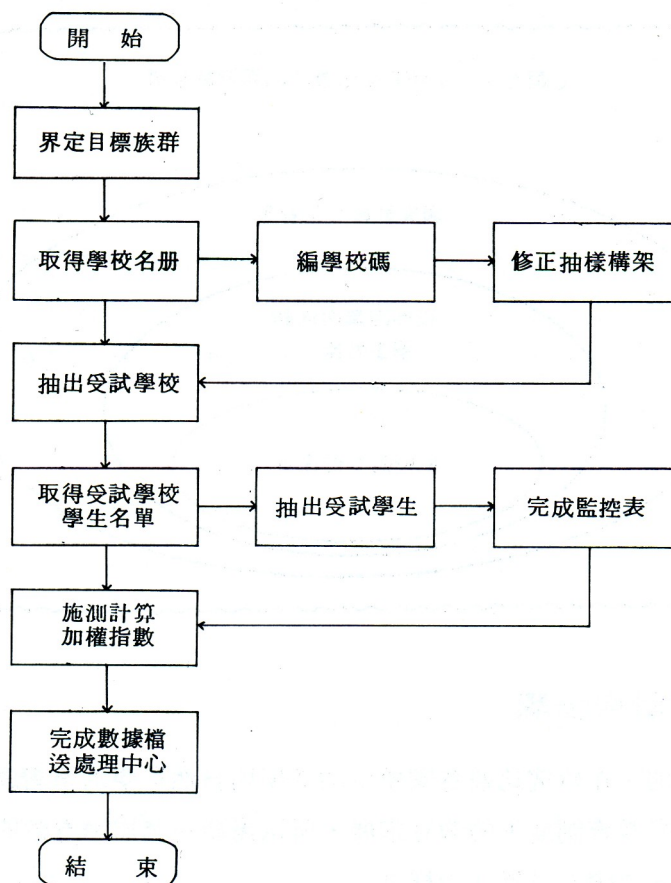


IAEP 的抽樣設計與步驟

IAEP 主要目的, 在研究比較各國學生的數學與自然科學的學習成就, 同時探討各國在不同文化背景與教育制度下的教育進展, 所以需要一套周密有效的取樣方法, 以便抽取真正能代表其各該國 (母群) 的樣群。

1991 年 IAEP 的抽樣方法, 主要根據 1988 年的 IAEP 的經驗而發展。基本上是所謂「二階段分層叢集抽樣法 (Two-stage stratified cluster sampling) 。

在第一階段，先抽學校。每一所學校都有機會被抽選。較小型的學校，則可能由數個學校，先構成一個學校叢集（School cluster）來參加抽選，這一群學校，稱為「超級學校（Super school）」。基本上，學校的抽選，必須按其規模大小，根據其學生人數在母群中所佔之比率來參加抽選。結果，學校愈大其抽中機率也就愈大。所以第一個步驟就是將母群中所有的學校，按其學校規模大小（以學生數為準）、地區（城市、鄉鎮、邊遠）給與分層（Stratift），根據其學生總人數分為五十五個分段（Zone），使每一段的學生總人數相等。再由各分段中，隨機抽出二校為受試學校。由於各校學生人數顯殊，愈大型的學校在「段」裡所佔位置愈大，被抽出機率也就較大，反之愈小型學校，抽中機率愈小。但，由於較小型學校為數較多，結果，以每一位學生來說，被抽選的機率還是一樣。



〔圖表十〕 IAEP抽樣設計及步驟

總之在第一階段，先分五十五段，每一段抽出二所學校也就是說各年齡群（13 歲群和 9 歲群）各抽出 110 所為本次 IAEP 受試學校。

第二階段為受試學生的抽選，並非使受試學校內每一位學生都參加測驗。根據學校所提供「合資格學生名單」以隨機方式抽選 32~34 名學生為受試學生。再以隨機分配其中半數學生為測驗數學，另一半學生則為科學的受試學生。

當測驗完成後，還要根據實際參加測驗（學校與學生）的比率，計算其「抽樣加權指數（Sampling weights）」以做為「變異數估計（Variance estimation）」的依據。整個抽樣工作的流程如圖十。

IAEP 的施測——統一標準步驟與監控系統

由於參加國多達 20 國，各有其特有之社會文化系統與教育制度，為求比較研究之合理性，關於施測工作，IAEP 研擬了一套周密的標準化與監控系統。

（一）監考過程與技術：IAEP 印發極為詳盡的手冊，以規範整個施測過程，內容包括測驗前、測驗中，及測驗後，應檢查、應說明、應填寫、應執行每一項細則。我國則由省市廳局聘請各縣市教育局長為主試，各受試學校校長為主考，再由各校主考聘請各該校適當教師為監試員。再由教育部委請台灣省國民中學教師研習會及國民學校教師研習會，分別在豐原與板橋各該會，各舉辦一天的本計畫工作協調會，除由國教司林司長來發說明本計畫的主要性之外，由科教中心詳細說明本計畫的背景及目標，以及施測的過程細則，監考規則，主監考的職責，並協調各有關事宜。

（二）觀察員：除主監試之外，IAEP 亦規定各國均應設置觀察員（QC observer），以抽樣方式到達各校監控並記錄其施測情形。我國則由省市廳局聘請各三十位督學等各省縣市視導人員，到各受試學校視察並照 IAEP 所印發“品質管制手冊”紀錄並由本中心收齊後送 ETS 彙集備查。除本國所聘觀察員之外，尚有 ETS 所派來的國際觀察員（Site observer）Susan Chyn 女士來台觀察，由本中心人員陪同訪視台北市內的受試學校。

正式施測與評分、數據處理

照 IAEP 研究工作進度，各國都在 1991（民國 80）年 3 月份分別舉行筆試。我國則經本計畫指導委員會決議，於 3 月 15 日（星期五）上午 9 時，9 歲及 13 歲群的數學與自然科學，同時舉行筆試。

施測過程在步驟上有許多規定，凡是教育評量都必有的標準化過程。但，由於事先在豐原及板橋教師研習會所舉辦的工作協調會，已取得共識，又有詳細的施測有關手冊，施測過程相當順利。從當天下午起，各校評量資料，都從全省各地區陸續回本中心。

收回的題本中，數學科先要經過閱卷程序。因為數學科的試題中有部分開放試題 (Open-question)，即，非選擇題。均由本中心聘請數學研究所及數學系專家，依照 IAEP 評分手冊評分初閱。然後再由另一位評分者複閱，才鍵入電腦建立數據檔 (Data file)。自然科學題本，則全是四選一的單選題，可直接鍵入電腦。鍵入的過程都是二次 (First entry 和 Second entry)，如果第一、二次鍵入資料有所不同，電腦會有警告聲，以便檢查修正。

實作測驗照 IAEP 工作進度表，應在筆試之後，四月底之前施測。並依照規定應由筆試的受試學校中，隨機抽出其四分之一，即 27 校為實作測驗的受試學校，再由各該校的各科筆試受試學生中，隨機抽出數學、自然科學各 12 位為實作測驗的受試學生。

由 IAEP 總部取得本項測驗所需之標準器材共有四套。本中心由 3 月底起，至 4 月中分批施測。施測者都事先根據 IAEP 實作測驗監考及評分手冊共同研習，以求一切步驟之標準化。施測結果，亦經記分 (Scoring or rating) 及複閱手續，再經過專家小組開會審核後才鍵入電腦檔案。

同年五月中，我們將全部資料輸入電腦，建立數據檔 (Data file)，在 IAEP 所規定時限之前，就以 DHL 送至設在加拿大的 IAEP 數據處理中心 (Data Processing Center)。這個中心負責收齊各國的數據檔後，一一統計，在同 (1991) 年 9 月，邀請各國代表到美國華盛頓 (或紐約) 分析討論。

IAEP 結果討論

ETS 在今年 2 月 5 日召開傳播媒體記者會，正式發表的成果只包括 13 歲及 9 歲級數學與科學的筆試部分。其他還有實作測驗，因數據處理較複雜，又非原來計畫中的主要部分，將在本年內再做後續研究後再發表。

本文中也先討論這些筆試的結果。

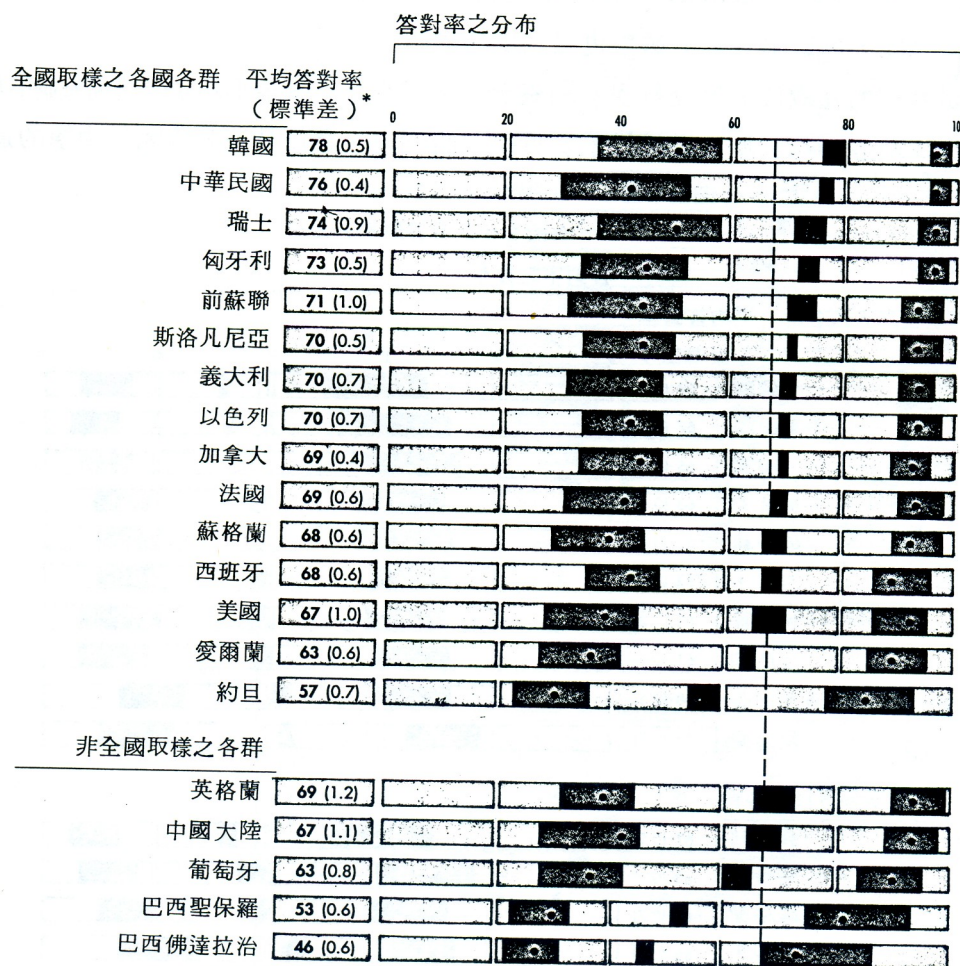
如圖表十一，IAEP 將二十個國家分為兩群來報告。上面由韓國至約旦為全國取樣的一組，共有十五國，評量結果應可代表各該國的這一年紀層的學童。我們中華民國屬於這一組。另外的下面五國，則或因無法取得全國教育當局的合作，或因學籍資料不全，或因其他技術上的理由不能由全國各校來取樣，這一組為四國五族群，自成一組。中共

只以東部二十省市參加評量，所以屬於這一群。

(一) 自然科學與數學各科筆試平均答對率：

圖表十一～十四中，各國都有一長矩形的「答對率分布」。在中間以矩形表示部分為「平均答對率」，其寬度的信賴區間 (Confidence interval) 有的寬，有的狹。我們十三歲科學的平均答對率為 76%，標準差 0.4。韓國的平均答對率為 78%，瑞士為 74%，匈牙利為 73%，但這四國的信賴區間互相重疊，所以在統計上，這四國之間並無顯著差異。如說第幾名，這四國應該都並列第一名。

〔圖表十一〕 13 歲群自然科學
各國各受試群答對率之分布



* 標準差為 Jackknifed Standard Error

平均答對率的信賴區間

第 1~10 百分點，第 90~99 百分點；中間的點為第 5 與 95 個百分點

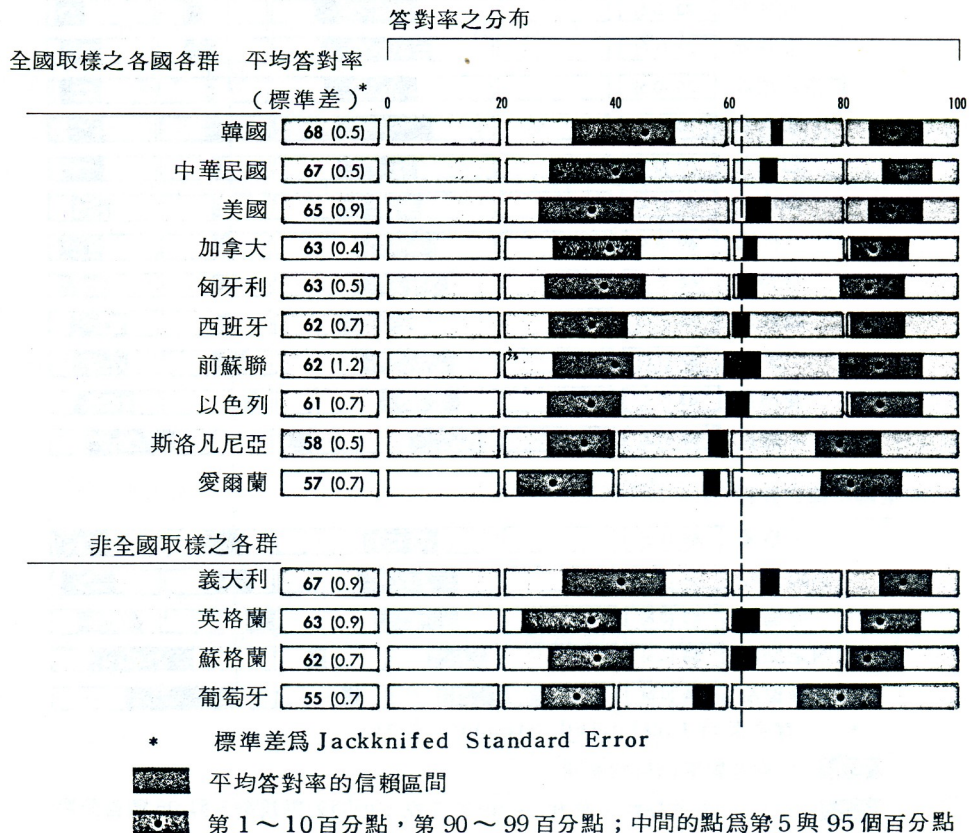
在十三歲級數學（圖表十三），我國以平均答對率 73%，和韓國並列第一，九歲級數學，以答對率 68%，與匈牙利並列第二，九歲級科學，則以 67% 的平均答對率，僅次於韓國（68%）名列第二，但，統計上都無顯著差異。所以由整體成績（Overall results）來看，我國的成績和韓國同列第一名，實值得大家欣慰。

再進一步觀察各國答對率的分布，先看圖表十一自然科學，各國分布右端的矩形表示其高得分群（前 10%）的分布，其中一個黑點就是第 95 百分位數（95th percentile）者的平均得分，如以高分群來比，無論科學和數學（圖表十三），我們的成就也最高。但，我們應該注意到左邊的矩形「低得分群（末 10%）」的分布。其中一黑點標明第 5 個百分位數（5th percentile）者的平均得分。由此表可見，我們的低分群的成績竟名列第十一位，倒數第五！（數學也名列倒數第四。）

可是我們的九歲群，無論科學（圖表十二）或數學（圖表十四）這種現象並不顯著，低分群的成績並不特別壞。我們似乎要重視這個問題。我們是否忽略國民中學的這一群

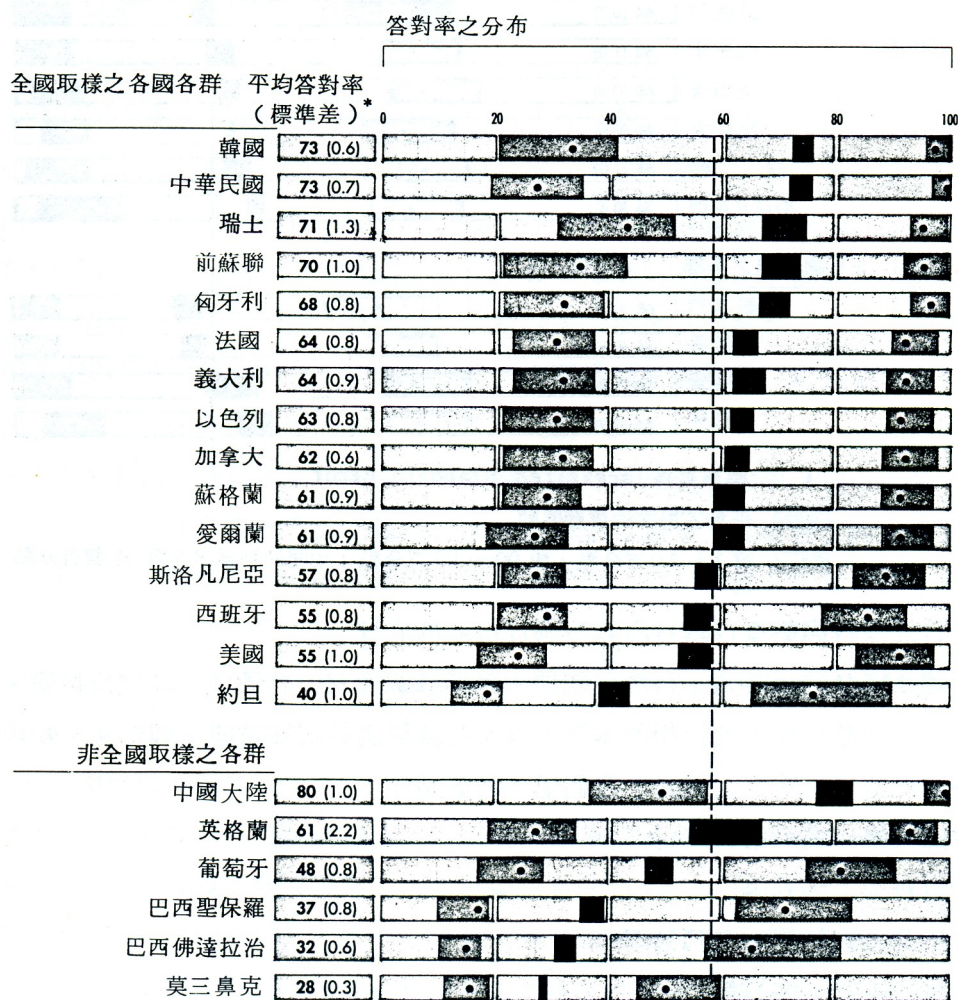
〔圖表十二〕 9歲群自然科學

各國各受試群答對率之分布



低成就的學生？從九歲到十三歲，學生的個別差異愈大，如果我們的制度，對這群學生有所忽略，這種累積（Cumulative ignorance）可能造成這種後果。對於低成就學生的學習輔導、課程教材，或其他因素？似應進一步深入研究。究竟這一群學生也是我們未來的公民，社會中的一成員，他們應有權接受更好、更有意義的、基礎的數學和科學教育。

〔圖表十三〕 13 歲群數學
各國各受試群答對率之分布



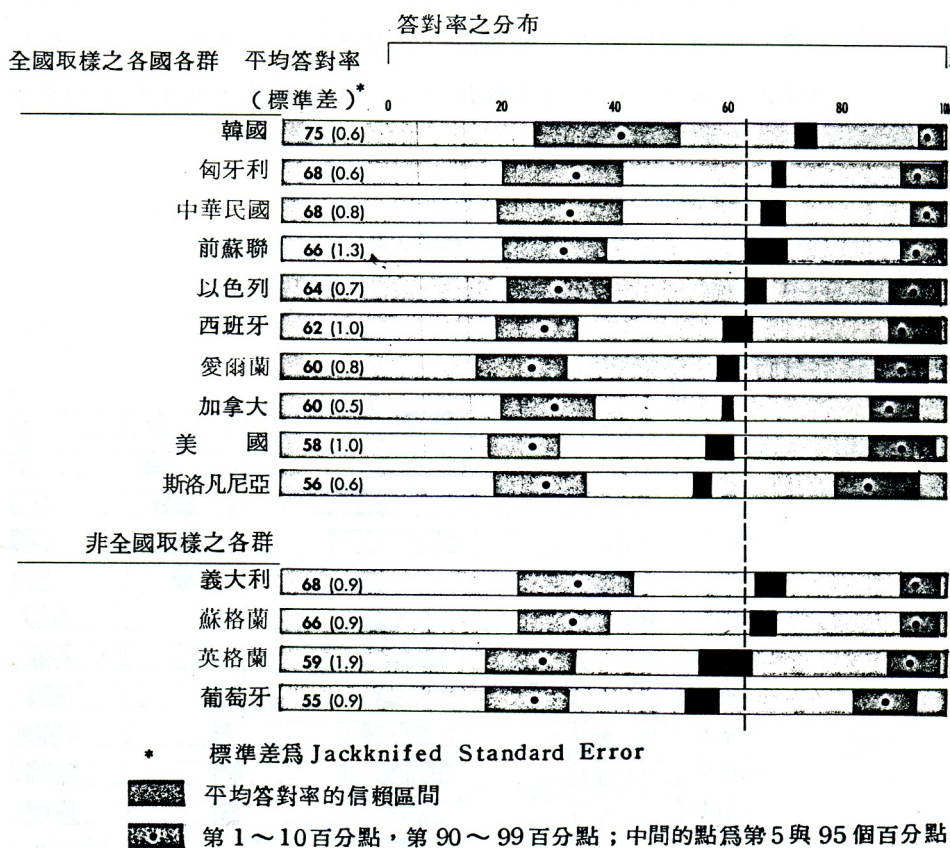
* 標準差為 Jackknifed Standard Error

平均答對率的信賴區間

第 1~10 百分點，第 90~99 百分點；中間的點為第 5 與 95 個百分點

〔圖表十四〕 9歲群數學

各國各受試群答對率之分布



(二) 自然學科與數學科教材內容與過程技能的成就

根據 IAEP 評量架構，自然科學題本包括四個教材內容領域，即生命科學、物質科學、太空及地球科學，及科學的本質，以及三個層次的認知技能，即知識、知識運用，及知識的統整等。在十三歲級，我們學生的成績如下：

生命科學：名列第二

物質科學：名列第二

太空及地球科學：名列第四

科學的本質：名列第六

我們十三歲的學生，大約有 3/4 為國中二年級，另 1/4 為國中一年級，幾乎沒有學過多少地球科學，其成績稍差，應可理解，但，「科學的本質」則不應那麼差。

所謂科學的本質，是指跨學科的科學素養，目的在於測量學生科學的思考技能，也

就是探討科學的能力，例如：觀察、分類、推理、解釋數據、形成假設、設計實驗、進行科學探討的能力等。我們學生的科學知識不差，但，科學過程技能則較差。

在數學科方面，根據評量架構，分為下面五個內容領域，我們十三歲群學生的成績如下：

數與演算：名列第二

測量：名列第一

幾何：名列第二

圖表、機率、統計：名列第二

代數：名列第四

我們分科成績並不差，但，數學技能就有一點問題。如圖表十五，我們的「概念理解」、「解題能力」都名列第二，而「過程知識」則名列第一，但，請注意瑞士，解題能力上的成績超過韓國與我國。關於這一點，瑞士的代表曾說過：「瑞士的學生表現以總分而言，不如韓國與台灣，但，我們在較難、較有深度的試題上，都超越他們（指韓國與我國）。假如，IAEP 不用“答對率”來統計，而根據試題難度，或其認知層次配分，例如單純記憶的試題配分 1 分，而較有深度、有關過程技能較難、較需要思考的試題，則配分 2～3 分，那麼前幾名的名次可能有所更動。」

再看圖表十五，我國學生中同意「學數學主要要靠記憶」的學生佔 54%，屬於多數的一群，但，同意「學會解決問題的方法，和求正確答案一樣重要」的也有 87%。顯然有一半以上的學生認為要學好數學，主要靠記憶，也要學會「解題方法」。

我們的學生是否只憑記憶來學數學和科學？科學的思考能力、探討能力、解決問題能力，和創造性思考是否被忽略了？都值得進一步探討。我們的孩子們都很聰明，也很用功，但，我們的教育方式是否很正確？

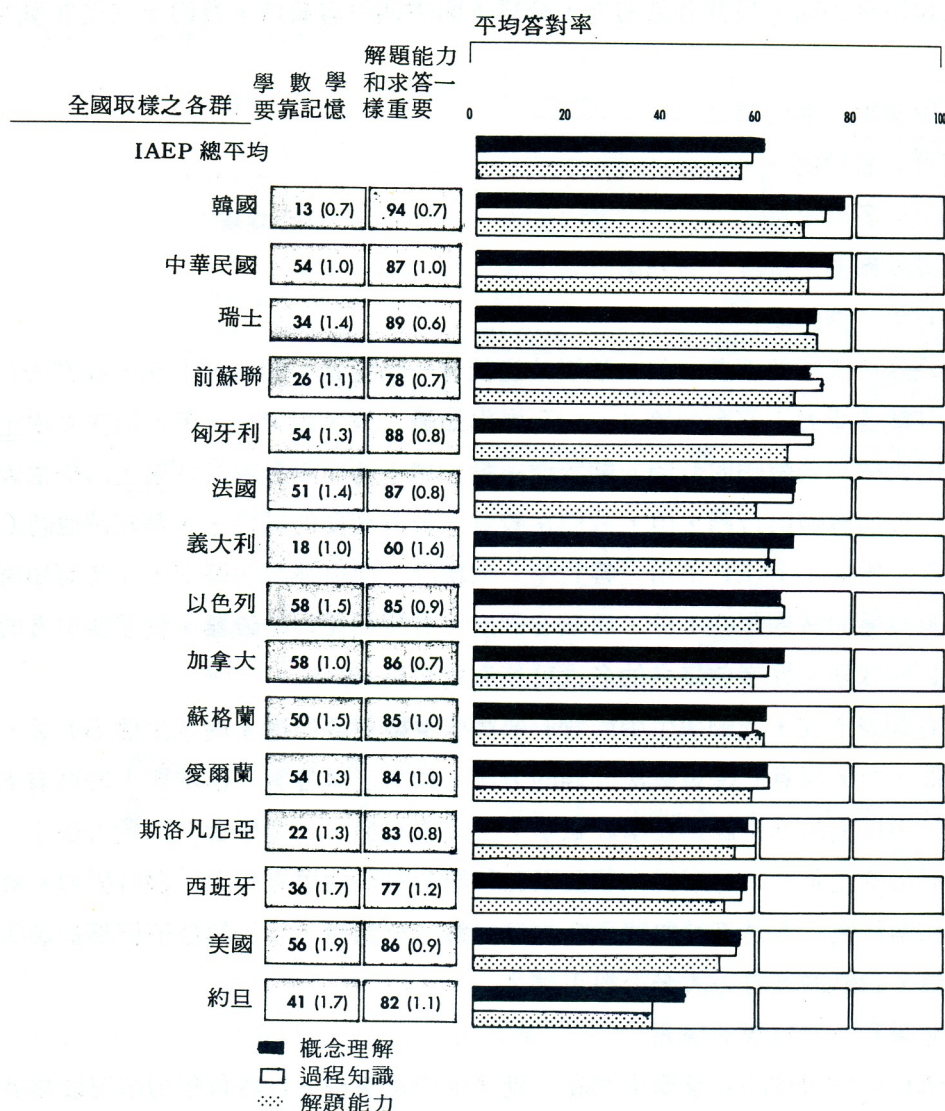
（三）數學和自然科學的學習

IAEP 根據學校問卷及學生問卷，就學生們在學校上自然科學的情況收集並統計一些資料如表十六。由表，我國每週上自然科學的時間為 245 分鐘。括弧內有三個星號表示其“Jackknifed 標準差”達 9.9 以上，各受試學校問卷所提供的數值有很大的變異。或許真的各學校上自然科學課的時數就有那麼大的差異？或許有些學校連一年級的生物、健康教育的時間都算進去？都需要進一步查證。但，在二十個受試國中，繼前蘇聯、中共、斯洛凡尼亞，我們上自然科學的時數排名第四。

第二欄表示每天都「聽老師講科學」的學生只有 25%，在 20 國中排第八。第三

〔圖表十五〕 13歲群數學

關於數學學習的看法百分率與各認知過程有關試題之答對率



欄表示「從未做科學實驗」的學生也只有 25%，比韓、瑞、匈、義、以、西、愛、約旦都好，但比蘇格蘭（只有 3%），加、蘇（13%）、英（2%）則較差。

第四欄問到「考試」和「家庭作業」。我們的學生之中，表示每週都至少有一次自然科學的考試者有 67%之多。韓國的學生則只有 21%，瑞（18%），義大利 9%……和我國同屬一「級」的有前蘇聯（88%）、美國（69%）、約旦（73%），做自然

〔圖表十六〕 13歲自然科學

平均答對率與教學情形

全國取樣之各國各群	平均答對率 (標準差)	學校所表示每週科學教學時數(分鐘)	表示每天都聽講的學生(%)	表示從未做實驗的學生(%)	表示每週至少有一次考試的學生(%)	表示每週都做四小時以上科學家庭作業的學生(%)
韓國	78 (0.5)	144 (2.8)	21 (1.0)	35 (1.7)	21 (1.6)	9 (1.0)
中華民國	76 (0.4)	245 (***)	25 (1.4)	25 (1.3)	67 (1.2)	10 (0.8)
瑞士	74 (0.9)	152 (***)	28 (1.6)	36 (1.7)	18 (1.2)	1 (0.4)
匈牙利	73 (0.5)	207 (***)	40 (1.3)	31 (1.7)	27 (1.6)	13 (0.8)
前蘇聯	71 (1.0)	387 (6.0)	80 (1.9)	13 (0.8)	88 (1.2)	59 (0.8)
斯洛凡尼亞	70 (0.5)	283 (7.0)	16 (1.1)	22 (1.5)	18 (1.0)	7 (0.7)
義大利	70 (0.7)	138 (3.1)	10 (0.9)	59 (1.9)	9 (1.0)	2 (0.4)
以色列	70 (0.7)	181 (***)	0 (0.2)	35 (1.4)	28 (1.9)	4 (0.5)
加拿大	69 (0.4)	156 (1.9)	21 (1.0)	13 (0.7)	26 (1.1)	4 (0.3)
法國	69 (0.6)	174 (8.1)	27 (1.4)	20 (1.7)	47 (1.4)	1 (0.2)
蘇格蘭	68 (0.6)	179 (4.5)	15 (1.2)	3 (0.3)	11 (1.0)	2 (0.4)
西班牙	68 (0.6)	189 (7.2)	38 (1.5)	51 (2.3)	42 (2.6)	12 (0.9)
美國	67 (1.0)	233 (7.9)	66 (1.6)	25 (1.9)	69 (2.0)	7 (0.8)
愛爾蘭	63 (0.6)	159 (4.1)	23 (1.5)	27 (2.1)	18 (1.1)	5 (0.7)
約旦	57 (0.7)	180 (0.7)	60 (1.8)	26 (1.4)	73 (1.8)	12 (1.0)
非全國取樣之各群						
英格蘭	69 (1.2)	194 (4.9)	11 (1.9)	2 (0.6)	8 (0.9)	2 (0.4)
中國大陸	67 (1.1)	331 (***)	23 (2.2)	29 (2.4)	42 (2.2)	16 (1.5)
葡萄牙	63 (0.8)	157 (3.4)	16 (0.9)	48 (1.7)	34 (2.0)	6 (0.7)
巴西聖保羅	53 (0.4)	178 (7.3)	12 (1.0)	35 (1.6)	45 (1.2)	8 (0.8)
巴西佛達拉治	46 (0.6)	124 (3.9)	10 (1.1)	44 (1.9)	55 (1.9)	8 (0.9)

科學家庭作業每週至少四小時以上的學生，我們有 59%，排名第一。在我們的學生中，每週考試次數與家庭作業時數與其 IEAP 成績，也呈正相關。

其他有關實驗設備、教具、計算器、電腦等的調查結果顯示我國各學校的設備並不比許多先進國家差。尤其電腦的持有率較高，但，「用電腦做作業」的最少，只有 6%，排名倒數第二。

根據以上資料來看，我國和各參加國相比，我們的學生上課時間似乎較多，也會做實驗，但，做不做實驗却與其 IEAP 成績無顯著相關（加拿大則呈正相關）。另一方面，考試、家庭作業都比人家多，學校設備並不差，在學習上電腦應用却最差。

表十七為數學科的情形，和科學情形相類似。

〔圖表十七〕 13歲數學

平均答對率與教學情形

全國取樣之各國各群	平均答對率 (標準差)	數學每週教學 時數 (分鐘)	表示每天 都聽講的 學生(%)	表示每天 都自己做 數學作業 的學生(%)	表示每週 至少有一 次小組解 題的學生 (%)	表示每週 至少有一 次數學考 試的學生 (%)	表示每週 至少做4 小時以上 家庭作業 的學生(%)
韓國	73 (0.6)	179 (2.0)	32 (1.0)	17 (1.0)	28 (1.6)	28 (1.9)	33 (1.1)
中華民國	73 (0.7)	204 (2.1)	64 (1.4)	32 (1.1)	38 (1.2)	87 (1.1)	24 (1.2)
瑞士	71 (1.3)	251 (3.9)	60 (2.3)	47 (1.9)	47 (1.5)	40 (2.5)	15 (1.2)
前蘇聯	70 (1.0)	258 (1.9)	62 (1.0)	40 (1.7)	54 (1.8)	52 (1.5)	33 (1.5)
匈牙利	68 (0.8)	186 (2.3)	40 (1.4)	37 (1.6)	55 (1.6)	17 (1.3)	11 (0.7)
法國	64 (0.8)	230 (1.8)	65 (1.3)	—	31 (1.2)	64 (1.3)	17 (1.3)
義大利	64 (0.9)	219 (3.6)	33 (1.3)	10 (0.7)	78 (1.1)	19 (1.6)	27 (1.4)
以色列	63 (0.8)	205 (3.6)	5 (1.2)	12 (1.1)	48 (1.7)	36 (2.2)	17 (1.1)
加拿大	62 (0.6)	225 (1.9)	51 (1.0)	50 (1.1)	40 (1.4)	53 (0.9)	15 (0.8)
蘇格蘭	61 (0.9)	210 (2.3)	23 (1.7)	48 (2.1)	27 (1.6)	17 (1.3)	4 (0.6)
愛爾蘭	61 (0.9)	189 (2.2)	67 (1.7)	54 (1.5)	42 (1.6)	19 (1.5)	17 (1.3)
斯洛凡尼亞	57 (0.8)	188 (4.3)	97 (0.6)	41 (1.4)	43 (1.5)	28 (1.5)	15 (0.9)
西班牙	55 (0.8)	235 (3.3)	58 (1.4)	39 (1.6)	63 (1.5)	31 (1.7)	22 (1.3)
美國	55 (1.0)	228 (5.6)	78 (1.4)	50 (2.7)	49 (2.4)	68 (2.1)	15 (1.3)
約旦	40 (1.0)	180 (0.6)	62 (1.6)	34 (1.4)	83 (1.1)	68 (1.5)	14 (1.0)
非全國取樣之各國各群							
中國大陸	80 (1.0)	307 (***)	74 (2.0)	78 (1.6)	68 (2.1)	63 (2.2)	37 (1.8)
英格蘭	61 (2.2)	190 (4.8)	17 (2.1)	21 (2.5)	44 (3.1)	28 (5.8)	6 (0.8)
葡萄牙	48 (0.8)	207 (2.7)	28 (1.2)	30 (1.6)	51 (1.6)	21 (1.8)	9 (0.8)
巴西聖保羅	37 (0.8)	226 (7.3)	34 (2.0)	35 (1.4)	60 (1.5)	44 (1.5)	16 (1.2)
巴西佛達拉治	32 (0.6)	230 (8.5)	26 (2.0)	31 (1.6)	69 (1.7)	56 (1.9)	18 (1.4)
莫三鼻克	28 (0.3)	217 (0.0)	63 (1.4)	62 (1.6)	79 (1.5)	94 (1.0)	11 (1.2)

(四) 學生在家學習活動

根據學生問卷及學校問卷資料，圖表十八（根據科學受試學生問卷）與圖表十九（根據數學受試學生問卷）也呈現一些有趣的事實。第一欄為自行閱讀課外書籍的學生。我們只有17%或19%的學生會「每天都看自己所喜歡看的書」，在二十國之中，只比韓國（11%）多一點。第二欄每天都做各科家庭作業二小時以上者44%與41%（排名第十一）；表示每天都看電視五小時以上者最少（只有7%或10%，美國、蘇格蘭、英格蘭最多22%~23%）。但，對於科學表示有興趣者我們却排名倒數第二（韓國最少只27%），數學也排名倒數第四！

〔圖表十八〕 13 歲自然科學

平均答對率與家庭活動

全國取樣之各國各群	平均 答對率 (標準差)	表示每天 都自行閱 讀課外書 學生%	表示每天 都做至少 二小時各 科作業的 學生%	表示每天 至少看電 視五小時 以上的學 生%	表示對科 學有興趣 的學生%
韓國	78 (0.5)	11 (0.8)	38 (1.5)	10 (0.8)	27 (1.3)
中華民國	76 (0.4)	17 (1.1)	44 (1.3)	7 (0.7)	51 (1.2)
瑞士	74 (0.9)	49 (1.2)	21 (1.3)	7 (0.6)	59 (1.5)
匈牙利	73 (0.5)	44 (1.3)	61 (1.5)	16 (1.1)	69 (1.2)
前蘇聯	71 (1.0)	48 (1.1)	52 (1.6)	19 (1.3)	66 (1.4)
斯洛凡尼亞	70 (0.5)	43 (1.5)	27 (1.4)	5 (0.6)	78 (1.2)
義大利	70 (0.7)	45 (1.4)	78 (1.2)	7 (0.8)	73 (1.4)
以色列	70 (0.7)	40 (1.4)	49 (1.4)	20 (1.2)	62 (1.6)
加拿大	69 (0.4)	36 (0.9)	26 (0.9)	15 (0.7)	62 (1.0)
法國	69 (0.6)	39 (1.5)	55 (1.6)	4 (0.5)	55 (1.3)
蘇格蘭	68 (0.6)	37 (1.4)	15 (1.5)	23 (1.3)	66 (1.2)
西班牙	68 (0.6)	34 (1.5)	62 (1.9)	11 (0.9)	78 (1.4)
美國	67 (1.0)	29 (1.4)	31 (1.6)	22 (1.7)	57 (2.1)
愛爾蘭	63 (0.6)	40 (1.3)	66 (1.6)	9 (0.9)	57 (1.4)
約旦	57 (0.7)	22 (1.0)	54 (2.0)	10 (0.9)	82 (1.0)
非全國取樣之各國各群					
英格蘭	69 (1.2)	36 (1.8)	26 (2.8)	23 (1.7)	66 (2.9)
中國大陸	67 (1.1)	28 (1.4)	35 (2.1)	2 (0.4)	74 (1.7)
葡萄牙	63 (0.8)	47 (1.2)	30 (1.7)	11 (0.9)	71 (1.4)
巴西聖保羅	53 (0.6)	31 (1.1)	48 (1.9)	18 (1.1)	69 (1.3)
巴西佛達拉治	46 (0.6)	41 (1.2)	50 (2.0)	20 (1.5)	74 (1.3)

我們學生之間，課外閱讀時間與其 IEAP 成績呈正相關（喜歡看書的學生成績愈好）；家庭作業時間多少和成績也呈正相關；看電視時間則呈負相關（電視看得愈多，其成績愈差）；並且表示對科學有興趣的學生，其成績也較好。

此外，還有些家庭環境有關資料顯示我們學生——

1. 家庭兄弟姊妹不算太多，大家庭顯然已不多。
2. 家庭藏書最少（不滿 25 本者達 35 %）。
3. 只有 20 % 表示父母親對科學或數學有興趣（排名最低）。
4. 只有 60 % 覺得家人關心他們的數學或科學課業（排名第九）。
5. 只有 45 % 覺得家人能幫助其課業（排名第七）。

〔圖表十九〕 13歲數學

平均答對率與家庭活動

全國取樣之各國各群	平均 答對率 (標準差)	表示每天 都自行閱 讀課外書 學生%	表示每天 都做至少 二小時各 科作業的 學生%	表示每天 至少看電 視五小時 以上的學 生%	表示對科 學有興趣 的學生%
韓國	73 (0.6)	11 (0.8)	41 (1.7)	11 (0.9)	71 (1.3)
中華民國	73 (0.7)	19 (1.2)	41 (1.3)	10 (0.7)	79 (0.9)
瑞士	71 (1.3)	51 (1.1)	20 (1.3)	7 (0.8)	85 (1.1)
前蘇聯	70 (1.0)	47 (1.3)	52 (1.6)	17 (1.0)	76 (1.8)
匈牙利	68 (0.8)	44 (1.2)	58 (1.3)	13 (1.0)	85 (0.8)
法國	64 (0.8)	40 (1.2)	55 (1.6)	5 (0.7)	81 (1.0)
義大利	64 (0.9)	47 (1.3)	79 (1.3)	5 (0.7)	86 (0.9)
以色列	63 (0.8)	40 (1.7)	50 (1.9)	20 (1.2)	90 (0.8)
加拿大	62 (0.6)	38 (0.9)	27 (1.0)	14 (0.7)	94 (0.4)
蘇格蘭	61 (0.9)	38 (1.5)	14 (1.1)	24 (1.3)	91 (0.7)
愛爾蘭	61 (0.9)	41 (1.3)	63 (1.9)	9 (0.9)	88 (1.0)
斯洛凡尼亞	57 (0.8)	42 (1.2)	28 (1.7)	4 (0.5)	83 (1.0)
西班牙	55 (0.8)	36 (1.3)	64 (1.5)	10 (0.8)	89 (1.0)
美國	55 (1.0)	28 (1.3)	29 (1.8)	20 (1.7)	90 (1.1)
約旦	40 (1.0)	24 (1.3)	56 (2.0)	7 (0.8)	77 (1.5)
非全國取樣之各群					
中國大陸	80 (1.0)	28 (1.5)	44 (1.8)	7 (0.5)	79 (2.1)
英格蘭	61 (2.2)	41 (3.2)	33 (2.8)	14 (2.2)	91 (1.2)
葡萄牙	48 (0.8)	44 (1.8)	30 (1.6)	11 (1.0)	84 (1.1)
巴西聖保羅	37 (0.8)	33 (1.5)	45 (1.9)	19 (1.2)	83 (1.0)
巴西佛達拉治	32 (0.6)	41 (1.3)	48 (1.8)	21 (1.5)	86 (1.1)
莫三鼻克	28 (0.3)	41 (1.6)	42 (1.8)	20 (1.2)	88 (1.0)

6. 有 90 % 表示家長希望他們學好數學（我們的家長們都希望望子成龍，但能幫助其學習的不如外國家庭者多）。
7. 我們的學生似乎都很“用功”，家庭作業做得最多，每天看五小時以上電視的最少，看自己喜歡看（與教科書無直接關係）的書者最少，而對科學（或數學）表示有興趣的，卻比外國學生少得多。韓國的學生也和我國完全一樣！

(五) 社會文化特性與教育制度

參加 IAEP 的二十個國家都有其不同的社會經濟制度、民族文化、學校教育制度。在這些背景下很難做「公平合理」的比較。有些社會文化背景的確能直接影響科學教育，但有很多社會因素並非直接相關。雖然如此，我們深信教育的成就表現絕非獨立單一因

素所能影響。

IAEP 在這次測驗之時，同時收集不少有關各國社會文化資料。除直接由各參加國的受試學校問卷，以及國家協調人的國家問卷（National questionnaire）收集資料之外，還引用 Britannica Book of the Year（1991），P. C. Globe Tempe, AZ：P. C. Globe, Inc.（1990）等文獻的資料。下面是大家可能有興趣的主要發現：

1. 我國都市化 74%，在十五國全國取樣國之中排名第七。
2. GNP 則在十五國中排名第十二。
3. GNP 用於教育經費上只有 3.6%，排名第十三位。
4. 識字者 92%（排名第十四位）。
5. 上課日數最多。
6. 每週教學時數多，排名第五。
7. 每班學生人數最多（韓 49、我 44，其他各國都只有 18-32）。

資料指出我國已經是相當工業化的國家，若以全部二十個參加國來看，我國的 GNP（國民生產毛額）排名第十三，但其用於教育上的經費却只有 GNP 的 3.6%，只得排名第十五名。我們的國民中學每班平均 44 名學生，除韓國 49 名，大陸（20 省市）48 名比我們多以外，其餘各國，連 GNP 比我們低的國家，都有較小的班級，例如匈牙利（27）、約旦（27）、葡萄牙（25）、巴西（38,32）等。事實上每班學生人數超過 30 名的參加國，也只有十九個國家中的五國，其餘各國每班學生平均人數都只有二十幾名。太大的班級，是否導致教師忽略或無法兼顧那一群“低成就”的學生？

結論—我們從 IAEP 學到什麼？

IAEP 有她自己的評量目標，但各參加國似乎也都有自己個別的參加目的。主辦國美國在評量工作告一段落，在做數據整理與解釋的一段時間內，都保留一切資料，等到今年二月五日正式發表時，就召開記者招待會。可見美國主辦單位 ETS 最主要的目的似乎是在引起全美各界，尤其輿論界的注意。無疑地，ETS 是想用科學方法收集資料，再做國際比較以顯出美國數學與科學教育的現況問題，以便訴求於科學教育界，並喚起民衆的關心，以謀促進科學教育改革之機緣。

其實，美國舉行其國內的教育評鑑（NAEP）已有數十年歷史，但這次國際比較，顯然才造成美國教育上的很大壓力。在二月五日記者招待會上，以及其後的媒體報導，我們看得出，ETS 這次是成功了，因為在傳播媒體的關心下，教育當局，包括美國教育

部、國家科學基金會都表示要編列龐大的預算以改進科學教育。

其他各國，當然有些國家將此國際性評鑑當做一種國際奧林匹亞，「力爭上游」，只以拿「第一」為目標；也有些國家想以 IAEP 這次所得資料為其國家發展教育的基準（Baseline），以為其教育政策、制度、實際問題探討的參考點。

我國受美國 ETS 之邀，正式參加這項國際性的教育研究計畫，筆者個人認為主要目的和美國 ETS 者相似。我們希望以這次國際性比較為我們研究科學教育的依據。像一面明鏡，透過這二十國的比較，更清楚地看看我們自己。

評鑑結果顯示我們的學生們表現並不錯，在二十個國家中名列前茅（無論數學與科學），但也顯示下面幾項問題值得研討。

（一）我們學生成就變異很大，尤其十三歲群，其得分之分布也最大。固然高分群的成績都比各國更好。但，另一方面，低分群學生的成績則比各國都差。對於國民中學低成就的學生，無論在課程教材、教法、評量以及學校內各種制度，似乎都需要做徹底的檢討，以便幫助這些低成就學生，在學校中能學得到一些生活上所需要的、基本的數學與自然科學。因為這群（佔10%）學生將來也是社會中不應被忽略的成員。何況只要是中華民國的公民，都應有權接受適當的教育。我們似應再深入探討此問題，究竟是因為班級學生人數太大，使教師無法顧及這一群？還是我們的單一課本引出的問題？或者我們的數學和科學教師缺乏某些重要的訓練？教學方法上是否還有必須檢討、甚至研究的餘地？

（二）根據 IAEP 所得資料顯示：我們大多數的學生，都表示他們的家人都很關心他們的家庭作業是否做好，也希望子女們都學好數學和科學，但，表示他們的家長能和子女談課業內容的學生却不多。比起外國，我們的學生家長之中，能夠（或願意）和其子女談科學、談數學的，都比較少。我們家長們似乎只關心學生是否考好？我們大多數學生的家長們，難道只知道給子女壓力，而沒有實質的幫助？至少在 IAEP 的問卷中，我們的學生確實有這樣的表示。事實是否如此？應再進一步研判。

（三）我們的教育投資，似乎不如許多先進國家的多。也許仍然有人會認為比上不足、却比下有餘。但，如果我們想一想，最近十餘年來中華民國所創造的所謂經濟奇蹟，這不就表示我們的社會在繁榮，也就是正在急速蛻變中。這種社會高速變遷中，如果沒有同步演進的教育，我們未來的公民要怎樣才能負起領導（至少也要適應）這種社會變遷的責任呢？教育投資和一般工商業投資一樣，其投資額和收回成效一定成正比，所不同的只是對於「教育成效」評估的方向與方法而已。

總之我們的學生在學科知識（knowledge level）上得分甚高，但在科學的本質（Nature of science）上以及解決問題的能力上，都有較差的表現。我們的科學教育提倡「科學方法」的教學已久，也都讓學生動手做實驗，然而學生的科學過程技能仍不如人。由 IAEP 其他數據看我們的學生實比人用功，在家不敢看太多的電視，也不太看書（和教科書和考試無關的書），家裡也不買太多「課外」的書，學校裡考試很多、功課重、家庭作業也多。這些數據似乎在暗示，我們學生求知的範圍以及生活圈都非常狹窄。除了應付考試似乎什麼都不重要了！在現在這種變遷快速、高度歧異的未來社會，人與人的關係日見複雜而微妙，科技亦日進月異的社會中，我們似乎更應該認真地檢討中小學的數學和科學教育的基本目標。我們所要培養的絕不是“考場英雄”，只會死記知識的學生，而似乎應該是能選擇、尋找自己所需要的知識，能運用所學知識，以解決問題，為自己以及其所處的社會，開拓新前途，能鑑賞這大自然的和諧運作，以及人類科技社會的互動機制，具有科學素養的未來公民。

我們似乎應早日設法為中小學生建立一個有意義、能充分發揮每一位學生潛能、充滿活力而能享受創造思考的學習環境，讓他們有一個真正快樂的童年生活，為開拓個人與國家社會光明的前途，而學習該學習的智能。