

Asia-Pacific Forum on

Science Learning and Teaching

亞太科學教育論壇

亞太科學教育論壇, 第九期, 第一冊

前言

科學課程學業成就調查的實施

楊寶山

中央教育科學研究所

中國 北京 100088

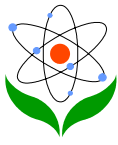
電郵: jks.yangbaoshan@cnier.ac.cn

內容

- [前言](#)
- [科學課程學業成就調查的基本框架](#)
- [科學課程學業成就調查的核心內容](#)
- [科學課程學業成就調查的主要方法](#)
- [參考文獻](#)

前言

國際的學生學業成就調查作為 IEA（國際教育成就評價協會）和 OECD（經濟合作與發展組織）兩個國際組織主持的大規模的評價專案。從調查取向來看，IEA 較為偏重於考查學生基於學校教育的學業情況；OECD 較為側重於考查學生離開學校教育的成年後的生存能力。從調查內容來看，兩者均包括國際上公認的閱讀、數學和科學等核心學科。所有參與的國家大都根據學生學業成就的世界排序，認真檢討各自的課程教學、教育制度、社會文化和家庭環境等眾多方面的影響。此外，美國的（NAEP）調查，英國的（APU）調查，日本的國研調查等都曾產生不同程度的影響。基於此，我們從調查的基



本框架、核心內容和主要方法三個方面探討科學課程學業成就調查的實施具有重要的意義。

科學課程學業成就調查的基本框架

從科學課程學業成就調查（評價）的實施來看，在宏觀層面上，一是直接參加 IEA 和 OECD 等組織的國際學生學業成就調查；二是如美國的 NAEP 調查、英國的 APU 調查、日本的國研調查等，定期組織不同年齡段學生（9、13 或 15、17 歲）科學學科的學業成就調查，建立國家常模，開展縱向或橫向比較研究。

在上述科學學業成就調查中，大都包括成就測驗和背景因素調查兩個部分。其中學業成就測驗分學科領域的核心內容和表現水準兩個維度。關鍵內容指課程內容標準中的核心內容；表現標準是要求學生達到的掌握水準。表現標準劃分的基本框架是布盧姆認知領域的教育目標分類（知識、理解、應用、分析、綜合、評價）。此外，加涅的學習結果分類和安德森等人提出的認知目標二維分類體系也在不斷影響著學生學業成就的調查研究。

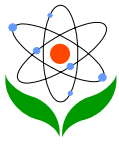
影響學生學業成就的因素，一般分為學習者、課程教學、學校及教育制度四個維度。學習者層面又包括性別、社會背景、自我認知、動機以及教育期望等；課程教學層面包括教學策略、課堂氣氛、班級大小、教師支持、教材使用、學校組織和結構等；學校層面包括學校類型、學校結構、學校資源、學校管理以及課堂實踐等。

TIMSS 的科學評估框架由內容維度和認知維度兩個部分組成。在內容維度中，涉及的領域主要包括生命科學、物質科學、地球科學等幾個部分；在認知維度中，主要包括領會、應用和推理等幾個部分，每一部分又包括不同層次能力的要求。科學探究能力作為主要評價指標，始終貫穿於整個評價的內容與過程。

PISA 的科學評估框架基於終身學習的動態模型。其基本理念基於兩個層面：第一，學生在校的學習是將來所需知識和技能的基礎；第二，學校的功能在於使學生具備終身學習的能力。在評估中，重點測評學生所具備的閱讀、數學、科學等基礎性知識、技能、態度、情感等方面在真實情景中的運用情況。其中科學素養是其測量中的一部分。

NAEP 的科學評估框架包括科學領域與認知要素兩個部分。科學領域涉及地球科學、物質科學和生命科學；認知要素包括概念理解、科學探究和實用推理三個要素。該評估框架規定了 4、8、12 三個年級學生學業的評定水準。

APU 的科學評估框架基於過程技能的評估。主要包括圖表、器材的運用，觀察、設計、分析、探究等幾個主項目。在評估中，重點測評學生所具備的科學過程技能。



在中國內地，有關科學素養的測評框架問題也曾進行過諸多的探討和實踐。特別是新課程實施後，科學素養作為科學課程教學的首要目標。在科學學科的學業成就評價中，更加突出了基於科學的知識與技能、過程與方法、情感態度價值觀三個維度的科學評估架構的科學素養的考察。

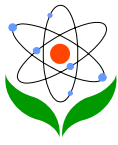
科學課程學業成就調查的核心內容

在 TIMSS 科學評估中，包括內容維度和認知維度兩個部分。在內容維度中，TIMSS 科學評估框架的每一領域包括幾個主題，每一主題涵蓋了多數參與國家的科學課程標準中一列目標，同時提供了每一主題相應的一組評估目標。在認知維度中，包括三個認知領域：一為領會，包括學生需要知道的事實、過程和概念；二為應用，集中在學生遇到問題時應用知識和概念理解的能力；三為推理，超出常規問題的解決方案以圍繞新情景、複雜關係和多步驟的問題展開。TIMSS 科學評估框架中的三個認知領域由不同層次行為區分構成，描述在不同認知領域中學生的技能或能力。領會中包括識記、下定義、描述、舉例說明、使用工具；應用中包括比較/分類、使用模型、關聯、說明資訊、揀出解決辦法、解釋；推理中包括分析/解決問題、整合/綜合、假設/預測、設計/計畫、得出結論、概括、評價、證明。另外，每個領域中列出了與其相對應的具體的行為。

有關學校、教師和學生的問卷調查，主要調查學生科學學習的情況以及對科學學習的信念；教師問卷主要調查教師對科學的信念和關於教學實踐的信念；學校問卷主要調查校長對學校政策和實踐問題的看法。在 PISA 的科學評估中，重點測評學生所具備的科學基礎性知識、技能、態度、情感等方面在真實情景中的運用能力。有關科學能力的測量，PISA 將科學能力定義為運用科學知識的能力、識別問題的能力和在事實基礎上做出決定的能力，幫助人們對自然世界的認識。科學能力被認為是衡量教育產出的一個關鍵部分。該定義並非意味大量知識的貯存，而是強調科學地思考面臨的現實問題。

例如，關於科學的概念。學生需要掌握一定數量的關鍵概念，從而可以理解自然世界中確定的現象和人類活動所導致的變化。關於科學的過程。對科學知識運用和理解能力的評估。亦即獲得、解釋、作用於現象的能力。此外，內容還涉及關於科學的條件和應用的領域等。

在 NAEP 的科學評估中，包括科學領域與認知要素。其中科學領域涉及地球科學、物質科學和生命科學，其中物質科學包括物理和化學；認知要素則細分為概念理解、科學探究和實用推理三個要素。為獲取更多有用的評估資訊，NAEP 設計了學生、教師、學校領導等問卷，用於瞭解學生的種族、家庭狀況、學校類型等相關資訊。



在認知領域中，概念理解要素重點考察學生對科學知識和概念的理解。其中，科學知識包括從學校教育和自然界中學到的各種事實、事件，以及用於解釋、預測自然現象的科學概念、定律和理論。科學探究主要考查學生使用科學工具的能力，包括制定計劃，使用工具獲得資訊，交流探究的結果等。實際推理考查學生在新的、真實世界中運用科學的理解等能力。

在最近十年三次測評中，該項科學評價非常強調對科學概念的理解，它在各個年級所占百分比都幾乎達到 50% 左右，2005 年 12 年級最高，達到 56%；同時還可看出，年級越低，越強調科學探究。隨著年級增加，對學生實際應用能力的要求也在逐步的提高。

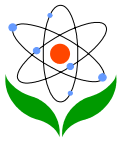
在 APU 的科學評估中，包括主要包括圖表、器材的運用，觀察、設計、分析、探究等幾個主項目。在評估中，為每一個主專案提供了相應的一組評估目標。例如，有關實驗器材的運用，主項包括：量度儀器的運用，物理數量的估算，根據提示進行實際操作等。又如，有關分析與應用的評估，主項包括：分析已經獲取的資料，分辨不同程度的推論，以及評量學生應用科學概念包括物理、化學和生物概念的能力。再如，有關設計探究的評估，主要涉及三種不類型的活動，可以是設計整項探究活動，可以是提出可測試的假設，可以是設計探究活動中的某些步驟。

在中國內地，有關科學素養的測評內容問題也曾進行過諸多的探討。新課程實施後，在科學學科學業成就評價中，更加突出了知識與技能、過程與方法、情感態度價值觀等方面的綜合考察。在科學知識的考查上，突出了密切聯繫學生的學習和生活實際，考查學生對科學知識的理解以及應用知識解釋現象、分析和解決問題的能力。在科學探究的考查中，注意從實際問題出發，關注學生的探究能力。在情感態度與價值觀的考察中，把情感教育滲透到具體的、生動的科學事例與活動中。

從科學學業成就調查（評價）的試題特點來看，在 TIMSS 科學評估的各種題型中，在領會方面，分別從識記、下定義、描述、舉例說明、使用工具等幾個方面入手展開問題；在應用方面，先後從比較/分類、使用模型、關聯、說明資訊、揀出解決辦法、解釋等幾個方面入手展開問題；在推理方面，分別從分析/解決問題、整合/綜合、假設/預測、設計/計畫、得出結論、概括、評價、證明等幾個方面入手展開問題。

例如，在領會方面的問題中，如說明期望學生能舉出恰當的例子支持或闡明有關事實或概念的陳述，識別或提供詳細實例來說明對一般概念的認識；在應用方面的問題中，如解釋則期望學生能為觀測或自然現象提供或識別一種解釋，展示對一些基本科學概念、原理、定律或學說的理解等。

在 PISA 科學評估的各種題型中，重點考察學生所具備的科學知識、技能、態度、情感等方面在真實情景中的運用能力。



在科學的概念方面，要求學生需要掌握一定數量的關鍵概念，從而可以理解自然世界中確定的現象和人類活動所導致的變化。PISA 從物理、化學、生物學、地球和空間科學概念中集中提問。更具體一點，即從這些論題中定義概念，包括生物多樣性、強度和運動以及物理變化。

在科學的過程方面，要求學生掌握對科學知識運用和理解能力的評估。也就是獲得、解釋、作用於現象的能力。PISA 測量五個過程：科學問題的知識、證據的鑒別、結論的提出、結論的交流和對科學概念理解的演示。

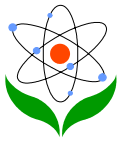
在科學的條件和應用的領域方面，命題時注意到科學能力的應用條件是指日常生活，而不是課堂或者實驗室，這裏包括對整體生活方向的論點，如個人直接關心的事件等。

可以看出，PISA 試題的一個顯著特點就是十分強調知識和技能與日常的生活相關。它雖然考核的是學生科學學科的成績，但卻十分注重從日常生活取材，在設計問題時跳出課程具體的知識內容，大量地運用學生生活的真實情景，注重學生在相應學習領域的基礎能力。學生必須理解關鍵的概念、掌握特定的過程，同時也要有在各種不同情景中運用這些知識與技能。此外，PISA 還特別注重收集學生態度以及學習方法等資訊。

在 NAEP 科學評估的各種題型中，在科學知識方面，要求學生從科學教育以及自然界中學習到的各種事實、事件，用於解釋、預測自然現象的科學概念、定律和理論；在科學探究方面，考查學生使用科學工具的能力，包括制定計劃，使用多種科學工具獲得資訊，交流探究的結果等；在實際推理方面，考查學生在新的、真實世界中運用其科學理解能力。

在 APU 科學評估的各種題型中，有關實驗器材的運用，在量度儀器的運用方面，利用量筒量度液體的體積或細小固體的體積，利用杠桿秤量度物件的品質，利用安培計量度電流等；在物理量的估算方面，估計箱子的體積、葉片的面積、鐵線的長度、包裹的重力、皮球的品質等；在根據提示進行實際操作方面，轉移指定份量的固體化合物和稀酸，將固體化合物和稀酸在試管中混合，然後利用本生燈將試管加熱，最後過濾試管中的液體等。

在中國內地，有關科學素養的測評內容問題直接反映在科學學科的學業成就評價中，更加突出了知識與技能、過程與方法、情感態度價值觀等方面的綜合考察。例如，在科學概念知識方面的考察上，突出了密切聯繫學生的學習和生活實際，考查學生對科學知識的理解以及應用科學知識解釋現象、分析和解決問題的能力。如將食用油、醬油、食鹽、酒精分別放入水中，請判斷哪一種不能形成溶液。實際上，在考察學生對溶液的概念把握的程度。借助這樣一個聯繫生活實際的情境，讓學生利用所學知識回答現實生活中常見的問題。



科學課程學業成就調查的主要方法

在科學學業成就調查(評價)的方法上，宏觀監測大都是以紙筆測驗和問卷調查為主。目前，以自適應理論為依據的電腦因人施測技術和成長記錄袋評價的運用逐漸增加。在問卷調查中，可選擇學生、教師、校長或家長等作為對象；有時進行必要的訪談、實地觀察及相關文本分析等，可進一步瞭解學生學業成就差異的成因。

在調查物件的選擇上，常採用分層抽樣技術。例如，在 PISA 調查中，設計為兩階段分層取樣，第一階段的取樣單元為含有 15 歲學生的學校個體；取樣之前，將在取樣範圍內的學校(學校的最低參與率為 85%)或內隱或外顯地分配到層；然後，以隨機抽樣的方式在每層抽取學校，各層之抽取率與各層學校多少及學校內的學生數量成正比。第二階段的取樣單元為樣本學校中的學生個體，從每一所樣本學校中等概率抽取 35 名學生(底線為 20 名)。所抽取考生的作答率要求達到 80%(底線為 50%)。為了保證樣本的總體覆蓋率，學校和學生層面的總排除率必須低於 5%。

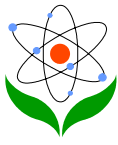
在 TIMSS 的科學知識的測試和調查中，除了測試和問卷調查之外，還包括課程分析、數學課堂的錄影、觀察和有關政策的研究。它的設計集中在學生的三個不同的學校階段：小學後階段、中學低年級階段和中學的末尾階段。由於每個國家學生的入學年齡不完全相同，學生被測試時必須考慮年齡和年級兩個因素。

有關學校、教師和學生的問卷調查，主要調查學生的科學學習的情況以及對科學學習的信念；教師問卷主要調查教師對科學的信念和關注教學實踐的信念；學校問卷主要調查校長對學校政策和實踐問題的看法。

在課程分析中，主要對科學課程的指導書和教科書進行比較，研究科目的內容、內容的銜接和對學生成績的期望等。他們通過對課堂教學進行觀察，對教育管理部門、校長、教師、學生和父母進行訪談。研究的內容包括教育標準、處理個性差異的方法、教師的生活和工作狀況以及學生生活等對學校教育的影響等。

NAEP 的科學評估方法基於相同的評價框架指導，使用相同的評估程式，以及整個評估使用了有效的評估工具：大型題庫和矩陣技術，並在不同年份的測試、不同年級的測試卷中特意安排了一些重疊的問題，與其他類型的測試共用部分相同的樣本，使得測試結果既有縱向可比性，又有橫向可比性。科學評估量表由學生問卷、教師問卷、學校領導問卷等組成。

在 NAEP 的科學評估中，在參加評估的州內根據人口統計學和地理組成進行抽樣。為保證樣本的均衡性，NCES 和 NAGB 規定州和地方學校的參與率不得低於 85%。通常在各州 4 年級和 8 年級各選取 100 所學校作為樣本，再在作為樣本的學校和年級選取 25 名學生參加每個科目的評估。



在測試時間方面，該組織制定了詳細的評估計畫表。全國評估通常與州評估和實驗性城市地區評估是隔年進行並且避免重迭，而全國長期趨勢評估則是四年一次。今年（2008 年）進行全國長期趨勢評估。

在 APU 的科學評估中，由於學生和考題都是以隨機方式抽樣，因此兩方面的樣本都應具有代表性。在學生試卷的設計方面，由於評估所牽涉的內容專案眾多，每次測試只能評估部份項目。每個子項目都設有一個試題庫，部份試題適用於兩個或三個年齡組別，以便於比較跨年齡的成就差異。在筆試前或後，會抽取部份考生進行實作測試。

在學校問卷的設計方面，包括科學教學的師資和資源，進行科學教學的目的，對科學教育的重視程度，分配予科學教學部門的財政資源等。在學生問卷的設計方面，主要考察學習環境與學生學業成就差異的關聯。如學生的性別，種族，修讀科目，職業取向，對有關科學議題的興趣；問卷亦問及學生課餘的興趣及嗜好等。

在中國內地的科學評估中，由於多種原因，我國大陸雖然未整體直接參加 IEA 和 OECD 等國際評價組織的學生學業成就評價專案，但是從 20 世紀 70 年代後期開始，我國的理論界先後參照布盧姆等的教育目標分類框架，從認知、技能、情感態度和學生品德等多方面地進行了引進、吸收和本土化改造。20 世紀 90 年代起，我國在學生學業成就現狀調查和影響因素分析方面進行了多次的探索和實踐。“十五”課題以來，特別是自新課程實施開始，一些專案組、課題組進行了廣泛的理論研究和實踐探索，並且已經取得了一些具有重要意義的成果。

參考文獻

1. 國際教育成就評價協會: <http://www.iea.nl>
2. 經濟合作與發展組織: <http://www.oecd.org>
3. 美國的 NAEP 調查: <http://nces.ed.gov/nationsreportcard/itmlr/>