科學教育學刊
2009, 第十七卷第六期, 459-479

Chinese Journal of Science Education
2009, 17(6), 459-479

中小學學生科學學習成果的評量架構與命題示例之研發

郭重吉[1]　邱美虹[2]　黃台珠[3]　張惠博[4]　張俊彥[5]
周進洋[3]　王國華[6]　陳忠志[7]　譚克平[2]　黃鴻博[8]

[1]國立彰化師範大學　理學院
[2]國立台灣師範大學　科學教育研究所
[3]國立高雄師範大學　科學教育研究所
[4]國立彰化師範大學　物理學系暨研究所
[5]國立台灣師範大學　科學教育研究所暨地球科學系（所）
[6]國立彰化師範大學　科學教育研究所
[7]輔英科技大學　應用化學及材料科學系
[8]國立台中教育大學　科學應用與推廣學系

（投稿日期：民國 98 年 8 月 10 日，修訂日期：98 年 11 月 24 日，接受日期：98 年 12 月 8 日）

摘要：本研究旨在參考評量教育進展的文獻和國內外實施大規模評量的經驗，針對中小學學生科學學習成果研擬評量架構和命題示例。評量架構是針對 4、6、9、12 年級學生，先從生物、化學、地科和物理的學科內容整理重要概念和原理，其次考慮學生科學實務表現，包括重要科學概念和原理的理解、應用、科學探究能力、量化推理能力、科學本質的認識以及態度與情意因素。從這兩個向度的交叉形成雙向細目表，針對選定的細目，敘寫學生在基礎、精熟、進階等三個層級的預期表現，據以發展試題。本研究係屬先期規劃，過程中邀請中小學教師和專家學者參與研發和諮詢；研究成果尚待考驗，但應可供規劃和發展全國性中小學學生科學學習成果調查之參考。

關鍵詞：科學學習成果、大規模評量、評量架構、命題示例

壹、緒　論

中小學生科學素養的提昇是科學教育的重要目標，就學生個人、課室、學校和全國的層次而言，學生科學學習成果的評量在促進科學教學的成效上扮演至為重要的角色。在本節中先就本研究的背景和研究目的加以介紹，在第二節中探討當前科學教育目標、科學教學和評量在理論與實務的相關文獻。第三節簡述本研究所採之研究方法和實施過程，在第四節中就各科列舉數例簡要說明本

研究所發展的評量架構和命題示例。最後，在第五節綜述研究結果的應用和限制，並對未來研究方向提出一些建議。由於篇幅限制，有關文獻探討和研究成果的部份，在本文中僅能簡略介紹，詳細內容請參見國科會研究報告（郭重吉等，2008a）。

一、研究背景

面對當前快速變遷的科技社會，提升國民科學素養已經普遍被列為世界各國教育重要目標之一，透過科學教育，學習者得以理解所處的自然環境，追求更美好的生活與未來，就國家社會而言，成功的科技教育養成優秀科技研究與相關從業人才，奠立國家經濟與競爭力的基礎，而成熟之民主社會也有賴於普遍具備科技基本素養、理性思考、善於溝通合作、不斷自我學習與面對科學相關社會議題作明智決定的公民，方得以實踐。

二十一世紀的來臨，宣示人類進入所謂「知識社會」，我們面對的是一個更複雜、變遷快速的世代，人類除面對石化燃料的逐漸匱乏、溫室效應等所引發的環境與生態危機，全球化也逐漸成為驅動社會變遷的重要力量。然而，在另一方面，地區化、本土化的呼聲亦不絕於耳。在科學教育領域，早期由美國所啟動新科學課程改革運動，曾被引介到世界各地，然而後續的研究卻發現，科學知識的認知或許是跨地區、跨文化，科學教育面對的是人，科學教育過程與成果深受到當地文化與學生所處自然社會文化所影響。近年來，在建構主義的風潮影響下，對於學習者科學知識概念建構、改變歷程、與影響因素有更深的理解，其肯定學習者所在區域文化、個人特性與影響因素的複雜性，在全球化浪潮下，正視科學學習本土化研究有其必要性。

至於跨國之間的比較，在追求相互理解、觀摩學習與謀求改善的基礎下，針對學生科學學習，過去幾十年來，國際間推動多次大規模的學生學習成就調查，包括 TIMSS（Trends in International Mathematics and Science Study）和 PISA（Programme for International Student Assessment），這些調查研究或各有其背景與目標，我國也曾以正式或非正式的名義參加，透過這些大規模的國際調查研究，國際間得以理解學生在各年齡層、科學各領域中之學習狀況及所存在的優勢與劣勢，並對影響學生科學學習之諸多相關因素有更精確的掌握（Guo, 2007）。針對調查結果，參與國家可以進一步針對本身缺失擬定具體科學教育改進之策略。然而，這些跨國研究計畫，或基於跨國比較之特性，所關注或調查的部分主要是以國際間普遍、共同性的部分為主，對於個別區域、文化特徵與特殊關注的議題，則較難著力，為進一步深化我國科學教育研究之本土內涵，有必要進一步建立本土化學生對科學的認知與學習之長期追蹤資料，以作為規劃未來規劃我國科學教育革新的重要依據。

二、研究目的與範圍

基於以上相關研究計畫背景，本研究主要目的如下：

(一)發展評量架構

評量架構是規範評量調查內容與後續工具發展的關鍵，本研究將參照國外既有經驗與成果，並參酌國內教學實際情況，擬定4、6、9、12各年級學生在自然科學方面所應知道的核心知識和具備的重要認知技能，進而設計中小學科學學習的評量架構。並選擇部分較具代表性的教學目標，把不同年段學生在重要認知能力和學習成果上的表現，區分成基本、熟練和進階三個等級，並分別敘明各等級學生學習表現的特性。

(二)設計評量實例

為使得後續調查研究計畫，在計畫執行、工具發展上有一致的觀點和可資參考的實例，針對學生在上述評量架構中的各個具體的預期表現的細項，選擇較具代表性之項目設計命題示例。考慮短期內評量實務的發展和實施之可行性等因素，評量實例之類型力求多元，但仍以紙筆測驗和實作評量為主。

本研究中強調從科學知識和認知技能的結合、學科間的關聯和不同年段之間的銜接，對中小學科學學習評量做比較整體且有系統的規劃。有關實際評量試題、任務和試卷的發展與實施，自應遵守測驗統計與心理計量方面的一般規範，而其相關的細節，例如，命題技術、試題及試卷設計、學生取樣、題本設計以及量尺分數等，這些固然都很重要，但不在本研究探討的範圍。詳細的規範可參閱 The Student Evaluation Standards(Committee on Standards for Educational Evaluation, 2003）以及 Beatty 與 Carlson（n.d.）的簡要說明。

貳、文獻探討

為期有助於規劃合適的評量架構並發展適當的試題，在本節中將簡述晚近學者對科學教育、科學教學和評量的觀點，並簡介一些國內外大規模評量的規劃與實施。

一、當代學者對科學教育目標、科學教學和評量的觀點

近數十年來，科學教育的實務、研究和理論，深受科學史、科學哲學、認知心理學、教育社會學、質的研究法、心理計量理論與技術以及網路資訊科技的發展，而有長足的進展。Bransford, Brown 與 Cocking 編著的 “How People Learn”（National Research Council [NRC], 2000）這本專書中指出，有關學習的最新科學研究的特徵之一，是強調理解性地學習（learning with understanding），也就是將焦點放在認知的過程。他們強調可以將人視為目標導向的行為者（goal-directed agents），自己會主動追求資訊。學生進入正式教育體系時，本身已經具備一些先前的知識、技能、信念和概念，學生注意環境的哪方面，如何組織和解讀這些資訊，都深受其影響。這也就影響他們記憶、推理、解決問題和求取新知的能力（NRC, 2000）。該書也提及，近代學習理論的一項基本宗旨是，不同的學習目標，需要不同的教學方法來配合；至於新的教育目標，則需要在學習機會上有所改變。設計學習環境，則與在學習、轉化和能力表現的過程中特別重要的一些議題有關。而這些過程則深受學習環境是否以學生導向、知識導向、評量導向和社群導向的程度影響（NRC, 2000）。以學習者為中心的環境強調：有效的教學是從學習者對學習情境所帶來的原有知識開始，這包括文化實踐和信念，以及對於教材內容的知識。以知識為中心的環境所注重的是：思考和解決問題能力需要容易提取和加以適當運用的知識。而支持學習的評量方面則強調形成性的評量，提供機會讓學生修正和改進他們思考和理解的品質，以及評量必須和學習目標相互呼應。至於以社群為中心的環境則強調學生在家、社區活動中心、和課外的俱樂部的活動，都可對學生的學術成就產生重要的影響。

以“How People Learn”（NRC, 2000）這本書為基礎，針對自然科學的教與學，在 NRC（2005）Bransford 與 Donovan 特別指出下列三個原則：

1.學生對世界究竟是如何運作有一些原先就有的體認、看法和構想，而他們是帶著這些來到教室的。教師在進行教學時，

如果未能使用學生原先就知道的，則學生可能無法掌握教師所教的新概念和資訊，或是，學生可能為考試而學習，一旦在教室以外的情境時，則又回到他們原有的想法。

2.學生探究能力的培養有賴：

(1)對於基本事實知識的充分了解。

(2)對於事實和想法的理解，是在一個概念架構的脈絡中進行。

(3)以有利於擷取和應用的方式，來組織知識。

3.採用強調後設認知的教學方式，有助於幫助學生經由設定自己的學習目標而學會如何控制自己的學習，並監測自己在達成目標上的進展。

針對中小學科學教育的目標，NRC（2007）在"Taking Science to School: Learning and Teaching Science in Grades K-8"的專書，指出以往的研究經常把精熟科學的知能和表現（Proficiency in Science）視為由彼此獨立的知識和能力組成，但其實理解科學是多方面的。當今的研究顯示，對某方面科學知能的獲得，是與在其它方面的知能密切相關（例如，一個人在熟悉的領域會有較好的推理能力）。因此，該書中強調當前中小學比較寬廣的科學學習目標，應包括下述彼此纏繞在一起的精熟科學的四股知能和表現：

1.知道、使用與詮釋對自然界的科學的解釋；

2.提出並且評鑑科學的證據和解釋；

3.理解科學知識的本質和發展；

4.積極而有成效的參與科學實務（scientific practices，或譯為科學實踐）和對話方面的活動。

在 NRC（2007）的書中強調對幼稚園到八年級學生的科學教育，應該建立在由過去30年在認知和發展心理學以及科學教育方面的研究，對孩童如何學習科學所提供的新洞察力和新的體認之上。基於這些對學習的洞察力，我們需要在科學教育的標準、課程、教學和評量方面做一些改變，讓他們是環繞著精熟科學的四股知能和表現組織而成，並且用逐漸增長的方式，讓學生跨越幼稚園到八年級所建立起的科學核心概念或想法。有關標準、課程、評量、教學和教師專業發展的構想、設計和實施，應該被當作一個彼此符合、相互一致的系統來進行。在此觀點下，標準和課程應該對透過縱貫幼稚園到八年級學校教育持續的教學而得以實現的重要的科學想法和實踐，訂定明確具體而前後一致的目標。評量應該針對學生逐漸形成的新的想法為教師和學生提供及時的回饋，以便支援教師改進教學的努力。師資培育和專業發展應該集中於發展教師在下列方面的知識：所任教的科學、學生怎樣學習科學、以及一些特定方法和科技以支援所有學生學習科學。

針對中小學學生科學認知與學習的評量的發展趨勢，近年來亦有許多專書分別從評量的角色與功能、評量的目的與類型、評量的原理和科學學習評量的研究等方面加以探討（郭重吉等，2008a）。從系統化的教育改革的角度而言，在國家科學教育標準（NRC, 1996）描述的願景中，指出評量是在科學教育系統上面的一個主要回饋機制。例如，對學生而言，評量的數據可提供關於他們達成教師和父母預期的回饋；對教師提供關於學生學習情形的回饋；對整個學區提供教師和學程的有效性的回饋，而對決策者則提供政策運作情形的回饋。回饋將刺激在政策上的變化並導致在科學教育系統的改變，引導教師專業發展，並且鼓勵學生改進他們對科學的理解。

NRC（2003）在廣泛回顧相關的文獻，指出一個理想的評量系統應具有下列的特

性：

1.完整：在一個完整的系統中，會運用許多方法來提供多種證據以支援教育決策。一個精心設計的系統包括形成性評量（支援學生進行中的學習並且幫助教師做教育上的決定）和總結性評量（評量學生在一個學習階段完成的成就水準），藉以促使學生達成一組做得到而且明確傳達的學習成果。可能也包括教學品質量的評量，並提供學生實際學習成果，而非學生應付考試技巧，有所增長的證據。

2.一致：一個一致的系統是：大規模或教室評量的概念基礎或者學習的模型是彼此相容。而且，經由此系統不同的評量所測量的內容、過程和技能也是彼此相容。對於一個系統來說，一致性的要求，要反映在科學教育標準、課程、教學、和專業發展之間的彼此相符，以致於這些元素都是針對相同的一組學習目標。

3.連續：在一個協調的評量系統中，學生的學習進展是持續被測量和評量的，這可以是涵蓋一整個學年、跨越幾個年級、或是包含學生整個在學的時期。評量是在和教學密切聯繫下持續進行的。

4.整合：一個整合良好的評量系統，是被仔細設計，以致於能和一個較大且一致的教育系統密切配合，此教育系統包括提供教師適切的資源和專業發展，以確保教師有能力達成科學教育標準對他們所期望的表現。

5.包括優質的評量：此系統所有的評量都應該具有優良的品質，首先意味著他們必須符合相關的專業標準，其次對優質的大規模和教室評量，在文中有更進一步的說明。

為協助各州設計、發展與實施高品質的自然科的評量，以達成聯邦政府通過的 No Child Left Behind 法案，NRC（2006）在 “Systems for State Science Assessment”一書中報導，他們組成了一個科學學習成就測驗設計委員會，邀集許多知名專家學者一起研究，強調以系統的取向來進行有關評量方面的設計、發展與實施。圖1所示為科學教育系統的示意圖（NRC, 2006），從圖中可以看出評量在科學教育改革中所扮演的角色包括：對教學、學程、教學方案等的改進；讓學生、家長、教師及行政人員了解新課程的目標、精神與特色；了解學生及班級的學習情形以及反應教師的教學績效等。

上述的這些發展趨勢，對於整個科學教育的政策、研究與實務皆有重大的影響。一個明顯的例子是在 Kali, Linn 與 Roseman（2008）書中倡導的宗旨：在設計與發展科學課程、教學和政策上，對於學生的科學學習和評量，宜重視統整的理解（integrated understanding）、知識的整合（knowledge integration）和課程的融貫性（curriculum coherence）。

二、國內外大規模科學評量

多年來，在國際間 TIMSS 及 PISA 等有關學生科學學習評量計畫的推動實施，其結果不但廣受重視，且對各國科學教育政策、研究和實務產生重大衝擊，而其評量架構和試題編製方面的理念和做法，也提供本先期

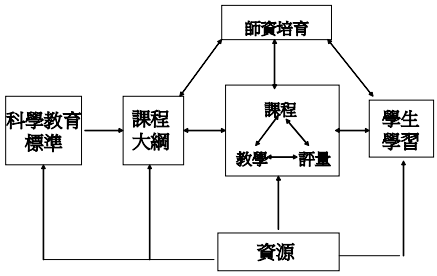


圖 1：科學教育系統示意圖

規劃相當有價值的參考資料。以下先針對TIMSS 及 PISA 可供參考的部份簡要加以介紹，其次簡介美國、加拿大、澳洲等國在這方面的努力，最後再對國內建置中的幾個相關的資料庫略加說明。

(一)TIMSS 2007

TIMSS 是由聯合國教科文組織所屬機構IEA（The International Association for the Evaluation of Educational Achievement）所推動執行的計畫，該組織成立於1959年，至今已有四十八年的歷史，IEA 過去關注國際間教育政策、實務與學生學習的表現，曾參與國家數超過六十個，針對數學與科學學習狀況的調查，過去約以四年為一循環的方式實施，最近幾次分別在1995，1999，2003，2007年進行。這幾次的調查不僅評量學生學習的成就，也對於影響學習的相關因素諸如學校課程、家庭背景、學習科學興趣、性別關連等進行調查與資料收集，而所收集資料經整理建檔，建置資料庫開放給各國研究者進行進一步研究分析。在 IEA 所出版 “TIMSS 2007 Assessment Frameworks”（IEA, 2005）中對於評量架構發展歷程、內容等有相當完整的資料。

TIMSS 2007科學評量架構由內容向度（content dimension）和認知向度（cognitive dimension）兩方面組成。在內容向度中詳細說明了科學評量中涉及的學科領域或教材內容（如生物、物理、化學、地球科學），在認知向度中描述了期望學生在學習科學內容時表現出的思考過程（亦即，知悉、應用和推理）。

TIMSS 2007科學評量架構的內容向度包含許多領域（content domains），每個領域還包括幾個重要的主題（topic areas），每一主題之下包括數項教學目標，這些教學目標是多數參與國家的科學課程標準所強調的。這些目標的敘寫方式是根據學生在作答時可能被引發的行為而設計的，它所代表的乃是對學生的理解和能力的期望。評量架構中還針對每一主題訂定了一組評量目標，如在上表物質科學內容領域中，針對力與運動這一主題，其評量目標包括：辨識使物體運動的常見的力（如落體所受的地心引力、推力或拉力）；描述如何利用天平來決定物體的相對重量；說明不同物體的重量和他們可以浮沉的能力的關聯。

要順利回答 TIMSS 的試題，學生除了必須熟悉科學學科內容之外，還需要運用許多認知的能力。在 TIMSS 2007科學評量架構中，基於學生面對評量中的各種試題時所應該知道和能夠做到的，將認知向度分為三個領域：一為知悉（knowing），包括學生需要知道的事實、過程和概念；二為應用（applying），集中在學生遇到問題時能夠應用知識和概念理解的能力;三為推理（reasoning），超越例行性問題的解決以至於涵蓋陌生的情景、複雜的情境和涉及許多步驟的問題。四年級和八年級的認知領域相同，只是三個領域所占比例，隨著學生從四年級到八年級，在認知能力、成熟、教學、經驗以及理解的深度和廣度的增加，而做相應地變化。

TIMSS 2007科學評量架構中的三個認知領域，本身已具有某種程度的層次，但在發展試題時，對於每一個認知領域會包含難易不同的題目。具體地說，在不同認知領域中，包含一些進一步描述學生的技能和能力的細項。知悉中包括回憶/認識、下定義、描述、舉例說明、應用工具與步驟；應用中包括比較/對比/分類、使用模型、關聯、詮釋資訊、尋找解決辦法、解釋；而在推理中則包括分析/解決問題、整合/綜合、假說/預測、設計/計畫、得出結論、概括、評鑑、驗證。對於上述的細項，在報告中還列出了與其相對應的具體行為。

有鑑於科學探究是目前許多國家科學課程強調的重點，在 TIMSS 2007的評量架構中，以科學探究作為橫跨整個評量工作的主軸，涵蓋學科內容和技能兩個部分，並與所有科學領域的內容互相重疊。科學探究的評量包括一些題目和任務，用以讓學生展示出在進行科學活動時對一些工具、方法和程序所必備的知識，應用這些知識從事科學探究，並基於證據應用科學的理解提出解釋。科學探究的過程能夠促進對科學概念、推理和問題解決能力更廣泛的理解。

(二)PISA 2006

PISA 創立於1997年，代表 OECD（Organisation for Economic Co-operation and Development）成員國對於所屬成員學生教育學習成就的關心，該計畫從創始以降，每三年為一循環；分別於2000，2003，2006舉行，各循環參與國家數分別43、41與57國，持續進行調查學生在閱讀、數學與科學方面素養的學習成就。PISA 調查對象為 OECD 成員國與部分非成員國十五歲的學生，主要在探討學生面對未來生活準備的狀況，在 PISA 2006以科學素養為主要調查項目，內容除涵蓋科學知識、技能外，並增加對科學態度的調查部分，該調查並不以熟練學校課程內容為主要調查目標，而是以成年生活所必須具備的重要的知識與技能為重心，並且強調能將各領域所學之知識、概念與能力運用在不同情境中。PISA 2006所發展評量架構（Assessing Scientific, Reading and Mathematical Literacy – A framework for PISA 2006（OECD, 2006），其中對科學素養的評量，在學科部份包括物理、化學、生物及地球科學。試題類型分成三大類：(1)形成科學議題（Identify science issues）：要求學生從所提供的資訊之中，擬訂可以透過科學方法解決的研究問題；(2)解釋科學現象（Explain phenomena scientifically）：針對日常生活中常見的現象，如石雕受酸雨侵蝕，解釋其發生的原因；(3)科學舉證（Use scientific evidences）：利用科學證據來支持本身的主張或論點。此外，PISA也收集學生的各項家庭環境與學校背景資料，以及各種有關學校的背景資料及其所提供的教育內容。這些資訊與各國十五歲學生在閱讀、數學及科學素養表現有所關連，將有助於進一步研究作為教育政策之參考（林煥祥、劉聖忠、林素微與李暉，2008）。

(三)美國、加拿大、澳洲的全國大規模科學評量

近年來，許多科技先進國家，如美國、加拿大、澳洲等國，其定期的全國大規模科學評量，亦直接或間接地受到 TIMSS 和 PISA 的影響，而有重大的改變。有關加拿大在這方面的調查研究可參見 The SAIP Science III Assessment（http://www.cmec.ca/pcap/science3/indexe.stm）以及 The PCAP-13 2007 Public Report（http://www.cmec.ca/pcap/2007/report.en.stm）。澳洲方面相關全國性調查為 The National Assessment Program: Science Literacy（ http://www.mceetya.edu.au/mceetya/default.asp?id=12181），其中包括每三年一次針對6年級學生科學素養的評量。底下僅以美國的情況為例，簡要加以介紹。

NAEP（The National Assessment of Education Progress, n.d.）為美國教育部所屬國家教育統計中心（The National Center for Education Statistics）負責執行的一個計畫，該計畫定期調查各地區學生學習成就，並提出報告藉以追蹤比對各地區學生學習狀況與預期目標間的差距，NAEP 收集學生成就資料已經超過三十五年歷史，過去曾針對學生閱讀、數學、寫作與科學及其他學科領域進行調查，調查取樣早期以學生年齡9，13，17歲為主，1983年後也有採取就學年級取樣4，

8，12年級為對象者。在最近三次有關科學學習成就調查分別在1996，2000，2005，約每五年進行一次。目前該中心正在發展預計在2009年執行的科學學習評量架構。NAEP 2009和 NAEP 2005的科學評量架構有所不同，有關 NAEP 2009的科學評量架構和命題說明雖然尚未正式出版品，但在 National Assessment Governing Board（NAGB）（http://www.nagb.org/）的網站可查到這兩份相關的報告，其中對於 NAEP 2009科學評量的規劃過程、評量架構、題目類型等有相當詳盡的描述。有關 NAEP 已完成的調查，該單位設有相關網站，提供數據和工具給教師等使用者使用，針對研究者所提供的進一步的技術上的細節則可進入 NAEP Research e-Center（http://nces.ed.gov/）參看。

全國性的 NAEP，係針對4，8，12年級學生的大規模且具全國代表性的評量，並非針對個別學生的測驗。在州的層級的閱讀、數學、科學和寫作方面的評量其施測對象為4和8年級的學生，閱讀和數學每2年實施一次，科學則每4年一次。在測驗題本的設計上面，針對全部取樣學生，在認知方面的題目考試時間為50分鐘，另外10分鐘用於讓學生回答相關背景方面的問題。此外，有一半的學生樣本，將經由動手實作或利用電腦互動（Interactive computer tasks, ICTs）的任務，接受為時30分鐘的測驗。

在 NAEP 2009科學評量架構中強調科學內容（content）和科學實務（practices）兩者之間的緊密連結，在科學內容方面主要包含物質科學、生命科學、地球與太空科學以及跨學科領域的部分；而科學實務方面則包含辨識科學原理、應用科學原理、應用科學探究以及應用科技設計等四大部分。NAEP 2009科學評量架構的發展，包括針對不同年級分別對於這兩個評量向度擬定一些重要的主題和次主題，並且進一步經由內容與實務之間的結合，敘明學生所應該要知道和會應用的科學知能，亦即所謂的表現期望（performance expectations，或譯為預期表現）。對於一些重要科學上的想法，例如物質的狀態或生物的遺傳，學生的理解會隨著年級的增加而變得更為深入、精緻與熟練。在研擬 NAEP 2009評量架構時，也注意到有關學生學習進程（learning progression）的問題，雖則承認目前的研究基礎仍然不足。

(四)國內科學學習大規模調查研究與相關資料庫的建立

上述國外大規模的調查中，雖能提供許多寶貴的資訊作為參考，但若欲了解我國特有的社會文化、價值觀及影響學習的因素，達到進一步改善我國的教育品質的目的，自有必要進行國內學生科學學習大規模調查研究，建立國內的教育和科學教育方面的資料庫。近年來國科會科學教育發展處把學生學習成果長期追蹤調查列為重點研究，鼓勵研究者提出申請。在學生科學概念學習研究方面，為深入瞭解我國學生科學概念發展的狀況，約自西元2000年由國科會推動「我國學生科學概念學習研究」之整合型研究計畫，有系統的結合各學科領域之科學教育研究人員，從發展工具到實施大規模施測調查，計畫執行情形詳見國科會研究報告（郭重吉等，2008b），調查報告和研究結果將被納入台灣師範大學科學教育中心正在積極建置的科學教育方面的資料庫。上述計畫的部分研究成果並已在重要國際科教期刊以專輯形式刊載（Chiu, Guo, & Treagust, 2007）。

近年來國內開始建立的教育資料庫，主要有台灣教育長期追蹤資料庫（Taiwan Education Panel Survey, TEPS）（http://www.teps.sinica.edu.tw）和台灣學生學習成就評量資料庫（Taiwan Assessment of Student Achievement,

TASA）（http://tasa.naer.edu.tw/default.asp）。TEPS 係由教育部、中央研究院、國立教育研究院籌備處和國科會等單位以「學校為基礎的科學學習環境相關因素」之理論架構共同規劃，主要是以問卷調查的方式，蒐集全國中等學校學生在身心發展、學業成就及其影響因素的資料。主要想了解學生的學習能力受到學習機會和努力程度，與其綜合分析能力、行為、身心健康的交互影響。綜合分析能力測驗包括：一般推理、科學、數學、語文。背景問卷則包括學生、家長、教師和學校等四個面向（台灣教育長期追蹤資料庫）。TEPS 資料庫的內容目前較為有限，且對數理概念也較為簡化，所以不易分析比較我國不同年級學生的數學與科學概念學習情形。

TASA 之建置，係在九年一貫課程實施之後，為提供教育部研訂課程與教學政策之重要參據，由國家教育研究院籌備處針對國民中小學學生學習成就所建立常態性之資料庫。TASA 2007年學生學習成就評量自然科評量目的主要在瞭解國小四、六年級、國中二年級、高中二年級及高職二年級學生自然科學習成就的現況及其關聯因素，望能提供教師、家長、教育行政人員和研究人員全觀而具體的參考資訊。該測驗以全國為研究對象，透過測驗上等化的技術來進行分析研究，期使學校與學生都能在最正常的學習歷程中，呈現自然科的學習成果。

參、研究方法

本研究之執行，主要透過邀請科學教育領域相關領域資深研究人員參與研究，經由文獻探討與相關資料之收集、分析，研訂研究的大方向並分別組成數個研究小組進行相關研究工作，包括草擬評量架構與設計評量實例，進而邀請專家學者與在職的優秀教師座談，經由座談會方式共同研討，使研究成果能益臻完備。

一、研究範圍

(一)規劃目標

本研究終極的目標在於建立我國科學教育長期追蹤本土基本資料，惟考量此一調查所涉及變項繁多，及國內研究人力有限，需要整體規劃，分階段逐步執行。本階段的研究，主要目的是針對中小學學生科學認知與學習成果之全國調查研究進行先期規劃。研究重點在於確認中小學學生在自然科學方面所應知道的核心知識和具備的重要認知技能，並且探討學生隨著年級增加在這方面知能的增長情形。

(二)調查對象

本研究參照過去國際調查之成例，以中小學學生為調查對象，並按就學年級分別以4，6，9，12年級學生為取樣之對象。

(三)評量的內容

本研究以規劃調查學生在二十一世紀面對未來生活世界所需具備之科學素養為目標，包含科學的學習成就和認知技能。為維持未來進行長期追蹤調查內容之穩定性，並得以跟國際調查研究成果相互參照比較，而依照一般慣例將以科學內容主要成分區分物理、化學、生物、地球科學等學科領域，以及跨學科的部份，進行規劃。在發展評量架構與命題示例時，雖參考我國現行課程在不同學習階段之結構與內涵，但並不與目前實施之課程做直接之連結。

(四)和國內外大規模科學評量的關聯

和前述 TIMSS、PISA、NAEP、TEPS、TASA 等國內外大規模科學評量相互比較，本研究所欲發展的評量架構和試題實例在理念、格式和內容方面，一方面是有密切關聯，

但另一方面是有所區隔。最主要的差別在於本研究終極的目標在於建立我國科學教育長期追蹤本土基本資料，而 TIMSS、PISA 是跨國的比較研究，NAEP 則是美國全國性和各州的大規模評量。當然，TIMSS、PISA 和 NAEP 三者在科學學習成果的評量理念、目標和方法並不一致，例如，TIMSS 比較注重學科知識內容，PISA 比較注重科學知識在日常生活上的應用，NAEP 強調與課程標準的搭配，三者施測對象的年級或年齡也有所不同。至於 TEPS 和 TASA 因起步較晚，目前仍在進行，其在科學方面的評量比較注重學習成就的評量，後續工作或許可參考本研究的一些初步成果。具體而言，本研究所發展的評量架構在結構上面係仿造 NAEP，但在內容上面，包括學科內容和科學實務這兩向度的主、次項目，其對應的表現期望以及示範試題等，係由研究人員參酌文獻探討和國內情況加以研擬。

二、研究項目

本研究之主要工作如下：

(一)發展評量架構

評量架構是規範評量調查內容與後續工具發展的關鍵，本研究參照國外既有經驗，並參酌本土特性，擬定出4、6、9、12各年級學生科學學習的評量架構。同時設法把不同年段學生在重要認知能力和學習成果上的表現，區分成幾個不同等級，並敘明各等級表現的特性。

(二)設計評量示例

為使得後續調查研究計畫，在計畫執行、工具發展上有一致的觀點，本規劃案將針對各學科之評量架構在4、6、9、12各年級，選擇具代表性之評量架構並設計評量實例。考慮短期內評量實務實施之可行性等因素，評量實例以紙筆測驗為主，涉及儀器操作的實作評量和電腦互動測驗等暫不考慮。

三、研究計畫參與成員之組成與運作

(一)參與研究人員之組成

本研究的進行分成兩個層級，其一為核心研究小組，成員共計十一名（含計畫主持人），均為科學學習相關研究領域之資深研究人員，負責整個研究計畫的規劃與執行。其二為各領域研究小組，負責各學科內容領域（物理、化學、生物、地科、跨學科）與認知領域的研究工作。各領域研究小組，分別由適當的核心小組成員主持，負責其中一至二個領域的研究工作，另再邀請數名專家學者、優秀中小學教師參與該領域研究規劃，在評量架構規劃草案完成前另邀請其他相關專家學者舉行座談，聽取意見進行修正，以求規劃案能趨於完備。

(二)運作與分工

本規劃案由研究計畫主持人負責整體規劃與推動，其他共同主持人、協同主持人協助計畫的規劃與執行。參考 NAEP 2009的評量架構，本計劃中擬定的評量架構簡介如表1。學科內容的向度區分為：物理、化學、生物、地球科學以及跨學科等五大領域，而認知向度方面則包含：科學知識的理解、科學知識的應用、科學探究的能力、量化推理能力、科學的本質、態度與情意因素等六大領域。各領域研究小組針對其負責的部份先經由與會人員認同取得共識，再將結果送至核心研究小組的會議中討論、彙整。

發展評量架構的步驟包括：針對4，6，9，12各年級學生重要科學知能，亦即主要科學學習目標，擬定各大領域的主題和次主題並加以簡明的界定；其次，如表1，對於內容領域和認知領域相互交叉所形成的格子，明確的敘述對學生表現的期望，並依難易或複雜程度進一步分為基本、熟練和進階

表 1：評量架構簡介

內容 \ 科學實務	P 物理	C 化學	B 生物	E 地科	I 跨學科
1.科學知識的理解	對學生表現的期望	對學生表現的期望	對學生表現的期望	對學生表現的期望	對學生表現的期望
2.科學知識的應用	對學生表現的期望	對學生表現的期望	對學生表現的期望	對學生表現的期望	對學生表現的期望
3.科學探究的能力	對學生表現的期望	對學生表現的期望	對學生表現的期望	對學生表現的期望	對學生表現的期望
4.量化推理能力	對學生表現的期望				
5.科學的本質	對學生表現的期望				
6.態度與情意因素	對學生表現的期望				

三個層次分別敘明，以做為命題的依據。由於各大領域中的主題和次主題為數不少，在本研究中對於上述步驟的所有可能的情形，無法逐一詳列，僅選擇比較重要、比較具有示範作用的一些情形，敘寫對學生表現的期望。同理，對於評量實例，也是選擇較具代表性的例子設計一些題目，以供參考。

(三)研究會議

核心研究小組大約每月一次聚會討論，並透過建立研究網站、電子郵件等管道進行聯繫與協調。各個領域的研究小組，在研究過程中至少召開六次領域研究小組會議，其中至少有一次邀請本領域小組成員以外之專家學者與優秀中小學教師參與領域規劃草案之研議。

肆、研究成果

本節先介紹評量架構中學科內容和科學實務的各主題和次主題，其次再簡述主要研究成果，包括評量架構中各項表現期望和命題示例。

學科內容和科學實務向度所擬定的主項目和次項目，分別如表2和表3。針對表2中的各個次主題，各科的研究小組分別確定不同年級學生應該學會的重要科學概念和原理，亦即 NAEP 所稱的內容陳述（content statement）。以生物科「生物體的構造與功能」中的一個次主題「構造與發育」為例，其所得結果如表4。同樣地，針對表3中的各個次主題，在本研究中亦分別針對不同年級學生應該學會的科學實務，訂出具體的評量指標，詳見（郭重吉等，2008a）。橫跨各個年級在學科內容和科學實務這兩向度的各個細項，學生應該具備的表現期望以及對應的命題示例，其具體項目繁多，例如，在附錄所列表5為不同年級學生對「生物體的構造與發育」應該具備的表現期望，表6和表7分別各舉出生物和化學的一個命題示例供做參考。

伍、結論與建議

本研究長遠的目標在於建立我國科學教育長期追蹤本土基本資料，藉以促進國內中小學科學教育的革新，當然，這有賴政策、研究與實務層面整體的規劃、密切的配合和長期的推動。以美國為例，National Research Council（簡稱 NRC）此一民間非營利機構，長期以來即致力於進行各項相關研究，以改進政府決策和公共政策，促進大眾教育與理解，以及增進在科學、工程、科技和健康方面知識的建立與傳播（ http://

表 2：學科內容主題與次主題

科目	主題	次主題
物理	物質	物質的特性、物質的改變
	能量	能量的形式、能量的傳遞、轉換和守恒
	運動	巨觀層次的運動、力與運動
化學	物質的形成	物質的組成、物質的構造
	物質的性質	物質的一般性質、固體的性質、液體（水溶液）的性質
		氣體的性質、物質的分類
	物質的變化	化學反應、化學平衡、酸鹼反應、氧化與還原
		有機化學反應
	能量	反應熱、能源形式及其轉換
	化學應用	生活中的化學、化學與永續發展、化學與先進科技
生物	生物體的構造與功能	構造與發育、物質與能量的轉化、生物間互相依賴
	生物體的改變	遺傳與生殖、演化與多樣性
地科	地質	地球的歷史、地震、岩石與礦物、地質作用、化石
	氣象與海洋	氣溫變化、風向風力、氣象與生活、天氣圖、颱風、全球氣候變遷、聖嬰現象
	天文	觀測月亮、月相的變化、宇宙的組織、太陽系、太陽的視運動、天體運動

sites.nationalacademies.org/NRC/index.htm）。本文所引述該組織近十多年來一系列的研究報告，從科學教育標準、學習理論、教學與評量乃至於中小學的科學教育，對於美國及其他世界各國的中小學教育或科學教育，皆有重大的影響。

當前中小學學生科學教學的趨勢，強調學生經由參與解決日常生活情境中的問題，來學習科學知識和科學過程、科學本質等科學實務。亦即科學的教學應讓學生在實際參與科學實務中，一方面了解重要的科學事實、概念和原理，一方面熟練觀察、紀錄、進行實驗探究等科學過程技能，同時也要了解科學知識的形成和發展過程。這些科學實務包括進行探究、與同學分享自己的想法、練習使用科學上特有的字彙和語法、建立模型（實體模型、數學模型或電腦模擬）、以及發展對一些現象的表徵。在這些科學實務的教學中，特別強調的是學生在知識、技能、情意和社會互動等各方面的學習，是相輔相成、密不可分而一起持續成長的。此外，以往在科學的課程、教學與評量，普遍流於過分強調對一些片段、獨立知識的傳授，疏於深入了解學生隨著年級、經驗、和心智發展的不同，而對重要科學核心概念和原理有不同程度的了解，因而在課程和教學活動的設計未能提供學生適切的學習環境，協助學生朝向更高層次的理解來發展；而在科學師資培育和專業發展上，也未能採行適當的因應措施（NRC, 2007）。相形之下，科學教育在教學和評量上面新的發展趨勢，是重視統整的理解、知識的整合和課程的連貫一致（Kali, Linn, & Roseman, 2008）。

如上所述，強調科學實務和學生的學習進程是當前科學教育兩大努力的方向，這樣的精神不但反映在 NAEP 2009的評量架構，也跟 TIMSS 和 PISA 等國際比較研究所強調的方向相符，而這對國內在規劃、設計長期追蹤學生科學學習的調查研究上面，亦甚具參考價值，在本研究中嘗試朝此方向作一些

表 3：科學實務的主項目和次項目

主項目	次項目
1.科學知識的理解：學生對於科學事實、資訊、概念的基礎	1.1. 事實（facts）：科學事實、科學專有名詞及科學定律。
	1.2. 關係（relationship）：科學事實、概念及科學定律中與之相關的屬性、結構、功能和關係之知識。
2.科學知識的應用	2.1. 生活的應用：科技發展的歷史及在人類文明所扮演的角色，瞭解科技可以改變人類的生活。
	2.2. 科技的應用：科學與科技的關係，及科學原理如何運用於科技之中。
3.科學探究的能力	3.1. 觀察：利用感官獲取關於物體或事件的訊息。
(1) 科學過程技能：科學家從事研究以獲取科學知識的歷程中所運用的手法與策略	3.2. 分類：根據物體或事件的特質或標準，以聚集或排序列出不同的類別。
	3.3. 測量與記錄數據：利用標準或非標準測量或估算，來描繪物體或事件的尺寸。
	3.4. 分辨與控制變因：辨識出能影響實驗結果的變因，控制到只操弄獨立變因，而其他均保持不變的地步。
(2) 科學探究的能力（問題解決的能力）：問題解決是個體利用認知過程去面對和解決不熟悉、更為複雜的真實情境，問題內容通常是跨領域的，需要學生進一步加以推論、分析以決定與問題相關之關係與概念。	3.5. 理解問題/提出問題或假設：學生是否瞭解文本、圖表或公式等資料來源，並能藉此從他們的背景知識中抽取適當資訊來理解給予的資訊。
	3.6. 表徵化：包括學生如何建立表徵、轉化表徵，或者如何應用給定的外在表徵以解決問題。
	3.7. 分析/找出特徵：包括學生如何分析與定義問題中的變數以及變數之間的交互關係。
	3.8. 提出並評鑑解答（含解釋或預測）：由給予的資訊預測趨勢、進行推論、提出解釋。並能評估不同解釋和問題解決的策略，以支持其想法。
4.量化推理能力	4.1. 歸納推理：指能觀察數據中的規律或特徵，並根據此規律作進一步的預測或推論。
	4.2. 演繹推理：指在處理數據時，能根據邏輯的理由推導出結論，並說服他人該推理在整體上的合理性。
	4.3. 比例推理：指能利用變數間比例的關係進行推理。
	4.4. 合併與分解：指能適當處理科學數據的合併與分解的推理能力。
	4.5. 單位轉換：指能進行單位的轉換，進而用單位協助推理。
5.科學的本質：在科學的知識和其發展中所蘊含的獲知的方法、價值與信念。	5.1. 科學的世界觀：科學家對他所做的以及他如何看待他成果的信念與態度，包括：世界是可理解的；科學的想法是可改變的；科學的知識是持久的；科學無法對所有的問題提出完全的答案等。
	5.2. 科學的探究：科學的特色也包括在其探究的特殊處：科學是需要證據的；科學是結合邏輯與想像的；科學進行解釋與預測；科學家試圖察覺並避免偏見；以及科學不是權威等特質上。
	5.3. 科學的事業：科學是包含有個人的、社會的以及機構的不同面向的事業，因此：科學是複雜的社會活動：科學被分成不同的學科內容，並且在不同機構執行；執行科學時也遵行一般性的倫理法則；而科學家也參與其特殊的或一般的公共事務。
6.態度與情意因素：於眾多的態度與情意因素中以對科學的焦慮感、動機及性格（外向型或內向型）為主要的探討要項。	6.1. 焦慮：在學習理論中，焦慮指一種次級的或條件性的內趨力，它驅動個體做迴避反應；焦慮包含了緊縮不安焦急憂慮擔心與恐懼等感受；焦慮有兩種類型，特質性焦慮與情境性焦慮。
	6.2. 動機：引起個體活動或維持已引起的活動並促使該活動朝向某一目標進行的內在歷程；各項動機中以成就動機、親和動機、合作動機、誘因動機與目標導向動機為主。
	6.3. 性格：個體在其生活中對人、對事、對己以至對整體環境適應時所顯示的獨特個性；性格展現於需求、動機、性趣、能力、態度或價值觀等方面；從情境論或互動論的觀點來看，性格與行為的關係有不同的解釋。

表 4：「生物體的構造與發育」的內容陳述

構造與發育：生物體的基本需要（四）；常見的動、植體構造與功能（六）；生物體的構造階層（九）；生物體的化學組成（十二）。

四年級	六年級	九年級	十二年級
B4.1 生物體需要食物（植物體需要光線）水和空氣；也需要排泄廢物的方法；也需要一可以生存的環境。	B6.1 植物的根、莖、葉、花、果實、和種子有不同功能；植物可由種子或根莖葉繁殖。	B9.1 所有的生物體（從單細胞到多細胞）都是由細胞所組成。多細胞生物，有特化的細胞執行特殊的功能。器官和系統是由細胞所組成，其可以供應細胞所需要的營養、空氣、和排除廢物。	B12.1 生物體是由複雜的分子所組成，它們含有碳、氫、氧、氮以及硫。
B4.2 植物體有根、莖和葉的構造（230-2a）。	B6.2 有的植物會開花，有的不開花（例如蕨類）230-3a。	B9.2 卵經過受精作用會進行細胞分裂，產生一團細胞，然後細胞特化成胚胎的不同構造和功能之組織。	B12.2 細胞的許多生化反應是靠不同的分子，特別是蛋白質所執行。蛋白質分子是由胺基酸分子所組成，蛋白質分子的特殊功能受其胺基酸分子的排列順序與分子的形狀影響。
B4.3 動物的構造。	B6.3 動物構造的功能。	B9.3 葉子的構造及功能；植物體內有輸導組織及功能。	B12.3 細胞的生理功能受其內在與外在的環境所影響。例如當多細胞個體發育時，細胞所處環境會引發細胞的分化。生物體細胞分化是受到不同基因所調節的。
B4.4 生物生長的過程（230-2b）。	B6.4 動物的生殖。		
	B6.5 陸生及水生動物形態及其運動方式不同，水生動物具有適合水中生活的特殊構造。		

初步規劃。配合科學學習環境調查資料的蒐集，本研究的成果可以作為建立本土化科學教育長期追蹤調查資料庫之開端，以利有關國內學生科學認知與學習研究之進行。另外，就學校和課室教學的層次而言，本研究的成果也可提供科學教師課室評量以及科學教師專業發展在實務和研究方面的參考。

當然，從比較長遠的眼光來看，學生科學學習成果的追蹤調查和比較分析，在評量的理論和實務層面仍有一些值得進一步商榷的問題，例如，資料蒐集的對象是個別學生、班級、學校、學區、縣市或全國？施測結果如何計分、加權和統計分析？試題內容和課程大綱的關聯為何？在不同時期如果課程大綱有所改變的話，對資料的蒐集和分析會產生什麼樣的衝擊，該如何因應？類似問題，在後續研究中宜及早規劃，而有些國外經驗亦可資參考，例如，美國的全國性評鑑 NAEP 始自1969年，迄今仍繼續運作；而英國的 APU（Assessment Performance Unit）在1980至

表 5：「生物體的構造與發育」學生應該具備的期望表現（橫跨年級）

主項目	次項目	細目	四年級	六年級	九年級	十二年級
科學知識的理解	科學事實知識		能指出生物體的基本需要	能指出常見的動植體構造與功能	能指出生物體的構造階層	能指出生物體的化學組成
	概念間關係		能理解生物體基本需要之簡單關係	能理解常見的動植體構造及其相互關係	能理解生物體的構造階層間的關係	能理解生物體的化學組成間的關係
科學知識的應用	生活應用		能理解生活中生物體基本需要的科學現象	能認識生活中常見的動植體構造的功能及其影響因素	能理解生活中生物體的構造階層所產生的現象	能理解生活中生物體的化學組成的相關應用
	科技應用		能理解生物體基本需要物質的簡單原理	能認識常見的動植體構造簡單原理所創造出的科技產品	能認識動植體構造複雜原理所創造出來科技產品的專業領域運用	能認識動植體構造較複雜原理所創造出來科技產品的專業領域運用
科學探究的能力	科學過程技能	觀察	能覺察生物體基本需要物質的特徵和屬性	能依規劃步驟觀察常見的動植體構造	能由不同角度觀察與比較生物體的構造階層	能由不同角度觀察與測試生物體的化學組成之性質
		分類	能依特徵將生物體基本需要物質歸類	能依據相似或相異性分類動植體構造並訂分類準則	能依差異性程度做生物體的構造階層分類	能依差異性程度做生物體的化學組成分類
		測量與紀錄	能運用感官測量生物體基本需要物質的特性	能選用適當工具測量動植體構造，並運用適當方法紀錄	能選用適當工具測量動植體構造，並運用適當方法紀錄，及解釋資料	能選用適當工具測量生物體的化學組成，運用較嚴謹適當方法紀錄，及解釋資料

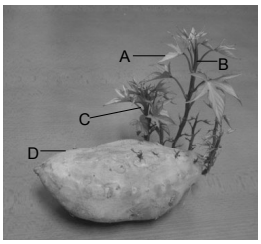
1985年間針對11，13，15歲學生的英文、數學和科學的學習每年進行調查，但隨著1988年通過的教育改革法案規定國家課程的實施，這方面的調查研究於1993之後即因納入國家課程評量計畫而不復獨立存在（Britton & Schneider, 2007; Gipps, 1990; Harlen, 2008; Scott & Owen, 2005）。

本研究所發展的評量架構尚屬初步嘗試，在學科內容和科學實務這兩向度的主、次項目的選定，從理論和研究的角度而言，仍有許多檢討、改進和尚待充實的空間。例如，許多研究顯示學生的後設認知在科學學習上面扮演至為重要的角色，但在本研究中由於種種實際因素，並未特別加以考量。另外，有關學生預期表現的部份，也未能進一步整體的加以檢視，確保各個學科和年段之間的銜接連貫和前後一致，尤其是對於學生不同表現和理解層次的區分，仍需進一步有理論的依據並經現場測試的考驗。另外，就建立長期追蹤資料庫而言，對於整個評量架構的細部內容，宜建立一套完整的編碼系統作為命題、分析、解釋和報導測驗結果的依據。

本研究所發展的命題示例係以紙筆測驗為主，對於動手實作、電腦模擬、電腦適性化測驗和網路科技在評量上面的應用，並未

表 6：生物科命題示例（六年級生：物體的構造與功能－常見的動植體構造與功能）
命題陳述 6.1.：了解植物根、莖、葉、花、果實、種子的構造與功能，並知道植物除了利用種子之外，亦可藉由根、莖、葉繁殖。同時能分辨有的植物會開花，有的不開花（例如蕨類）；

一、科學知識的理解：學生能辨識、指出植物的基本構造和功能

基礎性（basic）	精熟性（proficient）	進階性（advanced）
 上圖中，A. B. C. D 哪一個字母所標示的是番薯的根？ 答案：D	大花咸豐草的種子具有附倒鉤的小刺，這種種子可能的傳播方式是哪一種？ A. 風力 B. 水力 C. 動物 D. 彈力 答案：C	布鞋有了魔鬼氈就可以省去綁鞋帶的麻煩，這種構造是運用了哪一種自然界（生物）的現象特殊功能？ A. 非洲鳳仙花果莢的彈力可將自己的種子傳播到遠方。 B. 大花咸豐草種子的倒鉤小刺可緊黏衣物。 C. 仙人掌的針狀葉可減少水分的蒸散。 D. 荷葉表面的特殊構造可讓水珠聚集。 答案：B

二、科學知識的應用：學生能辨識、指出問題的原因與解決辦法。

基礎性（basic）	精熟性（proficient）	進階性（advanced）
下面哪一種植物可以用根來繁殖？ A. 番薯 B. 綠豆 C. 西瓜 D. 小蕃茄 答案：A	在觀察和記錄植物的成長過程中，可以發現到哪個現象？ A. 綠豆種子發芽後，子葉會變成綠葉。 B. 鮮艷的花朵是為了吸引蝴蝶蜜蜂來傳粉受精。 C. 蕨類植物開花後會形成孢子。 D. 果實的甜份是為了種子發芽的養分。 答案：B	請選出以下有關落地生根繁殖的敘述，哪一項正確： A. 落地生根可用莖繁殖，所以不會開花。 B. 用葉繁殖長出的落地生根，不會長根。 C. 落地生根用莖繁殖與用葉子繁殖的生長過程是一樣的。 D. 落地生根可用多種方式來繁殖，可以增加後代生存的機會。 答案：D

加以考慮。對於學生學習長期追蹤調查資料庫的建立而言，這是未來發展必然的趨勢，建議在後續研究中應朝此方向多加努力。另外，為期提昇中小學科學教育的成效，有關學生科學學習的評量應該以長期且系統化的角度來規劃和進行，考慮到各項科學學習目標的整合、跨越學生不同年級的學習、評量與課程和教學之間的一致性、以及課室（形成性）評量和總結性與大規模評量的聯結。這需要從科學教育政策、研究和實務之間做整體的規劃與推動，一些有關學生科學學習評量的研究與政策的制定，除了在技術層面要考慮到資料的可信、可靠和適用性之外，在過程中要結合教師、研究者和教育行政主管共同參與，而有關學生科學學習的評量的研究與資料庫的建立，整體而言，要能兼顧個別學生、班級、學校、縣市、乃至全國等不同的層級，期能提供教師改進教學實務和制定或修訂政策之參考。另外，從實務層面來考慮，太多的大規模測驗對學校教學難免形成干擾，因此，國內各相關單位正在規劃或進行的有關學生學習的長期追蹤調查，應該設法加以整合，例如，閱讀、數學和自然科學方面的測驗可統一規劃、協調實施步調。而如何使大規模的追蹤調查能與學校的教學評量連結，甚或融入學校常態的評量或

表 7：化學科命題示例

主題	物質的變化	次主題	化學反應
類別	科學知識的理解	次類別	科學關係
年級	9 年級	命題者	

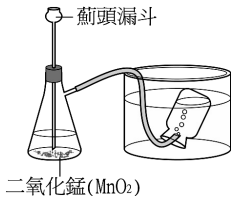
Content statements

P9.24：化學反應的快慢可以用反應速率來表示。而影響反應速率的因素有：反應物的性質、溫度、表面積和催化劑。如表面積愈大，反應物粒子相互碰撞的機率增加，所以反應速率愈快。

Illustrative Item

Basic

右圖為關於雙氧水置備氧氣示意圖，下列有關實驗的敘述，何者正確？



(A)雙氧水可一次加很多，可增加反應速率。

(B)二氧化錳是催化劑，反應前後不會減少。

(C)雙氧水濃度愈高，反應愈慢。

(D)二氧化錳是催化劑，反應後會減少。

Key: B

Proficient

小明以相同濃度的雙氧水做實驗，比較下表不同條件下的反應速率，試回答問題：

實驗	反應速率	溫度	是否加入二氧化錳
A	甲	20 °C	否
B	乙	30 °C	是
C	丙	40 °C	是

反應速率甲、乙、丙何者較大？為什麼？

(A)甲速率較大，因為溫度較低。

(B)乙的速率較大，因為溫度適中且有加入二氧化錳。

(C)丙的速率較大，因為溫度最高且有加入二氧化錳。

(D)乙、丙的速率一樣快，因為都有加入二氧化錳。

Key: C

Advanced

小明為了解反應速率的關係，他利用貝殼及鹽酸生成二氧化碳氣體，且分別測定二氧化碳產生的速率，他依下表分別測定二氧化碳產生的速率。

實驗	貝殼	鹽酸	溫度
甲	1 克整塊	40%，20mL	30 °C
乙	1 克碎片	40%，20mL	30 °C
丙	1 克粉末	40%，20mL	40 °C
丁	1 克粉末	40%，20mL	60 °C

請問甲、乙、丙、丁四次實驗，二氧化碳產生的速率由大而小依次為：

(A)甲＞乙＞丙＞丁

(B)乙＞甲＞丁＞丙

(C)丁＞乙＞甲＞丙

(D)丁＞丙＞乙＞甲

Key: D

考試，也是未來值得加以探討的問題。

致　　謝

本研究承蒙行政院國家科學委員會專題計畫補助經費（計畫編號 NSC 96-2522-S-143-002-LS），謹在此致謝。計畫執行過程中定期舉行的會議，多承郭允文研究員、林陳涌教授、林妙香教授、曾憲雄教授和黃國禎教授的參與，各分組的研究並曾邀請多位中小學教師、研究助理、碩博士班學生和專家學者參與，在此一併致謝。

參考文獻

1. 林煥祥、劉聖忠、林素微、李暉（2008）。**台灣參加 PISA 2006 成果報告**。行政院國家科學委員會計畫編號（NSC 95-2522-S-026-002）。花蓮市：國立花蓮教育大學。
2. 郭重吉、邱美虹、黃台珠、張惠博、張俊彥、周進洋、王國華、譚克平、黃鴻博（2008a）。**科學認知學習之全國性調查研究－先期規劃**。行政院國家科學委員會計畫編號（NSC 96-2522-S-143-002-LS）。台東市：國立台東大學。
3. 郭重吉、張惠博、陳忠志、周進洋、邱美虹、林陳涌、黃台珠、譚克平（2008b）。**2003 中小學科學學習研究**。行政院國家科學委員會計畫編號（NSC 92-2511-S-143-007）。台東市：國立台東大學。
4. Beatty, E. E., & Carlson, M. O. (n.d.). *Using the student evaluation standards: A field guide for k-12 math & science teachers*. Retrieved December 1, 2008, from http://www.wmich.edu/evalctr/jc.
5. Britton, E. D., & Schneider, S. A. (2007). Large-scale assessments in science education. In S. Abell & N. Lederman (Eds.), *Handbook of research on science education* (pp. 1007-1040). NJ: Lawrence Erlbaum Associates, Inc.
6. Chiu, M. H., Guo, C. J., & Treagust, D. (2007). Assessing students' conceptual understanding in science: An introduction about a national project in Taiwan. *International Journal of Science Education, 29*(4), 379-390.
7. Committee on Standards for Educational Evaluation. (2003). *The student evaluation standards: How to improve evaluations of students.* Newbury Park, CA: Corwin Press.
8. Gipps, C. (1990). National assessment: A comparison of English and American trends. In P. Broadfoot, R. Murphy, & H. Torrance (Eds.), *Changing educational assessment: International perspectives and trends* (pp. 53-64). London: Rotledge.
9. Guo, C. J. (2007). Issues in science learning: An international perspective. In S. Abell & N. Lederman (Eds.), *Handbook of research on science education* (pp. 227-256). NJ: Lawrence Erlbaum Associates, Inc.
10. Harlen, W. (2008). Science as a key component of the primary curriculum: A rationale with policy implications. *Perspectives on education 1*, 4-18. Retrieved November 8, 2009, from http://www.wellcome.ac.uk/About-us/Publications/Reports/Education/Perspectives/index.
11. IEA. (2005). *TIMSS 2007 assessment frameworks.* MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved December 1, 2008, from http://timss.bc.edu/TIMSS2007/PDF/T07_AF.pdf.
12. Kali, Y., Linn, M. C., & Roseman, J. E. (Eds.) (2008). *Designing coherent science education:*

Implications for curriculum, instruction, and policy. New York and London: Teacher College Press.

13. National Assessment of Educational Progress. (n.d.). *The main NAEP version of the NAEP data explorer*. Retrieved December 1, 2008, from http://nces.ed.gov/nationsreportcard/nde.
14. National Research Council. (1996). *National science education standards*. Washington, DC: National Academies Press. Retrieved December 1, 2008, from http://www.nap.edu/openbook.php?record_id=4962.
15. National Research Council. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press. Retrieved December 1, 2008, from http://books.nap.edu/catalog.php?record_id=6160
16. National Research Council. (2003). *Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment*. Washington, DC: The National Academies Press.
17. National Research Council. (2005). *How students learn: History, mathematics, and science in the classroom, committee on how people learn*. Washington, DC: The National Academies Press.
18. National Research Council. (2006). *Systems for state science assessment*. Washington, DC: National Academy Press.
19. National Research Council. (2007). *Taking science to school: Learning and teaching science in grades k-8*. Washington, DC: National Academy Press. (Free Executive Summary Retrieved December 1, 2008, from http://www.nap.edu/catalog/11625.html).
20. Organisation for Economic Co-operation and Development. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD. Retrieved December 1, 2008, from http://www.oecd.org/dataoecd/63/35/37464175.pdf.
21. Scott, E., & Owen, E. (2005). *Brief: Comparing NAEP, TIMSS, and PISA results*. Wahington, DC: National Center for Education Statistics.

Developing the Assessment Framework and Illustrative Test Items for Science Learning Outcomes of Elementary and High School Students in Taiwan

Chorng-Jee Guo[1], Mei-Hung Chiu[2], Tai-Chu Huang[3], Huey-Por Chang[4], Chun-Yen Chang[5], Ching-Yang Chou[3], Kuo-Hua Wang[6], Chung-Chih Chen[7], Hak-Ping Tam[2] and Hon-Bo Huang[8]

[1] College of Science, National Changhua University of Education

[2] Graduate Institute of Science Education, National Taiwan Normal University

[3] Graduate Institute of Science Education, National Kaohsiung Normal University

[4] Department of Physics, National Changhua University of Education

[5] Graduate Institute of Science Education & Department of Earth Sciences, National Taiwan Normal University

[6] Graduate Institute of Science Education, National Changhua University of Education

[7] Department of Applied Chemistry & Material Science, Fooyin University

[8] Department of Science Application and Dissemination, National Taichung University

Abstract

With reference to existing literature and experiences gained from both domestic and international large-scale assessments in science, this study aims at developing an assessment framework and some illustrative test items, which will provide guidelines and concrete examples for item and test developments, and the administration of a national survey in subsequent studies. The assessment framework, designed for the 4th, 6th, 9th, and 12th graders, consists of, first of all, a scheme to categorize students' major learning outcomes in content domains including biology, chemistry, physics and earth science. Within these subject areas, students' expected performances in terms of their proficiency in science practices, including understanding and applications of science concepts and principles, scientific enquiry skills, quantitative reasoning skills, nature of science, and attitudes and other affective factors toward learning science are specified. In order to follow up students' learning outcomes on a long term basis, numerous examples of students' expected performances, varying from basic, proficient to advanced levels are developed. Serving as guidelines for subsequent studies along this line, some general principles of quality assessment and specifications for test development in science, as

well as illustrative test items, are developed in this study. The study was coordinated by the first author, while the other coauthors respectively organized research teams consisting of science teachers, scientists, and science educators of different content areas in developing the assessment framework and illustrative test items. The work done in this study are of a preliminary nature, field tests and further refinement may be necessary.

Key words: Science Learning Outcomes, Large-Scale Assessment, Assessment Framework, Illustrative Test Items