

發展「真實性評量」來探討台灣中學生的科學素養

張淑女

真理大學通識教育學院

收稿日期：二零零七年十月廿五日(於二零零八年六月三十日再修定)

內容

- [研究背景](#)
 - [理論背景](#)
 - [測試方法](#)
 - [研究結果](#)
 - [結論與建議](#)
 - [致謝](#)
 - [參考文獻](#)
-

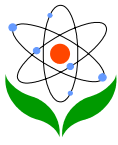
研究背景

自從科學素養一詞在 1950 年代被 Paul Hurd 所提出後，科學素養成爲近數十年來全球從事科學教育的研究人員與在職教師們的終極目標(Laugksch, 2000)。而伴隨著科學素養的科學教育目標，許多國家的教育改革潮流也隨之展開，台灣也不例外。

台灣最早在 1968 年開始實施九年國民義務教育改革，而最近的一次課程改革是在 2001 年開始實施所謂「九年一貫課程」(Chang & Chiu, 2005)。在這一次的教育改革理念指出，國民中小學課程應以人的生活爲中心，配合學生身心能力發展歷程；尊重個性發展，激發個人優良潛能；涵詠民主素養，尊重多元文化價值；培養科學知能，適應現代生活需要(教育部，1998)。根據上述的理念所發展出來的課程目標主要有三個面向(教育部，1998)：

- 「人與自己」－強調個體身心發展的關係，與促進個體的身心發展。
- 「人與社會」－著重於社會與文化的關係，與增進社會文化的參與。
- 「人與自然」－自然與環境的關係，正確認識自然並適切運用。

基於上述三大面向所產生的課程目標，可以進一步轉化成個體在九年一貫課程所應積極培養的十大能力，如表一所示。而除了此十大能力的轉化，課程改革人員也規劃出七



大學習領域來加以落實，從國民教育的一年級至九年級，均提供語文、健康與體育、社會、藝術與人文、數學、自然與科技及綜合活動等七大領域，用來取代過去升學考試的傳統科目，希望統整學科知識與學生生活經驗，避免知識的片斷性，希望強化課程之間的連貫性與生活經驗的結合；就七個學習領域之中的「自然與科技」領域而言，此領域所包括的學科包括自然科學（物理、化學、生物、地球科學）及科學應用（食品、織品、營建、運輸、製造等方面的科技和資訊傳輸、生物科技等）(教育部，1998)。

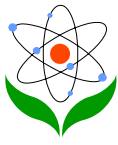
表一、九年一貫課程目標與十大能力（整理自教育部，1998）。

課程目標	十大能力
(一)「人與自己」的面向	
1. 增進自我了解，激發個人潛能。	⇒ 1. 了解自我與發展潛能 2. 欣賞、表現與創新 3. 生涯規劃與終身學習
2. 培養表達溝通、創作及審美能力。	
3. 獲得運動常識與技能，陶冶體育情操。	
(二)「人與社會」的面向	
4. 培養負責守法、積極參與、團體合作的民主法治知能。	4. 表達、溝通與分享
5. 發展互助合作的群己關係，重視倫理價值，提昇道德判斷能力。	⇒ 5. 尊重、關懷與團隊合作
6. 啓迪包容異己，尊重多元文化，恢弘國際視野。	6. 文化學習與國際瞭解
(三)「人與自然」的面向	
7. 啓發探究興趣、批判反思、創造發明等科學態度。	7. 主動探索與研究
8. 學習科學方法，理解科學概念，注重科學實驗，應用科技新知。	⇒ 8. 規劃、組織與執行
9. 了解自然環境，維護自然生態。	9. 獨立思考與解決問題
10. 掌握科技資訊，培養終生學習意願。	10. 運用科技與資訊

由上述的九年一貫課程改革的理念與目標，不難看出台灣對於科學發展與培養學生科學思考及科學態度的重視，以及促進學生能將科學知能與生活情境結合的決心，而這樣的思潮十分符合全球所提倡的科學素養理念。基於此教育改革理念，以下內容將藉由筆者過去在真實性評量的研究經驗來探討其對於促進科學素養的可行性，以及介紹台灣在2001年的九年級學生科學素養中的認知指標表現情形。

理論背景

就 Miller 所提出的科學素養觀點認為，在現今科學與科技快速發展的時代，科學素養應該要包含了解科學的方法、重要的科學概念與詞彙，及科學與科技可能對社會所帶來的衝擊 (Miller, 1983)。其他的一些學者也提到科學素養的提昇是爲了改善個體在現今科技主導社會中的生活，而終極目標是希望能夠提昇國家的國際競爭力 (Aikenhead & Ryan, 1992; Laugksch, 2000; Thomas & Durant, 1987)。然而，在最近一次的台灣教育改革思潮，亦符合了上述科學素養的意涵。在台灣教改所規劃出的七大學習領域中，在自然與科技的學習領域，教改委員們進一步規劃出六個學習指標來促進科學素養的提昇，



包括過程技能、科學認知、科學本質、科學態度、思考智能、科學與資訊的應用 (教育部, 1998)。然而，在這樣的學習指標之下我們要如何知道學生是否有達到科學素養的目標呢?在過去，有學者發展出選擇題來評量 11 及 12 年級學生 對於科學—技術—社會 (Science, Technology and Society, STS)的議題 (Aikenhead & Ryan, 1992)，或是利用是非題來評量學生對於科學本質，科學認知，與科技對社會衝擊的觀點 (Laugksch & Spargo, 1996)。然而，從 Champagne 和 Newell 的研究發現，美國學生在這些傳統的評量方式表現很差，認為應該要以真實性評量來評斷及促進學生的學習動機與技能 (Champagne & Newell, 1992)。同時有學者也強調，將科學概念融入真實情境是非常重要的 (Champagne & Newell, 1992; Yerrick, 2000)。而根據 Newmann 及 Archbald 的想法，真實性成就 (authentic achievement)的意涵指的是有受過訓練的探究、高層次的思考及問題解決能力，進而能夠將這些能力及所學的知識從學校轉移到真實的生活 (Newmann & Archbald, 1992)。這樣的觀念事實上符合於台灣教育改革理念以及科學素養的精神。因此，在西元 2000 年及 2001 年的一個教育部及國科會共同補助的兩年期研究計畫(計畫總主持人為邱美虹教授)，即以真實性評量的理論為依據來設計發展評量系統，探討 1,469 位台灣九年級學生在此六項學習指標的表現 (Chang & Chiu, 2005)。

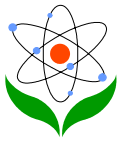
真實性評量(authentic assessment)是由 Wiggins 所提出之概念 (Wiggins, 1989)。在過去，標準化的傳統考試被視為是量化，而真實性評量較屬於另類質化的評量，這兩者在過去被視為是兩種不同的文化，但現在已經可以被整合來評量學生的高層次思考能力 (Dori, 2003; Wolf, Bixby, Glenn, & Gardner, 1991)。根據 Cumming 及 Maxwell 的真實性評量觀點，真實性評量具有評量四種能力的面向：(1) 探究評量—評量可以探究知識的能力；(2) 情境評量—特定情境的能力；(3) 問題導向評量—可以將學校所學遷移至生活的問題解決；(4) 價值評量—可以持有對職業價值觀點的能力 (Cumming & Maxwell, 1999)。本兩年期的研究計畫即為發展評量自然與科技學習領域六項學習指標的真實性評量試題，進一步根據上述四種真實性評量面向加以設計，皆是以情境評量為基礎，再分別整合探究評量、問題導向評量及價值評量。評量的形式分為兩種：紙筆測驗（選擇題與開放性試題）與實作評量。詳細分類與試題型式，請參考後述之測試方法中的試題發展內容。

測試方法

在這個段落，將呈現本研究的研究設計、取樣方法、及試題的發展。最後，亦將闡述本研究的研究資料分析。

(一) 研究設計

本研究是一個經由國科會與教育部共同補助的兩年期計畫，台灣師範大學邱美虹教授為計畫總主持人。研究主要分為三個階段：研究與工具的發展階段、實際施測階段、與資料分析階段。在研究與工具的發展階段，主要在發展具有核心情境的評量工具來評量科學素養的六個學習指標，包括過程技能、科學認知、科學本質、科學態度、思考智能、科學與資訊的應用 (教育部, 1998)。經由評量工具的信、效度檢驗之後，進行實際施測階段。在資料分析階段，進一步要了解真實性評量對於探討科學素養的可行性，及學生的表現為何。



(二) 取樣方法

本研究所採用的分層隨機抽樣方式是依據 TIMSS 的抽樣原則，從台灣的北部、中部、南部及東部，共抽樣了 44 所學校，1,503 位九年級學生（平均年齡 15 歲）參加此真實性評量，最後的資料分析有效樣本數為 1,469 位學生，而由於經費的限制，只有邀請六所學校參加實作評量的部份，總共 193 位學生參加。台灣的北部、中部、南部及東部各區的詳細參與人數，請參考表二。

表二、在各區域中依學校大小抽樣的班級數目

學校大小 區域	北區 班級數/人數	中區 班級數/人數	南區 班級數/人數	東區 班級數/人數	總和 班級數/人數
大	17/517	8/264	8/271	—	33/1052
中	3/93	4/134	3/100	—	10/327
小	1/21	1/34	2/55	1/14	5/124
總和	21/631	13/432	13/426	1/14	48/1503

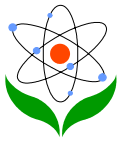
註：學校大小之定義，「大」為校內國三班級數超過 13 個班級；「中」為校內國三班級數超過 6 個班級，但少於 12 個班級；「小」為校內國三班級數少於 5 個班級。

(三) 試題的發展

本研究評量工具之發展是根據 Cumming 及 Maxwell (1999) 所提出的真實性評量理論為基礎，以調查台灣九年級學生的科學素養。為發展評量自然與科技學習領域六項學習指標的真實性評量試題，進一步根據 Cumming 及 Maxwell (1999) 所提出的四種真實性評量面向加以設計，皆是以情境評量為基礎，再分別整合探究評量、問題導向評量及價值評量。評量的形式分為兩種：紙筆測驗（包含選擇題與開放性試題）、與實作評量。詳細分類與試題型式，請參考表三。

表三、符合自然與科技學習領域六項學習指標的真實性評量設計說明。

真實性評量的特徵		六項學習指標	試題型式
情境評量	問題為本評量	科學認知	選擇題 開放性試題 實作試題
		科學與資訊的應用	選擇題 開放性試題
		思考智能	開放性試題
	探究評量	過程技能	開放性試題 實作試題
	價值評量	科學本質	開放性試題
		科學態度	李克氏量表



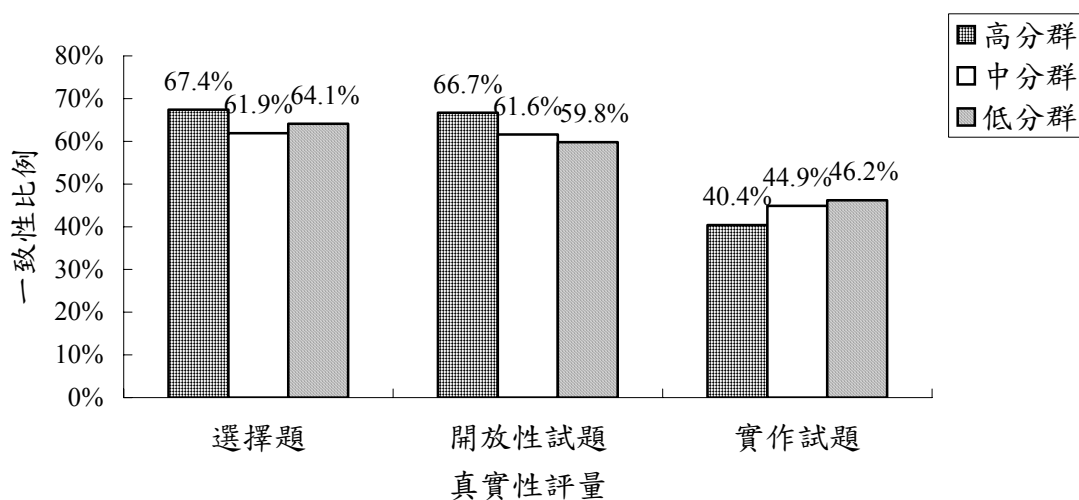
本文章主要著重在科學認知指標的設計與分析結果的陳述。在試題數目方面，認知指標有 21 題選擇題、4 題開放性問題 (60 分鐘的施測時間)，而實作試題只有一題 (30 分鐘的施測時間)。紙筆測驗的選擇題與開放性試題之題目的核心情境是以五位同學一起到陽明山露營為主，試題包括：學生們在組裝驅動車時，電學概念的應用；煮水餃時溫度與熱的概念應用；吃火鍋時的現象；咖哩雞料理包的營養成分標示等等。而實作試題則是提供學生們每人四小瓶的眼藥水滴瓶，裡面盛裝四種不同的透明溶液，同時提供學生酸鹼指示劑與透明膠片，讓學生在透明膠片上自己進行實驗，辨別出四種透明溶液分別為 $\text{HCl}(\text{aq})$ 、 $\text{NaCl}(\text{aq})$ 、 $\text{CaCl}_2(\text{aq})$ 、還是 $\text{CaCO}_3(\text{aq})$ 。詳細試題的內容，請參考 Chang and Chiu (2005)。

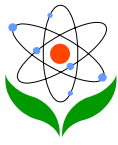
經由專家效度的檢驗之後，進行 Guttman 信度的測試結果，21 題選擇題為 0.63 (Lambda 2)，而 4 題開放性問題為 0.80 (Lambda 2)，實作試題為 0.85 (Lambda 2)。

研究結果

研究結果，以筆者所參與研發的科學認知評量為例，試題包括三個部份：21 題選擇題、4 大題 32 小題的開放性試題、與一題實作測驗。就紙筆測驗 (21 題選擇題、4 大題 32 小題的開放性試題) 之答對率而論，對於生活中常見現象的問題形式，學生成績表現不理想，只有一些記憶性問題 (如：酒精的成分為何?)，或問題與課本內容較接近的題目 (如：氣體的原子表現形式) 回答的較好。而針對實作測驗的結果，在 193 位學生的施測後，就每一試題之答對率而論，學生對於操作式問題形式表現頗理想，只有在回答「為什麼」的這類開放性問題表現較差。

進一步，結合台灣的高中入學考試成績 (基本學力測驗) 與本真實性評量的結果進行分析，發現高分群 (前 27%) 的學生，只有 66% 左右的學生在選擇題與開放性試題的表現也屬於高分群，而甚至只有 40.4% 的學生在實作評量的部份也屬於高分群，此結果說明了真實性評量的傳統標準化測驗的差距 (圖一)。此外，若依據真實性評量的總成績與實作評量的表現將學生分別分成高、中、低三群，進行相關性分析，發現不管選擇題、開放性試題或實作評量皆與真實性評量具有高相關性，其中實作評量與真實性評量的表現，以低分群學生表現的一致性最高 (表四)。





圖一、基本學力測驗與真實性評量表現的一致性。

表四、真實性評量總成績與實作評量部分表現的相關性分析。

實作試題	高分群	中分群	低分群
真實性評量			
總成績	0.414*	0.273*	0.507*

*.顯著水準為 0.01 (雙尾檢定)。

結論與建議

由本研究的真實性評量結果，除了證實了發展真實性評量於評量學生科學素養的可行性之外，從本文章所呈現之台灣九年級學生在科學認知指標的表現，更告訴我們 真實性評量可以測出一般傳統標準化測驗所不能測量的認知能力，傳統標準化測驗表現優良的學生未必能在真實性評量有同樣的表現，換句話說，在傳統標準化測驗 表現優秀的學生，是否踏出校園工作後，可以展現同樣的能力，值得深思！

此外，針對在學校的智育成績表現較差的學生，身為教師者應該利用各種多元的評量方式來了解學生的其他潛質。在本研究的施測過程發現，大部分學生對於實作評量的試題非常感興趣，在評量的過程，也都顯露出微笑，這些是傳統紙筆測驗無法觀察到的表情；而本研究結果也告訴我們，實作評量與基本學力測驗的一致性較高，因此，從簡單的實作活動來教導與評量低成就的學生是從事科學教育的我們，可以使力的方向。

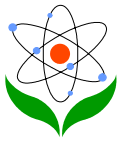
在現今科學與科技快速發展的時代，科學素養的意涵包括要讓學生了解科學的方法、重要的科學概念與詞彙，及科學與科技可能對社會所帶來的衝擊 (Miller, 1983)。綜合上述，從真實性評量的研究結果，相信藉由真實性評量的發展，除了可以較確切地評量到學生是否能將課堂上學習的成果展現於真實生活的現象與問題解決之外，更可以激發學生將學校學習的知識轉換至生活真實情境的動力；此外，筆者也認為，在知識爆炸的現代，獲得知識的管道十分多元且方便，因此，學生所需具備的能力應該從過去的知識導向轉變成知識應用與批判資訊的能力養成，相信從發展真實性評量的方向著手，對於了解學生的科學素養與促進科學素養目標的達成，是可以有實質的幫助與貢獻。

致謝

感謝國科會與教育部之研究補助 (NSC89-2515-S-003-012-X3 及 NSC90-2511-S-003-101-X3)，邱美虹教授在研究上的領導，以及研究團隊所有教授們所提供的寶貴建議。

參考文獻

教育部 (1998)。國民教育階段九年一貫課程總綱綱要。



- Aikenhead, G. S., & Ryan, A. G. (1992). The development of a new instrument: "views on science-technology-society" (VOSTS). *Science Education*, 76(5), 477-491.
- Champagne, A. B., & Newell, S. T. (1992). Directions for research and development: alternative methods of assessing scientific literacy. *Journal of Research in Science Teaching*, 29(8), 841-860.
- Chang, S. N., & Chiu, M. H. (2005). The development of authentic assessments to investigate ninth graders' scientific literacy: in the case of scientific cognition concerning the concepts of chemistry and physics. *International Journal of Science and Mathematics Education*, 3(1), 117-140.
- Cumming, J. J., & Maxwell, G. S. (1999). Contextualising authentic assessment. *Assessment in Education*, 6(2), 177-194.
- Dori, Y. J. (2003). From nationwide standardized testing to school-based alternative embedded assessment in Israel: students' performance in the matriculation 2000 project. *Journal of Research in Science Teaching*, 40(1), 34-52.
- Laugksch, R. C. (2000). Scientific literacy: A conceptual overview. *Science Education*, 84, 71-94.
- Laugksch, R. C., & Spargo, P. E. (1996). Development of a pool of scientific literacy test-items based on selected AAAS literacy goals. *Science Education*, 80(2), 121-143.
- Miller, J. D. (1983). Scientific literacy: A conceptual and empirical review. *Daedalus*, 112(2), 29-48.
- Newmann, F. M., & Archbald, D. A. (1992). The nature of authentic academic achievement. In H. Berlak, F. M. Newmann, E. Adams, D. A. Archbald, T. Burgess, J. Raven & T. A. Romberg (Eds.), *Toward a New Science of Educational Testing and Assessment*. Albany, NY: State University of New York Press.
- Thomas, G., & Durant, J. (1987). *Why should we promote the public understanding of science?* Oxford, UK: University of Oxford.
- Wiggins, G. P. (1989). A true test: toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.
- Wolf, D., Bixby, J., Glenn, J. I., & Gardner, H. (1991). To use their mind well: Investigating new forms of student assessment. *Review of research in education*, 17, 31-73.
- Yerrick, R. K. (2000). Lower track science students' argumentation and open inquiry instruction. *Journal of Research in Science Teaching*, 37(8), 807-838.