

科學教育評量的趨向

黃寶鈿

國立臺灣師範大學化學系

壹、緒 言

在科學教育方面，教師往往會想到下列問題：「教些什麼？」「如何教？」「教學的對象在未學習前的科學能力為何？」「學生在學習過程中，有那些優點和缺點？」「如何預測學生未來的可能成就？」等等。因此，評量是科學教育計畫中的重要活動之一，是不能忽視的。（圖一）

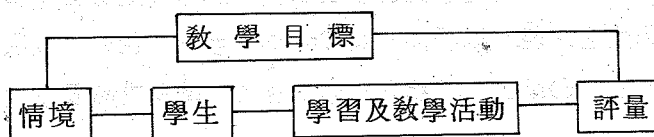


圖 一

科學教育的評量範圍包括很廣，諸如：學生對科學過程的了解，學生科學能力的甄選，多元天才（Multi-talent）的發現，科學態度的培養，實驗技能及科學態度的判斷，預定學習成果的評估，以及教師能力的自我評量等。但是，不管評量的對象是自己或他人，都是一件不容易的工作。

在介紹科學教育評量的趨向之前，首先說明評量之定義及其功能。

貳、什麼是評量？

阿曼（Ahmann）及古勞克（Glock）^① 對評量所下的定義是：「……評量是教師通常從各種不同來源來收集資料，並且分析資料的過程，這種過程是以達到價值判斷的目的。收集這些資料的方法，可利用測驗的方式，亦可由其他技術獲得，但不一定有數量（Quantity）之數據，……。」

一般科學教師認為評量就是「考試」或「分數」，是不十分正確的。評量（Evaluation）的英文字根是「Value」，文中意義是「價值」，所以，評量是對某一「事」或「物」

的評價；亦即「判斷」(Judge)或「決定」(Decide)某「事」或「物」是「好」是「壞」、「滿意」、「不滿意」或「尚可」等等。根據報載，行政院國家科學委員會曾於民國七十二年四月二十日舉行全國國中數學及自然科學抽測，其目的就是在於評定目前國中數理科的學習情況。

參、評量的功能

1. 評估學生的個人學習成就，以判斷其努力的程度，作為認定其資格及熟練技巧的方法。例如化學學習成就之評量⁽⁹⁾。
2. 因為評量可以提供教師回饋(Feedback)，所以，可以診斷學生個人或全班之學習困難，以補救教學，乃個別化教學的一種過程。
3. 評估課程例如教材及教法的教育成敗，以改進教學，有助於計畫的評鑑。
4. 歐開(Okey)⁽⁵⁾研究的結果表示診斷性(Diagnostic)評量的次數愈多，學習成就愈大。因此，評量可促進學生的學習動機，刺激學生努力學習。
5. 辨識學生之才能，以提供分類或安置之參考。例如：我國目前資優實驗班的學生是根據其個別智商一三〇以上，並參考其創造能力、學業成績及學生評量等資料，擇優選取而編班的。
6. 預測學生的未來成就或學習某一學科的潛能。例如：科學性向測驗，得分高的學生表示學習自然科學的能力較強⁽⁷⁾。

肆、科學教育評量之趨向

在美國已有一些科學教育者指出未來科學教育評量之特別趨向，尤其相關的科學教育目標顯示出未來的評量將是較複雜的，評量的範圍也將擴大⁽⁸⁾，因為科學對人類福祉及社會問題的關係是人類特別重視的教育目標之一。

美國國家科學教師協會(NSTA)，認定一些除了紙筆式的評量外，還應包括學生自我評量(Self-evaluation)，效標實作評量(Criterion performance evaluation)、高層次思考能力的評量、情意領域(Affective domain)行為目標之評量等，因此，評量之技巧方面也應考慮這些不同的方式，所以，一九八〇年代的科學教育之評量應接受這些挑戰。

科學教育評量之趨向可歸納成下表，左欄表目前或過去之情況，右欄表示未來的趨向。

由……	趨向於……
1.團體施測為主。	多種施測模式：包括大團體、小團體及個別的評量。
2.紙筆式測驗為主。	不同測驗模式：包括圖形測驗及實驗的操作測驗。
3.總結性之評量。	不同評量方式：包括前測、診斷性及形成性之評量。
4.低層次的認知領域之評量。	高層次的認知領域之評量（例如：分析、評鑑的能力，創造性思考能力之評量）及情意領域之評量（例如：態度、興趣、價值觀之評量）及技能領域的評量。
5.常模參照成就評量為主。	更包括效標參照評量，精熟測驗（Mastery testing），自我評量及同儕評量（Peer evaluation）。
6.評量科學事實及科學原理之了解。	所包含之目標及科學過程、科學本質、科技及社會有關。
7.評量學生之學習成就為主。	亦評量課程、教法、計畫之是否有效。
8.教師自編測驗為主。	靈活運用教師自編測驗，標準化測驗（Standardized test），研究工具及教師所收集之題庫。
9.以考慮評量之總成績為主。	亦注重分測驗成績、試題難度、鑑別度並利用機械或電腦來處理這些資料。
10.單向評量模式為主。（例如僅以分數、等第來表示評量結果）	多向評量方式，考慮學生對科學概念、科學過程、實驗步驟、班上討論及解決問題之技巧與能力。
11.評量對象只限於普通學生。	亦注重資優、益智及殘障學生之科學教育的評量。

現在大部分的評量著重於考試，其特點是使用同一種試題，只有一位老師主持考試，施測的對象是一群學生（一個班級或好幾個班級）。雖然，這種團體的施測方式仍然在使用，

，但也盛行個人或小團體的施測方式。例如：皮亞傑（Piaget）認知能力的面試方式。因為課程及教學方法目前很重視個別差異，所以，評量的方式也重視個別測試的方法，利用電腦的評量就是根據此原則而設立的。強生（Johnson）⁽⁴⁾、貝克（Baker）和弗列得力克（Frederick）⁽²⁾等人曾經利用電腦設計一些題庫，需用時，可從電腦選擇測試內容相當之不同題目。為配合個別化之教學，教師可從題庫中選擇適合之題目來測試學生。

雖然，目前大部分的考試仍依賴紙筆式測驗（Paper-and-pencil-test），未來的評量過程將採用多種不同的評量工具。例如：實驗室實作測驗，或圖畫式測驗（pictorial test）。圖畫式測驗可以減少因語文閱讀之能力的差異而影響測試結果。教師宜採用各種不同之測驗方式，以評量學生是否已學會某些科學原理、原則或科學過程。

學校中最常見的考試，是屬於總結性的評量（Summative evaluation），比如期末考或大學入學考試⁽⁸⁾等，科學教育的評量已趨向於其他的測量方式，例如：診斷性的考試，教某一科目之前的前測（Pretest）及形成性的評量（Formative evaluation）。科學教育的診斷性評量（Diagnostic evaluation）之重點在評估學生對於某特殊單元所需要的技巧或能力是否能夠達到預定之目標。前測是在教學之前進行施測的，且著重於評量與教學成果有關係之不同變因。多數前測用來評估科學之事實及原則，較少用來評估教材中有關科學方法的部分。教師若發現學生在某些單元方面有弱點，則需補救學生的弱點，以免造成學生的學習困難；若仍不成功，可能就要使用特殊教材或教法以進行個別化教學了。

大多數教師自編測驗及科學成就的標準化測驗，其大部分內容含低層次之認知（記憶性）領域，雖然，這些內容是某些科學課程的重要部分，但，並不是唯一的目標。高層次認知之評量題目則較不易命題，命題所費時間也較多，這些高層次之題目宜配合課程、教師、學生、教學目標等。例如：評量批評性思考，解釋數據及形成假說的題目應包括層次在「記憶」以上的認知領域。有些情意的（affective）教學結果也應包括在科學評量之範圍內，比如價值與科技之關係的題目，即屬於情意方面的領域。

過去或現在大部分的評量過程的參照標準是一種「常模群體」（norming group）。標準化測驗之參照標準為從母群體中隨機抽取之樣本。教師自編測驗之常模群體包含一班，或同一教師指導下之一些班級。個別學生之表現往往與其他同學的表現相比，並利用某一種常模群體。

雖然，科學原理或事實是科學教育的主要目標之一，但是，要培養學生成為具有科學素養的國民，比訓練學生成為科學家是更重要的科學教育目標。因此，未來的評量應包括科學過程、科學本質及科技、社會和科學相關方面之內容。所以，認知（cognitive）、情意（affective）及技能（psychomotor）三領域，皆應涵蓋於科學教育之評量。

傳統的測驗與評量多半著重學生的學習成就，而課程及教法等之評量也開始受重視。這

些多層面的評量之最後目的是更進一步評估各因素之彼此關係或是對整個評量之影響。

除了教師自編測驗之外，各種其他來源的評量工具，例如標準化測驗等也常用於科學教育的評量；尤其電腦及微電腦之普遍使用，對各種評量工具的使用也有很大的助益。

測驗成績往往以總分來表示，目前測驗結果之分析已強調各分測驗分數或各題分數，同時應用項目分析如難度、鑑別度之分析，並且利用電腦計分及存儲測驗結果之資料，以提供教師教學之參考。

學生學習自然科學之成績常以分數或等第表示，這只是單向的評量。而多向的評量尚包含更大的範圍，如概念、過程、實驗操作、解決問題之技巧、學生在班上之反應情況、及各種情意的變因。科學教育的目標已經更複雜化了，其評量範圍當然也更廣，並且也採用不同的評量方式，如自我評量、教師評量、父母評量及同儕評量。

評量的對象方面，以前只評量一般的學生；目前科學教育之評量也重視資賦優異學生或殘障學生之評量，當然，所使用的工具也因目的而各異了。

伍、結 語

科學教育之評量結果之處理，對教師、家長、學生而言，是一種挑戰，尤其電腦的廣為應用，這些成績或評量資料如何保密，尤其需要警覺。

科學教師宜發展及應用各種評量工具來評估，預測及診斷學生之學習，以達到教學目標、單元目標及學生之需要，並且要利用評量結果以改進教學方法，並謀求補救學生之學習困難，或發現學生之潛能。

參考文獻

- (1) Ahmann, J.S. & Glock, M.D. *Evaluating Student Progress*. Boston : Allyn & Bacon, 1981. p. 32.
- (2) Barker, F. & Frederick, R. Computerized test generation in secondary Science teaching. *Science Teachers' Bulletin*. 1976, 41 (2), 35-37.
- (3) Butts, D. *Designs for Progress in Science Education*. Washington, DC : National Science Teachers Association, 1969.
- (4) Johnson, D.H. CAGEE — An Approach to Computer — Assisted Generation and Evaluation of Examinations. Brockport, N.Y. : SUNY College at Brockport, 1973.
- (5) Okey, J.R. Diagnostic testing pay off. *The Science Teacher*. 1976, 43(1), 27.
- (6) 黃寶鈿：高中化學學習成就的研究，化學第三期，民68。
- (7) 黃寶鈿：國中學生科學性向與化學成就關係之研究。1981 美國NSTA年會發表。
- (8) 黃寶鈿：大學入學考試化學科成績相關之研究。中等教育，民67。