

教 育 論 壇

主題：落實國小自然科多元教學與評量

多元化評量之理念與方法

引言人：簡茂發

國立臺灣師範大學

教學是師生共同參與而產生交互影響的動態過程；而評量則是運用科學方法和技術，蒐集有關學生學習行為及其成就的正確資料，再根據教學目標，就學生學習表現的情形，予以分析、研究和評斷的一系列工作。在整個教學歷程中，評量是承接轉合的關鍵部份，而不是教學歷程的終點站，並非表示教學活動的結束。教學評量的主要目的，在於分析教學得失及診斷學習困難，作為實施補救教學和個別輔導的依據。

由於各級學校長期使用紙筆式測驗作為學習成就評量的主要工具，簡便易行，固然發揮了相當的功能，但沿襲已久，殊少變通，流弊所及，也成為當前教育改革中熱烈討論的重要課題之一。教育部在民國八十七年九月三十日公布的「國民教育階段九年一貫課程總綱綱要」中曾提及評鑑方法應採多元化方式實施，兼重形成性和總結性評鑑（教育部，民 87 b）。因此，多元化教學評量理念與方法之評介，在課程及教學的研究發展上，實為當務之急，有其必要。

壹、學習評量概念的演進

檢視過去和最近的相關文獻，學習評量的發展演進可由其所用的名詞與涵義，分為三個階段。美國在「八年研究」(The Eight-year Study) 時代之前，強調的是“Measurement”以量化的方法取得正確可靠的數據；到後來則認為應該從教育的目標、人格的發展各方面來進行評量，亦即除了客觀的數字之外，尚須有一些價值標準來加以衡鑑，而將“Measurement”提昇至“Evaluation”；晚近，學者們又將“Evaluation”提昇至“Assessment”，強調評量時應考量各種相關的整體情境，從各種可行的途徑，蒐集全面性、多元化的資料，再從各個角度和不同觀點加以比較

分析與綜合研判，進行整合性的詮釋，獲致充分的了解。

貳、教學評量問題的檢討

目前中小學教育由於升學競爭而導致教學未能正常化的結果，尤其是在教學評量方面產生許多的流弊，這些有所偏失的現象，值得大家關心、省思和切實的檢討。舉其要者，包括下列各項：

- 一、偏重智育或學科知識的評量：考查記憶性知識，忽略推理思考和過程技能（process skills）。
- 二、評量偏重學習結果，而忽略學習過程的了解。
- 三、考試次數太多，考試如上戰場。學生雖身經百戰，謀取致勝之道，但未必是常勝軍，無信心，沒把握克敵致勝。在個人方面，課業負擔重，心理壓力大，害怕考試，造成考試焦慮症候群；在人際關係方面，因惡性競爭而對立，存有敵意，猜忌懷疑，彼此疏遠，不能相互尊重與合作。
- 四、常用紙筆式測驗，以坊間普遍印行的測驗卷為評量工具，千篇一律，習以為常，覺得厭惡，又無法拒絕，真有強其所難之感。
- 五、無法充分了解考試分數的意義及其所隱藏的訊息，評量的診斷功能尚未發揮。人人競逐高分，考試淪為競賽的工具，學生成為考試的機器，結果失敗挫折者居多，考試的負面效應層出不窮，如作弊、逃學……。
- 六、考試領導教學，教學未能正常化。
- 七、不適當的比較，滋生許多流弊。
- 八、升學取向，惡性補習，戕害兒童及青少年身心健康。

參、教學評量的內涵與類型

教學評量包括三大部分，即教師的教學效率之評量（evaluation of teacher's teaching effectiveness）、學生的學習成就之評量（evaluation of students' learning achievement）、課程的設計與實施之評量（evaluation of curriculum program）。

從評量的時機和性能而言，教學評量可分為「形成性評量」（formative evaluation）和「總結性評量」（summative evaluation）；從評量資料的解釋方式而言，教學評量又可分為「常模參照評量」（norm-referenced evaluation）和「標準參照評量」（criterion-referenced evaluation）。

美國教育學者 R. K. Watkins 曾將學校中通常所採用的成績評量措施，歸納為下列九種方法：教師的評判（teacher's judgement）；口頭述誦（oral recitation）；論文考試（essay examination）；標準化客觀測驗（standardized objective test）；教師自編客觀測驗（informal teacher-made objective test）；學生作品的評定（the rating of sampling of products of pupils' work）；操作的評定（performance rating）；非正式記述的評量（informal descriptive evaluation）；機械記錄（mechanical recording）。

肆、多元化評量的理念

由於教學和輔導的範圍非常廣泛，除知識技能之外，尚包括學生的品行、人際

關係、學習態度、興趣、方法和習慣等方面，所以必須採用不同的考查方法，以配合學科性質和教材內容進行多種方式的評量。

一、成績評量是多方面的：

從前一般人以為學校的成績評量只限於知識技能的考查，事實上現代國民中小學注重德、智、體、群、美五育均衡發展，應該把評量的範圍擴大至教育的全面。當我們從事某一學科的教學時，雖然以該學科本身為重點，但也不能忽略與該學科有關的其他知識，而且透過各學科之學習，學生在其態度及理想上可能有所改變或進步，因而副學習（associate learning）與附學習（concomitant learning）之結果，也應納入成績考查的範圍，不要僅以主學習（primary learning）作為成績評量之唯一對象。最近美國心理學者認為學校教學包括： 認知方面—以學生心智能力之發展為核心； 情感方面—以學生之態度、興趣、對人對事之方式以及各種鑑賞能力為重點； 心理動作方面—以各種技能為主。凡此無一不屬於成績評量之範圍。

二、成績評量必須採用多種方式：

因為成績評量之範圍非常廣泛，所以應該採取不同的考查方法，以配合教材之性質與內容，進行多種方式的評量。評量時所採用的方法愈多，蒐集資料愈齊全，其結果愈客觀正確，愈能符合成績考查的要求。

伍、多元化評量的方法

以往傳統的學習評量，大都採用標準化紙筆式測驗（paper-and-pencil test）或教師自編的課堂測驗（teacher-made classroom test），為時已久，習以為常。由於紙筆式測驗較為通俗，使用普遍，易於被非專業的人所誤用或濫用，而造成許多不良的結果。因此，現今教學評量的研究發展趨勢更注重彈性的、變通的、多元化的評量，並且強調動態的過程，乃出現“Alternative Assessment”（另類評量、變通性評量）、“Dynamic Assessment”（動態評量）和“Performance Assessment”（實作評量）等新名詞。

教學包含教師的「教」和學生的「學」。過去只要求「教學正常化」，但現在更強調「教學卓越化」。換言之，希望在教師方面，能達到「有效率的教學」（effective teaching）；在學生方面，能達到「有意義的學習」（meaningful learning）。所謂「有意義的學習」，係指兒童及青少年所學的內容與方式力求與其日常生活情境切近且密切結合，同時有所謂的“Authentic Assessment”（真切性評量）。此種評量特別注重在真實的或模擬的情境中進行考查，強調評量的內容與方式應配合兒童當前的生活經驗，使其覺得自然親切，而不致格格不入，且能充分反映其真實的行為改變或成長，獲得全面的深切了解。

此外，尚有“Portfolio Assessment”（卷例評量），係指在學生學習過程中彙集相關的各項資料，針對評量的需要，找出一些切要的项目作成適當的組合，進行成套的評量，而非零碎個別项目的考核，具有持續累積的評量效用。由於評量不只限於靜態的作品，還要評量作品產生的過程以及作品的優劣，是一種整體系統性的評量，因而又連帶出現一種“Work Sampling System”（工作取樣系統），亦即從工作項目中找出重要的樣本，讓學生操作反應，再進行觀察和記錄。

上述各種評量方法都是相關的，主要強調的都是要讓學生在切近生活經驗的活動中實際操作學習，再由教師進行有系統的觀察、記錄和評分。

另外，針對學科的學習，現在有所謂的“Curriculum-based Competency testing”（課程本位能力測驗）。每一個科目或課程，都有其教材大綱及教學重點，同時也要考慮教學目標，而教學目標又分為情意、認知、技能三方面。因此，在評量時，課程本位能力測驗應考慮針對該科學習的活動，包括教材內容、行為目標、學習情境、以及學生個人的背景能力等各方面；亦即課程設計、課程實施、學習情境、個人等因素都應納入考慮。

最近五年來，國立臺灣師範大學科學教育中心與美國馬利蘭州教育廳合作進行學習評量的改進計畫－MSPAP（Maryland School Performance Assessment Program）。MSPAP 就是一種超脫選擇題紙筆式測驗的評量方式，採用先舉行一個活動，讓學生參與、觀察，然後詢問他一些相關的問題，讓學生有所表現，再根據其表現來評分。這是上述「實作評量」典型的實例，在心理計量學與教學評量領域研究發展上，頗受稱許和肯定。

教育部於民國八十七年八月二十六日修正公布的「國民中學學生成績考查辦法」規定：學校對國中學生成績之考查，應視學生身心發展與個別差異，以獎勵及輔導為原則，並依各學科及活動性質，得就下列十五種評量方式選擇辦理。

紙筆測驗：就學生經由教師依教學目標、教材內容所編訂之測驗考查之。

口試：就學生之口頭問答結果考查之。

表演：就學生之表演活動考查之。

實作：就學生之實際操作及解決問題等行為表現考查之。

作業：就學生各種習作考查之。

設計製作：就學生之創造過程及實際表現考查之。

報告：就學生閱讀、觀察、實驗、調查等所得結果之書面或口頭報告考查之。

資料蒐集整理：就學生對資料之蒐集、整理、分析及應用等活動考查之。

鑑賞：就學生由資料或活動中之鑑賞領悟情形考查之。

晤談：就學生與教師晤談過程，了解學生反應情形考查之。

自我評量：學生就自己學習情形、成果及行為表現，做自我評量與比較。

同儕互評：學生之間就行為或作品相互評量之。

校外學習：就學生之校外參觀、訪問等學習活動考查之。

實踐：就學生之日常行為表現考查之。

其他。（教育部，民 87 a）

例如：自然科學課程強調科學概念、科學方法和科學態度之教學，故考查學生的學習結果時，必須兼顧這三方面的評量，不可有所偏廢。科學概念的評量，可採用口頭考問、紙筆測驗或情境測驗等方式進行之；科學方法注重過程技能（process skills）之訓練，此方面的評量宜以實際操作方式實施之；至於科學態度方面的評量技術，較難達到客觀衡鑑的要求，必須同時並用行為觀察法、評定量表、自陳量表、項目檢核表及個別晤談等方式，設法蒐集多方面的資料，以便比較分析與綜合研判。

陸、因材施教與自我比較

教育最基本的理念與方法，就是「有教無類」和「因材施教」。學齡兒童及青少年來自不同的家庭社經背景，在身心發展及行為表現各方面，都有個別差異的現象。在當前以班級教學為主要的教育體制下，應先充分了解學生個別差異，以便普遍實施適性教育，促使學生各有所長的才華，都能因多樣化的學習情境，而獲得有尊嚴且快樂的成長。既然配合學生的個別差異，實施適性教育，把每個學生帶上來，理應把握「因材施教」及「因材評量」的原則，進行多元化評量措施，採自我比較的方式，解釋分析各方面整合評量的結果。

學校教師應該提供各種教育活動的機會，採用各種適性評量的方式，讓每個學生都能充分發揮才華，把努力學習的成果從各方面表現出來；再就各科成績作適當的比較分析，以顯示出相對的意義。從教育和評量的觀點來說，最好採用自我比較的方式，也就是以學生本人的學習潛能及實際的學習表現相互參照作合理的解釋分析。通常有三種方式，第一基於學生個人的潛能來解釋他在學科成績上的相對意義，可以看出他「努力」的程度；第二把學生個人過去和現在歷次考試評量所得的成績前後加以比較，可以看出「進步」或「退步」的情形；第三把學生個人在許多不同學科所得到的成績畫成側面剖析圖，從中可以清楚地看出各科分數的高低，可以了解他在各學科領域學習成果的相對優劣，也可以凸顯出整個學習結果在各方面的「長處」和「短處」，作為進一步學習輔導的依據。總而言之，我們應該注重因材評量，採用自我比較的方式。

柒、結語

學生學習成就的評量，必須顧及個別差異，符合因材施教、因材評量的原則。成績評量涉及教育各方面，因而必須採用多種方式。多元化評量是教學評鑑與心理計量學研究發展的最新趨勢，有其理論基礎與實務應用的意義。傳統的紙筆式測驗，因評分標準客觀，分數較少誤差，且施測方便，乃廣被採用，但有時過度僵化，並非最佳的評量方式。事實上，教學評量除了一般的紙筆式測驗之外，尚有許多其他變通的方式，可視學科性質、教學情境及評量重點等因素，酌採其中幾種不同的評量方法，作彈性的搭配運用，以發揮最大的評量功能。

參考文獻

- 王寶壩（民 84）。現代測驗理論。台北市：心理。
- 余民寧（民 86）。教育測驗與評量：成就測驗與教學評量。台北市：心理。
- 吳武典、簡茂發等（民 83）。我國心理與教育測驗彙編。台北市：中國測驗學會。
- 周文欽、盧欽銘等（民 86）。心理與教育測驗。台北市：心理。
- 教育部（民 87a）。國民中學學生成績考查辦法。台北市：教育部。
- 教育部（民 87b）。國民教育階段九年一貫課程總綱綱要。台北市：教育部。
- 路君約（民 81）。心理測驗。台北市：中國行為科學社。
- 簡茂發（民 86）。心理測驗與統計方法（修訂版）。台北市：心理。

- 簡茂發 (民 82) 。測驗的編製。載於黃光雄，簡茂發主編之教育研究法 (再版) 。台北：師大書苑。
- 簡茂發、何榮桂等 (民 81) 。我國心理教育測驗彙編。台北市：中國測驗學會。
- Aiken, L.R. (1997). *Psychological testing and assessment (9th ed.)*. Boston : Allyn & Bacon.
- American psychological association. (1997). *Standards for educational and psychological testing (revised ed.)*. Washington, DC: Author.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing (7th ed.)*. Upper Saddle River, NJ: Prentice-Hall.
- Aylward, G. P. (1994). *Practitioner's guide to developmental and psychological testing*. New York : Plenum Medical Book Company.
- Banta, T.W., Lund, J. P., Black, K. E., & Oblander, F.W. (1996). *Assessment in practice: Putting principles to work on college campuses*. San Francisco, CA: Jossey-Bass.
- Conoley, J.C., & Impara, J.C. (Eds.) . (1995). *The twelfth mental measurements yearbook*. Lincoln, NE: The University of Nebraska Press.
- Cronbach, L.J. (1990). *Essentials of psychological testing (5th ed)*. New York : Harper Collins.
- Cunningham, G. K. (1998). *Assessment in the classroom: Constructing and interpreting tests*. Washington, DC: The Falmer Press.
- Devellis, R.F. (1991). *Scale development: Theory and applications*. London : Sage.
- Friedenberg, L. (1995). *Psychological testing: Design, analysis, and use*. Boston: Allyn & Bacon.
- Gallagher, J.D. (1998). *Classroom assessment for teachers*. Upper Saddle River, NJ: Prentice-Hall.
- Gregory, R.J. (1996). *Psychological testing: History, principles, and applications (2nd ed.)*. Boston: Allyn & Bacon.
- Hopkins, K.D. (1998). *Educational and psychological measurement and evaluation (8th ed.)*. Boston: Allyn & Bacon.
- Kaplan, R M., & Saccuzzo, D.P. (1997). *Psychological testing: Principles, applications, and issues (4th ed.)*. Pacific Grove, CA: Brooks/Cole.
- Kline, P.L. (1993). *The handbook of psychological testing*. London : Routledge.
- Lindquist, E.F. (Ed.). (1951). *Educational measurement*. Washington, DC: American Council on Education.
- Linn, R.L.(Ed.). (1989). *Educational measurement (3rd ed.)*. New York: American Council on Education / Macmillan.
- Linn, R.L., & Gronlund, N.E. (1995). *Measurement and assessment in teaching (7th*

- ed.). Englewood Cliffs, NJ: Merrill / Prentice-Hall.
- Lyman, H.B. (1998). *Test scores and what they mean* (6th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Murphy, K.R., & Davidshofer, C.O. (1998). *Psychological testing: Principles and applications* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Oosterhof, A. (1994). *Classroom applications of educational measurement* (2nd ed.). New York: Maxwell Macmillan International.
- Phye, G.D. (Ed.) (1997). *Handbook of classroom assessment: Learning, adjustment, and achievement*. San Diego, CA: Academic Press.
- Popham, J. M. (1995). *Classroom assessment: What teachers need to know*. Boston: Allyn & Bacon.
- Sax, G. (1996). *Principles of educational and psychological measurement and evaluation* (4rd ed.). Belmont, CA: Wadsworth
- Thorndike, R. L. (Ed.). (1971). *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.
- Thorndike, R. L. (1982). *Applied psychometrics*. Boston: Houghton Mifflin.
- Thorndike, R. M., & Lohman, D. F. (1990). *A century of ability testing*. Chicago: Riverside.
- Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Walberg, H.J., & Haretel, G.D. (Eds.). (1990). *The international encyclopedia of educational evaluation*. Oxford: Pergamon Press.
- Wholey, J.S., Hatry, H. P., & Newcomer, K.E. (Eds.). (1994). *Handbook of practical program evaluation*. San Francisco, CA: Jossey-Bass.
- Worthen, B.R., White, K.R., Fan, X., & Sudweeks, R.R. (1999). *Measurement and assessment in schools* (2nd ed.). New York: Longman.
- Wragg, T. (1997). *Assessment & learning: Primary and secondary*. New York: Routledge.

小 啓

本刊第四十五期「從知識創新談教育效能之提昇」一文，係陳書梅教授於本館主辦之「現代教育論壇」現場演講討論的內容，該文亦同時刊載於書苑第五十一期（90年1月），65-71頁。

