

教學評量的基本原則

高博銓

台灣師範大學教育研究所研究生

摘要

教學評量是教師教學過程中，相當重要的一環，而其所發揮的診斷和回饋功能，則有助於改善教學的成效，促進教學目標的達成。基於此，本文以教學評量作為探討的主題，旨在分析實施教學評量時，所應秉持的基本原則。而根據相關文獻的探討，本文指出教學評量的實施，應該採取綜合性、全程性、真實性、有效性、以及多樣性的原則，寓評量於教學之中，打破傳統將教學與評量分離的思維模式，以獲致較佳的教學成效。

關鍵詞：教學評量

壹、前言

教學評量是教學過程中，相當重要的一環，也是達成教學目標的主要手段。一般而言，教師會經常利用評量的回饋作用，監控並診斷學生的學習概況，以適時地調整教學的策略，改善其教學的成效。不過，觀諸教育的現況，可以發現，教學評量事實上並未能充分發揮其預期的功效，尤其受到社會傳統文化、升學主義、以及文憑主義的影響，使學校、家庭、教師、以及學生幾乎將全部資源導向考試，忽略正常的教育目標和理想，影響所及，教學目標嚴重地受到窄化，許多評量的良法美意，常被扭曲、誤解，造成評量失焦的現象。值得注意的是，近年來，由於教育改革的呼聲不斷地升高，其中學校教育的僵化評量方式，備受抨擊，是以，如何改善現行的評量方式，發展符合評量原則的教學評量，已成為教育改革的重要課題。至於行政院教育改革審議委員會（1996）在其《教育改革總諮議報告書》中，所提出的多元入學方式，主張暢通升學管道，舒緩升學的壓力，以回歸教育的本質，則對於教學評量的改革，深具意義。

貳、教學評量的重要性

教學評量是瞭解學生學習情形，以作為調整教材教法和補救教學的依據，其對

於教師、學生、乃至於學校行政人員，所具有的回饋和決策功能，殊為重要。事實上，以評量為導向的教育改革已成為近年來提昇教與學的標準，促進有效學習，以及公共績效的重要途徑 (Black,1998)。進一步來說，教學評量在提供資訊，以作為決策之用，不過，其決策的範疇，主要是有關於學生個人和課程教學改進的決定，而不涉及有關行政工作調整的決定，後者如評鑑學校行政運作、教職員工作表現、成本效益等，應是屬於教育評鑑的範疇，所以，教學評量可以視為是教育評鑑的一環，旨在促成學生的安置與選擇決定，以及教學決定為主要功能，至於其重要性則可以從下述三項功能來加以說明（李咏吟、單文經，1998；何英奇，1992；郭生玉，1991；簡茂發，1996）：

一、教學決定的功能

教育學者 Glaser 曾指出，教學歷程應該包含四個要素：教學目標、起點行為、教學歷程、以及教學評鑑，而美國教學評量專家 Kibler 也在其出版的《教學目標與評量》(Objectives for Instruction and Evaluation) 一書中，提出所謂的「教學基本模式」(The General Model of Instruction, GMI)，指出教學的基本歷程可以分為教學目標、學前評估、教學活動、評量等四部分，進而闡述四者之間的交互關係，並特別強調評量的回饋作用及積極功能。基於上述，教學評量對於教師理應具有下列四項功能：

- （一）瞭解學生的起點行為。透過預備性評量，教師可以瞭解學生的起點行為，以作為訂定教學目標的參考。
- （二）建立確實可行的教學目標。教師瞭解學生的起點行為，對照原先設定的教學目標，據以提高、降低或維持教學目標。
- （三）確定教學目標達成的程度。透過形成性評量或總結性評量以掌握教學目標達成的程度。
- （四）改進教材教法，增進教學效果。評量的結果可以回饋給教師，作為改進教材教法之依據。

二、輔導與諮商決定的功能

對學生而言，教學評量的功能為：

- （一）增進學生瞭解教學目標。
- （二）激發學生的學習動機。
- （三）協助學生瞭解自己的能力、性向、潛能、以及人格特質等。
- （四）診斷學生的學習與適應困難。
- （五）輔導學生做最佳的升學、就業、生活方式等選擇。

三、行政決定的功能

除了教學決定的功能、輔導與諮商決定的功能外，教學評量也具有行政決定的功能，這主要是就教育或學校行政人員來說，其功能包括：

- （一）選擇決定。例如大學入學甄試、啟智班與資優班學生的選擇。
- （二）安置功能。譬如能力分班，把學生安置到適當的班級。

(三) 分類決定。指將個人分派到不同類別的工作、職務、學校或科系中。

(四) 課程與教育計畫決定。對於新課程、教法、教育革新計畫與研究等，需經實驗、評量後才能決定是否可行。

參、教學評量的基本原則

誠如上述，教學評量旨在瞭解學生的潛能與學習成就，也可以發現學生的學習困難，作為補救教學及個別輔導的依據，同時亦能評估教師教學的效率，以提供教師改進其教材、教法的參考。而近年來評量思維的轉變，也代表了學習、學校教育、以及教學方面，產生了典範的轉移。在此趨勢下，教師對於評量的判斷往往植基於學童的先備知識，而且會善用教學過程中，所取得的回饋，同時也必須致力於滿足學生不同的學習需求 (Tunstall & Gipps, 1996)。並且學習與同仁、家長、以及學生共同分享其教與學的決定 (Stiggins, 1996)。也就是要重新思考評量與教學的目的，思索如何利用評量和教學活動以有效地支持學習，同時對於可以達成的課程的目標、內容、與標準，預先加以評估 (Wiggins & McTighe, 1998)。此外，透過教學評量也可以知悉學生學習進步的情形，從而激發其學習的動機。然而，欲充分發揮上述的諸項功能，就應該確實掌握教學評量的基本原則，靈活運用，才能引導教學的進行，獲致理想的成效。至於其相關的基本原則，臚列如下：

一、綜合性

教學評量的綜合性原則是指教學評量的內容能夠涵蓋各種不同種類的學習內容。有關學習內容的分類，Bloom 認為，人類的學習可以分為認知、情意、技能三大類，認知領域包括了知識、理解、應用、分析、綜合、以及評價；情意領域包含接受、反應、評價、組織、以及品格；至於技能領域則包括知覺作用、心向作用、引導反應、機械反應、複雜反應、技能調適、以及創作表現（張春興，1994）。此外，Gagne 認為學習應該包括智能技巧、語言訊息、認知策略、動作技巧、以及態度等五大類型，並進一步將智能技巧分成區辨、具體概念、定義概念、規則、以及高階規則 (Gagne, 1985)。由是觀之，人類的學習具有不同的類別，據以因應生活世界的不同需求和挑戰，而教師的職責就是要提供學生習取各種類型能力的機會，以增進學生適應社會生活的能力，而教學評量的過程中，能夠涵蓋這些不同的學習層面，自然有助於教育目標的達成。

二、全程性

教學評量的全程性原則是指教學評量的實施兼重過程與結果，掌握學生全程的學習狀況。一般而言，從評量的時機與性能來看，教學評量可分為「形成性評量」(formative evaluation) 和「總結性評量」(summative evaluation)。形成性評量可以協助教師獲得教學歷程中連續性的回饋，隨時瞭解學生們學習的概況，適時調整其教學的策略，同時，亦可提供學生建設性的回饋，一方面可以強化本身優異的學習機制，另一方面也可以反思其學習的缺失。是以，形成性評量對於教師和學生所提供的回饋，可以有效地引導教學和學習活動，進一步提昇教學的品質。而總結性評量

則是在課程結束之際，用來檢視學生是否達成預期的教學結果，其實施的方式往往是單元學習的後測 (posttests) 或是期末測驗。除了上述有關學生的學習成效或效能之外，總結性評量也常常包括 (Kemp, Morrison, & Ross, 1994: 159)：

- 學習的效率（教材的精熟度／時間）。
- 發展方案的成本效益分析。
- 後續的投入和支出。
- 對課程或方案的反應。
- 方案長期獲益的情形。

是以，總結性評量旨在用來評估學生是否具有最基本的先備知識，以利後續的學習，同時也在評估學生未來學習情境中的學習成就潛能。也因此，從整個學習的歷程來看，總結性評量是另一階段學習的參照，也是一種形成性的評量，兩者循環不已，關係至為密切。

三、真實性

教學評量的真實性原則旨在強調，教學評量必須在班級教學的情境下，由教師設計，並且在教室課程的脈絡之下，針對學生的表現加以評量，才具真實性 (Calfee & Hiebert, 1991)。詳言之，評量應該在真實的生活情境中進行，才能評量學習者在真實情境中的表現過程。在此過程中，評量與實際的教學息息相關，教師所採取的評量活動與評量方式，皆源自實際的教學脈絡，因此，評量完全是由實際擔任教學工作的教師所設計，其評量與教學相互結合，教學即評量，評量即教學，兩者密切配合，尋求最佳的教學成效。當前一般所指稱的「真實評量」(authentic assessment)、「實作評量」(performance assessment) 或是「另類評量」(alternative assessment) 就是以此為目標。

過去，教學評量，尤其是最具關鍵性的教室評量，由於受到主客觀因素的影響，例如行為導向、標準本位、績效責任、時間考量、班級規模、教師專業知能等，常是以教師的觀點來設計評量，且與學生的實際生活脫節，缺乏真實性，因而備受質疑。Gardner(1999)就指出，「所有的評量都必須以行為作為評量的對象，我們無法評量一個人的內心呈像。」有鑑於此，學者乃紛紛提出改革之道，Hargreaves, Earl, Moore, & Manning (2001) 就從技術、文化、政治、以及後現代的角度，來探討教室評量的革新，指出評量的落實必須回歸到學生的層面，摒除制度或機構層面的束敷，並且重新定義或重新詮釋「成就」、「標準」的本質，以學生為本，真實為度，才能發揮評量的功能。

四、有效性

教學評量的有效性原則乃是強調，評量必須確實能有效地測量出學生的實際表現，這可以從評量的客觀性 (objectivity)、效度 (validity)、信度 (reliability)、以及標準化 (standardization) 等特性來加以檢視 (Mayer, 1987；單文經，2001)。所謂的客觀性是指評量的結果不會受到評量對象的性別、種族、群體、階級等個人、社會、或文化背景所影響，而仍能獲得可靠的結果。所謂效度是指評量的項目能確實反映

評量的目的與內容，其內涵包括了表面效度 (face validity)、預測效度 (predictive validity)、以及建構效度 (construct validity)。所謂的信度是指評量的結果可以讓人信任，所以好的評量方式，不論評量次數的多寡，其結果應該都是前後一致的。至於所謂的標準化或可參照性 (referencing) 則是指測驗的分數能被轉換成一個數字，可用來告訴學生他在團體中所處的地位。另外一種提供參照（或標準化）的方法，是透過標準參照。這些標準可以幫助學生和教師把原始分數，轉換而為學習的有用量數。準此以觀，有效的教學評量可以經得起客觀性、效度、信度、以及標準化的考驗

五、多樣性

教學評量的多樣性原則是指評量的方式採取多元的策略。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，乃出現所謂的「另類評量」（變通性評量）(alternative assessment)、「動態評量」(dynamic assessment)、「實作評量」(performance assessment) 等新名詞（簡茂發，1999：13），其多樣性的評量趨勢，殊為明顯。

誠心而論，評量應該採行多元的方式來實施，這主要是因為學習者的真實能力誠難以單一的評量方法或工具，就可精確測得。基於此，Foriska (1998) 曾提出多種評量的方法，來評量學習者的真實能力，其所提出的評量方法包括：教師觀察、學生日記、同學相互觀察、學生反省札記、表演活動、學生設計、觀察檢核表、同學互評、訪問、角色扮演、學生自我評量、小組完成任務、書面報告、解決問題任務、家長報告、書寫式情境測驗等各種評量方法，其目的旨在強調，能夠綜合而完整地評量學生的學習成就。至於 Gardner (1999) 所指出的八種智慧觀：語文智慧、邏輯-數學智慧、視覺-空間智慧、肢體-動覺智慧、音樂-節奏智慧、人際智慧、內省智慧、以及自然觀察者智慧，則更是推動多樣性評量的有力基礎。值得關注的是，講求多元雖然是評量的重要原則，但學習評量絕不能是為了多元而評量，必須實際考察評量的方法是否真能瞭解學習者的學習情形，也就是評量的方法必須與評量的目標密切結合，否則易淪為「打稻草人」的窘境。

肆、結語

歸結上述的討論，可以瞭解，教學評量的實施應該採取綜合性、全程性、真實性、有效性、以及多樣性的原則，使教學評量的活動能夠更趨於周延，更接近真實的生活世界，與學生的內心世界接軌，同時力求教學與評量的融合，相輔相成，才能進一步提昇教學的品質。

參考文獻

行政院教育改革審議委員會（民85）。教育改革總諮議報告書。臺北市：作者。

- 李咏吟、單文經（民 87）。**教學原理**（三版）。臺北市：遠流。
- 何英奇（民 81）。教學評量的基本原則。載於國立台灣師範大學學術研究委員會主編，**教學評量研究**，3~30。臺北市：五南。
- 郭生玉（民 80）。**心理與教育測驗**（六版）。臺北市：精華。
- 張春興（民 83）。**教育心理學**。臺北市：東華。
- 單文經（民 90）。**教學引論**。臺北市：學富。
- 簡茂發（民 85）。評量。載於黃政傑主編，**教學評量**，1~40。臺北市：師大書苑。
- 簡茂發（民 88）。多元評量之理念與方法，**教師天地**，99，11~17。
- Black, P.(1998). *Testing: Friend or foe? The theory and practice of assessment and testing*. Bristol, PA: Falmer Press.
- Calfee, R., & Hiebert, E.(1991). Classroom Assessment of Reading. In R. Barr, M.L. Kami, P. Mosenthal, & P. D. Pearson(Eds.), *The handbook of reading research*(Vol.2, pp.281~309). New York, NY: Longman.
- Foriska, T. J.(1998). *Restructuring around standards: A practioner's guide to design and implementation*. Thousand Oaks: Corwin Press.
- Gagne, R. M.(1985). *The conditions of learning and theory of instruction*. New York: Holt, Rinehart and Winston.
- Gardner, H.(1999). *The disciplined mind*. New York: Palmer & Dodge Agency.
- Hargreaves, A., Earl, L., Moore, S. & Manning, S.(2001). *Learning to change*. San Francisco: Jossey-Bass.
- Kemp, J. E., Morrison, G. R., & Ross, S. M.(1994). *Design effective instruction*. New York: Macmillan College Publishing Company.
- Mayer, R.E.(1987). *Educational psychology: A cognitive approach*. HarperCollins Publisher.
- Stiggins, R. J.(1996). *Student centered classroom assessment*. Englewood Cliffs, NJ: Prentice Hall.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching , learning, and assessment in the classroom*. Bristol, PA: Falmer Press.
- Wiggins, G., & McTighe, J.(1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum.