

多元評量：理念及其應用

彰化師範大學及朝陽科技大學兼任教授 張清濱

壹、前言

教學是一個複雜的歷程，它涉及教師、學生、目標、教材、教法、及評量。教學是師生共同的課題，沒有教學，學校教育就會淪為空談。然而，教學效果必須透過評量才能顯現出來。自從20世紀以來，評量方法從口試到筆試，從質化到量化，從簡答題到多重選擇題，其目的莫不在增加效率並使評量系統更客觀、更公平、更可行。時至今日，教學評量仍是大眾注目的焦點。教育部宣布自2014年實施12年國民基本教育，活化教學列為重要目標之一，多元評量更有其利用之價值。本文擬就多元評量產生的背景及其應用，敘述於後，以供參考。

貳、多元評量產生的背景

近數十年來心理學家及教育家們的學說及理念說明多元評量的重要性，也奠定多元評量的理論基礎，多元評量乃應運而生。

一、多元智慧理論

Gardner (1983) 指出人類的智慧至少有七種：邏輯－(logical-mathematical)、語文的(linguistic)、音樂的(musical)、空間的(spatial)、肢體－動覺的(bodily-kinesthetic)、知己的(intrapersonal)、及知人的(interpersonal)智慧。後來，他又提出第八種智慧－自然觀察的智慧(the naturalist) (Gardner, 1995)。

Gardner的多元智慧理論給學校教育人員一個重要啟示：3R's教育或核心課程雖有一席之地，藝能科目諸如音樂、美術、體育、及社團活動諸如交友及人際關係仍不可偏廢。人生中有許多的際遇與機會。擅長跳舞、歌唱、打球、烹調、機械操作者可以登

上世界的舞臺，揚眉吐氣。對於中途輟學或不擅於傳統紙筆測驗的學生，學校如果能改用另類評量，給予表現的機會，照樣會有揮灑的空間。

二、學習型態理論

學習型態(learning styles)係指學生學習偏好的方式(Kellough & Kellough, 2003: 29)。Riessman (1966) 從感官的觀點，發現學生的學習型態可分為三種類型：(一)視覺型(reading)：此類學生視覺反應敏銳，善於閱讀，一目十行，過目不忘；(二)聽覺型(listening)：此類學生聽覺反應靈敏，長於傾聽，輕聲細語，風吹草動，都可聽得一清二楚；(三)動作型(doing)：此類學生手腳靈活，敏於操作，喜歡打球、運動、吹奏樂器、做實驗、及操作機械等。

學生偏愛的學習型態取決於遺傳與環境因素。有些學生在正式場合學習成效最好，另有些學生喜歡在非正式、輕鬆的場合學習。教師教學應該採取多面向的角度看待學習型態。因此，教師應設計各種不同的情境，評量學生的能力與表現。

三、社會互動理論

依據建構論(constructivism)的觀點，瞭解(understanding)存在於環境的互動中(Savery & Duffy, 1995)。換言之，學習是在環境的互動中產生的。環境包括學生周遭的人、時、地、事、物。在學校的環境裡，影響最大的因素包括：(一)教師與學生之間的互動，(二)學生與學生之間的互動，(三)校園景觀與學生之間的互動，(四)學校設施與學生之間的互動，(五)學校組織氣氛與學生的互動，(六)生活經驗

焦點話題 >>>

與學生之間的互動（張清濱，2008：233-234）。

教師教學評量時往往著重課堂內的學習，而忽略課堂外，甚至校園外的學習活動。因此，教學評量應與生活情境相結合。

四、行為目標分類

行為目標 (behavioral objectives) 是一種操作型的目標陳述。它是以學生的行為，準確地描述教學後期望學生獲得可觀察的學習結果及標準（張清濱2009：150）。這些目標可依學習領域及層次分為：認知領域 (cognitive domain)、情意領域 (affective domain)、及技能領域 (psycho-motor domain)。行為目標的分類成為建立測驗題庫及測量各種教育目標的工具。

Bloom (1956) 等人於1956年出版教育目標的分類第一冊。他們把認知領域的目標分為六類及層次：知識、理解、應用、分析、綜合、評鑑。事隔45年，Anderson與Krathwohl 等人 (2001) 完成修訂目標的分類，把學習目標修訂為記憶、瞭解、應用、分析、評鑑、創造，稱之為修訂版的教育目標分類法。

認知領域的目標分類完成後，Krathwohl等人 (1964) 把情意領域的目標分為接受或注意、反應、價值的評定、價值的組織、品格的形成。Simpson (1972) 也把技能領域分為七個類別及層次：知覺、準備狀況、在指導之下練習反應、機械或重複練習、複雜的明顯反應、調適、創作。

行為目標的分類給教師們一個重要的啟示：學習是多面向的。雖然各學科的學習重點不盡相同，但都包含認知領域、情意領域、及技能領域。這三大領域又細分為許多層次。教學評量就應該包含主要的學習領域及層次。

五、五育均衡發展

國民教育以養成德、智、體、群、美五育均衡發展之健全國民為宗旨。九年一貫課程綱要揭示10項課程目標：(一)增進自我瞭解，發展個人潛能；(二)培養欣賞、表現、審美及創作能力；(三)提升生涯規劃與終身學習能力；(四)培養表達、溝通和分享的知能；(五)發展尊重他人、關懷社會、增進團隊合作；(六)促進文化學習與國際瞭解；(七)增進規劃、組織與實踐的知能；(八)運用科技與資訊的能力；(九)激發主動探索和研究的的精神；(十)培養獨立思考與解決問題的能力（教育部，2012）。

這些目標都需要透過各領域課程的實施，始能達成。譬如，要達成「發展尊重他人、關懷社會、增進團隊合作」的目標，教師宜多採用合作學習 (cooperative learning) 及小組教學 (small group instruction) 的方式，教學評量也要兼顧小組成員的自我表現及團體表現。

六、資訊科技發展

科學技術的發展，突飛猛進，一日千里。教育工學 (educational technology) 的發展改變教學的型態。課程與教學走向多元化、資訊化、網路化、數位化、及國際化。各種教學媒體諸如網際網路、社群網站、教育雲端線上作業、電腦化教學 (computer-based instruction)、電腦輔助教學 (computer-assisted instruction) 及多媒體在教學上的應用，如雨後春筍，相繼問世。

隨著資訊科技的發展，教學評量採用電腦化測驗 (computer-based assessments, computer-based testing)，更加普遍 (Oliva & Gordon, 2013: 366)。教學要以學生為中心，教學評量就應該以學生為本位，把每一個學生帶上來。電腦化適性測驗 (Computerized Adaptive Testing, CAT) 正可滿足這方面的需求。

電腦化適性測驗不同於電腦化測驗。電腦化測驗只是使用電腦作答，把電腦當作測驗的工具。電腦化適性測驗係依據考生的能力水準，循序作答的一套測驗。測驗時，考生坐在電腦機前依電腦軟體顯示出來的題目，依序作答。通常第一道試題難易適中，如果考生答對，則第二道試題難度升高。第一道試題如果答錯，則第二道試題難度降低。依此類推，直至電腦能判斷考生的能力為止，測驗即告結束（Straetmans & Eggen, 1998；張清濱, 2009：395-396）。在21世紀中，教育機構勢必投入更多的人力、物力，發展並設計各類科電腦化測驗。將來各類型考試也許將漸漸採用電腦化適性測驗。

參、多元評量的應用

以往的教學評量大都注重科學化的測驗與測量，強調客觀、具體、與效率。21世紀的教學評量有逐漸採用另類評量（alternative assessments）的趨勢。本文介紹檔案評量（portfolio assessment）、實作評量（performance assessment）、真實評量（authentic assessment）、及契約評分（grade contracts）如下：

一、檔案評量

檔案評量是一種多向度的評量包括認知、技能、情意領域，也兼顧歷程與結果，具有多元化、個別化、適性化、生活化、及彈性化等特性。檔案紀錄沒有固定、標準的形式，通常包括封面設計、目錄及頁碼、自我介紹、筆記、講義、學習單、課外研讀、日記、心得及感想、特殊優良紀錄、得意作品、研究報告及其他等。它可依自己的興趣，發展自己的潛能。譬如擅長電腦者可建立電子檔案紀錄；長於繪畫者，可以插畫、漫畫方式呈現；工於詩句者可穿插打油詩、現代詩、五言詩等；喜歡攝影者亦可穿插活

動照片、海報等。

檔案紀錄是一本個人學習的實錄，也是學習的寫真集，可以看出學生努力及成長的情形。傳統測驗較難測量的創造力、想像力及好奇心，檔案評量則可充分發揮出來。當然，檔案評量也有其缺點諸如教師評閱費時，有時不同的評量者評量出來的成績差距甚大。此外教師無法展示及囤放龐大的檔案紀錄（Marsh & Willis, 2003：294；張清濱, 2007：391-393）。

二、實作評量

實作評量著重學習的歷程而非學習的結果，旨在觀察學生表現課業的能力，尤其能否使用良好的判斷去選擇學習的內容（Marsh & Willis, 2003；Wiggins, 1998）。實作評量包含一系列的歷程，必須符合下列四個要領（Airasian, 1994）：

- （一）應有明確的目標：實作評量首應確定評量的目標是什麼，通常以行為目標敘寫，並且要包含主要的教學目標。
- （二）辨認可觀察的實作行為：目標確定後，教師應考慮以何種行為最能展示學習的歷程及結果，這些行為必須是客觀的、可觀察的、及可測量的。
- （三）能提供合適的場地：可觀察的行為必須在合適的場地進行，也許是教室、禮堂、實驗室或工場。教師應設計合適的場所，評量學生展示的行為。
- （四）備有預擬的評量或計分的標準：以演說的實作評量為例，演說的行為可細分為五個部分：1. 眉目傳神，2. 口齒清晰，3. 抑揚頓挫，4. 呈現論點，5. 總結論點。這些行為表現成為評量或計分的標準，評判可用來觀察並判斷演說的表現。

三、真實評量

真實評量係指學生的學習活動儘可能接近現實的世界，著重學生接受測驗時的反應

焦點話題 >>>

情境，亦即在現實的生活情境中產生。真實評量是標準參照評量，不是常模參照評量。它可以用來辨認學生的優點與缺點，但不排列名次相互比較。

真實評量類似實作評量，著重在學生的表現（performance）；學生要在現實的情境裡以合適的方式，展示自己的知識、技巧、或能力。譬如學生在模擬的超級市場學習英語會話，展現英語會話的技巧。依據 Wiggins（1998）的研究，真實評量具有六個特性：

- （一）評量切合實際：它反映出「真實世界」使用的資訊或技巧的方式。
- （二）評量需要判斷：它要解決非結構性的問題，答案可能不只一個。
- （三）評量要學生去「做」主題：即透過學習的學科程序。
- （四）評量在類似的情境中完成：相關的技巧在類似的情境中展現。
- （五）評量需要展現複雜問題的處理技巧：包括涉及判斷的問題。
- （六）評量允許回饋、練習、及第二次解決問題的機會。

四、契約評分

契約評分是一種標準參照評量。通常，任課教師在學期開始前，即明確擬訂學生一些基本的學習目標，另加上一些高層次的目標。學生在開學時必須按照自己的能力，會同任課教師，就A、B、或C三種等級，任選其一，簽訂契約，作為該生努力的目標。值得注意的是：此法沒有D及F級，因為教師不鼓勵學生失敗。如果李生簽訂契約C，學期結束，只要他（她）達成最基本的目標，即可得到C等成績。如果王生簽訂契約B，則他（她）除了完成最基本的目標外，尚須完成一部分高層次的目標。又如張生簽訂契約A，則他（她）必須完成最基本的目標及高層次的目標。此法的優點是：只要努

力，均有機會得到A；學習動機是自我導向；教學目標明確；培養責任感；個別化學習；自我鞭策，不與同學競爭；驅除壓迫感及恐懼感，建立自信心（Partin，1979：133）。

此法頗適用於低成就的班級，尤其適用於高智商而低成就的學生，重新建立自信心。但是，一些自不量力的學生可能好高騖遠，簽訂A級。學期結束，卻無法完成所定的目標，造成挫折感（張清濱，2007：395）。

肆、結語

多元智慧、學習型態、社會互動、行為目標分類、五育均衡發展、及資訊科技發展等理念建構多元評量的理論基礎。多元評量可視為評量的目標、內容、方式、情境、次數、人員、答案都是多元的。教育部要求中小學教師活化教學，教師應該善用有效的評量工具，評量學生的學習成就。

參考文獻

- 教育部（2012）。國民中小學九年一貫課程綱要。台北：教育部。
- 張清濱（2007）。學校經營。台北：學富文化事業有限公司，385-398。
- 張清濱（2008）。學校教育改革：課程與教學。台北：五南圖書出版公司，246-257。
- 張清濱（2009）。教學原理與實務。台北：五南圖書出版公司，335-338。
- Airasian P. W. (1994). Classroom assessment, 2nd. New York: McGraw-Hill.
- Anderson, L. W., Krathwhol, D.R., Airasian P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Rath, J., & Wittrock, M. C. (2001). A taxonomy for learning, teaching, and assessing: A revision

- of Bloom' s taxonomy of educational objectives. New York: Longman.
- Bloom, B. S. et al. (1956). Taxonomy of educational objectives: Cognitive domain. New York: David.
- Gardner, H. (1983). Frames of mind: The theory of multiple intelligences. New York: Basic Books.
- Gardner, H. (1995). Reflections on multiple intelligences: Myths and messages. Phi Delta Kappan, 77(3).
- Kellough, R. D. & Kellough, N.G. (2003). Secondary school teaching: A guide to methods and resources, 2nd ed. Columbus, Ohio: Merrill Prentice Hall.
- Krathwohl, D. R. et al. (1964). Taxonomy of educational objectives, handbook II: Affective domain. New York: David McKay.
- Marsh, C. J. & Willis, G. (2003). Curriculum: Alternative approaches, ongoing issues, 3rd. New Jersey: Prentice Hall.
- Oliva, P. F. & Gordon, W. R. (2013). Developing the curriculum, 8th ed. New York: Pearson Education, Inc.
- Partin, R. L. (1979). Multiple option grade contracts. The Clearing House, November, 133-135.
- Riessman, F. (1966). Styles of learning. NEA Journal, 3, 15-17.
- Savery, J. R. & Duffy, T. M. (1995). Problem-based learning: An instructional model and its constructivist framework. Educational Technology, 35(5), 31-37.
- Simpson, J. S. (1972). The classification of educational objectives in the psychomotor domain: The Psychomotor Domain 3. Gryphon House, 43-56.
- Straetmans, G. J. M. & Eggen, T. J. H. M. (1998). Computerized adaptive testing: What it is and how it works. Educational Technology, 38(1), 45-52.
- Wiggins, G. (1998). Educative assessment: Designing assessments to inform and improve student performance. San Francisco: Jossey-Bass.