

教學評量的理論與實用

劉安彥

佛羅里達州傑克遜州立大學教授

摘要

教學評量 (classroom assessment) 是教師在教學過程中所作的各種不同的評量，其主要目的在了解和增進學生的學習，而不是相互比較或是評定優劣高低。教學評量也為教師提供有關其教學效率的重要訊息，以作為教師改進教學之依據。他偏重 assessment for learning 而不只是 assessment of learning。多元化教學評量深受新近學習理論與動機理念之影響。常用的變通式評量包括：真實的評量 (authentic assessment)、以表現為憑的評量 (performance-based assessment)、檔案評量 (portfolio assessment)。但這些評量的方法與工具及其信度和效度，仍有待建立和加強。教育工作者必須加強教育社會大眾有關教育和教學評量的潛力和限制，才能在教育改革中化阻力為助力。

關鍵詞：考試 (examination)、測量 (measurement)、測驗 (test)、評鑑 (evaluation)、評量 (assessment)、教學評量 (classroom assessment)

壹、前言

一提到評量，很多人都會很自然地就想到考試，其實「考試」並非等於「評量」，這兩者之間基本上是有相當大的差別的。不過，在學校裡從事教學的教師，一般都會花很多時間在「考試」和「評量」的工作上。以其全部專業活動而言，大約有將近三分之一的時間是用在考試和評量這方面的。具高度效率的優良教師，大都會運用各種不同的方式來評量學生的學習，以做為其從事教學專業工作的重要依據，一方面去瞭解所教學生的學習進展和成就，另一方面則檢討個人的教學效率，

隨時為調整和改進教與學提供必要的資訊。由此觀之，評量：不只是考試和紀錄學生已學到之知識和技能而已，更重要的是，它還會直接地或間接地影響到學生目前的以及日後的學習和動機，教學評量在整個教學專業工作中所佔地位的重要性，是無可置疑的。

貳、考試、測量、測驗、評鑑與評量

考試 (examination) 是一種相當複雜的活動，通常是以試題或是其他的工具來測量 (measure) 應考者，以獲得如分數、等地和級別等某些結果。以入學考試為例，由於組別和科系的不同要求，應試者可能不只要參與學科筆試，而且還要加考實作和表演等藝術科考試。考試的結果一般是量化的分數。考試所需的試卷，通常稱之為量表 (scale) 或是測驗 (measurement)，用以測量智力的史比量表即為一例。心理學家運用多種量表來測量人格和態度，這些工具有時又稱之為 (inventory)。

各位可能研讀過名之為「心理與測驗教育」一類的專門書籍，這裡所謂的測驗具有更廣泛的內涵，它是一門專科，探討全盤性的有關於考試和測量的理論基礎、測量工具的編製和適當使用以及測量結果的解析和應用等重要事項。這在心理學的領域課程被稱之為 Psychometrics，如果是較偏重實用，有時又以 Psychological testing 名之。而在教育的領域裡，一般大多以 Measurement and evaluation 名之。也許各位也看過「心理與教育測驗彙編」一類的書，這裡「測驗」兩字指的是用來測量某種事物的工具-試卷、量表、或是如智力「測驗」一類的複雜多方面的測量設計。

從測量到評鑑 (evaluation)，可以說是從客觀到主觀。測量力求準確客觀，力求事項真象的準確認定。再以智力測量為例，我們應用如史比量表一類的智力測驗，針對某一個人，就其智力的高低進行測量，而得到所謂「智商」的測量結果。進一步就所得測量結果進行解析和鑑別，則有賴「評鑑」的運用。智商是智力的一個指標，至於此一指標稱之為高、低、優、劣，則必需依照某些主觀預定的條件來加以評鑑。又以教師給學生考試、打分數，然後再根據某些既定的標準來進行評鑑的工作，以為決定等第、級別的劃分（如甲、乙、丙、丁等）。

評量 (assessment) 則是目前最受大家中意的「新寵」。其實，評量同評鑑一樣，都深具主觀性，以測量結果為根本依據，再依既定的某些條件和準則來進行評量的活動。專案評鑑、教學評鑑和教授評鑑等都是大家時有所聞的。評鑑的總結，往往牽涉到優、劣、良、莠的區別以及適當與否的認定。各界所時尚的「教育評量」(educational assessment)，又具另一層意義。基本上，這裡所謂的評量乃是傳統式測量的延伸和多元化（多元化是目前相當受寵愛的辭語，如多元文化、多元入學、以及多元化評量）。所謂「引伸」指的是除了客觀的測量外，再加入評鑑的工夫。多元化則是運用各種不同的方法和方式來進行（不限於紙筆式的測驗），這種演變，有其哲學和理論上的依據，這稍後我們將有進一步的分析和闡釋。

參、教育與教學評量

教育評量乃是一種大規模的評量。例如：國小和國中學生的學力測驗即是，其

主要目的在檢測全盤性的教育成果。最近教育部公布一項計畫，將在數年後針對小三、四、六和國三學生進行全面性的基本學力檢測，以評量實施九年一貫課程與教學的教育成果，作為日後修訂改進此一重大教育措施的參考和依據。目前美國為了教育績效責任 (accountability) 的認定，也要求以州為單位，進行全面性的教育評量。教育評量一般是大規模的（包括整個學區或全州，為跨校區的），這種評量也是一種總結性 (summative) 的評量。

教育評量的結果往往牽涉到辦學和教學優劣以及適當與否的評鑑和認定，直接地影響到績效責任的追究以及賞罰的處置。這就是目前美國教育圈所常詬病的「高賭注」(high-stake assessment) 評量。雖然這種評量與一試定終身的入學聯考不可同日而語，但其對工作人員（包括校長和教師在內）和受教學生所造成的壓力，還是十分地令人頭疼，也多為各界有識之士所詬病、批評。

這種教育評量的弊端包括考試領導教學，致使一般教學無法正常化，又因此等評量所用工具又多為選擇題、紙筆式測驗，又偏重學科基本的、深具記憶性的知識，使得教師在教學上，未能強調思考和判斷、評鑑能力的訓練，加以不適當和有欠公平、合理的比較，顯著地增加了許多流弊和爭議。

簡單地講，教學評量 (classroom assessment) 乃是整體性、全面性、大規模的「教育評量」的一部分，這種說法主要是以評量的方式和範圍來加以界定。教學評量乃是教師在教學過程中所作的各種不同的評量，其主要目的在瞭解和增進學生的學習，而不是相互比較或是評定優劣高低。教學評量也為教師提供有關其教學效率的重要訊息，以做為教師改進教學的依據。由於這種評量通常是由教師自行在教室內以自選或是自編的測量工具來進行多方面的評量，所以有 classroom assessment 之稱。它偏重 assessment for learning 而不只是 assessment of learning。不過，這種評量在目前各方強調教育改革的聲浪中，並沒有受到該有的重視和強調。此一現象的繼續存在對教改的推行與效果深具負面性影響，為一嚴重偏差。

課程內容乃是教師教學的一大主軸。課程綱要的擬定主要是由專家、學者來主持，但更理想的安排，則應包括學校中教育人員（教師）的共同參與。課程綱要的擬定和頒布，只為教師提供其實際教學內涵的概括性資訊。一般教師都根據課程綱要，選用由學者參與所編輯的教科書，再配以個人自編的教材和教法，落實課程綱要的推行和成就，這也是一般人所認定的教學內涵。不過，課程設計與教材、教法之是否適當有效，則有賴各種評量來確認。由此觀之，教學評量同時具有三項重要功能：學生學習成就的評量、課程內容、設計與教材之是否適當的評量；以及教師個人教學效率本身的評量。基於此一認識，教學評量不能只是一種等到學習告一段落（如月考、段考或期末考）之後才進行的一種總結性評量。既然教與學兩者都是持續性的過程，妥善地配合教與學的持續性評量一形成式評量 (formative assessment)，更為重要。因此，教師在進行、從事教學評量時，應時時掛念如何去運用評量來提升學生的學習興趣與動機，以及如何不斷地增強學生的學習能力，以提高學生的學習成就。正面肯定又圓滿地解答這兩項重要考量，即是教師教學的優良表現。如何落實教學與評量統整的重要理念，有待我們更多的努力與參與。

評量的目的決定評量的設計、實施與運用。而評量運用之場合為評量之設計預立限制。評量既以其目的為主要取向，欲以某一特定評量來達成多種不同目標是有欠妥當的，它容易增加混淆和困擾，或為所謂多元化的評量提供有力的依據。

評量的設計包括三個基本層面，首先必須確定的是認知性的代表 (cognitive representation) 或是所要加以評量之事物本身之確認，其次則運用多元化的方法來進行觀察和測量的工作，之後再分析、解釋和評量所得結果，以求達到某一結論。試以評量學生的英語字彙為例，課程標準預定學生在某一年級（階段）即應習得之字彙數（或為兩百或三百不等），除非我們能夠考每個學生三百個字彙，我們就必須運用不同的方法來取得這三百個字的一個「代表」，（如包含四十個字或是五十個字的一份考題），再用這份試題來做評量，然後以某些預定的標準來做分析、解釋而獲得結論。

評量是以數據來進行推論，因為評量本身並非十分的或是絕對的準確，因此我們所能做到的，只是對學生的學力、成就以及所習得之相關技能，進行某種程度的評估而已，這種評量上的欠缺絕對準確性，在評量結果的運用和解釋時，必須加以注意。

課程、教學與評量乃是教師日常專業運作中的三大主軸。深入瞭解課程內容、有效地教導、輔助學生學習，以及適當有效地評量教與學的成果，乃是一位良好教師所不可或缺的要件。準確又有效地評量學生學習的進展與成果，不只是良好教學的首要條件，也是改進教育的一個很重要的環節，因為評量為教師、學生、以及其他人士（家長、決策人士以及社會大眾等）提供有關教育活動與成就的重要訊息，這是他們做決定和採取某些有關教育運作之行動的主要依據。

肆、教學與評量的整合

一九九〇年初期，美國的兩個教師組織（美國教師聯盟，AFT；和全國教育協會，NEA）會同全國教育測量諮議會（National Council on Measurement in Education），擬定了「教師教育評量能力準則」(The Standards for Teacher Competence in Educational Assessment)。該一準則指出：教學與評量的整合可依教學時程的三個階段來進行。而在不同的階段，教師評量學生的責任和內涵也隨之有所變化。

1. 教學前的評量

運用觀察、晤談和報告等方法來收集相關資料，並對此等資料進行解析，以期瞭解學生的文化背景、興趣、技能以及有關學習其他相關科目的能力。了解學生學習特定課程的動機和興趣。明定所期盼之學生學習成就。妥善設計針對個別學生或是不同組別之學生的教學方案與計畫。

2. 教學中的評量活動

這也就是一般所稱的形成式評量。這種評量的主要目的乃是監控學生邁進學習目的的程度，測定學生的學習成就以及學習困難。調整個人的教學活動，提供特定又適合的回饋給學生。提升學習的動機，制定學生達到其學習目標的程度。

3. 教學後的評量

描述個別學生達到短期與長期學習目標的程度。知會家長或是監護人有關其子弟的優點及需要改變、改善之處。紀錄和報告教學評量結果，以便進行全校性的分析，做為評鑑和擬定決策的依據。教師並以此就整體教學效率進行評量。另外也對課程設計與教材運用進行評量與檢驗。

隨時留意學生的學習與動機乃是教師專業的一大要件和挑戰。基本上大多數的學生是主動、好學的，學生透過發掘和建構為學習注入意義，也大多能設立學習目標的、計畫、執行來達到目的。透過自省、思考加以自控、自律及熱心，以內發動機為動力，積極參與學習活動，力求知識與技能的高度成就。許多優勢的學生本質則有較良好的教學評量來加以促進和增強。深入瞭解每位學生的學習動機乃是優良教學的基本條件，在考量學習動機與評量時，評量的設計與實施應以如何增進學生的內在學習動機以及主動又有意義地去積極參與學習為主要指導原則。公平又具挑戰性的評量有助學生的學習熱誠，太過艱難的評量則會增加學生的挫折感，也會不利地影響到學生的自尊心和自信，降低學生的自我效力 (self-efficacy)。過於簡單容易的評量則會使學生感到無聊，又因不必努力就可以得到好成績，而降低了學生的內在學習動機。從教學和學習活動到學習成果和表現的評量，回饋訊息的提供，以及由此引發的學生的知覺與感受，乃是一個循環式的過程。教學評量在這個歷程中佔有相當重要的地位，除了考核學習進展和成果外，又與學生的學習動機密切相關連。妥善適當的教學評量，可以增進學生的自我制約 (self-regulation)，自我監控 (self-monitoring) 以及自我評鑑 (self-evaluation)，還會提升學生的自我效力 (self-efficacy)。

伍、學習與評量理論的比較

評量既然是為了教學所需，其主要目的是在幫助學生學習，那麼教學評量的設計與實施，就必須考慮相關的學習理論，而且不只是一般性的理論，特定學科領域（如：自然科、語文科和社會科等）的學習理論也需要深入加以認識瞭解。這裡我就一般性的學習理論及其相對應的評量理論，向各位作一簡單的介紹。

根據 L .Shepard (2000) 的報告，傳統的與新近的學習與評量理論頗有差別，其對比如表一：

表一 傳統與新近的學習與評量理論

	傳統的	新近的
學習理論	個人的智力及學習能力是天生註定，固定的；聯結派 (associationist) 與行為學派 (behaviorist) 的學習理論。	智力是經由社會與文化歷程來形成認知 (cognition) 與建構式 (constructive) 的學習理論。
	1. 學習是刺激－反應連結的建立，知識為各種基本聯結之累積。學習是由簡而繁，架構分明的。	1. 學習在特定的社會、文化環境中進行，學生在社會場合中建構知識與理解新的學習受到既存知識與文化理念的影響。

	2. 學習的遷移有限，教學針對特定學習目的，明確地進行考試—教學—考試的方式來確保學習成就，考試與學習如影隨形。學習動機主要是依賴許多正增強(positive reinforcement)。	2. 學習與思考的自我追蹤，進級認知 (metacognition)，原則性的理解與學習，便於知識的轉移學業表現與能力依個人特質和自我認定而有別。
	社會效用 (social efficiency) 的課程	革新的課程
課程考量與設計	科學化的方法來經營管理學校，根據工作職務所需擬定詳細特定的教學目標。實用的課程內涵，理論性的內容只為少數優秀學生而設。詳細明確的課程標準。課程內容與實施依預期之社會職務角色而有別。	每個學生都能學習。學習學科內容強調思考與解決問題之能力。個別有差異，故其學習機會要均等，促進學生在不同學科討論和實作的社會化。學校內、外學習的真實配合；重要思考的態度的養成。學習場合應是一個相互照顧、民主化運作的社團。
教學評量	<ol style="list-style-type: none"> 1. 科學化、準確客觀的測量，用智力測驗鑑定個人能力而加以分類。 2. 以客觀的紙筆測驗來評量學業成就。 3. 偏重總結式的評量，考試頻繁、競爭激烈。 4. 試題嚴保機密，徒增學生之焦慮與壓力。 5. 分數嚴格計算，主要是在考學生。 	<ol style="list-style-type: none"> 1. 用具挑戰性的作業式差事來引誘思考，兼顧學習歷程與學習目標和個人訴求，配合教學持續進行，非一試決定。 2. 善用形成式評量 (formative assessment) 以增進學習效果。 3. 學生充分了解學習目標與預期成果，學生積極地從事自我評量，不只評量學生，也評量教師教學。 4. 兼顧學生個人之進度與目標，評量內容與標準公開又明確。

資料來源：Shepard, L.A. (2000). The role of assessment in a learning culture Educational Researcher 129 (7), 4-14。

陸、變通式的評量

民國八十七年，教育部公布了「國民中學學生成績考查辦法」（教育部，民 87）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」。

1. 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。
2. 口試：就學生之口頭問答結果考查之。
3. 表演：就學生之表演活動考查之。

4. 實作：就學生之實際操作及解決問題等行為表現考查之。
5. 作業：就學生各種習作考查之。
6. 設計製作：就學生之創造過程及實際表現考查之。
7. 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。
8. 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。
9. 鑑賞：就學生由資料或活動中之鑑賞領悟情形考查之。
10. 晤談：就學生與教師晤談過程，瞭解學生反應情形考查之。
11. 自我評量：學生對自己的學習情形、成果及行為表現，作自我評量與比較。
12. 同儕互評：學生之間就行為或作品相互評量之。
13. 校外學習：就學生之校外參觀、訪問等學習活動考查之。
14. 實踐：就學生之日常行為表現考查之。
15. 其他方法。

兩年之後所頒布的國民中小學九年一貫課程綱要「自然與生活科技」學習領域課程綱要（教育部，民 89），教育部特別提出：「教學評量不宜侷限於同一種方式，除由教師考評外，得輔以學生自評及互評來完成。其形式可運用如：觀察、口頭詢問、實驗報告、成果展示、專業報告、紙筆測驗、操作、設計實驗及學習歷程檔案等多種方式，以能夠藉此瞭解學生的學習情況來調適教學為目的，例如：教學目標若為培養學生的解決問題能力，則可採用成果展示及作業報告的評量方式，並不是只以紙筆測驗的方式評量，再者，評量的層面應包含認知、技能與情意」

教學評量多元化的訴求，我們很容易地可由上面引述的許多教育部頒定的原則和方法看出來。這兒我們所謂的變通式評量，是用來與傳統式的紙筆測驗和口試作對比的。時尚的多元化教學評量深受新近學習理論與動機理念之影響，這在前面已多做闡釋。底下所要介紹的是幾個主要的變通式評量。

（一）真實的（道地的）評量 (authentic assessment)

由於傳統的紙筆考試與實用世界缺少關聯，因而有真實評量的需求，力求評量反應真實的背景及其要求。所謂真實的評量於焉被提出並受到重視。這種評量強調在於可能類似實際情況與背景下，進行學習獲得知識與技能之評量。有些人把真實的評量與實作和表現的 performance-based assessment 相互交換使用。反對者則認為所謂變通式的評量並不比傳統式的評量來得高明，而且真實的評量之效度 (validity) 很難確定，加上對變通式評量能否妥當地去測量學生的知識與技能之置疑，有些反對人士對變通式的評量多存置疑和反對，是故 (authentic assessment) 也似乎隨之銷聲匿跡，不再多受留意。

（二）以表現為憑的評量 (performance-based assessment)

傳統的紙筆測驗偏重知識本身的測量，以表現為憑的評量重運用知識來實際操作，所以有時候又被稱之為實作評量。實作評量的許多特性包括有：學生構造答案，而不是選擇答案。直接觀察學生的操作行為；評量著重學生學習與思考方式之檢測，大都於實在情境進行。實作評量所用之問題幾無正確而客觀的答案，正如一

般較複雜的問題，往往沒有一個可用來解決問題的單一答案。這並不是說實作評量的題目沒有答案，不過，許多可能的答案可因背景環境的左右而頗有差異。實作評量是費時較多（準備和運用），而又大多缺乏客觀的信度和效度。自我評量常在課堂上論述個人的好惡及為人。除了自我評量之外，實作評量同時可以評鑑個人及其所屬之團體。例如：由一小組學生共同合作來解決一項難題，或是一起來作民意調查等類似的方案，評量可針對每位學生（小組成員）的個別貢獻，也可以由小組最後完成的結果來進行總結式的評量。

（三）檔案評量 (portfolio assessment)

Portfolio 有時又被翻譯成爲卷宗，故它又有卷宗評量的說法。檔案或是卷宗都是儲存許多資料的地方，這種評量方法就是隨時將學生學習過程中所習得之知識和技能等相關物件與成品，收集在一起，以之作爲教學評量的依據。基本上這些所收集的資料可分爲四項：成品 (artifacts)，複製品 (reproductions)，證據式證明書 (attestations) 和成果 (productions)。成果之有別於成品者，乃是成果爲學生特別爲其檔案所創作的一些文件，這些文件述說其學習目標，省思學習之歷程，以及標明檔案中的各項資料。成品則是在整個學習過程中，持續累積之片段資料。

善用檔案評量需注意下列事項：明定檔案目標，學生參與將目標之擬定，循計畫來執行，就檔案內容與學生共同檢核，預定日後評量時所用之準則、條件，打分數再加評鑑來完成。學生的學習檔案可以詳細保存其學習過程中的複雜事項和所學到的全部成果，也能夠由建檔來鼓勵學生學做決定和自我省思，不過，建檔相當費時、費事，以檔案來做評量也有同樣的困難存在。

變通式的評量目前受到較多的重視與贊同，不過這些評量方式的有效運用仍存在著不少的難題待解決。教師本身是否受過適當的訓練來推動這些較新的評量，這值得大家去瞭解和重視，教師們是否有足夠的時間來進行這許多費時又費事的評量工作，也是一大考驗。而且，這許多評量方法與方式，其工具性所反映之效度 (validity) 和信度 (reliability) 仍有待建立並加強。另外，這方面的相關研究目前並不多，這也有待學者專家們去努力。如果教學評量的結果將被應用到其他重大決策（如評量大型教育成果，或是學生入學申請之成績單），那麼更多、更重要的許多相關難題，將會是很令人傷腦筋的。

柒、結語

最後，我以請求大家重視建立良好公共關係 (public relations) 做總結。教育社會大眾（不只是學生家長）有關教育和教學評量的潛力和限制與限度是一個很重要的事項，教育工作者（不只是教師）必須加強這方面的努力，這對於一般教育措施與計畫的推行，以及教育改革的推動，都會產生具決定性的影響，唯有社會廣大群眾對這許多有關考試、評量、測量和測驗的理念和方法有較深入的瞭解後，我們才能夠去避免許多阻礙優良教育大業運作的困擾和抗爭。主動、積極地去化阻力爲助力，大家一起來爲幸福的未來盡心盡力。

（本文係劉安彥教授應財團法人李連教育基金會之邀返國所作演講內容）

參考文獻

- 教育部（民 87）。國民中學學生成績考查辦法。臺北市：教育部。
- 教育部（民 89）。國民中小學九年一貫課程綱要「自然與生活科技」學習領域課程綱要。臺北市：教育部。
- 簡茂發（民 91）。多元化評量之理念與方法。教育資料與研究。46，1，7。
- Madaus, G. F. and O'Dwyer, L. M. (1999). A Short history of Performance assessment. *Phi Delta Kappa*, 80(9), 688-695.
- National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. J. Pellegrino, N. Chudowsky, and R. Glaser, Eds. Center for Education, Division of Behavioral and Social Sciences and Education. Washington, D.C: National Academy Press.
- Santrock, R. (2002). *Educational Psychology*. New York, N.Y. :McGraw-Hill.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.

跨世紀的健全國民

跨世紀的九年一貫新課程應該培養具備人本情懷、統整能力、民主素養、鄉土與國際意識，以及能進行終身學習之健全國民。

～錄自教育部國民中小學

九年一貫課程綱要基本理念