

173-196

超越測量－評量典範轉移的探索與啟示^{*}

江文慈^{**}

評量正處於典範轉移之中，從心理計量模式走向更寬廣的教育評量模式，從測驗和考試文化轉向評量文化。評量典範的探討，有助於我們理解各種評量實踐的背後思維，讓我們更加審慎地選擇如何發現學生知道什麼以及可以做什麼的評量方法。本文的目的在探討評量典範的轉移，重新思索評量與學習的關係，以及對教育工作者的啟示。首先，說明評量典範轉移的脈絡背景，並依序探討「心理計量典範」、「脈絡化評量典範」和「個人化評量典範」。從評量典範的轉移中，可發現評量與學習的關係愈趨密切，且有更多的交融，進而延伸闡述三種評量與學習的關係取向，分別為「對學習的評量」、「促進學習的評量」，以及「評量即學習」等。過去，在教育現場中，多將焦點放在「對學習的評量」，亦即著重學習成果的總結性評量，而今則轉為關注「促進學習的評量」，目前更強調寓評量於學習之中，提出「評量即學習」的觀點。這三種評量取向在教學中各有其地位與功能，教師如何平衡地使用，便成為一個關鍵議題。最後，本文針對評量典範轉移所帶來的衝擊與挑戰，提出一些對教育工作者的省思與啟示。

關鍵詞：評量典範、典範轉移、評量與學習

^{*} 本文之初稿曾發表於【2006年第七屆海峽兩岸心理與教育測驗學術研討會】

^{**} 江文慈：世新大學社會心理系副教授
wendy@cc.shu.edu.tw

Beyond Measurement: Exploring Paradigm Shift in Assessment

Wen-Tzu Chiang*

Assessment is undergoing a paradigm shift, from psychometrics to a broader model of educational assessment, from a testing and examination culture to an assessment culture. This article explores three assessment paradigms: psychometric paradigm, contextual paradigm, and personalized paradigm. Understanding different assessment paradigms can help us carefully decide how we attempt to discover what students know and we can do. Furthermore, assessment and learning are intimately intertwined and often indistinguishable from one another. Three relations between assessment and learning are discussed: "assessment of learning", "assessment for learning", and "assessment as learning". These three approaches all contribute to student learning but in vastly different ways. All three assessment approaches have their place; to maintain the right balance is the key challenge. It is important to understand them all, recognize the inevitable contradictions among them, and know which one you are using and why. Finally, implication and reflection on assessment paradigms are discussed.

Keywords: *assessment paradigms, paradigm shift, assessment and learning*

* Wen-Tzu Chiang: Associate Professor, Department of Social Psychology, Shih Hsin University

超越測量－評量典範轉移的探索與啟示

江文慈

壹、前言

教育改革不斷地持續進行中，教學和課程典範的轉移有著許多的論述，那麼關於評量典範呢？Gipps 於 1994 年就指出了：「評量正處於典範轉移中，從心理計量走向教育評量模式，從測驗和考試文化轉向評量文化」，如今看來，這樣的轉變更為明朗了。

過去，教師使用自編成就測驗、課堂作業、觀察紀錄、口試和標準化測驗等方法來評量學生的學習成就。目前所採用的評量形式則更為寬廣了，不僅採用常模參照測驗而已，還有標準參照評量、形成性評量、實作評量和檔案評量等。除此，大家對評量目的的期許也更為擴大，期望評量能支持教學與學習、提供訊息給學校和師生、做為選擇和認證的機制、考核教學績效的策略，以及導引課程和教學等（Gipps, 1994）。這些評量的新形式及其目的，都明顯揭露出主導評量理論的傳統心理計量模式並不再適當，使得新的評量典範得以破繭而出。

處於評量典範的轉移氛圍中，近十年來評量改革在校園中陸續開展來，許多教師積極地實踐各種另類評量方法，在這過程中有些教師們也會思索：「實施這些評量方法，到底我的目的為何」？「花這麼多的時間與氣力在實施評量，究竟我要學生學些什麼？」，有些教師開始反思各種評量方法的目的與成效。的確，不同形式的評量，各自有其不同的哲學基礎；評量目的不同，評量方式也會有所差異。因此，我們有需要認識評量典範，以幫助我們更審慎地選擇如何發現學生知道什麼以及可以做什麼，也可以更清楚地了解這些評量實踐的背後思維。

另外，關於評量目的，以前多側重在如何正確測量學生的學習成就、分類與排名，而現在的評量則更關心如何運用評量來促進學習。許多學者也同聲提及，評量是用來支持學習的，而不是僅用來指出學生現在或過去的成就而已。評量應用來激勵學生的學習，透過評量，引導學生有更佳的表现，使其具備更多的能力（Glaser, 1990；Stiggins,

2005；Wiggins, 1998）。從文獻中我們看到了這樣的呼喚，在我國的評量政策也有重新的宣示，教育部於 2005 年修正公布的「國民小學及國民中學學生成績評量準則」的第二條中明確指出：「國民中小學學生成績評量旨在了解學生學習情形，激發學生多元潛能，促進學生適性發展，肯定個別學習成就，並作為教師教學改進及學生學習輔導之依據。」其中，強調評量的設計應以激發學生多元的潛能，促進學生適性發展，肯定個別學習成就，這隱喻著減少做個體間的評比，重視學生的自我比較。過去，教學評量發生在教學之後，針對一個課程、一個單元、一學期、或一學年，給予一個分數等第或對學習結果做簡要的評論。測驗和考試是課程中的例行事務，老師對學生學習成就的判斷較學習本質來得重要。現在評量改革所強調的是將評量統整在教學中，它比較不關心評比分類，而較著重透過對話來發展學生對學習的理解，激發學生的多元潛能。

評量由過去著重在學習成果的測量與考核，轉向促進與激勵學習。然而，這兩者同樣重要，教師在多重目的下，時有矛盾、時有衝突，因此我們有必要理解評量與學習的各種關係，進而思索如何去平衡各種目的，以達到更為周全的行動。評量總是深切地影響教育的內容和實施，評量的意識型態和政策作法將無可避免地衝擊到最基層的教師與學生，他們的感受最為直接、最為深刻。教育工作者、家長、政策製定者、社會大眾們，過去可能沒有機會去考量評量應如何做、該如何行，但在這個世紀裡，我們有機會去關心與發展對學生真正有幫助的評量，思考學生到底從教師、學校所提供的課程中學到什麼。我們有必要好好地仔細反省每位學生的真正成就為何，別再讓過去的舊思維，繼續影響學生的學習與成長。

綜上，本文主要透過文獻分析來闡述目前評量典範轉移的脈絡思維，並介紹三種評量典範，分別為「心理計量典範」、「脈絡化評量典範」和「個人化評量典範」。希冀透過對各種評量典範的認識，能更清楚地瞭解各種評量方法背後所蘊涵的意義，並重新思索評量與學習的各種關係，反思身為教師的我們如何平衡各種不同的評量目的。最後，討論處在評量典範轉移的氛圍中，教師與學校人員如何面對評量改革的浪潮，並針對評量典範轉移的挑戰，提出一些省思與啟示。

貳、評量典範轉移的歷史脈絡

評量典範的轉移其背後有著複雜的歷史脈絡，從鉅觀的層面來看，哲學思潮不斷再演變，哲學思潮對教學、學習本質的改變，進而影響評量本質的改變。知識論從結構主義、實證主義走向後結構主義、建構主義。在實證主義的知識論中，認為知識是客觀獨立存在的實體，因而評量強調以客觀和量化的方式來測量，重視評量結果的客觀性與權威性。然而，建構主義的觀點認為知識是個體建構而成的，評量目標則改變以往著重片段知識的記憶，轉而強調問題解決能力，以及師生及同儕參與的評量（吳毓瑩，1996；謝祥宏、段曉林，2001）。有關評量典範轉移的來龍去脈，Shepard（2000）曾從課程、學習理論和評量的發展歷史，來加以說明（如圖1）。

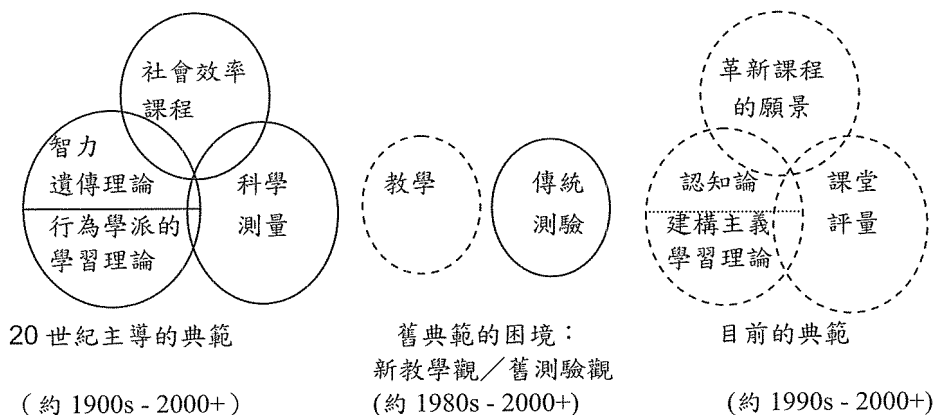


圖 1 新舊評量典範轉移的課程觀、學習理論與評量三者關係之演變

資料來源："The role of assessment in a learning culture," by L. Shepard, 2000, *Educational Researcher*, 29(7), 5.

從圖1可看出評量典範轉移的時間史，圖1左邊呈現二十世紀初的課程、教學與評量的觀點，用實線來表示。圖的中間，呈現了1980年代左右所浮現新的教學—社會建構教學（用虛線表示），它與傳統的評量（科學測量）無法契合，而有脫節的現象。1990年前後，在社會建構教學、新課程願景的氛圍下，課堂評量典範，因而醞釀成形。關於新的教學、課程與評量典範，則用虛線表示之，以和舊典範之間有所區別。

(Shepard, 2000)。以下分別依時間順序先後的歷史脈絡，說明演變內容。

20 世紀初，主導課程、學習和評量理論的主流典範分別是社會效率課程、行為主義的學習理論和心理計量的科學測量。當時，社會效率運動盛行，希望透過科學管理來解決工業化問題。於是，傳統課程注重社會效率，學校像工廠般的進行科學化管理，並藉由工作分析，詳細界定教育目標。課程內容著重實用性，並以精密的標準將課程內容和學生加以分類，透過精確標準的科學測量，提供社會適當的人才。在當時所盛行的學習理論是行為學派，主張學習是刺激－反應的聯結，知識和技能可切分成許多細小的步驟，透過嚴密的順序和階層來學習。要瞭解學生的學習情形，可透過「測驗－教學－測驗」的程序來確定，因此測驗與學習如影隨形。且當時的智力理論主張智力是天生不可變的，可藉由智力測驗的結果來區分學生的能力，並將之安置在合適的學習環境中。由於當時盛行心理計量的科學測量方式，在學校的教學中，就主張透過客觀測驗來評量學生的學習成就。

直到 1980 年代左右，在建構主義知識論的影響下，新的教學理論和課程理論於是興起，導致新教學與舊測驗系統無法配合，產生脫節現象（圖 1 的中間部份），於是新的評量典範從而蘊釀而生。

到了 1990 年代，主導課程、學習、評量理論的典範分別為革新的課程觀、建構主義及認知學習理論、課堂評量。革新課程主要在提供高層次思考和問題解決的內容，不只是記憶性的問題，並且加強社會化論述與學科的實務性，其目的在培養學生重要的心智能力，促使學生能關懷社會，實踐民主的行為。此時期的學習理論是採社會建構學習觀，學習是發生在社會脈絡中，而智力是在社會與文化的脈絡中發展而成的，是可以改變的。由於重視教育現場的學習，強調以學習者為中心，評量就轉向課堂脈絡的評量。課堂評量倡導以更多元的方式，來呈現學生的能力，提供具挑戰的作業以開展學生的高層次認知能力，並將評量整合在教學之中，透過評量的使用來支持學生的學習。從 20 世紀末到現今，新的評量典範逐漸成形，愈來愈明朗化了。

參、評量典範的探討與分析

上述介紹了評量典範轉移的歷史脈絡，而關於評量典範的探討，在相關文獻中，則有不同的論述觀點。有些學者從二元分立對照的觀點來論述，分別有：Lazear(1994)

的「舊評量典範」vs.「新評量典範」；Berlak (1992) 的「心理計量典範」(psychometric paradigm) vs.「脈絡化典範」(contextual paradigm)；與Shepard (2000) 的「科學測量典範」(scientific measurement) vs.「課堂評量典範」(classroom assessment)。

而指出新舊典範轉移過程中，歷經三個階段的則有：Gipps (1994) 提出從「心理計量」(psychometric)、「教育測量」(educational measurement)，到「教育評量」(educational assessment) 典範。Mabry (1999) 則歸納有「心理計量典範」(psychometric paradigm)、「脈絡化典範」(contextual paradigm) 和「個人化典範」(personalized paradigm) 等。另外，Serafini (2001) 則說明目前閱讀教學中的三種評量典範，他把課程視為是「事實」(fact)、「活動」(activity) 和「探究」(inquiry) 三者，相對應出評量是「測量」(measurement)、「程序」(procedure) 和「探究」(inquiry) 等三種評量典範。

有關評量典範的探討，雖然不同人所指稱的用語有所差異，但共通的是從科學測量的心理計量典範，走向關注教育脈絡的評量典範，近來則浮現出以學生為主體的個人化評量典範。以下引介「心理計量典範」、「脈絡化典範」和「個人化典範」等三種典範 (Mabry, 1999)，並分述其理念、做法、優點與限制。

一、心理計量典範

(一) 理念與做法

心理計量科學是從智力理論與智力測驗的編製發展而來，其中，心理計量有幾個重要的特色。一是「標準化」：為了使測得的結果能相互比較，必須讓所有受試者在相同情境下施測、計分及分數解釋，要求科學客觀。二是「測驗結果的普遍性」：亦即測驗分數對每一個人都有相同的意義。三是「單一向度」：也就是測驗在測量單一的屬性，透過因素分析，將與測驗總分相關低的題目刪除或修改，以編製一份能測量單一特質或屬性的測驗。四是「信度」：在不同情境或不同時間施測所得的結果應該是一致的。五、「常模參照的分數解釋」：個人表現需與所屬的參照團體做比較，採相對地位的解釋方式 (Gipps, 1994)。

在這個時期，學校的評量是以外量數（智力、性向、和成就測驗的結果），來做為不同教育決定的基礎。了解學生成就的基本策略是比較，將學生的表現與其他做測驗者的表現比較（常模參照解釋），或者與先前決定好的標準、決斷分數相比（標準參照解釋）。由測驗所測得的學生成就是靜態且成常態分配，可以互相比較，學生

的成就由等級來決定，而這些等級排名可做為篩選、分級、接受特殊課程或進入學校的依據。

（二）優點與限制

心理計量典範的評量方式主要是採用標準化選擇題型，在目前仍然是相當普遍，因為它具有客觀、效率的特色。所謂有效率是指實施容易、節省時間，而且有明確的數字提供清楚的評比訊息。可透過機器快速計分，減少人工計分的不一致和麻煩。特別是大規模的評量時，它更能省時省錢又省力，節省大量成本。此外，在短時間內可呈現很多的題目，不像論文題在相同的時間內只能呈現少數的題目。心理計量的評量方式之所以受到大家的喜愛，正由於它的好用，以及被認為是較有效度、信度與具公平性的（黃秀文，1996；楊銀興，2000；Gipps, 1994）。

心理計量取向的評量，雖有上述的優點，但是對於教學的負面影響也常見於文獻中，例如：對於學生獨特的成就，並無法提供良好的反映；對實際教學改進並沒有太多實質的幫助；還有並不符合新近的學習理論。因此，有些人就認為心理計量取向的標準化測驗並不適合做為教育評量（黃秀文，1996；Black, 1998；Gipps, 1994；Shepard, 1991）。另外，心理計量典範假設所測量的屬性是固定的、不可改變，這似乎是此典範的一大限制。因為教育評量是要協助個體學習、促進發展的，在於增能而不在限制。此外，常模參照的解釋並非自己的絕對表現，隨著團體的不同而有改變，對於看待學生的教育成就並不公平。而且，除了由常模、排序所造成競爭的困擾外，測驗的計分、題庫的安全性和錯誤報告也有過紀錄（Berliner & Biddle, 1995）。

再者，一些教育者和教育研究者也察覺到，採用標準化學生成就評量對教育現場，可能會有一些不當的後果（楊銀興，2000；Berliner & Biddle, 1995；Broadfoot, 1996；Shepard, 1991）：

1. 對課程的負面後果，例如：窄化課程為僅由選擇題可測得的科目、主題、和技巧；使課程表面化，將教學簡化為教導學生記憶大量的事實。

2. 對學生的負面後果，例如：學生因測驗成績的不佳而被標籤化，導致墮落沮喪；或者錯誤將學生安置在低學業情境，導致拒絕入學；學生被動學習，缺乏批判思考能力。

3. 對老師的負向後果，例如：考試引導教學，教師去專業化；學生成績低時，老師可能被指為教學不當而被解雇。

4. 對學校的負向後果，例如：公開低成績的困窘；學校喪失自主監督權；讓待在

低分學校的教師和家長感到挫折。

5.對教育政策的負向後果，例如：由於測驗的偏誤，導致錯誤鑑別學生，教育資源錯誤分配；成績報告錯誤，誤導學生成就和學生的品質，造成不適當的教育政策措

施。

心理計量取向的評量對教育現場可能造成一些負面的後果，但是，當時的教育界卻未能抗拒這種排名分類的做法。Wood（1990）就曾提出這樣的感嘆：「教育對於心理計量的襲擊和影響是相當脆弱的，缺乏有自信 and 不同觀點的人來思考評量的目的，還有理解各種測驗的適當與不適當。教育只是個雇客，看起來無助，僅能被動地接受心理計量的教義和實施」。他發出上述的感嘆，並呼籲要採取教育評量，因為心理評量無法仔細呈現學生真實成就的廣度、深度。原因是，知道學生的排名順序並無法完全知道學生知道了什麼的內涵，教育不是要將人排序分類，而是需要了解每位學習者的獨特性，幫助每個人有其獨特的成就。

二、脈絡化評量典範

（一）理念與做法

由於心理計量取向的評量無法適用於教育現場，加上新教學理論與舊評量無法適配的困境，基層教師們再也無法坐以等待測量研究者和心理計量專家來發展適於學校的評量典範。教師們經驗到標準化測驗的一些危機，且受到大型測驗也轉向實作評量的影響，因而喚起改變評量方法的行動（Mabry, 1999）。

來自基層的反省與力量，學者們也加入支持的行列，1980年代後期，許多教育者已深深意識到，無論好或壞的測驗，都控制了教室中的教學，而我們的教育目標若要遠離過於瑣碎或背誦式的學習，那麼勢必要改變評量的方式。G. Wiggins 於1989年發表「真正的測驗—邁向更真實、更公正的評量」一文，積極倡導「真實評量」（authentic assessment）的做法。又，H. Berlak（1992）也反對標準化測驗的做法，提出支持脈絡化評量的理論性評論，其主張是：

1. 評量的內容應是課程本位和真實內容（這些內容是學生有機會習得的，並可應用在學校以外的世界）。
2. 測驗不應僅有選擇題，而應包含建構反應題等開放題型。
3. 學習情境和測驗情境應相符合，學生在教室和自然熟悉的情境中接受評量。
4. 對於評量學生的學習、努力與成就，教師有權使用他們的專業判斷。

5. 評量結果不僅提供教師和相關人員知道學生學到什麼，也提供課程計畫者和教學設計者一些訊息。

另外，Glaser（1990）也提出：「評量是用來支持學習的，而不是用來指出學生現在或過去的成就而已。評量應用來激勵學生的學習，透過評量，引導學生有更佳的表現，使其具備更多的能力」。因此，他認為評量應該有下列方式，包括：記錄實際成就表現的檔案、引發問題解決行為的情境、評量學生對各種教學反應的動態測驗、紀錄學生推理過程和產品的計分程序等。換言之，Glaser 認為，我們應該有一套更廣泛的評量策略來評量學生更寬廣的認知層面，而不是只有測量學生在學科上所獲得或保留的知識，評量要評的應該是那些能實際應用到生活中的知識。在這個時期，真實評量與實作評量被廣泛的運用與倡導。

（二）優點與限制

脈絡評量典範的特色，在於評量的內容應是學生實際經驗到的課程，而不是與學生課程學習無關的標準化測驗內容。評量用途在反應學生的實際所學，而不僅僅是排名、分類而已，評量的回饋，對於教師、學生、家長都有實際的用途。其優點在於教師具有專業判斷權力，並提供機會給學生學習複雜的知識和技巧。

然而，脈絡化評量也有其限制。例如：規劃和實施實作評量或者真實評量的建構反應題，較為費時、費力、耗錢的。而且在一個評量時段，只能實施少數的題目，所能涵蓋內容的類推性，也引發人們質疑它的內容效度。另外，雖然對於教育內容和目的未有普遍的認同，有些人認為某些事必須學習，而對某些內容可能有所忽略，因此只與一些特殊課程相結合，並非普遍的主題和技巧。此外，常發生的困擾是，計分較為費時費力，也較為主觀。當評分者間的評分不一致時，就容易有糾紛，產生評分者信度的問題。整體來說，可行性、效度與信度，是脈絡化評量遭遇到較大質疑和挑戰的問題（黃秀文，1996；Mabry, 1999）。

三、個人化評量典範

（一）理念與做法

延伸脈絡化評量典範對促進學習的關注，個人化評量典範也逐漸浮現，由內容主導轉為學生主導，將更多的評量權力釋放給學習者個人本身，強調學生參與。個人化評量典範與建構主義的學習理論、個別化教育和多元智能理論是一脈相連的，其基本主張是：每個學生以不同的方式來了解相同事物，因此評量應多元化、個別化，允許

學生使用多種方式來展現所長。加上每個人有不同的興趣、目的，因此，評量應使學生預備達成其個人目標的成就（Mabry, 1999）。

個人化評量更關注學習者本身的參與，藉著適當地為個人量身訂製評量內容、實施和品質標準，擴大了認識和確信學生知道什麼的機會，而非透過和別人比較或與事先界定好的標準來了解學生的成就。在個人化典範的評量中，學生有更多的投入和權力，他們可以和教師協商那些評量內容與他們的個別成就有關，什麼格式可提供他們展現最佳的知識和技巧。學生可協商何時、何地來進行測驗。Mabry（1999）歸納說明，在個人化典範中，其評量的特色是：

1. 評量內容是由學生選擇或師生共同協商；
2. 評量的時間和方式也是由學生選擇或師生共同協商；
3. 著重描述性的評量或判斷，而非量化數字；
4. 提供修正、改進、擴展學生學習的建議，和個別化的回饋；
5. 自我評量的標準是由學生根據自己的狀況設定或由師生協商而來的。

綜合言之，個人化評量相信學生是獨一無二的，具有學習的能力，目的在培養學生的多元才能，讓他們發揮潛能。學生參與評量的過程與內容，自由度很高，學生的自我評量是這個典範的核心。

（二）優點與限制

個人化評量的內容、格式、實施情境較為彈性，常需根據學生的不同而有所改變，評量作業的複雜度和評量標準，也會為學生量身訂做。在個人化評量中，人的專業判斷是相當需要的，因為表現過程與結果都是多樣的，事先界定好的標準可能會對個人評量有所限制。因此採個別化標準，平衡考慮每一位學生的背景、機會、目標和興趣喜好等。

個人化評量的特色是採自比式評量（ipsative assessment）策略，它不在進行常模參照比較，而是詳細檢視個別學生的成就和缺失，興趣、進展、教育機會、校內和校外的努力，對學生的所學和能做，進行整體的理解。藉著彈性評量他們的技巧，提供更清楚和詳細的學習成就描繪，讓老師與父母更清楚了解學生知道什麼和能做什麼等有用的訊息。

當然，個人化評量也有一些限制。例如：即使是想瞭解個別學生的成就訊息，通常也需要一些外在參照，比方說某個學生的所做與能為，在其他同年齡者（或同班同學）相較，其表現為何。僅僅將焦點放在學生獨特的成就和技巧，並無法回答這些問

題。另外，許多脈絡化評量的困難，在個人化評量中情況愈形嚴重，以計分為例，採人工評量則會有許多的不一致和變動，特別是當學生用不一樣的作業和標準來進行評量時，變動更大；而且要設計一份符合個人化的考題和計分標準，是相當費時費力的；另外，主題的選擇或每個人被期待所能勝任訊息的涵蓋性，也都不同，因此問題也不少（Mabry, 1999）。

談到評量內容所涵蓋的適當度，在脈絡化評量典範中，實作評量者用少數的複雜問題來涵蓋全部內容，這種問題在個人化評量中更形嚴重。Mabry（1999）指出，課程本位的評量可能忽略其它主題領域，個人化評量也出現這種內容效度的問題。另外，個人化評量系統必須能適用於學生和教學脈絡，對於實施者也有新的挑戰。在一個學校或地區，個人化評量系統需要老師對結構化的評量架構有一致性看法，對於基本的原則和實施方式要能同意，並且也要有足夠的彈性讓教師能將評量與自己的教學風格、科目、年級和學生的需要相連結。由於個人化評量系統不是靜態的，它必須持續地適於學生和教室裡變動的需要。在此情況下，結構的一致性和強調動態、脈絡的彈性，有時候是很難同時一致性的達成。從這些討論，可看出個人化評量典範所面臨的困境了。

四、三種評量典範的綜述與比較

上述探討了心理計量典範、脈絡化評量典範與個人化評量典範，Mabry（1999）指出，這樣的評量典範轉移並非單獨發生的，而是與許多的教育變革同步共變而成的，包括了：1.學習理論的轉變：著重建構主義、發展歷程，而非行為主義。2.師生角色的轉變：視學生是主動的學習者，非被動的接受者；教師是學習的促進者，非知識權威的分配者。3.課程教學的轉變：著重探究批判、統整知識、合作學習，而非單獨科目的事實記憶。4.學校組織的轉變：學校管理從科層控制管理轉變為學校本位管理，強調教師專業的賦權增能。5.教育研究的轉變：透過質化的敘述了解複雜的脈絡現象，而非單一孤立事件的量化調查。6.學習評量的轉變：基於上述種種的轉變，學習評量朝向個人化，在於了解學生的個別成就，而非標準化測驗所界定的成就。以下茲將三種評量典範的評量策略、關注焦點、實施方式、與優缺點，整理如表 1 所示。

相對於心理計量的舊評量典範，新近的評量典範實則包涵了脈絡化典範與個人化典範，在脈絡化典範中，重視教育的真實情境，關注學習者的學習。而個人化典範，更趨向關注學習者，由課程內容主導，轉向學習者主導，將更多的評量參與權力，交

給學習者本身，評量與學習之間的關係則有更多的交集。

表 1 心理計量、脈絡化評量、個人化評量典範的比較分析

項 目	心理計量典範	脈絡化評量典範	個人化評量典範
1.評量策略	標準化評量	課程本位評量	個別化評量
2.關注焦點	哪一位學生知道的多	這位學生在教室裡學到什麼	這位學生知道了什麼
3.實施情境	標準化的施測情境	一般的教室情境	依據學生的時間和情況有所調整
4.題型	封閉題型	封閉題題型、開放題型	開放題型或由學生自行選擇
5.評分者	通常用機器	教師	教師或相關人員
6.評鑑	沒有自我評量	自我評量是重要的	自我評量是核心
7.過程與結果	總結性的報告和運用結果	主要運用形成性評量，也採用總結性評量	主要運用形成性評量，也採用總結性評量
8.優點	1.有效率 2.有公信力 3.良好的測驗發展技術 4.具鑑別、分類功能	1.提供教學訊息 2.教師賦權增能 3.指出學生學得多好	1.提供教學敏感訊息 2.學生賦權增能 3.教師賦權增能 4.詳細指出學生知道什麼和能做什麼
9.限制	1.限制課程的發展 2.對學生的所知所能提供較少的訊息	1.可能窄化課程。 2.未接受評量訓練的教師可能會有錯誤。 3.無法指出與他人表現的比較。	1.評量者非常需要專業判斷和彈性，未接受評量訓練可能會有錯誤。 2.對於學生在課程上或與他人表現比較，提供很少的訊息

資料來源：整理自 "Portfolios plus: A critical guide to alternative assessment (pp.23~45), " by L. Mabry, 1999, California: Corwin Press, Inc.

討論了三種評量典範，可瞭解到每一種評量典範皆有其優點與限制。某一種評量典範對於某些目的可能優於其他典範。例如：對於進行成就鑑別、甄選分類的大規模評量時，心理計量的評量系統就很適合；對於要促進學習，進行教學診斷，此時脈絡化評量就較為適當。此外，各種評量典範對於實施方法有不同的應用，甚至是相同類

型的評量技術，它們有各自的主張與運用方式。有時我們會發現，許多評量系統錯誤地將目的配對典範或技術。一個評量政策若混淆或誤配目的與技術，會產生不想要的系統，因而限制了這種評量方式的優點，未能完全達成原來的評量目的。以個人化評量系統為例，採檔案評量是希望能看到學習者的獨特成長與反思，若要將之標準化，進行比較，則會減低此評量方法的特性與功能（Gipps, 1994; Mabry, 1999）。所以，若我們能瞭解各種不同評量取向的功能與差異，就較能發展一脈相呈密實結合的評量系統，並將評量統整於教學中。

肆、從評量典範轉移再思索評量與學習的關係

評量典範從心理計量取向的科學測量，轉向關注教育脈絡的評量，更延伸至以學生為主體的個人化評量。在這個轉變歷程中，評量的角色與目的有所轉移，評量與學習的關係也有新的詮釋，評量不僅止於成績考核，給學生打成績排名次而已，評量更需要幫助我們正確診斷學生的需要，幫助學生學習，激勵學生追求卓越，希望能培養自我反思、自我改進的自主學習者。

評量的目的從著重正確測量學習結果的評量，轉向關心促進學習的評量。從這樣的轉變，可看出評量與學習的關係愈益密切，有更多的交融，在文獻中，可看出有三種評量與學習的關係型態，分別是「對學習的評量」(assessment of learning)、「促進學習的評量」(assessment for learning)，以及「評量即學習」(assessment as learning)。以下將分析這三種評量取向：

一、對學習的評量

在學校裡，評量的目的大多在評量學習的結果，即「對學習的評量」，它的目的是總結性的，意圖能證明學習結果，並向家長和學生告知學生在學校的進展。呈現的方式通常和其他學生比較或排名，並給一個相對位置的成績。通常在一個單元、一個課程、一個年級或一個重要階段結束時做，採用這段期間上課內容的測驗或考試。評量結果會以一個分數或符號（甲、乙、丙或者 A、B、C）來表示，最後，綜合各科成績給一個平均成績，向家長報告孩子的學習情形。

這類的評量仍然主導著大部份的學校評量活動，特別在中學階段，教師的工作多

在命題與評分。教師使用測驗或考試來評量學生學習的多寡與正確度，教師花在評量的工作大多是在打成績。其重點在比較學生的成就高低，給學生的回饋是成績等第，較少提供改進的意見。這種總結式的評量有其功能，提供我們瞭解學生的學習結果與在團體中的等級排名，著重正確測量學習結果。

正確的做好總結性評量本身就是一項挑戰。排名、打成績，在教育中有著長久的歷史，它被家長和大眾所廣泛接受。有人支持打成績，也有人懷疑它的公平性或正確性。教育研究者和理論者批評傳統打成績措施已經好一段時間了，在測驗理論裡，成績是被質疑的。為什麼？因為教師在打成績時，通常考量很多因素而不是只有學業成就；教師加權各項評量成績各不相同，他們錯誤解釋單一分數來代表複雜的表現（Marzano, 2000）。當教育扮演成功未來的重要一部份，就會更注意成績如何被計算，他們實際反應了多少學生所能做，然而這些成績是否真的站得住腳？除了分數的效度受到質疑外，更值得我們反思的是：「到底評量的目的是在“測量”或者在“學習”呀？」（Broadfoot, 1996）。而 Darling-Hammond（1994）也曾指出：「學校教育的目的是將學生教育的更好，而不是選擇一些聰明來學生來準備讀書；學校不該淪為分類學生的機構，而是要指導學生學習，幫助學習是學校教育的基本目的」。

二、促進學習的評量

那麼到底評量的意義為何呢？assessment 這個字源於拉丁字“assidere”，意思是“坐在一旁或相伴相隨”之意。進一步引申為老師坐在學生旁邊，來真正了解學生學習時，發生了什麼困難和有什麼困難，在於引導學習過程並促使進一步的學習（Wiggins, 1993）。但這樣的評量意涵卻與學校制度中的評量角色有所遠離。因此，近幾年來，一些評量學者大力倡導「促進學習的評量」，亦即運用評量來促進學習，以還原評量本來的面貌（Black, Harrison, Lee, Marshall, & William, 2003；Carless, 2005；Stiggins, 2002）。

有人認為「對學習的評量」和「促進學習的評量」可簡要比喻為總結性評量與形成性評量¹，在「促進學習的評量」中，教師所蒐集一些資料是為了幫助學生學習，以

¹ Stiggins（2002）對於將「促進學習的評量」等同「形成性評量」一詞有所微詞，他並不同意這樣的看法。他認為除包含形成性評量提供回饋訊息給教學外，更重要是學生必須參與在此評量歷程中，強調幫助學生學習，讓學生更有自信的學習，且促使學生自我評量與自我提昇。Stiggins（2005）倡導學生參與之促進學習的評量

進一步修正教學外。評量作業的設計乃在瞭解學生知道什麼和可以做什麼，使用這些觀察來設計下一個教學步驟。教師在教室中使用觀察、學習單、問卷等方法來獲得對教學有用的訊息。打成績不是用來比較學生，而是觀照每個學生的優缺點，並提供他們未來學習的回饋。

顯然的，教師在這裡是核心人物，其角色與前一個取向不太一樣。在「促進學習的評量」取向中，教師使用對學生的瞭解、課程內容、和評量目的，來確定學習需求。評量與學習是交相互動的，評量就在學習之中，教師提供協助也是評量的一部份，教師提供回饋以幫助學生來鷹架下一個步驟，這時候需要教師的診斷技巧來達成。而評量的記錄主要採用檢核表或學習檔案等，以追蹤學生的進展。總之，在這個取向中，評量的主要目的乃在改進教學、促進學習。

三、評量即學習

繼續延伸促進學習的評量，也有學者提出「評量即學習」的看法（Dann, 2002；Earl, 2002），乃意圖增強和擴展促進學習的評量角色，其中特別強調的是“學生的角色”。在這個取向中，學生本身不僅對評量和學習歷程有貢獻，而且被視為是個重要環結，讓學生主動參與評量，使學生檢視自己所學，並使用這些回饋來調整、適應，甚至改變自己的學習。讓學生自己成為的最佳評量者，是「評量即學習」的終極目標，這也是個人化評量典範的核心主張。

在某些方面，學生需要自我激勵，運用自己的才能和知識來做決定以及解決問題，而不是等待別人告訴他們什麼才是正確答案。有效的評量可使學生賦權增能（empowerment），問一些反省問題，考量一連串的學習策略。以學習檔案為例，當學生可以運用個人知識來建構意義時，能自我監控瞭解自己知道什麼和不瞭解什麼，並且決定下一步該怎麼走時，那時候，學生又向前跨進了一步。

成績記錄在這個取向中是個人事務，學生和教師一起決定什麼是學習的重要證據。學生定期反省他們的工作，並判斷如何做得更好，那麼與別人比較就沒那麼重要了，重要的參照點是學生先前的工作和抱負，以及持續的學習目標。總之，自我評量是這個取向的核心。

上述三種評量與學習的關係取向對學生的學習皆有貢獻，只是重心有所不同，茲將這三種取向的評量特徵歸納整理如表 2 所示。

（Student-involved assessment for learning），實則涵括了評量即學習的觀點。

表 2 三種評量取向的特徵

評量取向	目的	參照點	主要評量者
對學習的評量 Assessment of learning	安置、獎勵、晉升、學歷、資格成就的判斷	其他學生	教師
促進學習的評量 Assessment for learning	提供教師教學決定的訊息	外在標準或期許	教師
評量即學習 Assessment as learning	自我監控和自我修正、自我調整	個人目標和外在標準	學生

資料來源：”Assessment as learning: Using classroom assessment to maximize student learning (p.26),” by L. M. Earl, 2002, California: Corwin Press.

四、不同評量取向間的重構與平衡

上述三種評量與學習關係取向的討論，說明了每一種評量皆有其價值與重要性，而非誰優誰劣。然而，教學評量是個複雜的工作，身兼多重目的，在不同的情境中面對不同的聽眾，有不一樣的意義。評量的目的是多重的，有時支持了這個目的，卻和另一個目的產生衝突，那我們該怎麼辦呢？面對這樣的矛盾，Earl（2002）曾建議，根本之道是去瞭解各種評量取向間的意涵，瞭解他們之間無可避免的衝突與矛盾，知道你正在用什麼，為什麼採用，而且能更有智慧的運用。其中，一個要掌握的訣竅是如何去取得正確的平衡。

以目前的狀況看來，傳統環境中的教學評量大部份是總結式的評量，即「對學習的評量」，主要目的是在教學之後用來測量學生的學習成果，並使用成績來分類學生，向他人報導這些結果。也有一些教師使用「促進學習的評量」來建立診斷過程，進行形成性評量和回饋；然而，系統性的「評量即學習」卻是很少存在的。值得反思的是，學校教育是否總是一再地使用「對學習的評量」，而讓其他的評量方式沒有容身之地。Stiggins（2002）就指出，目前的評量仍多偏重「對學習的評量」，非常缺乏「促進學習的評量」，這也是目前評量的危機，我們需要努力加以平衡。

至於如何平衡呢？Earl（2002）將傳統評量中三種評量取向的關係繪製成一個金字塔（如圖 2-1），提供我們一個檢視方式。在傳統評量中，「對學習的評量」為主要，佔了大部份。他認為現在需要重構這三者之間的關係，如圖 2-2 之所示。

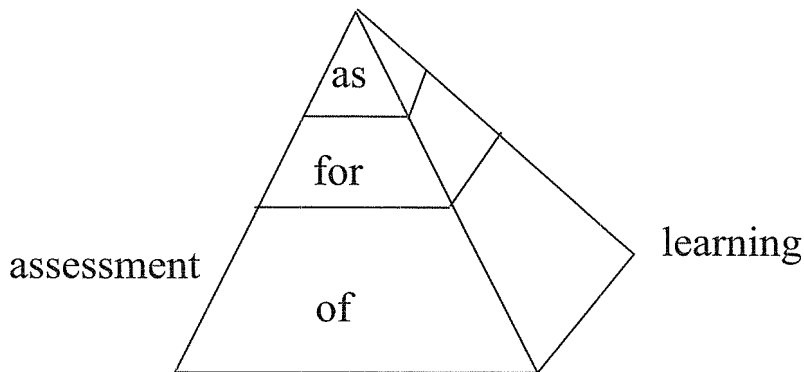


圖 2-1 傳統評量的金字塔結構

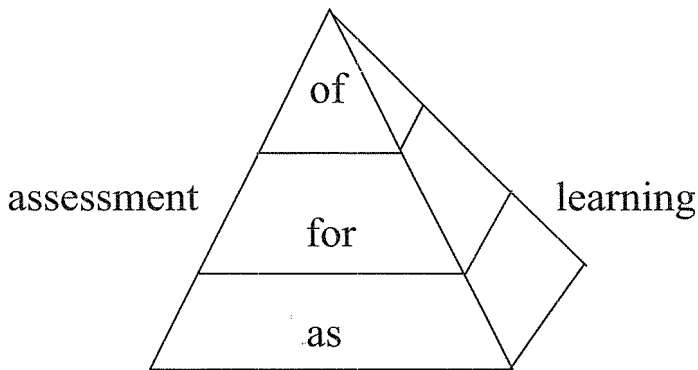


圖 2-2 重構評量金字塔的結構

資料來源：”*Assessment as learning: Using classroom assessment to maximize student learning* (p.27)”, by L. M. Earl, 2002, California: Corwin Press.

圖 2-2 重構了一個新的評量金字塔，來達成新的平衡。在新的結構中，強調增加「促進學習的評量」和「評量即學習」二者的份量，因為課堂評量的焦點是學生的學習，對教師而言是「促進學習」，對學生來說則「即是學習」。在這裡「對學習的評量」仍有其角色要扮演，當決定是需要做總結判斷時，但相較其他二者，此時的比重相對是較少的。

那麼為何要以評量即學習為基底呢？在課堂評量中，所關心的是促進學生的學習。若從評量典範的轉移來看，評量從心理計量取向的科學測量，轉向關注教育脈絡

的評量，更延伸至以學生為主體的個人化評量。在這個轉變歷程中，評量的角色與目的有所轉移，評量不僅止於成績考核，更重要的是幫助學生學習，進而激勵學生自我改進與主動學習。過去，學生被動地使用評量訊息來改進學習，而現在，則鼓勵學生自我評量，主動運用評量來管理與改進自己的學習，設定目標，為自己的學習做決定。評量的終極目標在培養自我改進、自我反思的自主學習者。因此，評量即學習就成為課堂評量的主要根基。

伍、評量典範轉移對教育工作者的啟示

上述討論了「對學習的評量」、「促進學習的評量」和「評量即學習」等三種學習評量取向，看來似乎可分別對應為是「心理計量典範」、「脈絡化評量典範」和「個人化評量典範」的核心理念。我們可瞭解到，超越測量，運用評量來促進學習，寓評量於學習之中，是目前新評量典範所強調的。如何運用評量來促進學生學習，進而培養能自我評量、自我反思的自主學習者，是一個重要任務。然而，要實踐新評量典範的理念，並非一蹴可幾的，必須考慮評量技術與文化等多方面的改變，同時也挑戰著教育工作者的信念。以下針對評量典範轉移所帶來的衝擊，提出對師資培育課程、評量文化以及教師評量信念的啟示。

一、檢視師資培育評量課程的內容

面對新評量典範實踐的需要，教師必須具備適切且足夠的評量素養。評量在教學工作中佔著很大的份量，但是，有些老師沒有修習過評量課程，憑藉著自己過去的經驗來進行。在台灣，目前師資培育課程中，教育測驗與評量課程是列為必選，並不是必修（教育方法課程 5 選 3 即可），因而有些師資生並未修讀評量課程，就進到教學現場了。

而在職教師，由於並未接受過足夠的新評量設計訓練，以致在面對新評量思維時，未準備好面對這樣的需求。Stiggins（2002）就指出：「要發展出促進學習的評量，需要教師們高度的專業知能，因為這其間牽涉的不只是評量方式的改變，而是整個教學內容、教學方法以及師生互動關係的改變，才能導引學生進入更深層次的學習與評量活動。缺乏專業知能的教師，評量的可行性及實施效果會大打折扣」。因此，教師

的評量素養的發展是相當需要的，包含在職教師和職前教師評量知能的養成，兩者必須雙管齊下。

有關教師評量素養的發展，我認為師資培育機構的教師有必要反思與重新調整評量課程的內容。Stiggins（2001）曾檢視美國的評量課程內容，他發現許多的評量教科書都在教導大家如何計算測驗的信度，評估測驗的效度，要學生進行試題分析，算出試題的難度值和鑑別度指數，焦點多放在測驗編製的心理計量部份。他指出這個現象，並不是反對要學習這些，而是要反思這些內容，教育現場的老師們在課堂的評量中會經常遇到嗎？師資培育者是否更需要關心教師們在教室裡的實際評量狀況，導引他們如何做好教室裡的教學評量，這才是實際且重要的課題。隨著教育評量典範以及實際教學評量的需要，師資培育的評量課程內容宜有所調整，需要將焦點多放在關注教室裡的學習評量。在台灣，民國 86、87 年之前的評量教科書，的確也呈現此種狀況，而到了民國 90 年之後，教育測驗與評量的教科書中，也漸漸增多教學評量的課題，內容也有所調整與改善了！除了教科書內容的改變，師資培育者也應不斷檢視所設計的評量課程內容，是否合乎時宜，還有是否符合現場教師的需要！

二、重塑新的評量文化：從考試文化走向學習文化

處在評量典範轉移的氛圍中，評量改革也大幅的推動。然而，評量改革是同時發生在微觀與鉅觀層面的，不僅發生在教室裡，也發生在學校裡，甚至是整個國家的教育政策。評量改革攸關許多人的利害，教師面對評量改革，除了面臨技術上的困難，也牽涉到文化與政治層面的挑戰，這些都需要花費許多的智力與情緒能量（Hargreaves, Earl, & Schmidt, 2002）。因此，在面對評量派典轉移時，除考量時間、技術支援系統、新舊系統間的轉換與適應、績效等技術問題外，也應關注評量文化的問題。

長久以來，由於升學主義的籠罩，使得考試扭曲了學生的學習動機，致使許多學生讀書的目的在於求得高分，為著文憑或"錢途"，或者用以取悅教師和家長，卻忽略了學習的原本意義和目的（盧雪梅，2000）。考試也宰制了教師的教學，教師的教學常因為考試與進度，遭到嚴重壓縮，使得教師的專業難以施展。學生的考試成績成為教師之間評比의 依據，讓校園隱伏著爭鬥與不安的氣氛。

Shepard（2000）曾語重心長的提出：「教師應該協助學生改變以分數結果為導向的評量文化，重塑以學習目標為導向的評量文化。幫助學生體驗學習活動本身是有意義和價值的事，而不是為了求取高分來取悅別人」。然而，在這種習慣於統一考試和

標準答案、用考試分數來說明學生的能力全部，以分數來衡量一個教師的認真程度、評斷教師的教學成效的考試文化下，要倡導新評量觀點似乎有困難，就學生部份，分數似乎是學生在校園中生存的籌碼，追求分數成為學習生涯的唯一目標。要學生對分數釋懷、轉向致力於學習，是需要教師、學校行政人員與家長，同步共變的。

此外，從學校的發展歷史來看，學校像個分類機構，評量的重心多放在打成績，排名次。而現在將重心轉移到學習，強調評量和學習是緊密結合，這個理念的確是吸引人的。但是要將這樣的想法用在學校上，可能有些隔隔不入。想像這樣矛盾，老師會說評量的最困難的部份，因為他們總是在監控學習和分類學生，而且也要根據評量，教導學生，並在種種矛盾的責任中奮戰。的確，這樣的緊張關係存在著，需要重新組合這樣的平衡。教師可能實施新評量的做為，但在其他的社群間卻沒有大的變動，特別是家長和學校的期望，家長總是對自己孩子有所期望，而學校也滿心期待著有亮眼的升學成績。

在台灣，分數掛帥的考試文化是難以擺脫的現實，我們常聽到家長問孩子在學校考了幾分，而不是在學校學得什麼。分數成了學習的唯一指標與內涵。張稚美（2000）就曾指出：「台灣家長的心智習性是習慣看自己和其他人孩子的"分數"，而不習慣去分析、鑑別孩子的各項學習或關鍵能力」。而現在要扭轉家長這種長久的心智習性，的確不容易。在參與評量改革的過程中，有機會與現場教師談及新評量的實施，教師認為家長的價值觀是讓他們挫折、無奈與灰心的重要來源之一。這的確是個大挑戰，不過，我也鼓勵教師，由於家長的求學成長過程中並沒有這樣的評量體驗與基模，總是會因為不瞭解而有所誤解，希望老師們能耐心的與之溝通，當家長可以看到評量對學生的學習有所促進時，他們會從質疑和摩擦中，轉為關心孩子學習的聯盟。因此，與家長溝通和說明現今的評量理念，是重塑評量文化的必要之務。

三、調整教師的評量信念與心智習性

重塑新的評量文化，首要之務乃教師要改變自己的評量信念與心智習性。以下就教師對評量角色的認知、運用評量激發學習動機、以及設計評分標準等評量信念，進行討論。

（一）評量角色的認知：「公平的裁判」vs.「拓展心智的窗」

在實踐新評量理念時，教師首先要面對的是把自己的評量角色定位在那裡？是「公平的裁判」或者是「拓展學生心智的一扇窗」（Shepard, 2000）？過去在考試文化

中，相當強調公平與客觀，因此「公平的裁判」是評量者所認同的角色。然而在以促進學習的評量中，評量角色必需有所轉換，否則會窒礙難行。

教師在進行課堂評量時，需要建立評量是動態的觀點，動態評量（dynamic assessment）乃基於 Vygotsky 最近發展區（zone of proximal development）的看法，評量可以是互動的，可以提供協助的。過去評量者的角色是獨立的中立者，不可以提供協助的。而在動態評量過程中提供協助是評量的一部份，教師可從評量過程中瞭解學生的學習策略和如何鷹架學生的學習，這種評量方式在平常的教學中即可進行，特別是支持學生學習下一個階段的學習。

另外，評量的歷程可以是探究的、問題解決的、腦力激盪等等。Serafini（2001）提出評量是探究的歷程，教師使用各種質與量的技術來瞭解學生的學習歷程，在這過程中，教師不再只是中立的施測者，學生不是被動接受，師生間可以是溝通互動的關係，教師運用問題來促進學生學習。此時，教師的評量角色，是拓展學生學習的一扇窗。因此，教師在教學過程中所進行評量的角色，必須與實施標準化測驗時的角色，有所區別。

（二）評量對激發學習動機的信念：「懲罰恫嚇」vs.「引導改進」

評量對學生所造成的情緒影響，是我們需要觀照的議題。Reineke（1998）就指出：「教學觸及心智，評量則觸動心靈」（instruction touches the mind, assessment touches the heart）。的確，不管是正式或非正式、例行或偶發的評量活動，皆對學生有深深的影響力。學生感受到的評量氣氛是教學的重要部份。學生的評量經驗通常伴隨他們好常的一段時間，持續影響他們對未來學習的看法，評量的訊息有著情緒的訊息。也因此，Stiggins（2001）和 Weeden, Winter, and Broadfoot（2002）等人，都不約而同的提到，教師需要修正評量對激發學生學習動機的舊信念，他們主張，評量是用來瞭解和幫助學生的學習，而不是用來懲罰、恫嚇學生的。

回溯過往我們的求學過程中，時常會遇到老師運用評量、打成績來威脅我們，如果考低分的話，就會有不好的後果。少幾分，打幾下，是過去年代的學校記憶。許多教師認為要加速促進學生的學習，就是要增加學生的焦慮。如此一來，評量往往成為最大的威脅恫嚇者。這也是很多人每每聽到評量或評鑑，就覺得是很危險害怕的事，總是讓人覺得無助與脆弱。

有些老師甚至認為，恫嚇太小是起不了作用的，因此給了很多。他們所持的觀點是：「若是愈高度利害者，則愈有危機感；愈有危機感，則學生會更努力；面對愈困

難的挑戰，他們會愈努力；愈努力的話，他們就會學得更多。」但這種觀點似乎並不全然正確，若從 Covington（1992）的自我價值學習理論中來看，學生在考試中若付出極大努力而後仍然失敗時，會感到羞愧痛苦，而且可能因懷疑自己能力不如別人而喪失自尊自信。若我們要擴大學生的學習，必須使之拋掉心中的恐懼。我們希望學生不再恐懼，而能深深致力在評量活動中，來增加他們的成功。

許多的評量經驗，都讓學生感到無助，並預期失敗，這種信心的喪失使他們停止嘗試，停止學習，更進而導致失敗，讓學生感受到挫折而沒有勇氣。當然，這並不是指所有的評量結果都是簡單正向的、讓學生都受到激勵。其實，如果學生沒有達到標準，評量也必須正確反映這個事實。但當這個結果反應缺乏成功，我們必須努力去改變教學來預防這種失敗變成慢性的蔓延，亦即避免造成習得無助感。而身為教育工作者的我們必須發現另一種方式，讓學生有成功的希望。因研究習得無助感聞名的心理學家 Seligman（1998）曾指出：「要增加學生的學習，是要增加學習者的信心，讓他們相信自己是可以學習的，提高自我效能感」。

因此，教師可設計具有挑戰的任務，稍微提高門檻標準，協助他們有信心的達成。另外，也可採用學生參與的評量方式，讓學生知道達到卓越的過程，讓學生在心中有有所路徑達到成功。亦可採用自我評量，讓學生持續地紀錄自己努力的過程和成果（例如：運用檔案評量來反映個人的學習成長和反思歷程）。使用評量、持續紀錄與溝通，讓評量成為激發學習動機的不絕源泉，而不是威脅恐嚇者。如此，則可培養學習導向的學習者，而非分數導向的學習者。

（三）評分規準的設計：「分類工具」vs.「鷹架支持」

Wiggins（1998）曾指出：「評量不應只是學習結束後評核表現的工具，它應該扮演更積極的教育功能，亦即以系統化的方式培養學生自我檢核與自我改進的能力」。那麼要如何達成呢？設計適切的評分規準（rubric）就具有這樣的教育功能，特別是對所要評量的作品或表現以漸層的方式列出，讓教師知道學生已明確達成何種層級的目標，也讓學生瞭解該如何努力才能達成何種層級的目標。

評分規準將評量的標準明顯化，促使學生反省這些標準，並用標準作為學習的基礎。在此過程中，學生是學習評量的積極參與者，而不只是對一系列考試的回應者。評分規準的設計對於提昇實作評量的品質有所助益。但令人擔憂的是，許多人將焦點放在如何把學生的表現做等級分類，沒有發揮評分規準將實作規準具體描述的機能，反而僅僅對這些描述轉換為一個數字符號，學生很快地被標籤為“2”或“3”；或者“A”

或”B”等，這樣的結果產生了反作用。把一個對學習有用的工具淪為分類的工具，對於促進學習和教學的價值不見了，教師不得不戒慎！

再者，Earl（2002）曾提醒，一個好的評分規準仍然不太夠，學生需要對什麼是優秀的作品或表現能夠看到或者意想到。評分規準的訂定宜指出何謂優良的表現或作品，讓學生能清楚明瞭。雖然優良的表現或作品不會只有一個模式，但要讓學生能夠看、聽或想像這些優秀表現的各種樣子。當學生正努力要朝一個困難目標前進時，特別是需要統整新學習和不同技巧的複雜作業時，他們需要看到如何做，如何達成的過程。讓學生知道要去那兒，多久可達成，如此具有鷹架作用與激勵效果，不會漫無目標。亦即，評分規準不僅是幫助教師判斷學生實作表現的品質，也應幫助教師進一步協助學生獲得應學得的技巧與內容，激發學習動機，達成促進學習的積極目的。

參考文獻

- 吳毓瑩（1996）。評量的蛻變與突破－從哲學思潮與效度理論思考起。教育資料與研究，13，2-15。
- 教育部（2005）。國民小學及國民中學學生成績評量準則。台北：教育部。
- 黃秀文（1996）。從傳統到變通：教學評量的省思。國立嘉義師院國民教育研究學報，2，1-26。
- 張稚美（2000）。落實多元智慧評量是對心智習性的一大挑戰。載於郭俊賢、陳淑惠（編譯），落實多元智慧教學評量（頁9-12）。台北：遠流。
- 楊銀興（2000）。傳統評量與新式評量之比較及國小教師對實施新式評量相關問題覺知情形之研究。國立台灣師範大學教育研究所博士論文，未出版，台北。
- 謝祥宏、段曉林（2001）。教學與評量：一種互為鏡像關係。科學教育月刊，241，2-11。
- 盧雪梅（2000）。多元教學評量的理念與實務。教育研究月刊，76，57-66。
- Berlak, H. (1992). Toward the development of a new science of educational testing and assessment. In H. Berlak, F. M. Newman, E. Adams, D. A. Archbald, T. Burgess, J. Raven, & T. A. Romberg (Eds.), *Toward a new science of educational testing and assessment* (pp.181-206). Albany: State University of New York Press.
- Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Reading, MA: Addison-Wesley.
- Black, P. (1998). *Testing: Friend or foe? The theory and practice of assessment and testing*. London: The Falmer Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.
- Broadfoot, P. (1996). Assessment and learning: Power or partnership? In H. Goldstein & T. Lewis (Eds.), *Assessment: Problems, developments and statistical issues* (pp.21-40). New York: Wiley.
- Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education*, 12(1), 39-54.
- Covington, M. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. New York: Cambridge University Press.
- Dann, R. (2002). *Promoting assessment as learning: Improving the learning process*.

- London and New York: Routledge, Falmer.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64(1), 5-30.
- Earl, L. M. (2002). *Assessment as learning: Using classroom assessment to maximize student learning*. California, Thousand Oaks: Corwin Press, Inc.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Falmer Press.
- Glaser, R. (1990). Toward new models for assessment. *International Journal of Educational Research*, 14(5), 475-483
- Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, 39(1), 69-95.
- Lazear, D. (1994). *Multiple intelligence approaches to assessment: Solving the assessment conundrum*. Tucson, Arizona: Zephyr Press.
- Mabry, L. (1999). *Portfolios plus: A critical guide to alternative assessment*. California: Corwin Press, Inc.
- Marzano, R. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Reineke, R. A. (1998). *Challenging the mind, touching the heart: Best assessment practices*. Thousand Oaks, California: Corwin Press, Inc.
- Seligman, M. (1998). *Learned optimism: How to change your mind and your life*. New York: Pocket Books.
- Serafini, F. (2001). Three paradigms of assessment: Measurement, procedure, and inquiry. *Reading Teacher*, 54(4), 384-393.
- Shepard, L. A. (1991). Psychometricians' beliefs about learning influence testing. *Educational Researcher*, 20(7), 2-16.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5-15.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Stiggins, R. J. (2005). *Student-involved assessment for learning* (4th ed.). New York:

Merrill prentice Hall.

Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment: What's in it for schools?* London and New York: Routledge, Falmer.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 703-713.

Wiggins, G. (1993). *Assessing student performance*. San Francisco: Jossey-Bass.

Wiggins, G. (1998). *Educative assessment*. San Francisco: Jossey-Bass.

Wood (1990). The agenda for educational measurement. In T. Horton (Ed.), *Assessment debates* (pp.48-56). London: Hodder & Stoughton.

投稿收件日：95 年 11 月 2 日

接受日：96 年 1 月 30 日