

教學評量改革的難題分析 ～技術、文化、政治與後現代觀點*

江文慈

九年一貫課程自 90 學年度起，在國民中小學已經分階段開始實施，成為這一波教育改革的重大工程。為了落實課程改革，教學評量也必須跟著共變。然而，教學評量改革涉及的層面廣遠，攸關許多的利害關係人，有必要從不同的視角來審思其中的複雜度。本文從技術、文化、政治和後現代等四個不同的觀點，來分析目前教學評量改革所面臨的難題與挑戰。技術觀點在檢視發展新評量技術的策略、組織和結構等問題。文化觀點則關注新評量理念，是如何地被解釋與統整到學校的社會文化脈絡中，還有新舊評量文化轉換間的衝突與兩難。政治觀點則著眼於各種不同團體對評量期望的協商和權力角力，還有評量歷程中，有關人際互動的權力關係和控制行為的探討。後現代觀點乃在探討後現代社會的特徵，帶給教學評量可能的契機與危機，包括：確定性消失的自由與無所適從；高科技模擬的真實與虛幻；多元繽紛的實質與表象；無邊際自我的增能與浮誇等。希望透過從不同的視角來省思此波評量改革可能的契機與危機，或許可以尋繹到轉機，以俾更審慎地執行對教育價值的追求，並促使在推動評量改革時，能更嚴謹、均等和持續的發展，得以超越形式之爭，回歸教育本質。

關鍵詞：教學評量、評量改革、多元觀點

本文作者現為世新大學師資培育中心副教授。

* 本文係世新大學學術研究專案【編號：9218】之部份研究成果，特向世新大學致謝。

壹、課程改革中的重要環節 - 教學評量

九年一貫課程自 90 學年度起，在國民中小學已經分階段開始實施了，為了能落實課程改革，教學評量也需要跟著改變。於是，教育部在 90 年 3 月 29 日公布了「國民中小學學生成績評量準則」，以因應九年一貫課程的實施。如同陳明印（2001）的分析所言，課程改革要達到目標，絕不能將改革的力量，停留在課程設計的表面層次，必須將改革植基在實際的教學活動，同時深化在評量的運用上，方能帶動課程改革的效果。

在諸多的評量改革項目中，大型的評量（例如：基本學力測驗、學科能力測驗案）總是受到較多的議論與矚目，因為這些測驗是高風險的評量，它攸關著學生未來的前途。但是，教室裡的教學評量改革亦是不容忽視的，因為它主導著教師的教學與學生的學習。鄭富森（2001a）就曾提及，教學評量是教育改革成敗的關鍵，若教學評量不改，其他的教育改革措施都將白費力氣。

考試領導教學，似乎是難以改變的樣態。所以，與其責難考試領導教學，不如積極改變考試的體質，讓它對教學產生良性的影響，這是新評量派典的主張，也是此波評量改革的訴求。其中，與過去評量方法較大不同是，倡導一些超越傳統紙筆測驗的另類評量(alternative assessment)方法，包括實作評量(performance assessment)、檔案評量(portfolio assessment)等等。這些另類評量的理念與做法，是希望能激勵學生為自己的學習負責，將評量統整到教學中，並融入實際活動，以促使學生有創新、應用知識的能力，而不是只有記憶基本知識和發展基本技巧而已。其目的，在於培養具有生產力、反思力的學(Anderson,1998；

Mabry,1999；Reineke,1998)。

此次的教學評量改革承載著教育改革的重責大任，希望對創新教學、快樂學習、培養學生帶著走的能力，能帶來正向的助益。評量改革所倡導的各種另類評量也挾帶大量的優勢，在校園裡大舉的推動。然而這樣的評量改革究竟在校園裡實施的狀況為何，有何困境呢？又，評量改革所倡導的新方法總是如倡導者所言般的運作，較能真實反映學生的學習結果，激發高層思考能力？它總是人性化、適性地用來支持學生的學習和發展，還是也有意想不到的後果？這些關於評量改革可能面臨的深層問題，都需要進一步地省思與檢視。

貳、教學評量改革的難題分析：多元觀點

改革本來就是非常複雜的過程，涉及的層面廣遠，同時也攸關許多的利害關係人。因此，在理解改革時，有必要從多個視角來思考，以避免單一向度的狹隘和侷限。House(1981)曾從技術(technological)、文化(cultural)、政治(political)等三個觀點，來解釋學校改革的複雜本質，頗有系統與見解。但在考量社會變遷快速的衝擊下，後現代(postmodern)的觀點，也提供我們另一視角來理解改革的脈絡。Hargreaves,Earl和 Schmidt(2002)就曾經從技術、文化、政治與後現代等四個觀點來解析另類評量的實施。本文就奠基在這些基礎上，從技術、文化、政治與後現代等四個觀點，來審視台灣目前此波教學評量改革可能遭逢的難題與挑戰，以更深入地反思評量改革的實踐。

一、技術觀點

教育創新是以理性分析為本的技术性工作，創新的本身、特色、內容和介紹，

都是一項技術，其結果是可預期的。創新的目標是既定的，它的問題就在於如何做最佳的執行(鄭燕祥，2001； House,1981)。在評量改革場域裡，技術觀點著眼於發展新評量技術的策略、組織和結構，包括：教師在獲得各種新評量技巧所面臨的挑戰，設計有效評量之困難，還有時間和資源等問題。

(一) 教師評量專業知能的不足

就技術層面來看，評量改革的成功與否與教師的評量知能有高度的關聯，老師在執行新評量策略時，必須具備有新式評量的知能，包括：實作評量、檔案評量…等，並熟練這些方法，以成為專業的評量者(Stiggins,2001)。然而，評量改革通常假設教師有能力執行；但實際上，卻有相當大的落差。目前，有些研究調查教師在教學中使用另類評量的狀況，發現這些另類評量對老師而言是嶄新的世界，大部份的老師都只受過些微的訓練，而且缺乏基礎評量知識。楊銀興(2000)的研究發現：國小教師對新式評量的認知情形不佳，大約有四成的教師並不了解，非常了解者極少。而鄭富森(2001b)也指出，目前國內在推動多元化評量時，所面臨的難題之一即熟悉度過低，因為過去多使用紙筆測驗，大家只熟悉這種評量方法，而不知道其他的教學評量技術。而訪談的教師，就道出這樣的心聲：「另類評量，對我們而言，是全新的經驗。辦了幾場研習，就假設我們瞭解並知道如何運用了。雖然台面上熱烘烘，但台下的我們仍然霧煞煞，幾個評量名詞，至今只聞其名，但仍不知其內涵，如何做，就各憑本事了」(訪談 900912)。

另類評量的實施需要有具備教學與評量專業知能的教師，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。Stiggins(2001)曾提及，另類評量的實

需要複雜的技術，需要精密的專業訓練。例如：修正實作評量的評分規準，以捕捉學生表現的複雜度。因此，教學評量改革的挑戰是不僅要協助教師發展一些有意義且公平的評量技術，也要幫助教師理解這些技巧的理念，並能整合實作評量、檔案評量、自我評量等評量技術，將之運用到實際的教學情境。教師需要更多的專業訓練以熟悉各種教學評量的知識和技術，深入了解各種評量方法的優點、限制與適用時機，以避免一窩蜂地將闖關活動的評量形式，套用於所有學科評量的趕流行現象。

(二) 評量改革配套措施的不夠

1. 評量政策宣導不周

鄭富森(2001b)的研究中提到，許多教師反應到教育改革政策倉促，配套措施不完整。加上改革政策的相關資訊缺乏，且宣導或諮詢管道閉塞，讓基層教師即使想努力研習新政策，也經常是求助無門，最後造成大家對新政策只是一知半解。基層教師除了對評量改革政策缺乏瞭解外，也對政策與自己沒信心，認為自己未必有能力去落實這些評量改革措施，而且對於未來也缺乏信心，認為未來未必會比現在好。因此，對政策執行持保留與觀望態度，導致政策無法落實與深耕。

2. 評量支援系統欠缺

除了教師評量專業知能之外，教師未能實施新評量策略，有時候來自許多制度面的限制，例如：沒有充裕的時間、資源、專業發展和諮詢服務提供給教師。部份教師反應：「課程統整與教學創新已耗盡我們所有的心力，想要在評量繼續努力，已經力不從心」。在有限資源與心力下，只好回到熟悉的紙筆評量。張郁雯(2001)就指出，紙筆考試的評量之所以歷久不衰的原因之一是簡單易行，如果要採用多元化評量，

主題文章

則教師需要投入更多的精力與時間於評量的研發與使用，如此，必然會排擠其他教學相關時間。在缺乏強而力有的評量支援系統，則可能教師明知道多元評量的優點，也有能力與知識發展，卻仍舊依賴紙筆式的評量。因此，為提高多元評量的可行性，評量方法的開發與研究，發展具體可行的評量工具，提供教師參考，是相當需要的。否則，這些評量改革所推動的方案，也將樹之高閣，隨著改革浪潮的平息，而聲消跡匿。

(三) 評量技術本身的難題

1. 能力指標評量的困惑

目前，九年一貫課程已開始實施了，教師首先面臨的評量技術挑戰是，能力指標的評量。如何將一項項的能力指標轉化為評量指標，並選用合適的評量方法，進一步發展可信又有效的評量工具，來評量這些學習結果呢？然後，又要如何呈現這些評量結果？如何將這些評量結果讓學生與家長能瞭解呢？一些基層教師對於這些種種，感到有些困難，也有些焦慮，甚至無所適從。

實際作業上，從能力指標→單元教學目標→單元教學評量指標→設計教學活動→進行教學活動→學生實作→教師觀察評定後，才評定出“能力指標評量細目表”的其中一個項目。東風（2001）指出：依據十大能力所發展出的能力指標，再依能力指標設計評量方式，能評量出什麼結果，具有何種意義，令人質疑。以國小一年級生活課程為例，能力指標有九項，其中(7)發展科學過程技能(8)提昇科學認知(9)涵養科學精神。每一項都要評量，評量方式以 3、2、1 呈現。教師要如何評量這三項十分貌似的能力指標，就算十分努力的評量出來，學生又能從評量結果中，知道什麼，家長又能從這樣的評量敘述中了解什麼意義，了解自己的孩子會什麼和不會

什麼。部份教師對能力指標的評量感到困惑與納悶。

2. 另類評量實施的限制

目前評量改革，倡導多元評量的實施。其中，各種另類評量方法也非全然好用，實際上，它也有一些令人困擾的技術問題，例如：耗時費力、評量標準難訂定、信度與效度待檢核等等(黃秀文，1996；McMillan,1997)。各種另類評量的發展需要高度教學資源的耗損，常需花費教師大量的專業能量、體力和教學時間。鄭富森(民2001b)的研究指出，實施多元評量將大幅增加評量成本，造成學校相關預算沉重的負擔。以目前教師常設計的闖關評量活動而言，需要較多的人力和物力。又，檔案評量的實施，有許多教師頗贊同的，但它需要較多的文字描述與回饋，在時間壓力下，一般教師的反應是「我喜歡這種評量方式，但討厭它佔掉我太多的時間」。因此，常見到學生整本厚厚的學習檔案，卻看不到教師隻字片語的回饋。此外，實作評量的實施較費時，也因班級人數較多，而影響教學進度，他們表示，經常處在天人交戰之中。

3. 評量結果紀錄的困擾

實施多元評量後，教師向家長溝通學生學習成果的困境是，傳統的成績單無法適用，亦即大家非常辛苦地實施多元評量，最後卻無法把評量結果詳實紀錄下來，沒有讓學生和家長瞭解這些學習成果。所以設計出適合多元評量的成績報告單是當務之急。目前各縣市國民中小學的學期成績通知書有大幅的修正，有些學校也發展出學生學習成長護照，皆在努力改善這樣的困境。但是，這樣的改進是否也有困境呢？

以台北縣國民中小學成績評量作業要點中的學期成績通知書來看，各領域的能

力指標評量細目表，頗為繁雜，教師要做文字敘述的部份，由原有的一個欄位，變為 13 個欄位。以科任教師為例，如果擔任十個班級，一班 35 人，則該教師必須評量 350 人次的該領域細目表，連學生名字都未必叫得出，卻要評量多如牛毛的能力指標評量細目表，與描述其學習情形，教師憂心其評量效度，也不知意義何在。東風（2001）說道：這樣看似周延的評量表格，著實是煩瑣複雜，窒礙難行。光是紀錄，就人仰馬翻，不要說如何評量了。他憂心著，這樣的評量紀錄方式，只是徒具形式，欠缺實然，反而限制了多元的精神，落進了形式的陷阱。

4. 質化評語撰寫的難處

90 年頒布的「國民中小學學生成績評量準則」特別強調「評量紀錄應兼顧文字描述及量化紀錄，文字描述應依評量內涵與結果詳加說明，並提供具體建議」。質化的評量描述是目前評量所強調的紀錄方式，用以說明與補充量化數字、等第的意義與不足，針對表現程度的描述與評價，讓學生與家長清楚明瞭學生學習的情況，與可改進的地方（林怡呈，2002）。但以精密的文字反映學生特質的評量方式，無疑是教師很大的挑戰與考驗。與家長溝通評量結果，向來在親師間存有緊張的關係，基於鼓勵原則，所描述的語詞，大多採正面的敘述，訪談中的教師表示，寫來寫去都是那些相似且中性的語句，對於改進的陳述，則語多保留。加上每個領域，那麼多的欄位皆要寫評語，如何忠實的描述學習歷程與成果，實在存有相當的難度與挑戰。評語的撰寫是頗費腦力心力的，為了解決評語思考的費力，也有學校提供許多電腦評語可供勾選或參考，然而這種科技選擇的評語是否符合實際狀況？它是否會流於樣板，而回到過去成績單上的"八字箴言"，則有待評估了。

評量改革所倡導的多元評量，深具理

想色彩，但所花費的時間、成本、人力、心力都較以前來的沉重。在現實因素的考驗下（時間與資源有限），其命運又將如何呢？費時費力的另類評量是否又將棄守，重回經濟省力的傳統式紙筆測驗的懷抱中呢？教師又如何能在兩者之間取得平衡呢？將是教學評量改革的一大技術挑戰。

二、文化觀點

改革若僅限於技術改革，不涉及成員價值信念的相應改變，則往往流於表面，如果改革效果並非無效，也是短期而且有限的生命。成功的改革，不僅是技術的改造，也同時是文化的改革，才能達到預擬的目標，並且產生持久又有效的改進（鄭燕祥，2001）。House（1981）指出，改革或創新，不只是在技術層面，而是從改變現有關係到建立溝通、促進彼此的了解。改革是新觀念與學校的文化歷史進行互動融合。在評量改革的脈絡中，文化觀點著重在瞭解新評量的發展是如何融入原有的文化中，還有教師處在新舊評量派典轉換間，其評量價值和信念轉換的衝突。

（一）分數掛帥考試文化的扞格難入

評量改革要導向多元學習，不再以考試分數為依歸。但在台灣的教育現場中，你說分數不重要，那是自欺欺人的。定期考試時間一到，教師們還是得努力地去幫學生複習功課，因為它不僅牽涉到學生的學習，還有教師教學成效的檢核、同學年班級的評比、家長對孩子的期待、以及學校的辦學績效等等。一位剛任教的小學教師，就道出她對學校定期評量的震撼：「學校的習慣是在考前一週就要把考試所有的範圍教完，考前空出來的一週，可想而知就是要幫學生複習再複習，所有的課都停下來，讓學生不斷的練習考卷，這不就是

主題文章

又回到過去的時代?! 學生很在意成績, 因為他們的爸爸媽媽很在意; 學生因為一兩分計較老半天, 因為他回家會因為那一兩分被處罰, 這就是我萬分痛苦的地方。事實上在開學之初的班親會, 我就已經先幫家長打預防針, 希望家長要關心孩子的平時努力, 而不要以一次的成績來判斷孩子, 家長似乎也都明白, 也都希望孩子快快樂樂的成長, 可是一到了考試全都變了樣……, 我則是盡量在挽回頹勢, 每當訂正練習題時, 我都會一再提醒學生真正了解題目的意思, 比考幾分來的重要, 但是徒勞無功……, 你不得不承認, 在教學現場中的定期評量, 對一間學校、一個學生、一對家長、一個教師竟有如此大的影響」(訪談 1039011)。

這樣的校園寫真案例顯露出, 分數掛帥的迷失, 是台灣評量文化難以擺脫的夢魘。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。張稚美(2000)就指出, 台灣家長的心智習性是習慣看自己和他人孩子的"分數", 而不習慣去分析、鑑別孩子的各項學習或關鍵能力。而長久以來, 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的依據, 讓校園隱伏著爭鬥與不安的氣氛。

如果我們的評量文化, 依舊習慣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, 那麼評量改革要落實, 仍然還有一段漫漫長路要走的。一般人很少去質疑這些分數的效度和

意義, 使得它已逐漸控制了今天的教育。當然這並不是說要全然反對測驗, 而只是希望大家能夠瞭解, 考試分數只能反映學生才能和知識的一部份。再者, Shepard(2000)也指出: 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, 而不是為了求高分來取悅別人。評量改革要落實, 評量文化的重塑乃必要之路。

(二) 不同評量價值間的衝突與混淆

此波的教學評量改革, 強調評量原則適性化, 評量功能積極化、評量方法多樣化, 評量標準透明化、評量人員多元化、評量結果紀錄質量兼重等。在此新舊評量轉換過程中, 教師實踐新評量有何困難, 在評量信念和價值上產生了那些一些衝突, 進而對新評量的看法產生了質疑與困惑, 是值得關照的。

1. 快樂與深度學習的平衡難題

過去, 教學評量發生在教學之後, 針對一個課程、一個單元或一學期, 給予一個分數或等第, 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。現在評量改革主張將評量統整在學習中, 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。評量方式從有固定範圍課程和標準答案的評量, 轉向學生感興趣和重要主題探索的開放式評量(Anderson, 1998; Black, 1998)。評量是用來加強、深化和擴展學生的學習, 而不在於指出他們的錯誤或失敗(Lazear, 1999)。在台灣, 此波課程改革的訴求之一是快樂學習, 在快樂學習的基調下, 講求學習樂趣的教師可能花較多的時間, 絞盡腦汁設計生動有趣的教學活動, 而不在於分析考試的結果, 發現學生

的困難，找出改進之道。這樣的方式，使教師覺察到，強調多元能力展現，看起來多元多樣，但因為較少進行錯誤診斷，並針對錯誤加強練習，而導致寫、算等基礎能力削弱了。部份教師抱怨，現在的學生雖然能言善道，但卻錯字連篇。他們省思著，熱熱鬧鬧的評量活動下，學生究竟學到了什麼？多元豐富的評量活動真得深化了學習嗎？他們也擔心著，這樣的評量方式可能會削弱學生面對未來升學的競爭力。

2. 適性與公平的取捨掙扎

評量應適性化，是教學評量的基本原則之一。但適性化，常在公平的前提下，被徹底犧牲了。國內長久以來存在著評量即是考試、評比的觀念，也因此，公平性就被置於優先地位，凡事先求公平再談。公平有其重要性，但是過度地凸顯，有時候會淹沒掉評量應該是幫助教師了解學生的學習狀況，和透過評量輔導學生學習的真正目的。導致這樣的困境，其原因之一來自教師對不同類型的評量目的，有所混淆。張郁雯（2001）發現，教師無法區分入學相關考試的性質，與教室裡的教學評量性質有所不同。升學篩選的考試主要是用評量結果作為入學的依據，因為考試結果對學生影響重大，所以家長與學生都很關心此類的考試，最基本的要求是公平。而教室裡的評量的目的，不在篩選學生，而是用來瞭解學生、輔助教學，因此需要適性化和多元化，可針對不同學生發展不同的評量方法與內容。但由於傳統聯考觀念的根深蒂固，未能清楚區分其目的，常使得教室裡的教學評量也淪為區辨學生程度，且處處講求公平一致的狀況，無法針對個別差異，進行適性化的教學與評量。

3. 主觀與客觀評分的質疑脆弱

目前學校實施多元評量的一大困境

是，家長質疑另類評量分數的客觀性與公正性，並指出定期考試的成績較為客觀，報告或檔案的成績摻雜著許多教師主觀的意見與偏見，因此可能使評量結果不夠客觀公平，會失去比較的基礎。當家長或學生對評量結果產生質疑時，部份教師也會對自己評量的公允產生了懷疑與信心動搖，人們期待教師能做公平的判斷，認為老師是教室裡公平的裁判，但他們無法確知。部份原因來自過去採用的紙筆測驗，有所謂的標準答案，其分數的客觀性較少受到質疑。然而，目前的實作評量、檔案評量，需要一些專業判斷。若教師未能發展實作評量的評分規準，一旦有人質疑給分的標準，有時候教師難以細說清楚，且不太確定所給的分數是否非常科學、公平與客觀時，會導致對新評量方法的實踐有所遲疑。

4. 激勵與仲裁的價值擺盪

評量本來就存有多重功能，此次的成績評量準則特別強調獎勵和輔導的原則。過去教師習慣於評量是打成績、排名次，目前他們對於評量是在仲裁或在激勵之間有所猶豫。教師努力地評量，力求公平，但最終的情況是造成有部分學生認為他們並不好，評量造成學生對自己的不滿意，連帶也對教師產生不滿。得到較低分的學生認為教師對他有偏見，因而疏遠教師，甚至忿忿不平。評分仲介著師生關係，學生得到成績的高低，影響著師生關係的密切。

一位教師說出了評分的難處：「對於成績評量，有時候我會想做個所謂“營養老師”，讓大家皆大歡喜，何樂而不為？基於鼓勵學生，讓他們高興，至少師生關係和樂不緊張，不會有紛爭。但是我總不能昧著良心，睜眼說瞎話，站在維持品質觀點，總是要公允、客觀且精確地區分？但標準為何？嚴格寬鬆如何拿捏？如果你仔

主題文章

細地去衡量他們的學習成果高低，學生會認為你對分數斤斤計較，嚴苛，因而討厭你？我好像是決定學生生殺大權的判官，是菩薩、是閻王似乎一念間？這種矛盾經常交織著？」（訪談 3029103）

成績是激勵學生不可或缺的，分數到底要如何公允的拿捏，要嚴，或要寬？評量在激勵，或在把關？評量乃必要之惡，還是善？評量是敵，或是友？一些教師在評量功能的不同價值中擺盪，感到困擾。

5.新評量舊教學的不搭窘境

在目前的評量界裡，正處於派典轉移的變化之中，從心理計量測驗到教育評量模式，從考試文化轉向評量文化(Gipps, 1994)。對今日適逢評量派典轉移的教師而言，其須懷抱新的派典、新的基本假設和新的技術好像是無庸置疑的。當舊的方法已不適用時，教師就須使用不同的策略來評量學生的學習。教師必須持續地關心如何讓每位學生皆有機會學習，並藉著評量來使學習更有意義 (Mabry,1999)。但是，當從傳統走向創新時，難免會有許多的困惑與負荷，因為當我們處於目前派典下來思想、感受和表達一個截然不同的派典時，當覺得完全無法理解。縱然有機會偶然領悟新的不同派典的概念，我們仍然很難將其完全轉換及應用到目前的教學與評量上。以檔案評量為例，有時候發現教師採檔案評量，但裡面的內容全是考試的測驗卷，也沒有多元的學習資料，沒有學生的省思，沒有同儕的對話，也沒有教師的回饋，教師的教學仍是傳統的教學。Stiggins(2001)指出，要發展出良好的另類評量，需要教師們有高度的教育專業知能，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。

從上述的分析，瞭解到新評量的實施在舊文化的桎梏下，顯得難以伸展。而教師處在新舊評量轉換之際，在不同的價值體系間存有衝突與混淆，使得對新評量的執行有所遲疑與保留。也因此，重新地思考教學評量的本質與目的是需要之路。

三、政治觀點

改革本來就是非常複雜、崎嶇的過程，充滿了權力爭鬥、競逐和妥協的痕跡 (歐用生, 2002)。House (1981) 指出，所有的教育革新涉及到權力、權威和不同利益團體間的競爭運作和協商談判。在評量改革的場域中，評量改革的政治觀著眼於瞭解各種不同團體對評量期望的協商角力，還有評量歷程中，有關人際互動的權力關係和控制行為之探討。

(一)不同評量期望角力下的左右為難

跨社群協商評量的期望，是評量改革的一大政治挑戰。因為每個人的立場不同，對評量的期望也有不同，教育政策製定者、教師、家長和大眾之間力量的拉扯，使得評量改革如同是個錯亂的活動 (Hargreaves, Earl, Moore, & Manning,2001)。由於各方觀點不一，要求改革的人又多，學校更欠缺明確的作法，以及家長和社會大眾認知的相關資訊都不同，讓評量成了一項複雜難解的教育習題。當政策制定者和廣泛大眾無法取得共識時，很難期待老師能順暢地運用所倡導的評量方法。

如果我們仔細的觀察與思索，就會發現評量政策的執行，有許多的矛盾就存在於政策本身之中。這些兩難的理由是數不盡的，有時候必須考量教育政策、傳統學校結構、課程統一進度、定期考試壓力、家長壓力及有限的資源等等。以目前的狀況來說，課程改革要教師能自主的設計統

整課程，但教師的現實困難是如何多元自主的教學，又要同時達到共同標準的績效要求。教育改革的績效評鑑常用傳統的成績報告來說明學生的學習進展和成就，這似乎與學習過程的彈性多元相抵觸。教師對於過程要多元，結果要統一，著實有些為難。「想採用評量改革的多元方式，但有統一考試的壓力？」「要彈性自主的多元評量，培養學生帶著走的能力，但最後的學力測驗，卻用選擇題型式的紙筆測驗來衡量？」一些教師被這種矛盾感到無奈與激動。在評量改革過程中，教師被要求採用他們並不全然了解或贊同的新作法，有些老師也會擔心，一旦學生成績不理想，甚至可能影響到老師的未來前途。因為有些學校將學生考試成績列為教師評鑑的項目之一。就在這些微政治力量的拉扯中，有時也讓教師感到不知該何去何從。

目前，九年一貫課程實施了，教室裡的教學評量賦予教師開放的空間，採用多元的評量方式來進行評量，教室裡，教師所需要的評量訊息是學生有否達到教學目標，那裡沒有達到標準、什麼學習問題需要解決，並培養學生具有自我反思與自我評量的能力。然而，在國中階段，「基本學力測驗」的動向仍然主導著所有家長和社會大眾的關注，他們所關心的評量訊息是學生或學校之間的學業成就高低比較。教室裡允許個別差異、著重自我發展，但社會上卻有一定的標準，強調競爭比較，教室裡和教室外的要求並不同調，也讓教師產生為難。

誠如吳毓瑩與吳麗君（2002）的研究中指稱：「評量改革的外在脈絡比評量改革本身更為重要，它支配並解釋評量改革」。在國小階段，學生的成績好壞，與他的未來升學前途，沒有那麼直接的關連緊要，因此各界尚能以平常心泰然處之。在國中，學生成績的評量攸關他們未來的前途

出路（雖然目前高中入學並未採計國中在校成績），而學生面對評量仍然戰戰兢兢、如臨大敵，因為他們都要面臨未來的大考－基本學力測驗。大多數的家長多關注這高風險的評量，因為不僅關係到自己孩子的前景，更影響著他們望子成龍、望女成鳳的期望，也牽動到家庭聲譽。而學校更是以之為辦學績效的重要指標之一。國中教師要施展多元評量，除了面對技術上的挑戰，文化面與政治面的箝制，更是讓他們感到腹背受敵，進退兩難！

（二）評量權力重新分配的多頭馬車

過去老師使用權力去判斷和分類學生，其標準和方法是不公開的。現今的評量主張評量標準要透明化、公開化，讓學生明確知道評量的標準，並合作共同發展，且力主評量人員的多元化。除教師評量外，亦輔以同儕評量、自我評量與家長評量(Lazear,1999；Mabry,1999)。評量權力的下放，使班級的權力得以重新分配，從學生到學生、從學生到老師、家長到老師，而評量權力的轉移，也使得學生、家長和老師之間，有更多的微政治關係。

然而，評分大有學問。就同儕評量部份，其立意在於讓同學間可以相互觀摩、學習，相互成長；然而也有反對的聲浪，反對者認為，不甚公平，除了質疑學生的專業判斷能力外，他們認為，同儕之間的關係經常左右著評分的高低，因為關係好者，則相互受惠（好朋友互相提攜）；關係不佳者，則會挾怨報復（有過節者要你好看）。另外，關於小組成績的評定，也有質疑，因為小組合作學習經常有人搭便車，賣力者有時無法突顯，搭便車者又坐享其成，不甚公平。導致小組有紛爭與怨言。在自我評量部份，贊同者認為讓學生自己評分是不錯的想法，若學生能誠實地檢

主題文章

討，能夠讓他們更了解自己的問題和所擅長與不足的地方。但反對者認為每位學生對自己的要求與標準不同，難以公平判斷。加上，學生會顧及自己的成績，當然會為自己打高的分數。還有，家長評量部份，家長也關心自己的孩子成績，誰會將自己的孩子打低分呢？又，當學生在自我評量中反映出教師教學不當，此時，也可能引發師生間的緊張關係，帶來一些冒險。

新評量倡導評量者的多元化，讓各方的聲音得以呈現，以便評量更客觀、公允與周延。但是，當所評是自己的成績，攸關自己的利益，那又要如何的公允誠實的評量呢？當誠實反應所帶來的負面後果，又將如何處理？評量人員的多元化固然帶來更多層面的關照，但評量人員間的權力角力，也在其中蘊釀發酵。

(三) 賦權增能策略質變為行為控制黑

手

目前評量主張透過不斷持續的評量，包括使用自我評量、同儕評量和檔案評量，讓老師更瞭解學生。在此過程中賦予學生參與評量權力，亦增進教師的專業判斷能力。然而，這樣的做為，也可能使學生隨時遭到監督，讓教師使用這些資料來進行分類與排名。特別是評量情意部份，老師似乎變相地運用這些評量方法，來對學生所做的每件事進行監控，做無止境的判斷。情意評量儼然成為一隻看不見的黑手，操控學生的行為。Hargreaves 等人(2001)就指出：情意評量本身要比其它領域（認知、技能）的評量有更多的主觀和偏見判斷，而且這部份的評量特別容易侵犯學生的隱私。評量學生的情意和行為的方式也加強了順從與控制，這是教育領域中最受爭議的，它觸及的並非只有心智而已，而

是人的心靈，其影響深遠，不得不慎。

將鏡頭拉近一點看情意評量，大部份的教師多將情意評量的焦點，聚在學生順從學校規矩的行為，例如：準時上下課、上課認真聽講，而不是愛發問、質疑、克服困難、自我肯定或創造力。因此，有許多的危險潛藏在情意評量中，教師不經意地使用評量來懲罰、監控或控制行為，學生也慣常接受這種順從教室行為，在教室裡就常聽到這樣的語言：某某某，你再講話，我就給你扣分要評量學業成就以外的東西，教師通常會面臨許多的挑戰。在進入評量學生的情意領域，教師總是習慣於評量學生的努力程度。在台北縣政府公佈的台北縣國民中小學成績評量要點中，就指出要對學生的努力程度與學業成就等第並列，將努力程度分為甲、乙、丙三級（台北縣政府，2001）。有教師質疑，努力程度的評量有賴教師的觀察、同儕評量與自我評量，它充滿許多的自由心證。許多教師描述他們評量情意的方式，似乎等同於使用評量來控制學生的行為（執行對學生所做每件事的行為監督，做為持續判斷的參考）。教師評量情意的項目一般有：上課參與、出缺席、課堂上的專心、對科目的態度、上課的肢體語言、作業是否準時完成、同組成員的評論等等，最後就這些內容做一個混合評語，給學生打成績，例如：理解概念佳，批判能力強，但卻沒有團隊合作的能力。所以從甲的等第中酌予扣分。另外，也有教師採用同儕評量的方式進行，要同學寫出分工明細表，並給成員回饋，它有其功能，不過卻需要相當的誠實和有價值的回饋。可惜的是，同儕評量也可能淪為打小報告、告密之用。這些原本要賦權增能的評量策略，反而可能變調為新的行為控制策略。

(四) 評量意識型態的極化戰爭

此外，目前另類評量的推動方興未艾，瞬間一躍成為教學評量的要角，這些評量新秀承載了大幅教改的能量與希望。相對地，傳統評量，特別是強調紙筆選擇題式的測驗，則遭到重大的批判，它好似教育的萬惡之源，人人得以起而終結之（吳鐵雄、洪碧霞，2000）。評量改革在倡導新的評量理念時，總是一味宣導新評量理念的好，並對照舊評量的壞；不時突顯新評量的善，同時挑起舊評量的惡，刻意加劇兩造的對立。結果引發兩造倡導者的彼此反對，互揭對方的瘡疤，甚至造成評量意識型態的極化爭執，如同正義與邪惡兩方的戰爭。這種危機，似乎超過另類評量的實施範圍了（Hargreaves, Earl, Moore, & Manning, 2001）。不過，令人關注的是，新評量的強力主張，是否也淪為更左派的權力灌輸呢？而在極左極右的擺盪後，是否又會是歷史的別無選擇呢？也就是評量改革幫助教師和學生走出這個“牢籠”，但會不會又帶他們進入另一個“牢籠”呢？這也是值得反思的。

總之，政治觀點將焦點放在評量歷程中的權力關係和控制行為。原本是採取分享增能、開放協商的新評量措施，在競爭群體的權力遊戲下，將使得評量更難以捉摸，也可能造成更左派的權力形式灌輸，不得不讓自己在賣力推動評量改革時，慎思與明辨自己的篤行呀！。

四、後現代觀點

後現代社會的特徵是快速、多元、複雜和不確定。其中，後現代的幾個特徵及其弔詭的爭議，也值得我們來反思這波評量改革可能的契機與危機。

（一）確定性消失的自由與無所適從

後現代社會引發科學知識和信仰體系的不確定性。由專業分化日益嚴重，專家

們之間相互競爭，觀點各不相同，知識的更新也越來越快，這種情形使人們擺脫了對某種知識或迷信的依賴，亦使人們感到無所適從（張文軍，2000）。再來，高科技的通訊交流技術，壓縮了空間和時間，使人們能知道世界發生的變化，也影響了原來的知識基礎的穩定性。加上，電子科技產生了大量充沛的資訊，它挑戰著“到底什麼內容最值得教”的假設。在此時，快速成長的多元文化、後殖民主義等議題，挑戰著長期學術取向所主導的霸權課程，目前學校教育強調學校本位、具有地方色彩的鄉土教育課程，也使得教師處於反對統一標準課程內容的不確定中（Hargreaves, Earl, & Schmidt, 2002）。過去學校裡，有一套清楚界定的知識體系，學生只要能精熟或複製內容即可，現在的課程內容其確定性消失了，伴隨而來的，評量什麼和如何評量的確定性被粉碎了，有些教師反而不清楚要評量什麼，以及如何評量，甚至不確定如何做才是對學生是最好的，他們對於自己的評量效度感到懷疑。一位國小教師在訪談中，說道：「過去按照課本上課，評量的內容很明確，也很少爭議。現在課程改變了，內容由教師自編設計，雖然打破知識權威的壟斷，但是，我們自行選擇的內容，是否正確？是否對學生就較好？我難以確定！也因此，對於要評量什麼，心裡也不確定……雖然我認真的準備，但是很心虛！」（訪談 1059011）。教師對於確定性的消失，雖然感到自由，卻也有無所適從的感受。

（二）高科技模擬的真實與虛幻

後現代社會的另一個顯著特徵是：充斥著高科技生產出來的影像－電視、錄影、傳真、虛擬實境等等。事情的呈現方式遠比事情本身重要。年輕一代生活在各種影像中，書本、學校生活對他們的影響不敵各種多彩多姿的視聽媒介，教育必須與充滿了各種影像的外部世界競爭，以贏

主題文章

得學生的注意力。同時，還需要引進先進技術，實行教學改革。影像的繁榮使人們的生活更為激動人心，參與性和娛樂性更強了，但對產生嚴肅持久的道德話語、深思熟慮的價值判斷卻起了阻礙作用（張文軍，2000）。高科技模擬世界為學校教育提供了更多的素材，製造了逼真的模擬實境，但也增加了真偽判斷的難度。在教育中，電腦科技、e-mail 和網際網路使得學生、家長或學校有更多的學習資源可運用，但也可能受到電腦控制或者欺騙。電腦網路世界中的真實與虛幻，令人捉摸不清。電腦科技時代給予學生從多元管道來撿選立即的訊息，透過按滑鼠下載圖片來編輯，在電子時代裡，很難辨別什麼是學生自己的，還有確定這些資料的來源是真實的、正確的。評量的“真實性”在後現代社會中遭到了質疑與挑戰（Hargreaves, Earl, & Schmidt, 2002；Terwilliger, 1997）。

（三）多元繽紛的實質與表象

多元化評量提供學生多元學習的表徵，給予最大的聲音和視野，來呈現多元活動和成就，在實作活動或學習檔案中，可透過書寫、口頭、數字、視覺、科技、戲劇媒體、混合型式來呈現他們的成就，讓學生從多元的視角來看成就，允許能力的多元複雜，也使得教師對學生有充分的了解，這是多元化評量的契機。然而多元化評量也潛藏著一些危機，以檔案評量為例，檔案的內容可能充斥造假或模擬的資料，並非實質的成就（Hargreaves, Earl, & Schmidt, 2002）。多樣絢麗的色彩圖片，使人注重表面的美，而忽略內容的質與量。師生的價值可能縮減為物質、影像、彩色封面、圖片，它們可能掩蓋了內容與分析。這種現象，使得檔案評量可能平庸化，減低了學習的本質，扁低至表面的外表和偽裝設計的真實，處在一種不嚴謹，感覺舒

服的世界。

（四）無邊際自我的增能與浮誇

在後現社會中自我的無邊際，有其建設性的一面是，人們對專家和權力的依賴減輕，人們能夠掌握自己的命運，作出更多有意義的決定。其不利的一面是：它容易導致自我陶醉和自我中心傾向。強調個人，忽視群體。人們認為沒有什麼做不成的事。而由於它拒絕界定自我的交往情境，不能認清機會和侷限，自我陶醉和無邊際的特質使個人的力量不僅未能得到加強，反而削弱了（張文軍，2000）。目前的評量方式，強調自我評量，其原本用意在促使學生做自我探索，自我反省。但 Hargreaves, Earl, 和 Schmidt (2002)指出，自我評量也可能導致自戀、自大，自我中心的人格。在自戀的人格中，個體可能持續的自我反射成為無止境的自我誇大，過度評估自己的能力。這種缺乏誠實的自我批判，可能導致無法誠實的考量他們本身成就的品質與價值。反而無法賦權增能，且可能削減了自我的競爭力。

參、慎行評量改革的價值追求，回歸教育本質

本文從技術、文化、政治與後現代等四個視角，來分析目前教學評量改革的困境。四種不同的觀點各有其著眼點，呈現出評量改革的不同問題與挑戰，也反映出評量改革的複雜度。技術觀點在檢視發展新評量技術的策略、組織和結構等問題。文化觀點則關注目前倡導的新式評量理念與方式，是如何地被統整到學校與社會的文化脈絡中，其挑戰是有關評量文化的重塑。政治觀點則著眼於各種不同團體對評量期望的協商和權力角力，還有評量歷程中，有關人際互動的權力關係和控制行為。後現代觀點在探討後現代的幾個特徵

及其弔詭的爭議，例如：1.確定性消失的自由與無所適從；2.高科技模擬的真實與虛幻；3.多元繽紛的實質與表象；4.無邊際自我的增能與浮誇等，帶給教學評量可能的契機與危機。

每一種觀點都指出，目前的評量都較過去需要更精密的技術、有更多的賦權、和更多的合作與反省。每一種觀點也注意到評量改革可能出現的危機，例如是否減少了深度的學習和班級的關懷，是否導致全面的監督，或者質變為自戀式的自我沉醉等。評量改革選擇正面的宣導，不應只是讓我們產生無窮盡的希望。誠如吳麗君(2000)所說：每個國家的評量政策，都可以視為各項競爭價值間暫時性的妥協結果，在這個暫時性的平衡上欲增加某一面

向的利益，則必須忍受另一方面的優勢遭到剝奪或加深一些既有的困境，因為任何一個評量系統都是在交互損益的情境中進行抉擇所建構成的。而種種的兩難困境都考驗著評量改革推動者、執行者與接受者的智慧。本文中所提出的困境分析，並不在詆譏評量改革的努力，而是希望透過從不同視角來省視此波評量改革可能的契機與危機，或許可以尋繹到轉機，以俾更審慎地執行對教育價值的追求，並促使在推動評量改革時，能更嚴謹、均等和持續的發展，得以超越形式之爭，回歸教育本質。

主題文章

參考書目

- 台北縣政府 (2001): **國民中小學學生成績評量說明手冊**。台北: 台北縣政府。
- 東風 (2001): **不要荒腔走板**。未出版手稿。
- 吳麗君 (2000): 交互損益見真章—談英國國家課程評量的來時路。**課程與教學季刊**, 3(1), 79~94。
- 吳毓瑩、吳麗君 (2002): 從比較教育的取向討論測驗評量在教育銜接中之意涵—一個可能的研究途徑。**國立台北師範學院學報**, 15, 313~336。
- 吳鐵雄、洪碧霞 (2000): 教學評量質變的反思與建言。**測驗與輔導**, 159, 3323-3324。
- 林怡呈 (2002): **教學歷程中多元評量之迷思、神話、與活化 - 以一個國小教師為例**。國立台北師範學院課程與教學研究所碩士論文。
- 教育部 (2001): **國民中小學學生成績評量準則**。台北: 教育部。
- 陳明印 (2001): 九年一貫課程國民中小學學生成績評量準則探究。**研習資訊**, 18(2), 25-44。
- 郭俊賢、陳淑惠譯 (2000): **落實多元智慧教學評量** (D. Lazear, 原著, 1999 年出版)。台北: 遠流。
- 張文軍 (2000): **後現代教育**。台北: 揚智。
- 張郁雯 (2001): 邁向教學評量專業化之路。**台灣教育**, 604, 33-39。
- 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、陳淑惠譯: **落實多元智慧教學評量, 專文推薦**。
- 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。
- 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。
- 鄭富森 (2001a): 教學評量: 教育改革成敗的關鍵。**台灣教育**, 604, 2-14。
- 鄭富森 (2001b): 多元化評量的十大難題。**測驗與輔導**, 164, 3452-3457。

- 鄭燕祥 (2001)：學校效能及校本管理發展的機制。台北：心理。
- 歐用生 (2002)：披著羊皮的狼？ - 九年一貫課程改革的深度思考。載於中華民國課程與教學學會主編：「新世紀教育工程：九年一貫課程再造」(頁 3-24)。台北：揚智。
- 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。
- Anderson, R. S. (1998). Why talk about different ways to grade ? The shift from traditional assessment to alternative assessment. **New Directions for Teaching and Learning**, 74,5-16.
- Black, P. J. (1998). **Testing :Friend or foe? The theory and practice of assessment and testing**. London : The Falmer Press.
- Gipps, C. V. (1994). **Beyond testing : Towards a theory of educational assessment**. London : The Falmer Press.
- Hargreaves, A. , Earl, L., Moore, S. & Manning, S. (2001). **Learning to change : Teaching beyond subjects and standards**. California : Jossey-Bass Inc.
- Hargreaves, A., Earl, L. & Schmidt, M. (2002). Perspectives on alternative assessment reform. **American Educational Research Journal**, 39(1),69-95.
- House, E. (1981). Three perspectives on innovation: Technological, political, and cultural. In R. Lehming, & M. Kane(Eds.), **Improving schools : Using what we know**(pp.17-41). Beverly Hills: Sage Publications.
- Mabry, L. (1999). **Portfolios plus : A critical guide to alternative assessment**. California : Corwin Press, Inc.
- McMillan, J. H. (1997). **Classroom assessment : Principles and practice for effective instruction**. Boston: Allyn and Bacon.
- Reineke, R. A.(1998). **Challenging the mind, touching the heart : Best assessment practices**. Thousand Oaks, California : Corwin Press, Inc.
- Shepard, L . (2000). The role of assessment in a learning culture. **Educational Researcher**, 29(7), 4-14.
- Stiggins, R. J. (1997). **Student-centered classroom assessment(2nd ed.)**. New York : Merrill.

Stiggins, R. J. (2001).The unfulfilled promise of classroom assessment. **Educational Measurement : Issues and Practice**,20(3), 5-15.

Terwilliger, J. (1997). Semantics, psychometrics, and assessment reform : A close look at "authentic" assessment. **Educational Researcher**, 26 (8), 24-27.

Conundrum of Classroom Assessment Reform: Technological, Cultural, Political, and Postmodern Perspectives

Wen-Tzu Chiang

Educational reform is indeed pervasive right now. The grade 1-9 curriculum has been implementing in 2001. In order to achieve the goals of this curriculum, it needs to extend to the reform of instruction and assessment simultaneously. Classroom assessment is a multifaceted phenomenon, and many issues are at stake in reforming it. Arresting ardor for classroom assessment reform, we need step back and reflect on it critically and carefully. This article examines classroom assessment reform from four perspectives: technological, cultural, political, and postmodern. Each perspective highlights different issues and problems in the phenomenon of classroom assessment. The technological perspective focuses on issues of strategy ,organization, and structure in developing new assessment techniques. The cultural perspective examines how new assessment are interpreted and integrated into the social and cultural context of schools. The political perspective views assessment issues as being embedded in and resulting from the dynamics of power and control in human interaction. Last, the postmodern perspective is based on the view that in today's complex and uncertain world, authentic assessments is fundamentally questionable. By drawing thoughtfully and critically form all four perspectives, we should and can exercise greater vigilance in pursuit of educational values that well move new assessment reform in educationally rigorous, equitable, and sustainable directions.

Key Words : Classroom Assessment 、 Assessment Reform 、 Multiple Perspectives

Assistant Professor, The Center for Teacher Education, Shis Hsin University in Taipei.