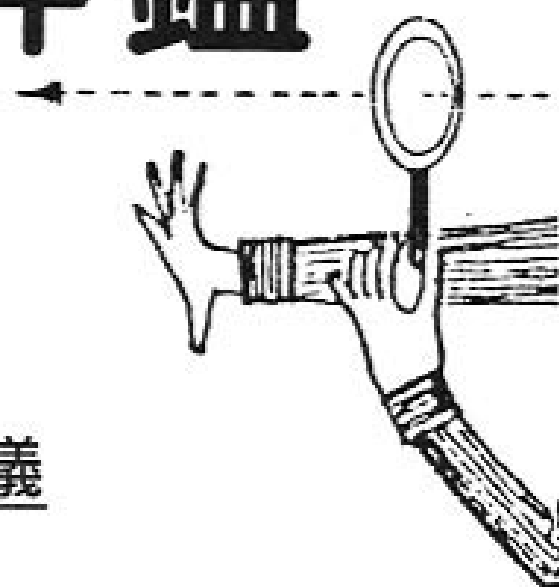


教育評鑑

的新取向



周珮儀

壹、前言

一九七〇年代在教育評鑑的領域興起了一種新的評鑑方法——質的評鑑，並與傳統量化評鑑的方式展開激烈論戰，前者源自人文學科的派典，強調整體的、質的資料及詮釋的方法，後者植基於自然科學的派典，強調實證、量化的觀察，以便利數理工具的分析，研究者的任務在建立因果關係，並加以解釋（張鈿富，民83）。

在這場論戰之中，除了消極的批判傳統科學式評鑑的缺失，也積極的由人文和藝術的觀點提出質的評鑑取向，將其理論與方法轉化到教育的領域。

目前我國的教育評鑑仍然受到量化典範的支配，量化典範固然有其優點，也有其限制，必須加以批判反省。本文將分析量化典範的形成背景和重要問題。並簡介質的評鑑的形成背景、因素、特徵和優點，期能為教育評鑑領域提供一種可行的變通方案。

貳、傳統科學式評鑑

傳統的評鑑方法是以自然科學為基礎的科技模式，誠如 Habermas 所言，自然科學只能展現較低層次的理性，亦即工具理性，在既定目的之下，工具理性只考慮手段的有效性，而忽略目的的本身所蘊涵的價值及意義問題（沈清松，民74）。在工具理性支配的科技模式下，評鑑使用量化方法和系統化的程序來建立可預測，可控制的共同法則。然而，傳統科學式的評鑑是否能幫助我們了解真實的教育情境？教育必然要關懷意義和價值的問題，而傳統評鑑方式卻無法處理這些問題。以下將追溯傳統量化評鑑的發展背景及其形成因素，並進一步對這種方法所產生的種種問題提出檢討。

一、形成背景

探討目前的教育評鑑必須追溯其歷史脈絡，因為在教育評鑑發展的初期，有一

些重大的事件與思潮，塑造了評鑑的基本形式與概念，影響至為深遠。1904年，Thorndike出版「心理及社會測量理論」(Introduction to the Theory of Mental and Social Measurement)一書，此後量表、測驗等測量工具，如雨後春筍般地出現，廣泛的用來了解教育現象、解決教育問題。1918年他又於著作中主張：「凡存在的東西，必有數量」；「具有數量的東西，必可測量。」(黃政傑，民79)他的主張，增強了測驗工作者的信心。

1911年，科學管理運動之父Taylor出版「科學管理的原則」(The Principles of Scientific Management)一書後，科學管理理論成為工業界和教育界的主要理論，它強調系統化、標準化和效率等三個重點，其中尤以效率的要求最為重要(江啟昱，民82)。

1932-1940年間，Tyler受託領導著名的「八年研究」(Eight-Year Study)，這個研究對教育評鑑產生很大的影響：

1. 研究方法專注於陳述清晰的目標，並把評鑑界定在確定目標是否達成上，這些目標並進一步發展為行為目標，提供課程發展和評鑑發展所需要的依據。
2. 強調直接測量學習的成果，而不是間接去測量教學品質、圖書藏量、資料範圍，與社區參與等投入，因而避免專業判斷或認可制度的主觀成份。
3. 將預期目標與實際結果作內在比較，不必如以往從事高成本的實驗組和控制組的高成本比較。
4. 目標導向的評鑑不必太注重學生分

數之間的信度差異，這樣的測量所涵蓋的結果變項反而比常模參照測驗來得廣(江啟昱，民82；黃光雄，民78；黃政傑，民79)。

綜合上述可知，早期教育評鑑的領域，將教育結合了效率運動(efficiency movement)，相信經由科學發展的課程和科學化的管理，能夠確保學校產出想要的結果，而對成就的測量是唯一能客觀地決定學校是否有生產力的方式(Eisner, 1985b)。這種評鑑觀念，深植人心。因此，典型的報導學校教育效能方法有兩種：一是評分制度(grade system)，二是標準化測驗(standardized testing)，而判斷學校或個人成功與否，在於他們是否獲得了好的分數。

造成這種情形，除了歷史因素的影響，還有以下幾種重要原因：

1. 為了滿足對於評鑑精準性的要求，必須將學生的表現量化，接著轉換為等級。當複雜的學習形式被化約為一個單一的分數或字母時，似乎就達到了一種精準的客觀性。
2. 為了向家長和社會大眾溝通，必須將學生在學校的表現化為一個簡明而準確的指標，卻未曾深思其是否會誤導。對於把評鑑視為連續性、質的觀點，往往會使家長感到複雜、不確定和沒有安全感。
3. 為了在競爭性的教育市場中作比較，必需產生一個一致性的等級。如果學生想要向上教育流動，他們必須獲得好的等級，這些等級所應達到的規準，在學校教育的階層中產生選擇的功能，增強評分系統的

使用。

4. 在標準化測驗和評分制度之外，未曾思索其他的變通方案（Eisner，1985b）。

這種將評鑑視為測驗和評分的科技模式，表面上看起來，雖然能夠滿足精準、客觀、簡明、易於了解、溝通和比較的需求，但是它卻存在著一些重大的缺失，往往會干擾更重要的教育目標。

二、問題檢討

(一)認識論上的問題：

科技模式強調的是「預測」與「控制」，以自然科學的認識論為基礎，經由技術的應用模式，獲得相當有限的結果。但是教育現象十分複雜，實際上是無法完全控制與預測的，而且社會現象的法則和自然現象是不一樣的。當科學的評鑑方法流行時，一切不屬於科學的完全被拋棄，教育評鑑乃成為片面的，嚴重限制了對教育現象的認知。

(二)尋求共同法則，忽略個別特質。

科學的假設和研究方式主要在尋找法則（laws），或是像法則的推論（law-like generalization），傾向於將特殊情境的特質作為「工具」來處理。在尋求共同的特質時，個殊的特質經常被視為雜訊而被忽略，無法包括在共同範疇之內的將會被去除。

(三)將特質化約為數量，不能說明複雜的教育現象。

將複雜的特質量化成為數字的差異，

會扭曲真實的情況。為了追求數字，凡是不易量化的都被拋棄掉。例如：情意目標和高層次的認知目標，在教育上很重要，教育的長期效果也比短期效果值得重視，但它們都不易測量，所以大多數的評鑑者就不會設法去取得這些資料。然而，在教育評鑑上更重要的往往並不是某種環境的量有多少，而是它究竟是什麼樣子，具有什麼意義。

(四)重未來而輕現在，重結果而輕過程。

科技的假設認為教育在達成特定的、可測量的目標，因此，不只要為每個學生設定特定的行為目標，還要對教學實施科學化的測驗程序，藉以確保目標的達成。當目標主導教學活動時，往往會為了未來犧牲現在，忽略教學活動過程中師生的互動和他們所作的決定與表現，未蒐集與分析影響結果的條件、過程和互動因素，因此，對學習結果的改進，幫助不大。

(五)強調主體和客體的分離，不能說明個人主動建構知識的事實。

科技模式強調客觀，他們所謂的客觀是指對事物描述和解釋的表徵形式，必須和所欲表徵的真實情況相符應。因此，研究的結果和真實的現象雖然是同形的，然而卻又是彼此分離獨立的。所以在這種研究中，常常以一些修辭的設計來說服讀者：研究者下結論或解釋資料時，沒有加入個人的觀點，例如：以「我們」、「研究者」等用語來代替「我」；而暗示、隱喻、類推、洞見等認知方式被視為是不客觀、不重要的。

然而，有許多學者認為：我們所見、所解釋的世界是我們自己建構的；對這個建構的世界的表達，受到我們的表徵形式和使用這些形式的技巧所影響，暗示、隱喻、想像、類推、洞見都是有用的表徵技巧和方法，缺乏這些技巧與方法，將會使得人類的文化資產與學術成果流失大半！

(六)學習變成標準化的模式。

因為測驗可以客觀地測量出學生在共同目標上的表現成就，所以無可避免地控制了課程的內容和形式，影響教學方式和教師在教室中建立的優先順序。為了尋求控制人類行為的共同法則，常會為行為設定一套共同的目標，再加上為了要求明確，常常以量化的方式來操作這些目標，因此，就會使用標準化的測驗。標準化測驗要求一致性，所有的學生所面對的工作和所應達成的目標都是相同的。即使是「個別化」，也被視為只是速率和步調的差異，而不是目標、內容和表徵模式的不同。此外，為了確保測驗程序上的客觀，常常採用有標準答案、只須機械式簡單反應的題目，例如：選擇題。這些題目去除學生下判斷的可能性，使他們沒有機會去做預先決定的正確答案之外的反應，結果會遺漏一些教學和學習過程中重要的特質（Eisner，1985a，1991a）例如：藝術作品所追求的最重要的成果之一——培養「有生產力的異質性」，在科技導向的教育實施中就無法達成。

(七)實驗處理的問題：

1. 方法決定問題，手段決定目的。將教育問題化約為適合研究派典的形

式，而不去找出適合這個問題的研究和評鑑程序。例如：相關、變異數分析、隨機選擇和分派，這些觀念的影響力很大，因此我們在構思問題時，常常受這些觀念的侷限，而忽略一些新的問題和解決方式。

2. 很少去描述和評估影響學生行為的環境因素。
3. 評鑑未能分辨學生未來能做什麼和現在能做什麼。在教育上不只要關心立即的效果，也要關心長期的效果。測驗的表現往往被視為是未來的指標，然而事實上對於依變項的測量只能顯示在實驗期間實驗處理是否有效，無法告訴我們這種效果能夠持續多久。
4. 實驗處理的短暫性：實驗處理的時間越短越容易達到嚴密的控制，控制可減少混淆，提高解釋的能力，然而，矛盾的是：實驗控制越多就越不易推論到實際教室中（黃政傑，民79；Eisner，1985a，1991a）。

參、質的評鑑方式

一、形成背景

1968年，Jackson發表了「教室中的生活」（life in Classroom）一書，他認為教學是一種非常複雜的過程，充滿機遇的事件，教師和學生都不能預測即將發生的行為，因此不能以輸入（input）、產出（output）等名詞來表示，也不是量化的結果所能解決的。1970年，他又於「學校教育的結果」（The Consequence

of Schooling) 一書中主張：「教育領域不再需要更好的測驗或更精密的設計，而需要以新的方法來考察一些舊的問題，例如：學校教育中那些非學術性的結果。」

1972年，「康橋會議」(Cambridge Conference) 檢討傳統科學式的評鑑方法，認為它們不能闡明複雜的教育現象，而且它們所提供的資料無助於大眾與專業人員教育決策上的需要，因此，提出了種種非傳統式的評鑑模式。1975年，舉行了「第二次康橋會議」，以「個案研究再探」(Rethinking Case Study) 為主題，具體地探討非傳統式評鑑模式的內涵(黃政傑，民79；歐用生，民78)。

在這種氣氛的刺激下，Parlett 和 Hamilton 在1972年提出了「闡明式的評鑑」(illuminative evaluation)。Stake 在1975年提出了「感應式的評鑑」(responsive evaluation)。1978年，Parlett、Jenkins、King、Macdonald 和 Hamilton 共同發表了「超越數字遊戲」(Beyond the Numbers Game) 一書；而 Willis 也主編了「質的評鑑」(Qualitative Evaluation) 一書。

因此，在一九七〇年代以後，這些思潮逐漸匯流起來，成為一股「質的評鑑」的勢力。就大學中從事質的研究的學者而言，對於評鑑方式侷限於量化的描述，以及自然科學派典在教育研究佔據主導地位的情形，感到不滿。他們反對「只有經由量化和科學的方法才能獲得客觀」的假設，主張讓問題決定方法，而非方法決定問題(Eisner, 1985a)。他們認為在很多人類活動的重要領域，存在許多無法量化或以科學來分解的特質，事實上，這些

無法測量的特質經常是最重要的；此外，他們也強調事件的情境因素和歷史因素在理解時的重要性。對一般教育人員而言，傳統的標準化成就測驗也不能滿足所有教育者的資訊需求，因為它們不能讓他們了解教師實際所教和學生實際所學的情況。因此，在各個教育階層都浮現了對質的評鑑的需求與探索。

另外，對「潛在課程」研究文獻的劇增也加深這種趨勢。由於潛在課程與價值、態度、動機等有關，因此對學生學習所產生的意義的理解，遠比事實的發現更重要。此外對這些經驗學習過程的了解（而不只是結果的認識），量化的研究方法常力有所不逮。所以「質」的研究，可用來處理潛在課程的問題(陳伯璋，民76)。

值得注意的是：近年來質的研究和評鑑雖然比較容易被接受，它們卻經常被定義為一種教育的俗民誌。然而，Eisner 認為任何社會科學(社會學、心理學、政治學) 的學者都能從事本質上屬於質的研究，人類學對質的研究沒有霸權，質的研究也不必然只能以人類學的方法來定義。此外，質的研究不一定是科學的，例如：教育鑑賞和批評就是植基於藝術。

二、形成因素

(一)政治方面

因為政治動機而採用質的方法的人，經常傾向社會主義或馬克斯主義的哲學觀點。有些人認為學校是一種有組織的陰謀，使兒童變得依賴、無知和容忍愚蠢，以便他們長大成人時能夠適應現存的社會

秩序，所以學校中一些無用的特質並非失敗，而是成功。採用批判方法的人對這種觀念大加抨擊，他們認為不能完全接受不合理的社會現況的再製（reproduction），而必須做更根本的重建。他們企圖提昇公眾批判性的知覺，給予情感的衝擊，引導公眾尋求去改革學校教育的結構和目標，而質的研究方法是比較能夠符合他們目的的工具。

（二）方法論方面

許多人對傳統的評鑑方法不滿，他們認為：教育心理學所使用的實驗室研究程序，並不適合研究教室生活，而去尋求較為彈性化和自然式的研究方法。例如俗民誌就是一種較為可欲和適切的變通方案，藉由進入情境成為一個整體和沒有實驗者干預的自然觀察，能對教育生活更有效的了解，發展出更有用的理論。這種動機本質上不一定是政治的，因為他們關注的是拓展評鑑的主要觀點與合法程序，而不一定要尋求學校或社會的激進改革。

（三）認識論方面

有些人認為科學認識論和量化的方法，無法完全表達我們需要了解的教育現象，為了補充這些評鑑方法，評鑑者需要尋求教室中遍在的特質、學生在學校中所擁有的經驗和學生作品的特質。觀察這些特質需要有敏銳的觀察力、應用理論解釋觀察事項的能力、和一套用以判定觀察事項在教育上的意義的教育價值。

將這些特質傳達給他人的能力也很重要，這需要使用具有藝術性的批判方法，例如：教育批評。教育批評者要能以某種

方式創造、轉譯、描繪和展露所描述的事件，使讀者有感同身受的參與感；他們所使用的表意語言，強調情感在理解時所扮演的角色，希望幫助讀者了解學校經驗的深層意義，這和社會科學所強調的情感大不相同。

三、特徵

質的研究有不同的理論基礎和研究派典，因此在教育的應用上，有許多不同的形式，也適用於各種不同的情境，但是仍可歸納出下列共同的特徵（Eisner，1991b）：

- （一）以實地為焦點。
- （二）研究者即研究工具。
- （三）具有詮釋的特質。
- （四）使用表意的語言。
- （五）注意特殊事項。

（六）質的研究的可信度來自於它的凝聚力、洞見和工具的實用性。

四、啓示

質的評鑑者對教育情境的觀點是更複雜、更個殊的，這被某些人視為缺陷；然而，我們應該深思對教育品質只下一個簡單的結論是否會付出更大的代價，質的評鑑在教育方面給予我們帶來了下列的啟示（Eisner，1985a，1992c）：

- （一）質的評鑑對實際的教育現象和它的結果提供了複雜而完整的意象。

傳統量的評鑑將教室所進行的種種的交互作用和所達成的教育成就，變成高度簡化的指標，根據這些指標所獲得的資訊來制定教育政策，常常會忽略重要的事

項，對於需要縝密而持續考慮的教育問題，以不切實際的方式來解決，因此我們必須以質的方法補偏救弊。

(二)質的評鑑形式能拓展我們的概念和語文的模式。

質的評鑑的目的，在於使個人真實地感受到他們想要了解的實體：質的評鑑經常使用表意的語言來達到這個目的。表意的語言把語言當作藝術的媒介，表達一些基本上屬於非語文的特質，這需要充份的想像和具備引起適當感情的表徵形式，利用語文的音律、節奏、特質、隱喻、和表意的類推，而非只是單純的作描述。這種表意形式的發展，肯定了藝術的、質的表徵形式和科學的、量化的表徵形式之間具有同等的智性地位，從而拓展對學習的內容領域和學科的觀點，在決定課程和教學內容時為它們爭取合法性的地位。

(三)質的評鑑重視教室生活中動態的過程。

傳統的評鑑重視教育的產出，它所呈現的通常是總結性的成就測驗的結果，而很少去處理教育實施的過程，例如：語文的特質、師生的關係等事項。傳統的評鑑方式雖然對誘導師生爭取高的分數很有效，但是對教育活動如何進行、問題癥結何在，並無法提供解答。此外，傳統教育研究的形式有一種類似游擊隊式的趨向：在最短的時間內進入教室、蒐集資料並離開，不免使人懷疑研究者所見的事物、懷疑這種實驗相對於教室生活經常的情況的代表性、懷疑研究者能提供教師何種回饋。

相對於量的評鑑，質的評鑑特別注重教育的過程，給予教師對他們在教室中的作為具體的回饋，使教師明智地改變或繼續他們正在做的事情，並幫助大眾和教育者更了解學校中進行的事物。

質的評鑑的假設能夠補充量化和科學方式，而對於兩者都有助益。

正如 Shulman 所言，教育研究的危機在於單一的研究取向或派典，其研究結果無法涵蓋整個複雜的教育情境及不同的層面，因此融合不同取向的評鑑方法，可以使讀者們對於人類社會現象有效廣泛的了解，增加選擇行動的可能性，促進教育評鑑領域的健全發展（郭玉霞，民83）。

肆、結語

每種評鑑方式都有優點和盲點，無論是質的評鑑或是量的評鑑，都只是一種評鑑的方式，而不是唯一的評鑑方式。質的評鑑方法的發展提昇了它在學術研究領域的合法性，使得更多研究人員勇於從事這種評鑑方式，得到許多量化和科學評鑑方式所無法觸及的研究成果。對於傳統量化的評鑑方法而言，質的評鑑方法是一種有活力的補充，一種可供比較的變通參照架構，在相互對照之下，一方面彰顯出量化評鑑的目的和特質，另一方面也促使其自我反省、自我改進。

總之，雖然質的研究不是教育的萬靈丹，它們對方法和時間的需求很高，推論困難，而且結論的確證很複雜。然而，它們提供了另一種觀點和變通方案，擴展了研究的視野，充實了評鑑的方法。

（作者：本會編審）