

我國教育評鑑的問題、 發展與展望

蕭霖

【作者簡介】

蕭霖，台灣屏東人。國立政治大學教育系博士。曾任德霖技術學院財務金融系助理教授。現任國立暨南國際大學教育政策與行政學系助理教。

摘要

效率的追求與品質的管控的觀念已被接受，教育評鑑遂成為達成此一目的的重要方式。教育評鑑要為組織所接受，首先，教育評鑑就必須為真評鑑。除此之外，教育評鑑要能夠提出可行的建議，排除組織的自我防衛，重視評鑑所使用的方法，評鑑者本身的角色立場，都能使教育評鑑更能達成預期目的。

此外，評鑑的發展可分為六個時期：改革時期、效率與測驗時期、泰勒時期、純真時期、擴張時期、專業時期，其發展有其社會變遷的需求及歷史背景。至於教育評鑑的展望，使用更科學的方法，更多元的評鑑方式，以及受評者對於評鑑更積極的態度等等，都會使教育評鑑更貼近其功能與需求並獲得更充分的發展。

關鍵詞彙：教育評鑑、教育政策、指標、全面品質管理

Key Words: Educational evaluation, Educational policy , Indicators, Total quality management

前言

「十年樹木，百年樹人」一語道出了教育的任重道遠，也間接的指出教育的成效是不可能立竿見影的；於是「只問耕耘，不問收穫」的埋頭苦幹，似乎成了傳統教育人員的規準。時至今日，資源稀少的經濟學概念已經大量的引用到教育事務的領域，教育不能再「不計成本」的投資，效率的追求與品質的管控的觀念已被接受，教育評鑑遂成為達成此一目的的重要方式。我國自民國六十四年起實施大學評鑑，相當程度受到美國教育評鑑的影響（盧增緒，1995），由高等教育逐漸向下發展，中小學的教育評鑑也逐步展開，教育評鑑遂成為教育行政中相當重要的一環。

由於評鑑是發展組織與人員效能最普遍、最基礎、最必要的長期性歷程，不但每一個過程都需要力求公平、合理、卓越、符合專業的標準，且每一步驟均需以促進組織與個人雙方永續發展為宗旨。評鑑與組織及個人關係密切，教育評鑑如果對組織或個人是公平的、具說服力的、有一套穩固的評鑑標準，方足以供專家、行政人員及其他監督委員透析評鑑結果，發展完善的改革策略，導政現存教育活動的缺失，使教育專業得以順利發展，並面對各界對教育效益的質疑與挑戰（Perterson, 2000；Shinkfield & Stufflebeam, 1995）。

基本上，評鑑猶如人體的健康檢查一樣，主要有兩種功能，一是證明（to prove）的功能：了解學校的運作情形及其表現；另一是改進（to improve）的功能：幫助學校改善其運作缺失。謝文全（2004）指出：教育評鑑包含下列三項要點，一、評鑑是對事務加以審慎的評析；二、評鑑是量度得失及其原因；三、評鑑旨在決定如何改進或重新計畫。李大偉（1986）指出：自動自發是評鑑的起點，自我改進是評鑑的本質，自我管理是評鑑的精神，激勵成長是評鑑的目的，自我研究是評鑑的核心。康自立（1989）指出：教育評鑑係針對某依特定的教育方案，有系統的蒐集方案特質及其相關影響因素，對方案的實際結果與預期結果間作價值的判斷，以便作為決定或是改進評鑑運作的一種歷程。

壹、教育評鑑的問題

一般而言優良的評鑑需合乎四項基本原則（黃德祥，薛秀宜，2004）：1.適當性（propriety），包括教育服務方向、正式評鑑的方針、利益衝突的處理、人事評鑑報告的運用、評鑑項目的周詳與專業程度。2.實用性（utility），評鑑方向的建構、清

楚的標準定義、可信賴的評鑑委員、報告合理且正確清楚表達、評鑑結果可使人遵循改進。3.可行性 (feasibility) , 務實的程序、政策與財政的可行性。4.準確性 (accuracy) , 參予評鑑者之角色定義、工作環境與程序文件的要求、有效可靠的測量程序 系統資料的保密性 避免偏見 嚴格監控評鑑系統(Perterson, 2000; Shinkfield & Stufflebeam, 1995) 但是評鑑的實際過程, 果真能如前述的如此客觀與標準程序, 則不無疑義, 也因此, 教育評鑑仍存有相當的困難與問題。

教育評鑑本身並無目的, 而是以教育評鑑的結果成為行政行為修正的依據為其重要的目的; 也因此, 教育評鑑的結果是否為行政人員所採用遂成為教育評鑑的重點。

一、教育評鑑真的是評鑑嗎

何以評鑑不被行政人員所接受, 首先, 究竟我們所做的是否真的就是教育評鑑? 教育評鑑的真偽的辨別就不能不釐清。從廣義的角度來看, 有所謂的「偽評鑑」、「準評鑑」以及「真評鑑」, 分述如下(林天祐, 蔡菁芝, 2001; 曾燦燈, 1989; Stufflebeam & Webster, 1983)。

(一) 偽評鑑 (pseudo-evaluations)

一般而言, 以政治為導向的研究, 多屬此類, 包括政治控制的調查(或稱「秘密研究」), 與公共關係 - - 授意的研究 (public relations-inspired studies)。前者評鑑的目的在透過對於評鑑資料使用的控制, 已達成增加當事人的影響層面; 其評鑑的目的在透過對於評鑑資料使用的控制, 已達成增加當事人的影響層面; 其評鑑資料的公布是選擇對當事人有利的部分, 甚至扭曲事實的公布, 已增加當事人的聲望、權力或金錢等等。後者與前者類似, 其目的在幫助評鑑的委託人, 替學區、方案或有關之過程, 創造積極的形象。此類評鑑, 除方法令人質疑外, 目的也都偏頗, 徒具評鑑的形式, 故稱為偽評鑑。

(二) 準評鑑 (quasi-evaluation)

準評鑑則會有一特定需要評鑑的問題, 而且能在方法論上做適當的運用, 目標本位的研究、績效責任研究、實驗或準實驗設計研究、測驗計畫和管理資訊系統, 均屬此類評鑑。這些評鑑研究, 通常導致無用的結論, 或因鎖定的資訊範圍太窄, 無法作為判斷一個方案是否有價值的基礎。所以此種評鑑, 儘管在方法論上適當, 但卻不一定能提出契合問題所需的評鑑資訊, 故僅能稱為準評鑑, 而不等於評鑑。

(三) 真評鑑

相較於準評鑑，真評鑑主要是設計來評定某一目標的評鑑均屬之，包括鑑定（或證明）研究、政策研究、決定導向的研究、消費者導向的研究、當事人中心的研究、鑑賞家為基礎的研究。真評鑑是屬於價值導向的評鑑，且評鑑的過程具有方法論的基礎，而評鑑的結果能做為決定、選擇、改進或評定的準據。

由此觀之，若教育評鑑的本質為偽評鑑，則評鑑淪為行政管理或控制的手段與工具，不僅評鑑本身毫無意義，也可能成為扼殺評鑑發展的元兇。準評鑑若能適當的調整其蒐集資訊的範圍，使得所獲得的資訊能夠真正的描繪出受評者的全貌，則與真評鑑同為有效的評鑑，而能達到評鑑預期的目的。以目前所推行的校務評鑑觀之，大部分的措施都朝向準評鑑與真評鑑的內涵去實施，只是，許多學校反應，校務評鑑仍有相當程度對校長領導學校的績效進行評鑑；導致學校老師配合執行意願低落，認為校務評鑑結果的好壞，只是校長個人的業務檢查或績效考評而已，而置身事外。

二、評鑑的功能是否達成

根據林文達 (1992) 指出，教育評鑑結果所以不被行政人員採用，大致有下列原因：1.評鑑者地位不能影響到行政工作，評鑑人員與受評者若無隸屬關係，或是無法發揮影響力，評鑑的結果可能純粹為了研究，則評鑑將毫無效果可言。2.機關首長以評鑑為政治手腕，不以合法方式運用評鑑結果，像是以評鑑來拖延決策時間、推卸責任，或是塑造形象。3.組織產生自我防衛機構，墨守成規，形成行政行為的慣性而抗拒變革。4.評鑑結果未能提出具體改進建議，若是評鑑結果僅止於資訊的提供，證明策略執行的缺失，而未能提出改進缺失的具體建議，使得受評者無所適從。

從林文達的分析可以看出，純粹的研究不能稱作教育評鑑；淪為機關首長的工具也非真正的評鑑（其實就是前述所稱的偽評鑑）；若是組織成員抗拒變革而否定評鑑結果，則需要適當的引導與教育成員的心態；倘若評鑑結果未能做成積極正面的建議，也不能算是完整的評鑑。

另外，秦夢群 (1998) 指出，教育評鑑最引起爭議的問題有三：1.評鑑的功用何在，若是評鑑只重視過程，而不據以作為改進的目標，或是不重視評鑑之後的追蹤，則評鑑功能大打折扣；2.評鑑人員如何選任，圈內人懂內情但可能為人情牽絆，圈外人較無利害糾葛但可能流於浮面判斷，孰優孰劣尚無定論；3.評鑑方法應為質

化或量化，質化方法的主觀疑慮與量化方法見樹不見林的漏失，也必須得到適當的解決。

秦夢群所指出的評鑑爭議有三：評鑑不能只注重過程，完整的評鑑應該包含追蹤以及再評鑑，評鑑本身無意義，真正的意義在評鑑之後的追蹤輔導。評鑑的人選存在圈內與圈外孰優孰劣的爭議，但是真正的專業評鑑者應該可以排除出身的考量。評鑑方法不應限定於某種固定形式，而是視資料性質而定，遷就質化或量化方法而刻意的作資料性質的轉換，轉換過程所造成的資訊扭曲，可能使得評鑑的功能喪失殆盡。

楊振昇指出(2000)，國內辦理教育評鑑有待商榷處有六個方面：1.評鑑者與受評鑑者對於評鑑目的並未建立正確的認知，也未形成共識，因而也影響到評鑑的成效。2.評鑑人員的適切性，以圈內人（可能有利害關係或私人情誼者）或是以圈外人為之，抑或是委由民間評鑑團體辦理，尚無共識。3.評鑑時間，評鑑時間多為半日至一日，受評者咸以評鑑時間太短，難窺學校全貌而對評鑑結果不盡滿意。4.評鑑指標是否周延客觀，評鑑人員與受評者未必有共識，甚至評鑑人員之間意見都有出入，使得評鑑結果公信不足。5.評鑑方法，量化方法可能忽略某些無法量化的評鑑項目，質化方法可能流於主觀判斷，在評鑑方法的採用上需要更進一步的謹慎與平衡。6.評鑑報告，僅列舉優缺點而缺乏具體建議，實用性不高；陳漢強（1997）更指出，將評鑑結果量化、排行，則是誤用評鑑的另一種現象。另外，吳俊佑（2002）指出，教育評鑑的缺失：1.過多和過於瑣碎的評鑑；2.目標不清楚的評鑑；3.沒有反省評鑑本身。

三、校務評鑑促進或是妨礙校務推動

校務評鑑的功能無非是希望推動校務發展，然而許多學校卻為了校務評鑑而影響正常校務推行，純粹為了評鑑而評鑑，甚至為了評鑑而衍生出許多衝突對立，失去了原先制度的美意。至於校務評鑑如何與校務推動緊密結合，以下有幾點可供參考。

首先，評鑑的指標應更具彈性，容許不同的教學方式與因地制宜的行政措施，評鑑的指標只需訂定可接受與不可接受的標準（陳世佳，2004），使得教育措施不至於因教育評鑑而失去教學多元化的面貌。其次，實作評量最能反映出現實狀況，只是，如此一來，評鑑的時間抽樣必須增加，評鑑結果的數量化呈現必須更加謹慎，評鑑者間的一致也要更加注意。再其次，評鑑者本身也需要充分的訓練與協調，評鑑建議具有時效性、可行性、實用性，使得評鑑結果為受評者所誠心接受，並有助

於校務推展。最後，校務評鑑的指標必須隨著時代潮流、教育思潮、社會變遷而隨時修正，避免流於「用舊的指標，評鑑現在的學校，去面對未來的社會」這類認為教育評鑑不合時宜的看法。

貳、教育評鑑的發展

評鑑的概念向來有之，但是行諸於文字與制度的做法則是近代教育蓬勃發展之後的產物。評鑑的發展可分為六個時期：改革時期、效率與測驗時期、泰勒時期、純真時期、擴張時期、專業時期，其發展有其社會變遷的需求及歷史背景。

一、教育評鑑的沿革與發展階段

根據Madaus&Stufflebeam (2000) 將教育評鑑分為六個時期：

(一) 改革時期 (age of reform, 1800-1900)

為教育評鑑的萌芽時期，教育評鑑為社會改革的一環。

(二) 效率與測驗時期 (age of efficiency and testing, 1900-1930)

效率為此時的追求的目標，隨著測驗理論的建立與開展，以測驗就成為達成此一目標的唯一手段。

(三) 泰勒時期 (Tylerian age, 1930-1945)

泰勒的「八年研究」，擴大了前一期侷限於測驗方法的範圍，並開爾後大規模教育評鑑之先河。

(四) 純真時期 (age of innocence, 1946-1957)

此期適逢美國國力達於顛峰，學校數量與規模大幅擴張，反倒使教育評鑑的技術停滯不前。

(五) 擴張時期 (age of expansion, 1958-1972)

由於太空技術被蘇聯超越，引起美國國內一片檢討之聲，教育評鑑的技術發展迅速。

(六) 專業時期 (age of professionalization, 1973 至今)

秦夢群 (1998) 指出：至 1973 年開始，教育評鑑開始蛻變為一種專業，與教育的其他領域如教育哲學、教育心理一樣自成體系。在此之前，評鑑的身份不明，有的被認為是測驗專家，有的更被譏為是補懂教育但與當權者掛鉤的打手。評鑑的專書與期刊闕如，大學中也多半將評鑑的概念侷限於教育測

驗的課程。此種模糊與片段的情況，在 1970 年代開始有所改變。

首先，各種專業期刊如「教育評鑑與政策分析」(Educational Evaluation and Policy Analysis)、「評鑑研究」(Studies in Evaluation)、「評鑑評論」(Evaluation Review)紛紛發行，提供最新的資訊與評鑑策略。此外，相關學會如美國「評鑑研究學會」(The Evaluation Research Society)則提供會員與社會人士在專業資料上的交換。

在大學方面，「教育評鑑」已漸成為一門獨立學科而被教授。新的模式如後設評鑑(meta evaluation)的出現，以再評鑑之形式評鑑原有之評鑑，使其品質更有所保證。不過儘管如此，以上成果並不代表教育評鑑之發展已臻完美之境，問題如評鑑應採用量化或是質化之爭，理論與實務上之落差等，均需列入未來努力解決的部分。可以預期的是，由於爭奪教育資源日漸白熱化，任何方案或是策略的提出，均需經過各方之嚴格檢驗；而評鑑之地位自當日形重要。

二、教育評鑑獲得重視的原因

教育評鑑運動于1960年代開始蓬勃發展，主要的原因有(張鈿富，1999)：

- 1.學校教育的保守、缺乏效率，使社會大眾普遍不滿，要求對學校教育成效進行客觀的評鑑。
- 2.政府對教育部門補助日益增加，為了決定對新方案補助與否，有必要實施評鑑。
- 3.在 70 年代初期，由於第一次全球石油危機，造成經濟衰退，政府財源短絀，教育經費因此大幅縮減，為使教育經費的運用合乎經濟原則，應對學校教育加以評鑑。
- 4.教育行政分權化的結果，使學生家長及關心教育的社會大眾，得以有機會參予教育系統的運作，紛紛開始要求評鑑教學方案。
- 5.由於教育工學的興起，很多新的教育輔助器材大量運用於教學上，促使教育人員開始研究這些視聽教材在教學上的運用及效果，並探討其對傳統教學的輔助或替代的可能。

秦夢群(1998)亦指出，促使評鑑興起的因素有四：

- 1.教育資源日漸短絀，由於受教人數激增，學生單位成本下降，對於績效成本的概念遂起。
- 2.對公立學校辦學績效的質疑，公立學校以公共預算支應，基於權利義務對等的概念，納稅人要求杜絕浪費。
- 3.課程與教學策略的更新，然而新舊教材與教法的優劣與差異，應由實際成效

的評鑑得知。

4. 教育自由市場化的要求，原先頗具優勢的公立學校，應與私立學校站在平等的位置上，接受檢驗。專業且具公信力的教育評鑑，因此應運而生。

我國教育評鑑的發展，約可追溯到1968年10月，台灣省教育廳所公布的「台灣省國民中學校務設施評量表」（楊振昇，2000）。真正大規模、全面性的教育評鑑，應是1975年開始實施的大專院校教育評鑑（張植珊，1979），教育部為提高大學教育水準，所辦理的大學教育評鑑工作，評鑑對象首先由大學院校之數學、物理、化學、醫學系及研究所開始。1976年起教育部又擴大辦理農、工、醫等學院之評鑑，1978年辦理商學院評鑑，1979年辦理文學院評鑑、師範學院評鑑，至此教育部先後完成我國大專院校各系所初次評鑑工作（陳漢強，1985）。

秦夢群（1998）指出，教育評鑑工作，不僅辦理大專院校的複評之外，並把評鑑的對象擴及至專科學校、中小學、及幼稚園等各級教育單位，至此教育評鑑開始迅速發展，在教育評鑑逐漸受到重視之際，教育評鑑的各項理論、模式大量由國外引進國內加以運用實施。

例如：在1992年，台灣省首次大規模實施對省立高級中等學校訓導與輔導工作的評鑑，使用的即是CIPP模式。評鑑準備工作由新竹師範學院負責，先由各小組制定項目評鑑表，接著再舉辦研討會，所有評鑑委員均必須參加以了解相關規定與原則。準備工作完成後，即由省教育廳安排各校評鑑，原則上小組由五人組成，以一天時間實地觀察。評鑑表已先寄給學校自評，委員到校後再填上意見。離開前評鑑小組必須與校方行政人員與教師進行綜合座談，報告成員所發現之優缺點，已達到回饋的功效。最後之報告由小組組長匯集成員之評鑑資料完成，以供教育廳作日後遷調校長的參考（秦夢群，1998）。

參、教育評鑑的展望

一、教育評鑑的實施原則

謝文全（2004）指出，實施教育評鑑的七項原則：一、方法要科學化，二、過程要民主化，三、兼顧過程與結果評鑑，四、兼作形成性與總結性評鑑，五、內部與外部評鑑兼用，六、應注意評鑑的綜合性及整體性，七、重視評鑑結果的應用。吳清山（2000）指出，學校教育評鑑若能參酌全面品質管理的理念與做法，相信更能發揮評鑑的效能。教育評鑑一方面可以了解教育的成效，一方面也可以幫助教育

缺失的改進，隨之全面品質管理的受到重視，教育評鑑應該參考全面品質管理的一些理念，逐漸揚棄過區考督和考核式的總結性評鑑，慢慢朝向形成性評鑑發展，以及鼓勵各校能夠定期進行自我評鑑，發現缺失立即改進，才是真正符合全面品質管理的精神。

賴志峰 (2002) 為教育評鑑及其評鑑人員提出四項建議，一、成立教育評鑑專業組織，加強評鑑研究與發展；二、建立適合國情的教育評鑑標準，邁出專業化的重要一步；三、發展教育評鑑人員專業能力基準，規劃評鑑人員養成方案；四、建立教育評鑑人員證照制度，提升評鑑為具有專業的職業。

二、多元評鑑的可行性

多元評鑑是教育評鑑的主要共識之一，由於多元評鑑包含了評鑑的重要精神，遂成為目前發展評鑑模式的重要依據，其特徵有：1.評鑑資料的多元運用，2.評鑑方式不斷改進，3.評鑑期限富彈性，4.評鑑項目可廣泛周延，5.提高評鑑的效度、信度與客觀性，6.周密的評鑑設計（黃德祥，薛秀宜，2004），分述於次。

總結性評鑑是建立多元評鑑資料庫的重要評鑑模式，總結性評鑑可提供整體總和的結果，但為了避免總結性評鑑對可能產生的評判過於偏頗，形成性評鑑正好可以補其缺失，相輔相成。真實評量是目前最受矚目的評鑑方式之一，因其所獲得的資料較為廣泛，且較具真實性，可以改善傳統的評鑑方法。評鑑時間具有彈性，比較可以在較長的時間採樣中，獲取更多更重要以及具代表性的資料，但是實際實施時，所耗費的人力、物力、時間，是否為各單位所能承受與支應，則是權衡利弊得失的另一個角度。

評鑑項目周延廣泛一直都是教育評鑑所追求的目標，只是相對帶來的評鑑成本，必須和評鑑的效果本身作折衷取捨。評鑑的結果，往往在做決定，牽涉學校的經費分配與人事安排，只有不斷的提高評鑑的信度、效度與客觀性，才能使評鑑結果令人心服，某種程度上將過程標準化，才能使評鑑信效度提高。最後，周詳的評鑑設計，甚至是後設評鑑，都是精進評鑑措施的有效方法。

三、結語

教育評鑑在我國行之有年，雖然各方評價不一，基本上多數採肯定的態度。評鑑的內涵，也隨著時代的潮流，從早年的督導、預算分配、主管考核，到現在的自我評鑑、診斷建議與追蹤，受評鑑者將評鑑視為行政運作的一部分，心態上也能持較積極的態度，教育評鑑的功能也才能充分發揮。至於教育行政主管機關，為教育

評鑑的發動者；除了因教育評鑑為行政監督上的重要措施之外，如何能夠以最少的干擾，為各級學校的校務推動作最大促進，保障並提升學生的受教品質，則是教育評鑑持續進展的重要使命。

參考文獻

- 吳俊佑（2002）。反省教育評鑑和教學評鑑的本質與問題。竹縣文教，28，49-51。
- 吳清山（2000）。全面品質管理在教育評鑑上的應用。北縣教育，35，27-31。
- 李大偉（1986）。技職教育測量與評鑑。台北：三民。
- 林天祐、蔡菁芝（2001）。教育評鑑的理念分析。教育研究月刊，91，36-44。
- 秦夢群（1998）。教育行政 - 實務部分。台北：五南。
- 康自立、李大偉、蕭錫錡等（1989）。專科學校教師教學評量手冊。教育部。
- 張植珊（1979）。教育評鑑。教育部教育計畫小組編印。
- 張鈿富（1999）。教育政策與行政 - 指標發展與應用。台北：師大書苑。
- 陳漢強（1985）。美國大專院校評鑑之研究。台中：台灣省政府教育廳。
- 陳漢強（1997）。大學評鑑。台北：五南。
- 曾燦燈（1989）。評鑑途徑的分析。輯於黃光雄編譯：教育評鑑的模式，53~84。台北：師大書苑。
- 楊振昇（2000）。教育研究與教育評鑑之反省與展望。暨大學報，4（2），27-50。
- 陳世佳（2004）。以教師專業成長為目標之教師評鑑。教育研究月刊，127，33-44。
- 盧增緒（1995）。論教育評鑑觀念之形成。輯於中國教育學會主編：教育評鑑，3~60。台北：師大書苑。
- 賴志峰（2002）。教育評鑑邁向專業化之途徑：評鑑人員證照制度。教育資料與研究，48，100-107。
- 黃德祥、薛秀儀（2004）。教師評鑑的模式與發展趨勢。教育研究月刊，127，18-32。
- Madaus, G. F., & Stufflebeam, D. L. (2000). Program evaluation: A historical overview. In D. L. Stufflebeam & G. F. Madaus & T. Kellaghan (2nd eds.), Evaluation models: Viewpoints on educational and human services evaluation (pp. 3-18). Boston: Kluwer Academic.
- Perterson, K. D. (2000). Teacher evaluation, New York: Corwin Press, Inc.
- Shinkfield, A. J., & Stufflebeam, D. L. (1995). Teacher evaluation: Guide to effective practice. New York: Kluwer Academic Publishers.

Stufflebeam, D. L. & Webster, W. J. (1983). An analysis of alternative approaches to evaluation. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.). *Evaluation models: Viewpoints on educational and human services evaluation.*, pp.23~43. Hingham, MA: Kluwer Academic.