



教育評鑑的本質—描述，與判斷？

郭昭佑／國立台灣藝術大學師資培育中心教授

郭工賓／台北市政府教育局督學

教育評鑑活動近來在台灣逐漸蔓延，亦日受重視。從過去一般最常見的中小學校務評鑑，及各法令規範的特殊教育、體育活動、資訊教育、民主教育、午餐、交通安全…以至於九年一貫課程實施後的課程評鑑，評鑑似乎成為教育人員的全民功課。在大學評鑑上也不遑多讓，自1996年全面性的大學評鑑後，台灣大學評鑑協會的學門評鑑及師資培育訪視才方興未艾，教育當局更以財團法人方式成立了「台灣高等教育評鑑基金會」，未來更將主導大學教育的評鑑體系。

而直指教育人員自身的「教師評鑑」，亦已在教育當局的「鼓勵」下開始試辦，大學校院部分更早在大學法修訂條文第二十一條中明訂：「大學應建立教師評鑑制度，對於教師之教學、研究、輔導及服務成效進行評鑑，作為教師升等、續聘、長期聘任、停聘、不續聘及獎勵之重要參考」。由此可知，評鑑進入學校教育，以至於教師生活中已是趨勢。

只是，台灣教育評鑑活動雖頻繁，但接受評鑑者多不明其意，只知道評鑑流程，如何準備，而不知何謂評鑑，停留在「知其然而不知其所以然」的階段。而過去對學校系統的評鑑大多以學校行政人員為主，但未來此一發展勢必將影響所有參與教育工作的人。因此，教育人員應有啟蒙的準備，試著瞭解評鑑，接受評鑑，並從評鑑中實際獲益。

對一般的教育人員來說，在教育評鑑的理解中，對教育活動較為熟悉，而對評鑑則較為陌生。究竟評鑑是什麼？它所從事的是什麼活動？這就是評鑑的本質。本文擬探究

教育評鑑的兩大本質：描述與判斷。事實上，對這兩個教育評鑑本質的認可仍有許多討論，尤其是判斷本質，更有些不同的爭辯，本文試圖從其不同的觀點，探究其被認可與否的軌跡。

一、「資料收集與分析」的描述本質普獲認同

原則上來說，評鑑的描述本質是較沒有異議的。自1960年代起，「為決策者提供資訊」即成為評鑑定義中廣為接受的要素之一（Alkin, 1969; Cronbach, 1963; Stufflebeam, 1969），先不論「為決策者所提供的資訊」是否包含判斷成分，描述本質早被認定於存在於此一評鑑定義的要素中。

過去有許多評鑑定義即聚焦在其描述本質上，Kemmis（1982）即指出評鑑是「整理資料訊息及理由依據的過程」；Thomas與Guskey則認為評鑑是一種有系統、有價值的調查（引自陳嘉彌、鍾文郁、楊承謙、柯瓊惠與謝元譯，2002）；McMillan（2000）亦曾定義評鑑為「收集資料，以利做決定的歷程」；Oliva（1997）更直指評鑑「是一種收集與解釋訊息的持續過程」。Torres（1996）在論及評鑑的描述本質時說道：「評鑑者無法確定自己是否有能力提供有關評鑑或方案的決定性資訊，因此只能儘可能的收集評鑑方案的內容與文化各個層面在過程中如何運作，並瞭解其當下環境及參與者所感受的意義與看法，以說明方案的情況」。

上述評鑑定義所呈現的描述本質，除了



對評鑑資料的「收集」外，亦可見須針對所收集資料進行解釋或說明其理由或依據之「分析」過程的需求，這種「資料的收集與分析」已然清楚詮釋評鑑的描述本質。而將所收集的「資料」經分析後，即可能成為評鑑後提供給決策者的「資訊」之一，Nevo即明白指出「描述」係收集系統資料而成為較客觀的資訊（Nevo, 1995）。客觀，亦是描述的重要特徵之一，即使是評鑑者的分析，亦須架構在所收集的資料上，因此是較客觀的。陳玉琨（2004）即指出評鑑須對事物的現狀、屬性與規律等資料進行量（或質）的客觀描述，其基本要求是它的客觀性，即真實地反映事物的本來面目。

由此可知，意涵「資料收集與分析」的描述確為評鑑的本質之一。在教育評鑑上亦同，據此，描述為教育評鑑的重要本質似乎已獲共識，而此一描述本質意涵著在教育評鑑過程中對教育客體的「資料收集與分析」亦應無太大疑義。

二、「判斷」是否應為評鑑本質的幾個爭議

評鑑的「判斷」本質並不似描述本質那般受到一致的青睞，許多不同來源的聲音質疑它成為評鑑本質的合法定位，不過這樣的質疑似乎未廣受認同：

（一）強調評鑑的發展改進目的同時拒斥判斷本質？

評鑑目的如化約為績效責任與發展改進，不同的立足點影響其對評鑑本質的認知。史丹佛評鑑協會（Stanford Evaluation Consortium）從發展改進的評鑑目的觀點，主張評鑑是「對一個現存方案的發生與結果之系統性瞭解目的在協助該方案及其它有相同目的之方案改進」（Cronbach, Ambron, Dornbusch, Hess, Hornick, Phillips, Walker & Weiner, 1980），基於此一理解，Cronbach

及其夥伴們清楚的拒絕評鑑的判斷本質，將評鑑者視為教育者，「其成功與否，係藉由其他人學習到什麼來判斷」（an educator success is to be judged by what others learn），並認為這樣的定義，比將評鑑者當作「受聘雇以決定誰對誰錯的籃球比賽裁判」來得適切。

這樣的觀點顯然把評鑑的判斷本質視為績效責任目的之同路人。然而，以發展改進為目的之評鑑難道就不須判斷？Hopkins（1997）雖強調評鑑的發展論述，但卻指出這並不意味著評鑑不關乎判斷，發展的成果可以經由清楚的分析與判斷加以促進；Stufflebeam 與Webster（1980）更以「協助某些群體判斷並改進某種教育的客觀價值」來定義教育評鑑。因此，即使僅從發展改進目的看待評鑑，亦或多或少與判斷相關。

（二）判斷應是決策者的事而與評鑑（者）無關？

有些學者定義評鑑為「為決策者收集資訊」以刻意忽視評鑑的判斷本質，將判斷移出評鑑活動的領域之外，在此概念下，評鑑者只須儘可能的收集與評鑑客體相關且可靠的資訊即可，而任由決策者使用這些資訊以判斷評鑑客體的品質並做適當決策（Nevo, 1995）。如此，判斷品質的工作即與決策同時落在決策者身上，而好像與評鑑者無關。

不過，Frado與Lee（1997）卻認為「評鑑是收集與應用資料，並提供資訊予教育方案決策參考，所提供的評鑑資訊應包含方案的功績與價值之判斷」；Nevo（1995）也指出評鑑者不應只是提供描述性的資料而不做判斷，而應描述評鑑客體並從各種觀點做判斷，以協助觀眾瞭解客體的本質與品質，這種判斷陳述（judgment statements）是評鑑者的權責；Brainard（1996）說得真切：「評鑑提供的是資訊與眼光，而不是最後的決



策」。從以上論述來看，依所提供資訊做決策雖然是決策者的事，但判斷似乎仍應與評鑑者相關，是評鑑活動的一部分，應屬評鑑的本質之一。

（三）判斷易引發憂懼而不宜成為評鑑本質？

Nevo（1995）認為一個描繪判斷本質的評鑑定義，可能使受評者產生憂慮與抵制，而不具判斷性的評鑑定義，如「提供決策者資訊」，則可能較容易被受評者接受，Scriven亦指出人們對評鑑的擔憂，常導因於其主觀判斷的不可靠而引發受評者的反對（引自陳玉琨，2004），為避免受評者憂懼並取得其信任與肯定態度，許多評鑑定義都試圖淡化評鑑的判斷屬性，甚至拒絕判斷成為評鑑本質。

然而，僅為了創造對評鑑的正向態度而忽視其本質是不實際的，發展對評鑑肯定態度的正確途徑，應該是在教育的各個領域中證實其建設性功能（Nevo, 1995）。陳玉琨（2004）也認為受評者的確可能因心理安全與自尊的需要而對評鑑採取抵制的態度，但因此而忽視其判斷本質是解決不了問題的，應該從受評者參與評鑑的需要著手，激發他們的內在動機，引導其積極與主動參與。由此可知，判斷所可能引發的憂懼不應成為排拒它成為評鑑本質的理由，不可因噎廢食。

三、判斷應為評鑑本質的重要論述

（一）評鑑的詞義與判斷活動密切相關

從字義上看，評鑑一詞係由詞幹“value”加上詞首“e”和動詞性詞尾“ate”組成，意即引出和闡發價值或進行價值判斷（霍力岩，2000）。在中國的文字中，「評鑑」是評定價值的簡稱，從本質上來說，評鑑是一種價值判斷的活動（陳玉琨，2004）。依據韋氏詞典，評鑑係指對事物的價值、品質、重要性、數量、

程度或條件加以檢視與判斷，與judgment、appraisal、rating、interpretation同義（引自Gove, 1981）。黃政傑（1993）更直指評鑑活動並不中立且隱含著價值系統的選擇。許多學者亦認同評鑑是充滿主觀的功績或價值評估的判斷活動（江啟昱，1993；陳玉琨，2004；Davis, 1981；Nevo, 1995），這也是評鑑之所以與研究不同的最大差異點。由此可知，評鑑除了「資料收集與分析」的描述本質外，與功績或價值評估的判斷活動密切相關。

（二）評鑑的判斷資訊有利合理決策

Mark、Henry與Julins指出功績與價值評估的判斷資訊將有助於決策合理化；例如在行動方針的適當選擇、或現有的方案與政策是否持續進行、或採行新方案、或從眾多政策擇其一等決策時均可參考相關的判斷資訊；當然，即使沒有評鑑仍可決策，而評鑑結果也不一定使人改變決策，但未參考功績與價值評估的判斷資訊所做的決策，思慮較不周詳也較不合理；當民主國家對其政策或方案難以取捨時，功績與價值評估能對評鑑工作產生最大的助益（Mark, Henry & Julins, 2000），有些評鑑人員甚至只用功績與價值評估來定義評鑑工作（Scriven, 1994）。由此可知，提供資訊予決策參考的評鑑歷程，其所提供的資訊除了對評鑑客體資料的收集與分析之描述外，如能包括功績與價值評估的判斷資訊將更有助於合理的決策。

（三）即使是描述亦時見判斷影子

功績與價值評估的判斷活動時見於評鑑過程中，即使在資料收集與分析的描述活動中亦常見其蹤。Davis（1981）即認為資料的收集和事件的描述，常是由「先在的」價值判斷所預設的潛在假定為基礎來收集與分析資料；Marsh與Willis（1999）也認為在開始評鑑前，評鑑者應該要察覺自己如何收集



與分析資料下所隱藏的不同價值觀；Wideen更從評鑑模式的分析中指出，不同模式的判斷成分不同，即使只是資料的收集與分析亦存有判斷成分，廣義的說，即使是評鑑本身的設計與模式選擇即是一種判斷，遑論應收集資料的選擇、資料的分析、資料呈現方式、以至於評鑑報告形式，在在都須要判斷，或多或少都帶著評鑑者的主觀價值認定（引自Hopkins, 1997）。由此可知，在評鑑歷程中，可能時時存在著評鑑者的主觀判斷，即使是資料收集與分析的評鑑描述本質，仍可能存有判斷本質的影子。

四、「功績與價值評估」判斷本質逐漸確立

由上述論述可知，判斷為評鑑本質之一已逐漸確立，Davis（1981）即認為一個評鑑者不僅要描述情境，也要做判斷，任何的資料收集、分析與報告如果沒有判斷，只能算是「準-評鑑」；Scriven（1991）亦指出「如果沒有評鑑的判斷過程，即無法區別出何者有價值，而何者無用」，他更表明「除非做判斷否則就不是評鑑」；Stake（1967）亦曾以反話說：「許多評鑑專家選擇不做判斷，那是因為他們沒有能力知覺並判斷其價值」；即使如Tyler所談的「目標達到的程度」，都將涉及判斷的議題（Bennett, 2003）。

近幾年來，評鑑者對於評鑑的判斷本質看法漸趨一致，並定位為「功績與價值的評估」（Eisner, 1979; House, 1980; Scriven, 1967; Stufflebeam, 1974）。Scriven於1991年出版的評鑑辭典（Evaluation Thesaurus）中提到「評鑑是在一連串評估事物的功績與價值過程中的產物」（引自McNamara, Erlandson & McNamara, 1999）。美國「教育評鑑標準聯

合委員會」（Joint Committee on Standards for Educational Evaluation, 1994）亦將評鑑定義為：「有系統的調查某一對象的價值或功績」。歸納來說，評鑑的判斷本質應可以「功績與價值的評估」來表示。

Marsh與Willis（1999）指出功績與價值的意涵不同，功績指的是做某件事物的成功程度，較適用於製作某件物體的手工藝，或某項活動實施過程的熟練度；價值指的是做某件事物的重要性，較適用在該物體或活動本身的價值上；Mark、Henry與Julines（2000）則以政策或方案為例以為區隔：「當我們使用功績這個語詞時，是在說明政策或方案的執行品質，而價值則在闡明政策或方案的執行所帶來的社會公益」；Scriven從測量上詮釋兩者間的不同意涵時亦有類似的看法：「功績的測量須依循專業標準，而價值的測量則應依循組織與社會需求」（Scriven, 1995）。我們常說一件事情做得好不好，和有沒有價值是兩回事，也就是這個意思。據此，「功績與價值的評估」為評鑑的判斷本質應無疑義，也更能提供「資料收集與分析」外的另一種決策資訊。

五、教育評鑑「描述」與「判斷」本質的初步共識一代結語

「系統的評鑑」（Systematic Evaluation）一書中即定義評鑑：「是一種選擇、收集並提供描述性與判斷性的資訊以利決策，並增進對研究現象的瞭解及符應績效責任的要求」（Stufflebeam & Shinkfield, 1985），湯志民（2001）亦指出評鑑是「一種以系統的方法收集和分析資料，並評估其價值，以協助決策的歷程」，黃政傑（1993）也認為評鑑不單是對現象質的與量的客觀描述，亦應包含價值或功績的判斷。



綜上所述，描述與判斷是評鑑的兩大本質，或有所偏重，但近來的評鑑學者趨向於結合兩者而定義評鑑為「一個包含描述與判斷的活動」（Gronlund, 1981; Guba & Lincoln, 1981; Nevo, 1995; Kemmis, 1989; Stake, 1967），大多數的評鑑研究亦致力於取得描述性及判斷性兩特點間的平衡，而在評鑑定義上均同時兼顧這兩方面的要素（Bennett, 2003; Nevo, 1995）。Gronlund（1981）更曾試著以一種極為簡潔的表述呈現評鑑的本質：

評鑑＝描述（質的或量的）＋價值判斷
金娣與王剛（2002）認為從這個意義上說，評鑑就是在量的描述或質的描述的基礎上進行價值判斷的活動，陳玉琨（2004）也認為Gronlund的這一公式雖然粗糙了一點，但它確實抓住了評鑑活動的本質。

評鑑的描述與判斷本質在教育評鑑上亦同。黃政傑（1993）認為從教育評鑑上來說，評鑑既是價值判斷的活動，而教育

又是充滿價值成分的，如何能排除評鑑活動呢？Worthen 與 Sander（1987）定義教育評鑑是對教育客體，包括教育方案、成果、計畫、歷程、目標的資料收集與價值判斷，陳玉琨（2004）亦指出教育評鑑是一發現、判斷與提升教育價值的過程。PDK（Phi Delta Kappa）於1969年所成立的「全國評鑑研究委員會」的委員們也認為教育評鑑應包括資料收集和價值判斷（引自江啟昱，1993）。

Stake從課程的角度清楚的主張「評鑑的兩個基本行動」是描述與判斷，而這兩者對於想要瞭解一個創新的課程是相當可行的（引自Davis, 1981）。由此可知，描述與判斷均為教育評鑑的重要本質，雖然兩者之間的特性不同，卻同時存在於教育評鑑活動中。描述本質意涵對教育客體進行評鑑資料的收集與分析，提供較為客觀的描述資訊；而判斷本質則在教育評鑑歷程中進行功績與價值的評估，提供較為主觀的判斷資訊；而這兩種資訊都將有利於教育決策。

參考文獻

- 江啟昱（1993）。CIPP評鑑模式之研究。台北：國立台灣師範大學教育研究所碩士論文（未出版）。
- 陳玉琨（2004）。教育評鑑學。台北：五南圖書出版有限公司。
- 陳嘉彌、鍾文郁、楊承謙、柯瓊惠與謝元譯（2002）。專業發展評鑑（Thomas 與Guskey原著）。台北：五南圖書出版公司。
- 湯志民（2001）。台北市國民中小學校務評鑑之評析。發表於國立新竹師範學院初等教育學系、國民教育研究所主辦，「第八次教育行政論壇」。
- 黃政傑（1993）。課程評鑑。台北：師大書苑。
- 霍力岩（2000）。學前教育評價。北京：北京師範大學出版社。
- Alkin, M. C. (1969). Evaluation theory development. *Evaluation Comment*, 2, 2-7.
- Bennett, J. (2003). *Evaluation methods in research*. London: Continuum.
- Brainard, E. A. (1996). *A hands-on Guide to School Program Evaluation*. Bloomington, Ind.: Phi Delta Kappa Educational Foundation.
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672-683.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornick, R. C., Phillips, D. C., Walker, D. E., & Weiner, S. S. (1980). *Toward reform of program evaluation: aims, methods, and institutional arrangements*. San Francisco: Jossey-Bass.
- Davis, E. D. (1981). *Teachers as curriculum evaluators*. London Boston: George Allen & Unwin.



- Eisner, E. W. (1979). *The educational imaginations: On the design and evaluation of school program*. N.Y.: Macmillan.
- Frado, S. H. & Lee, Okhee. (1997). Teachers' voice in program evaluation and improvement: a case study of a test program. *Teaching teacher education*, 13(6), 563-577.
- Gove, P. B. (1981). *Webster's third new international dictionary*. Springfield, MA: Merriam-Webster.
- Gronlund, N. E. (1981). *Measurement and evaluation in teaching*. N.Y.: Macmillan.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco : Jossey-Bass.
- Hopkins, D. (1997). *Evaluation for school development*. Great Britain: Open University Press.
- House, E. R. (1980). *Evaluating with validity*. Beverly Hills, CA: SAGE.
- Joint Committee on Standard for Educational Evaluation (1994). *The Program evaluation standards: how to assess evaluations of educational program*. (2nd). CA: Sage.
- Kemmis, S. (1982). Seven principles for program evaluation in curriculum development and innovation. *Journal of Curriculum Studies*, 14(3), 221-240.
- Mark, M. M., Henry, G. T. & Julines, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving public and nonprofit policies and program*. San Francisco: Jossey-Bass.
- Marsh, C. J., & Willis, G. (1999). *Curriculum-alternative approaches, ongoing issues*. (2ed.). U.S.A.: Prentice-Hall, Inc.
- McMillan (2000). *Educational research: Fundamentals for the consumer*(3rd). N.Y.: Addison Wesley Longman.
- McNamara, F. F., Erlandson, D. A., & McNamara, M. (1999). *Measurement & Evaluation: Strategies for school improvement*. N.Y.: Eye on Education.
- Nevo, D. (1995). *School-based evaluation: A dialogue for school improvement*. Tel Aviv, Israel: Masada.
- Oliva, P. F. (1997). *Developing the curriculum* (4rd). N.Y.: HarperCollins.
- Scriven, M. (1967). *The Methodology of Evaluation*. In Ralph W. Tyler, Robert M. Gagn'e, Michael Scriven (Eds.). *Perspectives of Curriculum Evaluation*. (pp. 39-83). Chicago: Rand Mc. Nally & Company.
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park, CA: SAGE.
- Scriven, M. (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation*, 21, 111-129.
- Scriven, M. F. (1994). Evaluation as a discipline. *Studies in Educational Evaluation*, 20, 147-166.
- Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record*, 68, 523-540.
- Stufflebeam, D. L. (1969). Evaluation as enlightenment for decision making. In W. H. Beatty (Ed.). *Improving educational assessment and an inventory for measures of affective behavior*. Washington, DC: National Education Association.
- Stufflebeam, D. L. (1974). *Meta-evaluation*. Kalamazoo, MI: Western Michigan University Evaluation Center.
- Stufflebeam, D. L. & Webster, W. J. (1980). An analysis of alternative approaches to evaluation. *Educational Evaluation and Policy Analysis* ,2(3), 5-20.
- Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic evaluation*. MA: Kluwer-Nijhoff Publish.
- Torres, R. T. (1996). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations*. London: Sage.
- Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidences*. N.Y.: Longman.