



教育評鑑專業化及評鑑模式的省思

賴志峰／逢甲大學師資培育中心暨公共政策研究所助理教授

一、前言

1887年至1898年間，J. Rice進行有關練習在拼字教學上價值之研究，被公認是第一個正式的教育方案評鑑，教育評鑑之發展歷經改革期（1792-1900年）、效率與測驗期（1900-1930年）、泰勒期（1930-1945年）、純真期（1946-1957年）、發展期（1958-1972年）、專業化時期（1973-1983年）、擴張與整合期（1983-2001年），到了現在，教育評鑑理論發展已漸趨成熟，但仍處於動態的、未成熟的專業（Madaus & Stufflebeam, 2000）。而我國教育評鑑發展，自六十四年大學評鑑實施以來，各級學校教育評鑑、校務評鑑及幼稚園評鑑相繼實施，教育評鑑儼然已成為教育制度重要的一環，但是教育評鑑如何再提升品質，才能符合專業化的規準，是值得探究的課題。

二、教育評鑑的專業化

《韋氏第三新國際字典》（Webster's Third New International Dictionary）將「專業」（profession）的定義為：「一種稱為專門知識的要件，且需長期和徹底地在科學、歷史、學術的原理基礎下，進行技巧和方法的養成，藉由組織的力量或協議的高標準來進行和維持，驅策成員不斷從事研究，其主要目的在於服務大眾。」（Gove et al., 1981）。而Scriven（1991）認為專業是高度的技能，相反的，業餘（amateur）為非全部時間且較少高度的能力；專業是工作以獲得報酬，業餘則是無酬的。專業是不同於生產線的工作者，並有三個特色：（一）高層次的專門技術，需經過訓練與天賦；（二）持續不斷的工作；

（三）為大眾的生活或幸福而工作。

Worthen（1994）曾提出成熟發展的專業必須符合以下九個特性：（一）對於評鑑專家的需求；（二）發展出獨特的專門領域內容（知識、技能）主體；（三）發展出人員養成方案，培養獨特知識、技能的實踐者；（四）符合資格的實踐者能有固定的職業機會條件；（五）建立證照制度的判斷程序；（六）成立學會以致力於實踐者未來的專業成長；（七）專業學會發展出成員入會的規準；（八）專業學會對於人員養成方案有影響力；（九）發展出引導評鑑實施的標準，Worthen同時評定美國的教育評鑑尚未符合上述第5、7、8規準，尚不能稱作是完全成熟發展的專業。Worthen、Sanders與Fitzpatrick於1997年又增列「評鑑功能建制化（institutionalization）」一項特性（引自潘慧玲，2004）。上述幾項規準，可說是評鑑邁向成熟專業（full-fledged profession）的重要指標，亦是建立評鑑公信力不可或缺的元素。此外，Scriven（1991）認為後設評鑑（meta-evaluation）是評鑑專業化的必要條件，Sander於2002年更提出內建機制化（mainstreaming）的概念（引自潘慧玲，2004）。我國教育評鑑發展至今，專業化已成重要課題，這些規準正是未來教育評鑑努力的方向。

教育評鑑如要晉升為真正的專業，評鑑必須具備的獨特技能和知識為重要基石，才能使評鑑的公信力能獲得大眾得信賴。在專業的領域中，評鑑必須具備的獨特技能和知識，Mertens（1994）提出以下三種：（一）研究方法論的知識和技能；（二）評鑑所需的知識和技能（但從其他領域借用）；（三）



獨特學科的知識和技能。Scriven認為評鑑最適當的概念是跨學科的（*trandiscipline*）。因此，評鑑必須具備的獨特技能和知識，可說是評鑑專業化的基礎建設，也才能發展出其他幾項專業化的特質。

三、評鑑模式的幾項省思

（一）評鑑模式的意義

雖然評鑑已廣泛運用，但是評鑑迄今卻沒有一致的定義，Scriven曾舉出十六種評鑑的類似名詞，包含裁判（*adjudge*）、評價（*appraise*）、分析（*analyze*）、評量（*assess*）、評論（*critique*）、檢視（*examine*）、評等（*grade*）、檢查（*inspect*）、判斷（*judge*）、評定（*rate*）、分級（*rank*）、批評（*review*）、評分（*score*）、研究（*study*）、測驗（*test*）等，這些名詞可能會產生混淆，但Scriven認為多樣化評鑑名詞之使用，不僅顯示評鑑在日常生活中的重要性，也代表研究領域的快速增加（引自Fitzpatrick, Sanders, & Worthen, 2004）。

就模式的意義而言，模式為一種概念性架構或理論性的組織體系，因為社會科學所用的任何一種模式都是表徵性的，而不是社會實體或社會現象的本身，其次，模式可說是由一組變項所組成的結構，這些變項之間，具有特定的關係（黃炳煌，1995）。以管理的觀點來看，McLeod（1995）認為模式是抽象的事物，代表許多物體或活動（實體entity），模式可分為以下四種：（一）物質的模式（*physical models*）：以三度空間來表示其實體，如汽車模型；（二）敘事的模式（*narrative models*）：用口頭或書面的來描述其實體；（三）圖形的模式（*graphic models*）：用抽象的線條、符號或形狀來代表其實體；（四）數學的模式（*mathematical model*）：數學的公式或等式即是數學模式。

評鑑模式是一種描述評鑑中各項要素（如基本假定、支持性理論、原理原則、方法、策略、架構、實施程序等）及其相互關係的示範性方法，可以具體表現整個評鑑的本質與特色，用來幫助各類評鑑關係人對評鑑工作和結構的認知，常見的評鑑模式或取向，計有目標導向、參與導向、專家導向、管理導向、消費者導向、抗詰導向等評鑑取向，以及目標本位評鑑、無目標評鑑模式、反應式評鑑模式、差距評鑑模式、CIPP評鑑模式、方案邏輯模式、賦權增能評鑑模式、認可制度、教育鑑賞與評論模式、績效責任評鑑模式、成本效益分析等（曾淑惠，2004）。如以模式的定義來看，目前許多已出現的評鑑模式，可能不能稱作真正的模式，而只是取向或導向而已。

Neurath、Cernap和Morris於1995年指出，數十年以前，任何領域的進步經常直接與「單意語言」（*univocal language*）的存在有關，也就是每一個概念只有一個或唯一的名詞，每一個重要的概念都是如此命名（引自Fitzpatrick, Sanders, & Worthen, 2004），「單意」（*univocal*）表示只有一個意思的，意義上並不模稜兩可的。評鑑領域在語意成份（*semantic*）的發展不全，常需要多餘的贅詞（*redundant verbiage*）來作澄清。大部分的評鑑模式並沒有明確的運用指導原則和程序，是如此模糊，另外，許多評鑑的書面作品（*writings*）並不是明確的、歸納的知識主體，無法發展、操弄、預測和解釋的作用，而且未經過實證的考驗，因此，可說是個人對於評鑑領域的概念，形成某種邏輯的架構（*logical structure*）（Fitzpatrick, Sanders, & Worthen, 2004）。因此，就模式的意義、單意語言、成效的實證考驗等方面來看，現行的教育評鑑模式並不一定符合上述判斷的規準，只能稱為取向、導向或邏輯的架構，但是社會科學的情況不完全相同於自然科學，



不過，這是教育評鑑邁向專業化的道路上，可以努力的課題之一。

（二）CIPP評鑑模式的目的

CIPP評鑑模式強調，評鑑最重要的目的不在於證明，而在於改進，這個想法原來是由Guba在Phi Delta Kappa評鑑研究委員會服務時所提出，Guba說：「評鑑的目的在於改進，不在於證明」（The purpose of evaluation is not to prove, but to improve.），後來，Stufflebeam加以修正這個概念為：「評鑑最重要的目的在於改進，不在於證明」（Evaluation's most important purpose is not to prove, but to improve. Or the most important purpose of evaluation is not to prove, but to improve.），這個改變除了強調評鑑著重改進以外，保留了評鑑作為證明、因果功能之空間（Stufflebeam et al., 1971; Stufflebeam, 2003a）。Guba為1989年出版的《第四代評鑑》（Fourth Generation Evaluation）一書的作者之一，係與Lincoln合著，該書第四代評鑑的觀點常為國內所引用。

Stufflebeam為何會將評鑑的目的作如此界定，可以從CIPP模式中找到答案。CIPP模式使用背景、輸入、過程、結果評鑑，以引導方案、計畫和教材的發展和改進，此為形成性功能；提供績效責任的資訊，為總結性功能（Stufflebeam, 2003b）。就績效責任的功能而言，有溝通、控制、激勵和改進等四種（吳清山，2000），就績效責任的利害關係程度而言，又有高度利害攸關（high stakes）、低度利害攸關（low stakes）之別，高度利害攸關以酬賞表現佳者，表現不佳者給予壓力、懲處，低度利害攸關則只是提供表現的資訊（賴志峰，2003）（高度利害攸關取自於「高度利害攸關測驗」（high-stakes testing），中文翻譯係請教臺大師資培育中心田芳華教授而來）。簡單的說，控制和激勵可大致歸類為高度利害攸關，溝

通和改進則為低度利害攸關。所以，績效責任並不能完全排除「證明」意涵的存在，這也是為什麼Stufflebeam為認為，評鑑「最重要」的目的在於改進，不在於證明，強調評鑑著重改進以外，也保留了評鑑作為證明、因果功能之空間，並未完全排除評鑑作為證明的用途。但是如果評鑑要作為高度利害攸關或證明的用途，應特別審慎，嚴格檢驗評鑑制度本身是否符合專業化的要求，公信力是否足夠，才不致於造成評鑑的誤用。

此外，國內許多文獻常引用，Stufflebeam說：「評鑑的目的在於改進，不在於證明」，恐有誤植之處，此話應該是Guba所言，事實上Stufflebeam所說的是應該是，「評鑑最重要的目的在於改進，不在於證明。」，兩者的差距只有「最重要」（most important）三個字（英文為兩個字）。

（三）評鑑模式的運用

CIPP評鑑模式是目前我國應用最廣的教育評鑑模式，但是對於CIPP模式理念的發展、評鑑的核心價值、四類評鑑在執行上的邏輯關係、四類評鑑適用時機和用途、採用的方法、蒐集資訊的來源與型態等方面的區別，較少加以關注，甚至限於現實的考量，未能落實CIPP模式的理念與理想，實務上可以稱作真正落實運用CIPP模式者，有如鳳毛麟角。我國在移植和大量運用CIPP模式之餘，仍應深切地思考，如何從國外評鑑模式建立的經驗中，建立起適用的評鑑模式和理論典範，考量各項評鑑技術的正用與誤用，方為教育評鑑立論與實務成熟發展深耕與生根的正途（曾淑惠，2004）。

目前盛行於國內之評鑑作法採美國的認可制，此模式在Stufflebeam認為二十一世紀仍可適用的九個模式中，所得的評價最低，Stufflebeam以教育評鑑標準聯合委員會（Joint Committee on Standards for Educational Evaluation）所發展的效用性（



utility)、可行性(feasibility)、適當性(propriety)和精確性(accuracy)作為依據，針對二十個評鑑模式作評估，指出其中的決定/績效(Decision/Accountability)、消費者導向(Consumer Orientation)、認可制(Accreditation)(以上三種適用於改進和績效責任)、應用導向(Utilization-Focused)、案主中心(Client-Centered/Responsive)、民主導向(Deliberative Democratic)、建構導向(Constructivist)(以上四種適用於社會議題和主張)、個案研究(Case Study)、結果監控/附加價值(Outcome Monitoring/Value-Added)(以上二種適用於問題與方法)等九個模式，在二十一世紀最具希望被持續應用的，其中同屬於改進和績效責任用途的決定/績效、認可制，分別名列為最高與最低分(潘慧玲，2004；Stufflebeam, 2000)，因此，評鑑模式本可多元，不當僅限於認可制的採用(潘慧玲，2004)。

目前我國教育評鑑的運用，可說是CIPP模式和認可制的組合，各級學校教育評鑑制度的發展，大多未跳脫這種型態，評鑑模式較為單調，相對的也擠壓了其他評鑑模式或取向的發展空間，其他模式大都停留在文獻建立階段，鮮有實際實施與驗證的機會。

四、教育評鑑未來的發展（代結語）

從國際上的觀點來看，學者曾預測評鑑未來的發展將有以下方向：（一）評鑑在

社會上的效用性將會持續增加；（二）評鑑將擴展至其他領域或學科；（三）評鑑將會更制度化；（四）評鑑將會快速全球化；（五）評鑑的生涯發展機會將會逐漸增加；（六）由於內部評鑑的益處，內部評鑑將變得更加重要；（七）評鑑專業協會和活動將緩慢但穩定的成長；（八）評鑑的量化、質化文獻將會增加(Fitzpatrick, Sanders, & Worthen, 2004)。至於國內教育評鑑之發展方向，早在八年前，行政院教育改革審議委員會的總諮議報告書，已對於教育評鑑提出多項建議，尤其提到教育行政當局，須逐步的從管理角色調整為監督立場，教育評鑑就成為非常重要的工作，在民間興學、教育視導、教育專業自主及教育資源分配之發展上，都有賴於一個良好的教育評鑑制度之建立，評估教育行政機關和學校推動教育改革的成效，同時，建議設置專責機構或委由民間學術單位進行各級學校及各種教育評鑑，以落實評鑑功能(行政院教育改革審議委員會，1996)，而教育界寄予厚望的國立教育研究院，也將規劃成立測驗與評量研究所，教育評鑑研究為其功能之一，以建立教育評鑑的標準和模式，包括學校的評鑑、教師的評鑑、教育動態和改革成效的評鑑等。回顧這幾年來，我國教育評鑑雖然不能稱為突飛猛進，但至少是非停駐於問題原點的過程(nonhomeostatic)，可以看到緩慢但持續的進步，當初教改會所規劃的教育評鑑藍圖，仍然是我國教育評鑑邁向專業化亟待努力的方向。

參考文獻

- 行政院教育改革審議委員會(1996)。教育改革總諮議報告書。臺北市：行政院教育改革審議委員會。
- 吳清山(2000)。學校績效責任的理念與策略。學校行政，6，3-13。
- 曾淑惠(2004)。教育評鑑模式。臺北市：心理。
- 黃炳煌(1995)。試為我國建立一個適切可行之課程發展模式--兼評國內當前之課程決策。國立



- 政治大學教育系舉辦邁向二十一世紀我國中小學課程革新與發展趨勢學術研討會。
- 潘慧玲（2004）。邁向下一代的教育評鑑：回顧與前瞻。國立臺灣師範大學教育研究中心舉辦之教育評鑑回顧與展望學術研討會。
- 賴志峰（2003）。美國2001年No Child Left Behind Act之測驗本位績效責任制度。初等教育學刊，14，135-160。
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). Program evaluation: Alternative approaches and practice guidelines (3rd ed.). Boston, MA: Pearson .
- Gove, P. B. et al. (Eds.)(1981). Webster's third new international dictionary. Springfield, MA: Merriam-Webster.
- Madaus, G. F., & Stufflebeam, D. L. (2000). Program evaluation: A historical overview. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), Evaluation models: Viewpoints on educational and human services evaluation (2nd ed.) (pp. 3-18). Boston, MA: Kluwer Academic .
- McLeod, R. (1995). Management information systems (6th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Mertens, D. M. (1994). Training evaluators: Unique skills and knowledge. New Directions for Program Evaluation, 61, 17-28.
- Scriven, M. (1991). Evaluation thesaurus (4th ed.). Newbury Park, CA: Sage.
- Stufflebeam, D. L. (2000). Foundational models for 21st century program evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), Evaluation models: Viewpoints on educational and human services evaluation(2nd ed.) (pp. 33-83). Boston, MA: Kluwer Academic .
- Stufflebeam, D. L. (2003a). The CIPP model for evaluation: An update, a review of the model development, a checklist to guide implementation. Retrieved July 28, 2004, from Website: <http://http://www.wmich.edu/evalctr/pubs/CIPP-ModelOregon10-03.pdf>.
- Stufflebeam, D. L. (2003b). The CIPP model for evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), International handbook of educational evaluation (pp. 31-62). Boston, MA: Kluwer Academic.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. M. (1971). Educational evaluation & decision making. Itasca, IL: Peacock.
- Worthen, B. R. (1994). Is evaluation a mature profession that warrants the preparation of evaluation professionals? New Directions for Program Evaluation, 62, 3-15.



教育與發展

