

《教育行政與評鑑學刊》

2006年12月，第二期，頁19-42

## 當評鑑遇上教育—教育評鑑意涵探究

郭昭佑

### 摘 要

本文旨在從文獻中歸納教育評鑑的意涵與要素，並探討「教育」與「評鑑」兩個領域在彼此相對位置中的特殊性，以期理解教育與評鑑間的關係及融合後教育評鑑的初構圖像，主要內涵包括：教育評鑑的多元意涵、教育評鑑之意義與要素、評鑑在教育行為中的角色、教育評鑑的特質。

關鍵詞：教育評鑑、評鑑、教育

---

郭昭佑：國立臺灣藝術大學師資培育中心教授

電子郵件：chaoyu@mail.tpc.edu.tw

收件日期 2006.9.14；修改日期：2006.10.30；接受日期：2006.11.23

# **When Evaluation Meets Education: The Exploration of the Meaning for Educational Evaluation**

Chao-Yu Kau

## **Abstract**

This article concluded the meanings and essence of education among the documents, and explored the specific characteristics between education and evaluation. The aim of this article was to understand the relationship between education and evaluation and the composition of education evaluation, mainly including definition and elements of educational evaluation, roles of evaluation plays of educational evaluation in education and characteristics.

Keywords: educational evaluation, education, evaluation

欲瞭解一個事物，必先釐清其概念，否則在不明究理的狀況下執行，易誤解其義，尤其在教育評鑑這個表面即帶有威脅性的領域，甚可能未蒙其利，先受其害。1980年代澳洲在推展課程評鑑時即產生質疑評鑑根源情形：「近來有許多困惑與不安圍繞著整個課程評鑑，其擔憂源自於有關評鑑意義和功能等困惑」（Davis, 1981）。臺灣近年來在教育現場上，評鑑的實踐與影響已甚為巨大，除了在中小學校既有的校務評鑑及各項方案評鑑外，即將登場的大學評鑑已如火如荼展開，而教師評鑑更是蓄勢待發，同樣的不安與困惑氣氛亦籠罩著所有的教育伙伴。然而，在此充滿評鑑氛圍，多數教育工作者都須實施並面對評鑑的同時，又有多少人真正瞭解教育評鑑的意涵，尤其是評鑑遇上教育時的特殊生成？

臺灣自1990年代大量實施教育評鑑以來，大家早已形成一種如何面對評鑑的制式模式—在自我評鑑中「填列訪評表並詳細準備資料」以迎接外部訪評，但又有多少人知道自我評鑑意義為何？如何實施？其次，似乎只要是「學者」便都能成為「合理的」評鑑委員？在扛著評鑑的大旗下，有時教師、甚至家長只要向著「評鑑委員」的光環，也順理成章的成為評鑑人員？而不管他是否理解學校教育？是否理解評鑑？總之，大家都知道「評鑑」怎麼做就可以了…。盧增緒（1995）在論述臺灣教育評鑑現況時，即指出我們的社會在處理教育問題時，似乎一向都停留在道聽塗說，人云亦云，和人人都是教育專家，以及「不知亦能行」的階段。由此可知，在臺灣教育評鑑實施如此頻繁之時，或許應再回頭審視並釐清其意義與內涵後再出發，在一個對教育評鑑適當理解的基礎上，也許在看待與面對教育評鑑的角度亦將有所不同。

為瞭解教育評鑑的基礎與原貌，本文主要以文獻分析方法，在研究架構上，首先從各種教育評鑑定義的文獻中歸納教育評鑑的意涵與要素，以期理解教育評鑑的基礎樣貌；其次，探討「教育」與「評鑑」兩個領域在彼此相對位置中的特殊性，包括「教育『評鑑』—評鑑在教育行為中的角色」（在眾多教育行為中，評鑑行為的特殊性），及相對的「評鑑『教育』—教育的評鑑特質」（在評鑑所應用的眾多領域中，應用在教育領域的特殊性），以期理解教育與評鑑間的關係及融合後教育評鑑的初構圖像。

## 壹、教育評鑑的多元意涵

### 一、字義層面

從字義上看，教育（education）一詞係由拉丁文“educere”而來，意指「引出」；中國說文解字中，「教，上所施，下所效也」及「育，養子使作善也」。綜合來看，教育須循循善誘，啓發受教者心能和人格發展以達至善境地的過程。

而評鑑（evaluate）一詞係由詞幹“value”加上詞首“e”和動詞性詞尾“ate”組成，其中“valu”代表價值，詞頭“e”的意義等同於“out”，即引出之意，故評鑑意即引出和闡發價值與判斷，其內涵與價值的概念密切相關（霍力岩，2000），Stake（2004）亦直指評鑑與價值的相關性。依據韋伯詞典（Webster's third new international dictionary），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）。Stufflebeam與Shinkfield（1985）在《系統的評鑑》（*Systematic Evaluation*）一書中即對評鑑有完整的定義：「評鑑是一種規劃、收集並提供描述性與判斷性資訊的過程，這些資訊包括評鑑客體的目標、設計、實施和影響之功績與價值，以利決策，滿足效能需要，並增加對評鑑客體的瞭解」。

與評鑑用語一樣，在教育評鑑上，亦有相當豐富多樣的語彙及可替換的詞語，像是教育評鑑（educational evaluation）、課程評鑑（curriculum evaluation）、方案評鑑（program evaluation）等，方案評鑑在美國較常用，英國的教育評鑑則幾乎與課程評鑑同義（Bennett, 2003），其定義更可能因時空脈絡而有不同形態。

### 二、各具代表性的定義

Tyler首將教育評鑑定義為「決定教育目標被實現程度的過程」，而後來大部分的評鑑研究者，尤其是北美洲的相關學者，都援引此定義或從這個定義出發，探討教育評鑑的意涵（引自Hopkins, 1997），當然，過去亦有許多知名學者在不同的年代，從不同面向與思維定義教育評鑑，豐實了教育評鑑的意義與內涵，更各具其代表性（詳如表1-1），而Bennett（2003）在檢視這些教育評鑑的定義後認為評鑑至少與下列幾點有關：(1)著重在新教學方案的介紹；(2)收集分析依經驗所得的資料；(3)採集有關這些資料的結論及鑑定；(4)把這些發現傳達給旁觀者；(5)利用這些發現輔助決策制定。

表 1 教育評鑑的相關定義

- 一個決定學校課程教學達成教育目標之程度的過程 (Tyler, 1949, p.105-106)。
- 收集並使用資訊以決定教育課程 (Cronbach, 1963, p.672)。
- 評鑑目的是檢驗學校課程是否達到目標或正朝著目標邁進 (Kerr, 1968, p.21)。
- 評鑑是取得在教學具體目標上的成就證據 (Bloom, 1970, p.28)。
- 衡量教學方法的價值，其過程就是課程評鑑 (Hamilton, 1976, p.4)。
- 教育評鑑是在描述所取得的有效資料，以提供決策參考 (Jenkins, 1976, p.6)。
- 評鑑是對內容加以解釋闡述 (Parlett & Hamilton, 1976, p.84)。
- 評鑑是有系統的檢驗當前的教學方案或隨之發生的事件—目的在協助改善受檢驗的教學方案或其他方向相同的教學方案 (Cronbach et al., 1980, p.14)。
- 評鑑是評估功績及價值的機制 (Nevo, 1986, p.16)。
- 評鑑能提供方法以實踐教育理念，並監督及加強課程發展 (Hopkins, 1989, p.3)。
- 教育評鑑與社會規劃、社會控制有關 (Norris, 1990, p.16)。
- 教育評鑑是在辨清某些變因的影響，這些變因改變了原來應有的結果 (Davis et al., 2000, p.253)。

資料來源：*Evaluation Methods in Research* (p.5), by J. Bennett, 2003. London: Continuum.

霍力岩 (2000) 亦在文獻概覽後提出界定教育評鑑五個代表性的看法，包括 (1) 教育評鑑是以教育為對象，並對其效用給予價值判斷；(2) 教育評鑑是利用所有可行的評鑑技術以評鑑教育所預期的一切效果；(3) 教育評鑑是對照教育目標，對由教育行為而產生的變化所進行的價值判斷；(4) 教育評鑑是人們依據某一社會的教育性質、教育方針和教育政策所確立的教育目標，對所實施的各種教育活動效果及兒童發展水準進行的科學判斷；(5) 教育評鑑是系統的、有步驟的從數量上測量或從性質上描述兒童的學習過程與結果，據此判斷教育是否達到所期望教育目標的一種手段。

### 三、詳細列舉要素的定義

除了各種不同且各具代表性的定義外，更有許多學者於定義教育評鑑時，詳細列舉並說明其要素：

蘇錦麗 (2004) 即認為教育評鑑係「有系統的採用各種有效方法，收集質與量的資料，對照評鑑準則 (指標或標準)，以判斷任一教育對象之價值或優缺點的

過程，並將其結果作為教育決策之參考」，其重點含括：

- (1) 評鑑係採系統化、科學化的各種方法，以收集資料的過程。
- (2) 評鑑資料應自多方面廣泛收集，且宜質量並重。
- (3) 評鑑必須依據準則（指標或標準），進行價值或優缺點的判斷。
- (4) 評鑑結果應作為相關決策與行動的參考依據，換言之，評鑑係以解決教育政策與教育實務之問題為導向。
- (5) 教育評鑑對象非常廣泛，包括：國家教育目標、政策、課程、教材、學生學習成就、學校校務、校長、教師、教育行政人員、教育專家、教育計畫、教育方案、後設評鑑等，上述教育範疇亦包括各級學校層級的評鑑。

金娣與王剛（2002）則認為「教育評鑑是指在系統的、科學的和全面的收集、整理、處理與分析教育資訊的基礎上，判斷教育價值的過程，目的在促進教育改革，提高教育質量」。其要點包括：

- (1) 教育評鑑的對象（即價值判斷的對象）可以是教育領域中的任意元素，即可以是教育的參與者（人物），也可以是教育現象和活動（事務）。
- (2) 教育評鑑的本質（即評鑑的主要性質）是對教育的價值作出判斷，是評鑑者的主體需要及被評鑑對象的客體屬性間一種特殊的效用關係與運動。
- (3) 教育評鑑的手段（即評鑑的方法和技術）是運用科學的評鑑技術和方法。
- (4) 教育評鑑的目的（即評鑑活動所要達到的境地）是為了促進教育改革，提高教育質量。

楊振昇（2001）則認為「教育評鑑係指針對教育現象或活動，透過收集、整理、分析、歸納等方式，提供給相關的決策人員，以作為價值判斷與決策之參考」。其要點包括：

- (1) 就範圍而言，教育評鑑包括相關的教育現象與活動，諸如校務評鑑、政策評鑑、方案評鑑、校長評鑑、教師評鑑、課程評鑑、與後設評鑑等等，均屬教育評鑑的範疇。
- (2) 就資料收集的方式而言，應重視多元的原則，尤應兼顧質與量資料的收集與彙整。
- (3) 就教育評鑑的人員而言，專業化的涵養、客觀、公正、超然的立場，以及適當的人格特質，均攸關教育評鑑成效的良窳，不可忽視。
- (4) 就教育評鑑的目的而言，則在藉由提供可靠、有用與充分的資訊給決策人員，使其能在各種變通方案中，做成合理可行決策，以達到促進教育發展的目標。

胡中鋒與李方（1999）則指出教育評鑑雖眾說紛云，但應有某些共同特點，其定義教育評鑑為「依據一定的教育價值觀或教育目標，選用可行的科學手段，透過系統的收集資訊、分析解釋，對教育現象進行價值判斷，為一不斷優化教育及為教育決策提供依據的過程」，並認為教育評鑑的共同特點為：

- （1）評鑑是一個過程。
- （2）價值判斷是評鑑的本質特徵。
- （3）以一定的教育價值觀為依據。
- （4）採用一切可行的科學手段。

鄭新輝（2002）則定義教育評鑑為「對教育方案或活動的表現品質，依預先設定的評鑑標準，有系統的運用科學技術，正確的收集與分析資料，客觀的判斷其所展現的價值與功績，並作適當處理決定的歷程」，其定義包括下列五個概念：

- （1）在「性質」上：是一種價值判斷的活動。
- （2）在「對象與內容」上：是以教育方案或活動為對象，包括該方案或活動的目標，歷程與結果的表現品質。
- （3）在「歷程」上：是一項有系統且複雜的歷程，其成效受多種因素所影響。
- （4）在「方法」上：是依設定的目標或標準，有系統的運用科學技術，正確收集資料的一種測量與分析。
- （5）在「目的」上：是客觀的判斷受評對象所呈現的價值與功績（merit），並作適當處理的決定。

#### 四、重點強調性的定義

上述學者透過詳細列舉要素的方式定義教育評鑑，並提出其內涵要素，極具參考價值。當然，亦有些教育評鑑的定義係採較為簡要的方式或在時空脈絡下特別強調某些要素。例如有些學者特別著重評鑑的本質，尤其是判斷。Hopkins（1997）即從評鑑的描述與判斷本質，詮釋教育評鑑「是一種對於教育目標有系統的描述，和/或對於他們的優缺點與價值的評估」。李緒武（1993）則特別強調價值判斷，認為「教育評鑑是對教育事業的價值作嚴正的品評，不僅是靜態資料數字的檢查，更是對一項教育計畫、教學過程、課程與教材，或教育政策的價值鑑定」。日本心理學家大橋正夫亦有相似的看法「教育評鑑就是依據教育目標，對教育行為產生的變化進行價值上的判斷」（引自胡中鋒、李方，1999）。

陳玉琨（2004）認為除了判斷本質外，亦重視社會與個體需求，定義「教育評

鑑是判斷教育活動是否滿足社會與個體需要程度的活動，是對教育活動已實現的（已經取得的）或潛在的（還未取得，但有可能取得的）價值作出判斷，以期達到教育價值的過程」。

方法，似乎是教育評鑑定義的另一個重點。秦夢群（1997）以方法與本質為主，精簡的定義教育評鑑為「對於教育現象或活動，透過收集、組織、分析資料，加以描述與價值判斷的歷程」。Nevo（1995）亦有類似的定義「教育評鑑為收集教育客體本質與品質的系統性資訊，並應用研究方法予以描述與判斷之活動」。林天祐與蔡菁芝（2001）則側重方法與目的，定義教育評鑑為「系統化地收集資訊，以為促進、達成教育目標的評鑑過程」。Popham（1993）在教育評鑑一書中從反向的非正式評鑑概念定義「系統化的教育評鑑」係「屬正式性質的評鑑，即對教育現象品質的正式評估」，品質評估，或可稱為價值的確定，而被評估的教育現象包括教學結果、教學計畫、教育產出或教育目標的達成等。

## 貳、教育評鑑之定義與要素

從以上林林總總的教育評鑑相關定義可知其複雜性，惟從這些定義中仍可窺見各要素的共識與分歧程度、其間的關係及有爭議的地方，本文嘗試依較有共識的要素、尚須討論的要素及仍待釐清的用語等三部分臚列，並藉以探究教育評鑑的概略定義，期更能貼近教育評鑑的真實意涵：

### 一、教育評鑑定義較有共識的要素

上述各學者對教育評鑑的定義各有聚焦重點，並未具有一致性，但仍可從其內涵中分析其要素，其中較有共識的要素包括：

（一）教育評鑑係一「歷程」

或稱為過程、或活動、或持續性、系統的歷程…。

（二）教育評鑑的歷程須「應用方法」

不管是資料的收集或分析，都須要方法的應用，或稱手段、評鑑技術、研究方法、系統的方法、科學的方法、有效的或所有可行的方法…，包含資料的收集、整理、分析、組織、歸納…等用語。

（三）「教育客體」的廣泛性



即所有的教育評鑑客體，或稱教育現象、教育活動、教育方案、教育事業…等，包括各相關客體的任意元素、客體之本質與品質…。

#### （四）教育評鑑須「提供資訊」

或稱資料，包括質的與量的、廣泛的、多元的、可靠的、有用的、充分的資訊…。

#### （五）教育評鑑最終須「予教育決策參考」

予決策者參考、行動參酌、決策制定、教育決策依據…。如依上述較有共識的要素組合，教育評鑑的簡要定義為「應用方法以提供教育客體資訊予決策參考的歷程」。至於提供什麼資訊、是否須有評鑑依據、可做什麼決策等，這些要素是否要放入教育評鑑的定義中，仍有討論的空間。

## 二、教育評鑑定義尚須討論的要素

除了上述較有共識的要素外，下列要素雖在許多教育評鑑論述中屢屢出現，但置於定義中是否適當仍有討論空間。

### （一）描述與判斷的本質應可置入

評鑑須提供什麼資訊予決策參考？這與評鑑活動的本質有關，評鑑在本質上是一個描述與判斷的活動，包括「資料收集與分析」的描述本質，及「功績與價值評估」的判斷本質，都是評鑑須提供予決策參考的資訊，然而判斷本質卻有些爭議，Cronbach（1963）「評鑑的主要意圖在改進而不在證明」的陳述間接否定了判斷本質，然而多數人仍同意描述與判斷為評鑑的重要本質（Gronlund, 1981; Guba & Lincoln, 1981; Kemmis, 1989; Nevo, 1995; Stake, 1967），只要敘明資訊來源，應可置入教育評鑑定義中。

### （二）評鑑目的是否置入仍須考量：

評鑑係提供資訊予決策參考，而主要的決策多含括在績效責任或發展改進兩個評鑑目的上，然而，在各教育概念的定義中，目的是否置入一直沒有定論，雖有些學者將目的置入其定義中（金娣、王剛，2002；陳玉琨，2004），惟多數學者最多僅列提供決策參考（胡中鋒、李方，1999；楊振昇，2001；蘇錦麗，2004；Alkin, 1969；Cooley & Bickel, 1986；Cronbach, 1963；Jenkins, 1976；Nevo, 1995；Stufflebeam & Shinkfield, 1985），並未清楚述及目的，因此評鑑目的置入定義仍須考慮。

### （三）資料的屬性在操作型定義時或可考量置入

收集與分析什麼樣的資料，或為質的與量的？或為多元、有效等概略性陳述？因資料的屬性向度甚多，無法窮盡，建議不置入定義中，但有關質的與量的資料，如係操作型定義時，為使閱聽者理解教育評鑑的操作定義，則可考量置入教育評鑑的定義中。

（四）評鑑參照在操作型定義時可考量置入，惟避免強迫性用語

有關評鑑時的參照或依據，如指標、規準、標準、準則、教育目標等，是正式評鑑實務中相當重要的元素，惟在許多評鑑歷程中不一定有清楚的參照依據，因此不一定要置入定義中，但如果在正式評鑑的操作型定義中應屬適當，只是仍宜以彈性用語表達，如「參酌」評鑑指標…，而不宜以「依據」、「依照」等強迫性用語，這將限制評鑑資料的來源，甚至遺漏重要的評鑑訊息。

### 三、教育評鑑定義中仍待釐清的用語

（一）資訊（information）與資料（data）

從許多教育評鑑的定義中，常見資訊與資料用語穿插其間，事實上，兩者宜釐清其概念，以免閱聽者無法區隔或誤解。一般來說，資料須經分析處理程序（process）才能成為資訊，兩概念須有適當的區隔。

（二）指標（indicator）、標準（standard）及規準（criterion）

在評鑑的參照依據上，指標、標準與規準經常混用。事實上，不同國家地區或不同學者即有不同的詮釋與使用，且在臺灣的翻譯上亦無固定譯法，類此語詞未有明顯的釐清，互相通用並沒有什麼問題。不過從西文原意、中文的可能對照及應用習慣上，可看出一些端倪：「指標」較屬評鑑的指引（index），即評鑑應關注什麼的問題，常用的指標系統則為指標間的組合，如某評鑑有六大項32個指標（甚至每個指標內涵數個參考重點），即是指標系統的用法。

「標準」依韋氏辭典其中一段論述「秉權責設定或建立有關數量、權重、範圍、價值或品質的測量『原則』」，（something set up and established by authority as a rule for the measure of quantity, weight, extent, value, or quality）（Gove, 1981），可知標準較屬後設的、原則性的評鑑參照，如 Stufflebeam 所領導的教育評鑑標準聯合會議在1981年即發展出一套目前最被認可的標準，包括有效性（確保評鑑提供實用的資訊需求）、可行性（確保評鑑是真實且謹慎的）、適當性（確保評鑑能正當且合理的執行）及準確性（確保評鑑資訊的適當揭露及有技巧的傳達）等四個面向30個方案評鑑的標準，用來判斷方案評鑑的良窳。

在韋氏辭典中「規準」的詮釋為「在判斷或決定時可依據的標準」(a standard on which a judgment or decision may be based)，其意涵解讀或有不同，但似乎較接近「在判斷或決定教育客體表現如何時可依據或參照的標準」，像是優、良、可、劣或待改進，類此評等的內容，例如表現在什麼條件下應為優，什麼條件時為良…，諸如此在判斷時所依據或參照的條件內容的標準，可以規準稱之。不過上述區隔僅提供參考或彼此溝通的平台，事實上在不同脈絡下即可能有不同用法，毋須強為一致規範。

#### 四、教育評鑑定義初探

依循上述討論，本文對教育評鑑的嘗試性定義如下：「教育評鑑係應用方法收集與分析教育客體資料，並評估其功績與價值，以提供相關資訊予決策參考的歷程」。依上述定義，在此歷程中有幾個要素：

##### (一) 方法應用的必要性

人們常常會因為直覺、漫談，有時甚至只是道聽塗說，或是以名氣來評論一間學校 (MacBeath & McGlynn, 2002)。Nevo (1995) 即認為評鑑不能只架構於直覺的基礎上，而是一個提供原則、方法與工具的系統歷程。在許多教育評鑑的定義上均提及研究方法的重要性，Austin 及其同事即指出「評鑑在下列兩種情況是有用的，其一是確認我們的懷疑或直覺時，另外就是讓我們終於弄清楚評鑑是如何運作的時候」，在這兩種情況中，我們要做的就是減少不確定的變因，而非妄下定論 (引自 Brainard, 1996)。評鑑雖無法做到全然與完整的理解，然而適當的應用方法，可讓資訊的收集、描述與判斷的過程與結果更接近評鑑客體的真值或全貌。

##### (二) 評鑑客體為教育客體

教育評鑑的客體為教育客體，實質上包括國家教育目標、政策、課程、教材、學生 (含學習成就)、學校整體或校務、校長、教師、教育行政人員、教育專家、教育計畫、教育方案、後設評鑑等，上述教育範疇亦包括各級學校層級的評鑑 (郭昭佑, 2000; 蘇錦麗, 2003)，林林總總，歸納而言，約可概略分為教育組織 (如學校整體或校務、教育局、教育部等)、教育人事 (如校長、教師、教育行政人員等) 與教育方案 (如教育計畫、課程、教學等) 三大部分。

##### (三) 描述與判斷的評鑑本質

描述與判斷是目前最常被討論的評鑑本質。描述，係收集系統資料經分析而成為較客觀的資訊；判斷，則多在若干個案中，依價值、社會規範及利害關係人

之偏好而評定之較主觀的資訊；評鑑被視為一種包含描述與判斷的活動是有相當共識的（Nevo, 1995），教育評鑑亦同，一為對教育客體「資料收集與分析」的描述本質，一為「評估其功績與價值」的判斷本質。

#### （四）提供資訊以為決策的歷程

Brainard（1996）指出在許多方案評鑑中缺乏廣泛資訊的提供，且所提供資訊又未能清楚描述方案活動，評論結果又不具說服力，那些未深入接觸方案的決策者如何做決定？Bennett（2003）從有關教育評鑑陳述中擷取共同概念，發現評鑑應收集與方案及決策的相關資訊。由此可知，教育評鑑是一種提供資訊以為決策的歷程，所提供資訊依評鑑本質約可分為二類，一屬描述本質的「資料收集與分析」資訊，另一為判斷本質的「功績與價值評估」資訊。

## 參、教育「評鑑」—評鑑在教育行為中的角色

自有正式教育以來，評鑑即扮演重要角色（Dochy, 2001）。在教育領域中，有相當多的行為，包括教學上的、行政上的各種行為，評鑑僅是其中之一，與其他教育行為比較起來，評鑑所扮演的角色自然有所區別：

### 一、管理的角色

Brandt 與 Tyler（1983）將教育作為區分為三種類型，包括教學的、支持的和管理的作為：教學作為目的在學習的獲得；支持作為目的在服務的提供；而管理作為目的則為管控的功能，例如擬定計畫，執行與評鑑，在此三種教育作為不同目的之結構上，評鑑將聚焦於學習獲得的測量（教育結果），質與量的測量之服務與傳遞（支持結果），及測量的品質及有效的管控功能（管理結果）。此一詮釋將評鑑定位在教育結果的測量、傳遞與有效管控上。

此一管理角色在教育行政上來說，完整的教育行政歷程普遍劃分為計畫、組織、領導、溝通、評鑑等五項歷程，而為確保教育行政作為能符合預期目標，因此對於行政運作的歷程與結果，須藉由系統性的評鑑來獲取各項執行資訊並不斷檢討改進，以強化目標達成的效能（鄭新輝，2004）。

當然，評鑑在教育領域中的管理角色不僅出現在行政上，在課程與教學上亦同。Tyler 由於目睹了1940年代美國課程發展的不良方法而感到灰心，因此於1949

年出版「課程與教學的基本原則」，提出一個依據四個問題的課程設計方法：目標、內容、組織、評鑑，其後被稱為目標模式；其邏輯方法是經由那些能判斷目標達成與否的行為改變（通常是學生測驗分數或考試結果且目標須以行為術語表示），以使得成就能夠被觀察或評鑑，是一種工具性和成果導向的評鑑形式（引自Hopkins, 1997），更是一種課程與教學的管理角色扮演。

## 二、導引的角色

評鑑除了在既定的目標下管理相關的教育行為以朝目標前進外，評鑑更有導引改進與發展此一具有即時性與未來性的角色扮演。陳玉琨（2004）認為評鑑能直接指向正在進行的教育活動，與實際活動緊密結合，導引教育活動之改進，有助於直接的、有效的提高教育與教學品質，更有利於教育工作者確立對評鑑的肯定態度，進行積極性的教育改革。

除了導引教育過程中的改進以符應教育目標外，評鑑更扮演導引未來發展與改革的角色。Chelimsky（1997）即認為從內部成員或受評鑑者的立場來看，發展性的評鑑可以是一個極具彈性且具有回顧及展望的工具，發展並增進學校效能。Bennett（2003）亦指出大多數對教育有興趣的人多贊同結合評鑑與革新的概念是重要的，且經由教育評鑑的歷程，使得教育客體的發展更為趨完善是教育評鑑的重要目的。Hopkins（1997）更直指評鑑必須扮演引導發展且與發展統整起來的重要導引角色。

## 三、工具與跨越學科的角色

Scriven在1994年主張評鑑學科除可以視為普通學科外，亦可以視其為工具學科與跨越學科（transdisciplines）。評鑑學科旨在發展一套工具，以供其他學科應用，故可視為一門「工具學科」，此外，它與統計學、測量學、邏輯學一樣，並不探討

某一特定領域的內涵，故為「跨越學科」。根據Scriven的觀點，成為跨越學科必須具備兩個條件：一是以服務其他學科或與其他學科合作為目的；另一是除了一般學科所須具備的核心成分（或稱學術核心，包括該學科理論的一般議題、方法論、本質、概念、範圍、變項間的關係與邏輯等）外，尚須包括應用領域，而評鑑的核心成分即成為其應用領域的基礎，基於科際應用的觀點，稱評鑑為一門「跨越學科」是無庸置疑的（引自蘇錦麗，2004）。評鑑在各領域間的應用相當普遍，即使在教育領域中，不管在行政或課程教學都能見其蹤跡，Halcolm《評鑑神諭》

(*Evaluation Oracles*)「特別的禮物」說得真切(引自吳芝儀、李奉儒,1995)

「請再次告訴我吧！師父！在混沌初開時，究竟是如何呢？」

「在混沌初開時，特別的禮物被贈給不同群體的人們。那些要付出關懷的人，被賦予對不幸者的憐憫、慈悲；工程師被賦予對事物的敏銳視察力；木匠被給予耐心，以畫出直線和完美的角度；工業技術人員被贈予勤勉，以使其能誠實不貳地遵循藍圖和他人的詳細指導；實驗科學家則被給予特定的信仰，相信世界可以依照他們的想像加以操控……。終於，剩下最後一個群體和最後一份禮物，這是探險家們。他們被贈予好奇心的天賦，使他們得以永遠探索這個世界，發現這世界上的許多奇蹟。」

「但是，評鑑者所擁有的是什麼呢？」孩子們好奇地問到，「你還沒有提到他們的特別禮物。」

Halcolm 微笑著。「親愛的孩子們，評鑑者散布在所有的其他各團體中，每一位評鑑者都被賜予屬於他們團體的特別禮物，而且每一位都以特別的方式來使用其天賦。」

「但是這難道不會造成評鑑者間的許多爭論：到底誰擁有了最特別的禮物？」

Halcolm 微笑不語。

## 肆、評鑑「教育」—教育的評鑑特質

除了瞭解評鑑與其他教育行為的角色區隔外，相對的，亦應瞭解評鑑應用於教育與應用於其他領域時有何不同。盧增緒(1997)即明白指出教育評鑑觀念之形成，當以教育觀念之形成為基礎，許多教育評鑑者，每競以評鑑概念之剖析為尚，全不悉教育為何物者，侈言正確教育評鑑觀念之樹立，直若緣木求魚，如果我們只談評鑑，而不談教育，就不再是教育評鑑了。

評鑑被廣泛應用在許多領域中，例如 Fisher 在1935年即成功的以「農業植物學」的設計，確定植物生長的因果關係，其基本觀念係以隨機抽樣的實驗結果推論母群體的狀態，當時許多教育學者開始援引 Fisher 的模式，在學校的教學及課程評鑑中應用類此實驗方法，其假設是：一種教學程序或課程應可透過隨機抽樣實驗加以測試，瞭解其應有的成效以為推論並做決定，就如同在農業上經由實驗

選擇種子或肥料一樣，即能獲得可預測的結果與利潤，然而，此一農業模式的援引取向對於教育研究來說是有問題的，特別是當它的結果要應用於個殊性相當高的教室裡時（引自 Hopkins, 1997）。事實上，評鑑在教育領域的應用，自有別於其應用在其他領域之處，主要約可歸納如下（王保進，2002；郭昭佑、閻自安，2004；黃玉幸，2004；鄭崇趁，1995；Hopkins, 1997；Kellaghan & Madaus, 2000）：

## 一、教育目的在人的成長

這是教育評鑑有別於其他領域評鑑行爲的特點，在許多領域，如企業界，在行政運作中或有評鑑行爲，但都只是爲達成企業目的之手段而已，對象主要爲事物標的；然而對從事教育工作的機構或人員而言，則應把人的成長之教育目的置於首要考量，離開了這一點，教育評鑑即失去它應有的意義，因此在教育評鑑中，除了追求績效責任或導引發展改進外，更應考量評鑑行爲可能引起什麼教育作用，對人產生什麼教育影響。

況且，教育不能像企業界中對員工的要求一樣，在評鑑後擇優汰劣甚至選擇成員，而應秉持有教無類，因材施教原則，不放棄任何受教對象，以保障學生的受教權益，這才是教育的本質與目的。陳善德（1998）即認爲像教育這種介入每個孩子成長過程中「無法重新來過」的事業，的確須要品質的確保，學校不是在提供教育訓練的「服務」而已，學校是在「培育」達到教育目標的孩子，必須「把每一位都帶好」。

## 二、不能要求規格化的教育品質

教育不在印出一個個相同模子產生的產品，而是一種形塑自由個體的創造性活動，期望營造最適切的學習情境，培養出具有不同思維的獨特個體。Mill 的理想社會在於每個人的自由獲得保障，在平等的自由氣氛下以討論的方式追求進步，個人個性才有機會發展，社會才有源源不絕的動力，進而達成民主社會的教育目標（引自黃玉幸，2004）。身爲教師，所關心的不應是班級或學校的總分數，重點在以適當的教學方法引領學生個別發展，以幫助他們突破潛能的限制，發揮與他人不同的獨特人格與能力。

Kells 認爲教育機構的目標經常是複雜且多元的（王保進，2002）。Stenhouse（1975）更直指「教師就如同園丁，以不同的方式來對待不同的植物，而不像大規模耕作的農夫，成爲一個盡可能以標準化方式來對待標準化植物的管理者」。據

此，教育是一個形塑自由個體的創造性活動，不能要求受教者呈現規格化的一致性品質，教育評鑑更不應以同樣的一把評鑑尺規，丈量每一個教育客體的表現。

### 三、教育影響因素錯綜複雜

過去教育評鑑者之所以受歡迎，或許是因為它被認為較容易使用，符合那些想瞭解目的一手段、因一果、線性和科層體制世界觀之政策制定者的需求，然而這個取向忽略了人類與其他生物（如植物）之間的基本差異；Lawton 即舉例說「人類的表現在觀察之下會有不同，但甘藍菜卻不會」（Hopkins, 1997）。Kells 亦認為大學並不像營利組織，具有成本控制、目標明確、線性、連續，以及可量化等單純特徵（王保進，2002），Hopkins（1997）更指出教育領域不像生產業界，可以預先設計好生產流程，電腦儀控，質量檢測，整個生產線是固定且可操控的。

由此可知，評鑑「教育」不可避免的將受到許多因素影響致無法精確的判斷，Kellaghan 與 Madaus（2000）即認為在教育評鑑中，即使評鑑者竭盡所能，對方案活動的描述仍非常困難，因為許多方案的影響脈絡極為複雜，Lawton（1983）更指出即使是具有高信效度的測驗，其結果的應用亦不免遭受實驗立論錯誤的質疑，因為教學過程及情境係植基於不同假設，因此無法真實比較。

在教育歷程中，影響產出的，除了學校教學因素外，學生個人、家庭、社會等環境因素都須考量，許多變項都是無法實驗操控的，如學生家長社經地位背景、整體社區文化、教師個性、學校風氣等，都會影響教育結果。而這些變項既不能忽視，且無法控制，更遑論為了實驗而隨機抽樣了，這在教育情境中是不許可的。況且上述變項並不具獨立性而是交互影響的，例如學校—教師、教師—學生，或學生—學生間都是彼此互動的，因此其教育結果並不只是單純的某種標準化的教學方式所產生的，而是、學校、教師與學生同時致力於有意義的教學行動所產生的學習或教育結果。這在評鑑上將很難釐清其間關係，藉以找出可能的問題癥結或判斷權責歸屬。

### 四、「教育」的專業慣性

評鑑係一看待人事物的過程，其中包括看待與被看待雙方。教育工作主要在教師的教學，而相對於受教者—學生而言，教學係一教人的專業工作，多為看待他人的過程，較少成為被看待者，此一教學專業所形成看待他人、評斷他人的職業慣性，一旦面對評鑑，立即陷自身於直接或間接被看待的情境；這種教育專業



慣性與其他領域人員比較起來，將更不容易接受被看待的過程，亦將存在著更多的不安與焦慮，感受來自他人對自己長期封閉專業知識的挑戰，甚而產生消極應對或拒斥的心態。陳美如與郭昭佑（2003）認為即使是「課程評鑑」這一個教師關心且較不直指教師自身的教育評鑑議題，仍然無法讓第一線教師對評鑑的畏懼釋懷。由此可知，當評鑑遇上教育的專業慣性，其間的火花將是耐人尋味的。

## 五、教育服務對象未成熟且多有代理人

教育是使人成熟的過程，其服務對象多是未成熟的個體，如以顧客的觀點而言，除了教學的直接顧客，共同關心教育的良窳（郭昭佑與閻自安，2004），因此教育評鑑除須對學生受教權利負責外，更須對其家長甚而社區負責，在評鑑過程中更應瞭解代理人的需求及其看法。

其次，由於學生係未成熟的個體，師生關係亦與一般社會組織中人的關係不同。在企業體中，組織與員工的關係清楚明確，所有規範亦可以現行法令與組織契約處理；惟學生與軍人相同，在行政法中都屬於特別權力關係，有些自然人得享的權利義務，並不見得適用於師生間的關係，例如學業評量中有關成績認定的教學專業權，或是學生行為考評中違規行為時以校規而未依法處理等，終究學校是教學的單位，在評鑑上須考量學生的未成熟性、可塑性及可教育性，不可動輒刑責相加。

## 六、教育成效常未能立竿見影

在國際品質系統標準 ISO 9002有一段詮釋：「凡製造結果無法經由隨後的產品檢驗與測試，得以完全查證者，例如製程不當所產生之缺失僅能在產品使用時方可顯現者，此等製程通常被視為「特殊製程」（special process）」，類此特殊製程的進行，必須嚴格執行下列管制以確保品質（引自陳善德，1998）：（1）製造過程與設備應事先檢定其能力並經核准；（2）由鑑定合格的專業人員執行；（3）製造的過程應有監控與管制；（4）製造過程作業（包括相關設備與人員）之任何驗證要求應予以規定；而教育正符合這樣的特性，教育品質的良窳無法在教育過程中完全得知的，有些甚至須等到孩子長大進入社會與他人互動後，才知道教育內涵在孩子身上產生什麼效果。

評鑑通常透過某些措施或手段瞭解並追求組織的最大利益，而這些利益是可

以經由實證檢測的。但如將教育視為特殊製程，有些成效並無法立即在教育現場呈現，即使透過測驗，成績亦僅能顯示部分教育成效，而有關學生的行為、認知、性格、態度等影響，不但無法立竿見影，且常須從未來表現中方見真章。Hopkins (1997)即認為有效能的教室或學校表現是很難建立準則加以判斷。陳善德(1998)亦指出教育成果的驗證，可能是教育活動中最難突破的瓶頸。正所謂「十年樹木，百年樹人」，教育成效實未能立竿見影，而教育評鑑的信效度亦終將受到考驗。

## 伍、當評鑑遇上教育—代結語

評鑑是人類存在以來，日常生活中很普遍的行為，在教育上的應用亦由來已久，且在1960年代以後廣受教育學者與社會大眾重視(秦夢群, 1997; Marsh & Willis, 1999)。然而，理解與行動相輔相成，人們對於教育活動的實踐應有基礎的理解，否則即可能失去方向與意義，尤其是「評鑑」這種具有表面威脅性的活動…。台灣的教育評鑑實施正面臨此種狀況，許多參與教育評鑑的人們並不真正瞭解其意涵，身在其中而不明其意。

因此，本文目的在釐清教育評鑑的意涵，並探討其角色及特質。然而教育評鑑的立論基礎人言人殊，且在教育現場的實踐上，亦因時因地而多元多樣，本文僅能究其梗概，略探一二。從上述有關教育評鑑的意涵、角色及特質，歸納而言，當教育與評鑑相遇，所激盪出的「可能」如下：

### 一、評鑑對教育的重要性

從定義上來看，教育評鑑係分析與評估教育客體以提供決策參考，可見評鑑結果極可能影響教育決策，甚為重要；其次，教育評鑑扮演導引的角色，具引領教育客體改進與發展的張力，對教育客體而言，其重要性不言可喻；第三，評鑑的客體既是教育，其特質包括服務對象未成熟且多有代理人，且目的係在人的成長，評鑑教育這種活動涉及人的可塑性且無法再重來，因此應極慎重為之。

### 二、評鑑在教育中隱涵的壓力及威脅性

從評鑑在教育中的管理角色來說，原本就具有壓力的性質；而在定義上，教育評鑑須評估教育客體的功績與價值以提供決策參考，這對受評的教育客體來

說 自身的表現將被清楚的檢視，更存有無形的壓力與威脅；尤其加諸於教育這個領域，教育有著長期的專業慣性，一向是觀察學生並判斷其學習成效之「教育他人」的專業角色，在教育評鑑中則須轉換為被觀察與判斷的客體，其壓力與威脅感是倍增的。

### 三、評鑑「教育」的複雜性

評鑑係屬工具與跨越學科的角色，在教育的應用上自有其不同的特質；教育評鑑定義中，首須收集與分析教育客體資料，但與其他受評客體不同的是，影響教育客體的因素錯綜複雜，有些甚至無法釐清，而教育成效亦未能立竿見影，無法獲得立即的產出，因此，如何評估教育客體的功績與價值是一件困難的工作，更何況教育的產出不像其他領域要求一致性的品質，教育的對象是人，其形塑須依循個體的需求而不應要求規格化的品質。由此可知，評鑑教育此一客體時，其複雜性是相當高的。

### 四、方法應用的關鍵性

由以上論述可知，教育評鑑雖重要，但相當複雜不易評鑑，且具壓力與威脅性。因此，教育評鑑定義中「收集與分析教育客體資料並評估其功績與價值」的「方法應用」即顯得格外重要。擴張其概念，方法包含資料收集與分析、功績與價值的評估、及制度的規劃與策略的應用。

在此，本文願借用教育評鑑聯合委員會（Joint Committee on Standards for Educational Evaluation）所發展的「方案評鑑標準」（The program evaluation standards），在評鑑方法的應用上，不僅應達到「精確性」以儘量反映受評教育客體不易釐清的品質；並具「有效性」，使教育評鑑確能提供所需要的資訊；「可行性」亦是重要的，以確保其實際、政治及成本效益之可行；除此之外，更應考量

「適當性」，以合乎倫理並確保評鑑參與者及受評鑑影響者的福利，降低教育評鑑的壓力與威脅性。

如此，當教育與評鑑相遇，將不再是一場黑洞抓黑貓，看不清真像，或是貓抓耗子，你追我跑的遊戲，透過精確的、有效的、可行的及適當的方法應用，將使教育評鑑成為教育發展的有效機制。

## 參考文獻

- 王保進（譯）（2002）。大學自我評鑑（原作者：H.R.Kells）。臺北市：正中。
- 吳芝儀、李奉儒（譯）（1995）。質的評鑑與研究（原作者：Q. P. Michael）。臺北市：桂冠。
- 李緒武（1993）。教育評鑑的意義與發展。載於伍振鶯（主編），教育評鑑（頁1-12）。臺北市：南宏。
- 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。
- 金娣、王剛（2002）。教育評價與測量。北京市：教育科學出版社。
- 胡中鋒、李方（1999）。教育測驗與評價。廣東：廣東高等教育出版社。
- 秦夢群（1997）。教育行政：理論部分與實務部份。臺北市：五南。
- 陳善德（1998）。良性互動機制下的基礎教育改革。教育資料集刊，**23**，313-340。
- 陳美如、郭昭佑（2003）。學校本位課程評鑑—理念與實踐反省。臺北市：五南。
- 陳玉琨（2004）。教育評鑑學。臺北市：五南。
- 郭昭佑（2000）。學校本位評鑑。臺北市：五南。
- 郭昭佑、閻自安（2004）。學校行政。載於秦夢群主編，教育概論（頁415-454）。臺北市：高等。
- 黃政傑（1993）。課程評鑑。臺北市：師大書苑。
- 黃玉幸（2004年5月）。淺析教育評鑑：彌爾自由論觀點。發表於國立臺灣師範大學教育研究中心主辦，「教育評鑑—回顧與展望學術研討會」。
- 楊振昇（2001年11月）。析論當前教育評鑑之困境與前瞻—鉅觀觀點之分析。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」。
- 鄭新輝（2002）。國民中小學校長評鑑系統之研究。臺北市：國立政治大學教育研究所博士論文，未出版。
- 鄭新輝（2004）。教育行政。載於秦夢群主編，教育概論（頁356-414）。臺北市：高等教育。
- 霍力岩（2000）。學前教育評價。北京：北京師範大學出版社。
- 盧增緒（1995）。教育評鑑初探。師大學報，**30**，115-148。

- 盧增緒 (1997)。教育評鑑的問題與趨向。載於伍振鷺主編，教育評鑑 (頁13-38)。臺北市：南宏。
- 蘇錦麗 (2003)。談教育評鑑之專業性專訪。教育研究月刊，112，31-36。
- 蘇錦麗 (2004年5月)。高等教育評鑑的趨勢與展望。發表於國立臺灣師範大學教育研究中心主辦，「教育評鑑—回顧與展望學術研討會」。
- Alvik, T. (1995). School-based evaluation: A close-up. *Studies in Educational Evaluation, 21*, 311-343.
- Bennett, J. (2003). *Evaluation methods in research*. London: Continuum.
- Brainard, E. A. (1996). A hands-on guide to school program evaluation. Bloomington, IN: Phi Delta Kappa Educational Foundation.
- Brandt, R. S., & Tyler, R. W. (1983). Goals and objectives. In ASCD (Ed), *Fundamental curriculum decisions* (pp.40-51). Alexandria, VA: Association for Supervision and Curriculum Development.
- Chelimsky, E. (1997). The coming transformations in evaluation. In E. Chelimsky, & W. R. Shadish (Eds.), *Evaluation for the 21st century* (pp.1-26). Newbury Park, CA: Sage.
- 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*. Boston: George Allen & Unwin.
- Gronlund, N. E. (1981). *Measurement and evaluation in teaching*. New York: Macmillan.
- Gove, P. B. (1981). *Webster's third new international dictionary*. Springfield, MA: Merriam-Webster.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.
- Hopkins, D. (1989). *Evaluation for school development*. Philadelphia: Open University Press.
- Hopkins, D. (1997). *Evaluation for school development*. Keynes, UK: Open University

Press.

- House, E. R. (1992). Future policy for school-based evaluation. In A. M. Indrebo, L. Monsen, & T. Alvik (Eds.), *Theory and practice of school-based evaluation: A research perspective* (p.77) Lillegammer, Norway: Oppland College.
- Kellaghan, T., & Madaus, G. F. (2000). Outcome Cvaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: viewpoints on educational and human services evaluation* (pp. 83-394). Boston: Kluwer Academic.
- Kemmis, S. (1989). Self evaluation in the context of program evaluation. *Evaluation Journal of Australasia*, 2(1), 2-30.
- MacBeath, J., & McGlynn, A. (2002). *Self-evaluation: What's in it for schools?* Lodon: RoutledgeFalmer.
- Madaus, G. F., Scriven, M. S., & Stufflebeam, D. L. (1983). *Evaluation models*. Norwell, MA: Kluwer-Nijhoff.
- Madaus, G. F., & Kellaghan, T. (1992). Curriculum evaluation and saassessment. In P. Jakson (Ed.). *Handbook of research on curriculum* (pp.119-154). New York: Macmillan.
- Madaus, G. F., & Stufflebeam, D. L. (2000). Program evaluation: A historical overview. In G. F. Madaus, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed.) (pp.3-18). Boston, MA: Kluwer Academic.
- Marsh, C. J., & Willis,G. (1999). *Curriculum-alternative approaches, ongoing issues*. (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Nevo, D. (1995). *School-based evaluation: A dialogue for school improvement*. Tel Aviv, Israel: Masada.
- Norris, N. (1990). *Understanding educational evaluation*. London: Kogan Page.
- Popham, W. J. (1975). *Educational Evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- Popham, W. J. (1993). *Educational Evaluation*. Boston: Allyn & Bacon.
- 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.

- Stenhouse, L. (1975). *An introduction to curriculum research and development*. London: Heinemann.
- Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic evaluation*. Norwell, MA: Kluwer-Nijhoff.
- Wentling, T. L. (1980). *Evaluating occupational education & training program*. Boston: Allyn & Bacon.

