

教育評鑑的概念與發展

吳清山、王湘栗

【作者簡介】

吳清山，台灣省雲林縣人，曾任台北市立師範學院校長、主任秘書、國民教育研究所所長、初等教育學系主任。現任台北市立師範學院國民教育研究所兼教育行政與評鑑研究所教授。

王湘栗，湖南省衡陽縣人，國立台灣師範大學教育研究所博士候選人。現任台北市國語實驗國民小學教師。

摘要

教育的永續發展，計畫、執行和評鑑，三者缺一不可。近十年來我國大力推動各項教育改革，為瞭解教育改革成效，確保教育品質，教育評鑑益顯重要。本文旨在探討教育評鑑的基本概念與近期發展。全文分為五個部分，首先，就教育評鑑的重要概念進行分析；其次，說明教育評鑑的學科性質、理論建構等評鑑理論的發展；再者，從教育評鑑的歷史回顧，闡述教育評鑑的起源，以及美國和我國教育評鑑的發展歷史；再次，析論我國教育評鑑所面臨的挑戰，包括本土化有待努力、專業化有待提升、後設評鑑有待提倡、評鑑倫理有待倡導及各項評鑑有待整合等；最後，提出八項因應挑戰的有效實施策略，以作為強化我國教育評鑑的建議。

關鍵詞彙：評鑑、教育評鑑、教育評鑑歷史

Key Words: Evaluation, Educational Evaluation, The History of Educational Evaluation

壹、前言

教育之能夠不斷發展，除了計畫和執行之外，評鑑亦屬相當重要的一環，透過有效的評鑑，一方面改進教育缺失，一方面提升教育品質。因此，評鑑遂成為教育永續發展的動力，長久以來，教育評鑑不僅為教育實務者所重視，而且學者們（如：K. Stufflebean, R. Stake, E. Eisner, M. Scriven...）亦紛紛致力於教育評鑑理論的研究，所提出量化評鑑或質性評鑑的實施策略及模式建構，皆有助於導引教育評鑑的實施。

教育評鑑理論具有不同類型，依 Scriven(2003)的看法，教育評鑑理論主要可以分為兩種類型：一是規範性（normative）教育評鑑理論，旨在探究教育評鑑應該是什麼，教育評鑑如何有效界定，偏重於應然層面。一是描述性（descriptive）教育評鑑理論，旨在描述教育評鑑是什麼、有哪些形式、如何做、做了哪些，偏重於事實層面。事實上，理想的教育評鑑，應該兼顧教育評鑑的事實面和應然面，才能讓教育評鑑收到良好的效果。所以探究教育評鑑理論或進行實際教育評鑑時，規範性和描述性的教育評鑑，都不能偏廢。

由於社會大眾對於教育品質重視有增無減，歐美先進國家也將教育評鑑視為提升教育品質的重要策略，除了實施校務評鑑、方案評鑑和人員評鑑之外，政府亦補助學者們進行教育評鑑相關的研究，以提升教育評鑑的品質。我國對於教育評鑑亦頗為重視，在民國八十五年的「教育改革總諮議報告書」的「好還要更好：提升教育品質」中，特別提出「強化教育研究與評鑑」的建議（行政院教育改革審議委員會，民 85）。此外，民國八十八年總統公布的「教育基本法」中的第十三條亦規定：「政府及民間得視需要進行教育實驗，並應加強教育研究及評鑑工作，以提昇教育品質，促進教育發展。」在在顯示教育評鑑是提升教育品質不可或缺的一股力量。

此外，我國經過一連串的教育改革之後，推動各項教育政策，實際收到哪些成效，目前並無足夠的實證資料可資佐證，主要原因在於缺乏一套嚴謹而客觀的評鑑機制所致。因此，王秀玲和林新發（民 93）特別提出「要建立反省與教育評鑑機制，確保教育改革成效」的建議。近年來，由於受到學校本位管理思潮與實務的影響，學校評鑑又再度受到重視，並成為提升學校教育品質，判定學校辦學績效之重要依據（林天祐，民 93）。所以，教育評鑑將成為未來教育行政的重大研究議題之一。

為了幫助大家了解教育評鑑的基本概念、理論發展、遭遇挑戰及未來實施策略，本文乃分別就教育評鑑的概念分析、教育評鑑理論發展、教育評鑑的歷史回顧、教

育評鑑的挑戰、教育評鑑的有效實施策略等方面說明之。

貳、教育評鑑的概念分析

一、教育評鑑的意義

要瞭解教育評鑑的意義，可以先從評鑑的意義著手。評鑑(evaluation)一詞代表一種價值判斷的動態過程，在不同的教育領域，用法不同。在教育心理、測驗統計學或課程教學上，多數以「評量」稱之；在教育行政領域上，則較多使用「評鑑」、「考核」或「評價」（謝文全，民 92），中國大陸多數學者使用「教育評價」（侯光文，民 85；陶西平，民 87；劉本固，民 89），台灣除少數學者使用教育評價外（李聰明，民 61；黃炳煌，民 72），近年來幾乎都使用「教育評鑑」，為使行文一致，除非引文需要，本文統一使用「評鑑」、「教育評鑑」。

要為評鑑下一明確周延的定義並不容易，由於不同學者所處評鑑發展歷史背景、評鑑取向的差異，而有不同的看法。學者們或從本質、目的、功能、方法等方面，或以獨立或以合併兩個以上的向度來定義評鑑，雖提出甚多看法，但終究沒有一項定義能獲得評鑑領域學者的一致認同（曾淑惠，民 91）。為瞭解各學者對評鑑所下的定義，依年代、國外、國內的順序，列舉數端如表 1：

表 1 各學者對評鑑的定義

學者	評鑑的定義
Stufflebeam(1973)	評鑑是為了判斷可用的決策所做的描述、獲取及提供有用資訊的過程。
Tenbrink(1974)	評鑑是獲得資訊進而形成判斷，並據以作決定的過程。
Chelimsky(1985)	評鑑是應用系統化研究方法以評估方案的設計、實施與效益。
Stufflebeam 和 Shinkinfield(1985)	評鑑不在證明，而在改進。評鑑是一種計畫、獲取和提供敘述性、判斷性資訊的過程。這些資訊涉及研究對象的目標、設計、實施和影響的價值及優缺點，以便指導如何決策，滿足教學效能核定的需要，增加對研究對象的瞭解。
Worthen 和 Sanders(1987)	評鑑是決定某件事情的價值。
Alkin(1990)	評鑑涉及系統化地蒐集、分析及報告一些可以用來改變態度或改進方案運作資料的活動，其中系統化指評鑑需經過規劃。
McLauaglin(1990)	評鑑是提供決策者有關於方案運作或效能的可信、有效、相關且有用的資訊。

House(1993)	評鑑是依據適當的標準來決定事物的程度或價值。
Worthen, Sanders 和 Fitzpatrick(1997)	評鑑是包含定義、說明、可辨證的參考準則等，以決定受評對象的價值、品質、有效性或重要性。
Douglah(1998)	評鑑是一個持續的探究過程，探究有關社會、經濟、環境條件與方案內部發展環境的過程，探究有關方案如何設計、方案傳遞給觀眾的意圖等問題，探究有關方案欲產生成果的相關問題。
Boulmetis 和 Dutwin(2000)	評鑑是一個蒐集與分析資訊的系統化過程，以決定目標是否達成或目標達成的程度。
Stufflebeam(2003)	評鑑是某一對象之目標、設計、執行和成果的優點、價值的描述性、判斷性資訊的陳述、獲得、提供和應用的歷程，以作為改進決策之指引，提供績效責任報告，制度化通報和傳布決定，以更瞭解所參與的現象。
張春興（民 78）	評鑑泛指對某種事物的價值予以評定的歷程，是在某種有計畫的活動實施之後，按照預定目標檢核得失的歷程。
吳芝儀、李奉儒譯 （民 84）	評鑑是任何藉由系統的、以資料為基礎的探究，來檢驗和判斷為了增進人類福祉而作的努力，俾便理解其達成的成就和有效性。
謝文全（民 92）	評鑑是對事務加以審慎的評析，以量定其得失及原因，據以決定如何改進或重新計畫的過程。
陳玉琨（民 93）	評鑑是一種價值判斷的活動，是對客體滿足主體需要程度的判斷。

根據表 1，可以瞭解評鑑是依據適當的標準，對某事項進行審慎、系統化的蒐集資料、分析和報告之價值判斷，透過價值層面與事實層面的比較，量定事務的利弊得失和原因，以獲取作決定的有用資訊，並提供改進的歷程。

至於教育評鑑的定義，各學者所提出亦有所差異。早在 1950 年，被後人尊稱為「教育評鑑之父」的泰勒（R. H. Tyler），曾將教育評鑑界定為：「決定教育目標實際上被了解的程度之過程」（引自 Nevo,1995）。Cronbach 在 1963 年時，曾將教育評鑑定義為「蒐集並使用資訊於教育方案的決定」（引自 Wolf,1990）；美國「教育評鑑標準聯合委員會」（The Joint Committee on Standard for Educational Evaluation,1994），則定義教育評鑑為：「有系統的評估某一對象的價值或優點」；Nevo(1995)歸納學者對評鑑的定義，認為教育評鑑是：「蒐集有關教育目標其本質與品質方面系統化的資訊之行動。」Worthen 和 Sanders(1987)則將教育評鑑定義為：「對一個方案、結果、計畫、過程、目標或課程之品質、效能或價值，做出正式的決定。」。

李緒武（民 82）指出：「教育評鑑是對教育事業的價值作嚴正的品評。評估的範圍甚廣，教師教學情況，教學計畫，教學成果，考試制度，課程教材，教育計畫，教育目標以及所有指導學生的學習活動，都可加以品評，鑑定其成效。」

吳清山和林天祐（民 88）認為：「教育評鑑係指對於教育現象或活動，透過有系統和客觀的方法來蒐集、整理、組織和分析各項教育資料，並進行解釋和價值判斷，以作為改進教育缺失，謀求教育健全發展的歷程。」

秦夢群（民 89）對教育評鑑的觀點是：「對於教育現象或活動，透過蒐集、組織、分析資料，加以描述與價值判斷的歷程。」

林天祐和蔡菁芝（民 90）認為：「教育評鑑為系統化地蒐集資訊，以作為促進、達成教育目標的評價過程。」

陳玉琨（民 93）在其「教育評鑑學」一書中提出：「教育評鑑是對教育活動滿足社會與個體需要的程度做出判斷的活動，是對教育活動現實的（已經取得的）或潛在的（還未取得，但有可能取得的）價值做出判斷，以期達到教育價值增值的過程。」

蘇錦麗（民 93）亦認為：「教育評鑑係有系統地採用各種有效方法，蒐集質與量的資料，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，並將其結果做為決策之參考。」

綜合上述，將教育評鑑歸納為：對於教育現象或活動，就其人員、方案、運作結果，對照評鑑標準，透過質與量的方法，採系統而客觀地蒐集、分析資訊，進行價值判斷，以了解教育成效、改進教育缺失和達成教育目標的歷程。

二、教育評鑑的類型

教育評鑑的類型很多，常見的分類方式可能從教育評鑑的目的、本質來加以分類，亦可從接受評鑑的客體、教育評鑑的方式，以及教育評鑑在教育決策上應用的情形來加以分類，茲扼要說明如下：

從評鑑的目的來區分，有形成性評鑑和總結性評鑑。此一觀念最早由 Scriven(1967)所提出來，形成性評鑑係指在教育活動進行過程中，為提供管理者、主事者改進資訊，藉以提升教育活動品質與效果，所進行的評鑑；總結性評鑑則指在教育活動結束的階段，為瞭解教育活動的實用價值與實施成果，所進行的評鑑。形成性評鑑的設計和使用，強調仍處於發展中階段受評對象改進的目的；總結性評鑑的設計，目的在於呈現有關受評對象的優點與價值的結論，並作為是否繼續維持、擴大或終止的建議（Joint Committee on Standard for Educational Evaluation, 1994）。

從教育評鑑的本質來區隔，有所謂真評鑑、準評鑑和假評鑑。Stufflebeam 和 Shinkfield(1985)依據美國「教育評鑑標準聯合委員會」(The Joint Committee on Standard for Educational Evaluation)在 1981 年所發展的三十項評鑑標準，將教育評鑑分為假評鑑(pseudo-evaluation)、準評鑑(quasi-evaluation)和真評鑑(true evaluation)三種。其中假評鑑包括秘密調查(covert investigations)、公共關係 - 授意的研究(public relations-inspired studies)，準評鑑包含目標本位研究(objectives-based studies)、實驗導向研究(experimentally oriented studies)，真評鑑則有決策導向研究(decision-oriented studies)、當事者中心研究(client-centered studies)、政策研究(policy studies)和消費者導向研究(consumer-oriented studies)。

從接受教育評鑑的對象而言，有所謂人員評鑑與方案評鑑。人員評鑑的對象包括教育事業中的重要角色，如：學生評鑑、教師評鑑、校長評鑑和行政人員評鑑等；另外，美國有對學區教育局長的評鑑；從評鑑的方案觀點而言，就美國的國情一論及教育評鑑，通常指的就是方案評鑑（潘慧玲，民 93），以我國的情況來說，校務評鑑、課程評鑑、教材評鑑，以及特殊教育、兩性平等教育、交通安全教育、營養午餐、生命教育、法治教育、學生輔導計畫、教訓輔三合一等評鑑，都可以歸類在方案評鑑的內涵中。

從評鑑的方式來分析，有所謂自我評鑑、內部評鑑和外部評鑑。為瞭解這三者間的差異。自我評鑑通常由受評鑑者根據自行發展或他人已發展出的評鑑檢核工具，填寫相關資料，自行進行逐項的檢核，以瞭解自身的實際表現和原先設定目標間的差距；內部評鑑通常是由受評鑑組織內自組評鑑小組，其小組成員除受評單位人員之外，亦可包括校外學者專家，兼採家長或學生的意見調查等方式進行；外部評鑑通常是由主管的上級機關所主導，其具體作法是邀集行政機關代表、學者專家、學校代表、教師代表、家長代表等，組成外部評鑑小組，進行評鑑工作。

從教育評鑑在教育決策上應用的型式上來區分，可以分為三種類型，分別是：知識性(intellectual)、社會性(social)和象徵性(symbolic)。知識性的應用以教育評鑑所提供的知識為基礎，用以協助決策或發揮教育作用；社會性的應用則有兩項功能，一是藉評鑑結果提高決策的合理性；二是利用評鑑結果尋求外在勢力、或組織內部對既有特定政策支持。象徵性的應用，藉由評鑑的程序和結果，來達成以下目的：一是當「行動」或「不行動」的藉口；二是來提升教育決策的專業形象；三是做為施展組織或個人威權的手段（吳明清，民 82）。

上述的教育評鑑類型並非絕對的，依照的不同分類角度，同一個評鑑可能被歸類到不同的類型中，也可能同時具有好幾種不同的類型特徵。

參、教育評鑑理論的發展

一、教育評鑑的學科性質

在建構教育評鑑的理論基礎之前，有必要先分析教育評鑑的學科性質。根據 Scriven(1994)的看法，評鑑學科除了具有普通學科的性質外，亦具備了工具學科與跨越學科(trans-disciplines)的性質。因為評鑑學科的目的在於發展評鑑工具，提供其他學科應用，因此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」的兩個必備條件為：

- (一)「跨越學科」係以服務其他學科或與其他學科合作為目的。
- (二)「跨越學科」除了包括一般學科具備的核心成分(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。

中國大陸學者陳玉琨(民 87)曾分析學者們的論述，歸納出教育評鑑的學科性質為「具有一定基礎理論性的應用綜合學科」。其理由如下：

- (一)「一定理論基礎性」係指認識教育評鑑中主體性與客觀性的關係、價值觀與價值尺度的問題、系統優化與評鑑一體化等基本問題，此類問題之研究屬理論基礎性質，並非可操作的。
- (二)教育評鑑的應用性質，指其具有較強的技術性質與實際應用的特性。
- (三)綜合學科則指教育評鑑不僅是兩種學科的交叉(意指交集、結合)，且將吸收與運用多種學科的理論、方法和技術。

根據上述學者的論述，可以瞭解基於科際應用的觀點，教育評鑑被視為一門跨越學科，亦即為一門具有其獨立理論基礎，同時具備應用性質的綜合學科。

二、教育評鑑的理論建構

教育評鑑雖在 1970 年代逐漸朝向專業化的方向發展，但誠如 Shadish(1998)所言，評鑑學者們都知道評鑑理論的重要性，因為任何一門專業，都有其獨特的知識基礎(knowledge base)，教育評鑑理論亦建立在其知識基礎之上。套用 Shadish 的用語，評鑑理論是界定「評鑑是什麼(Who We Are)」的關鍵。換言之，評鑑理論是評鑑有別於其他專業的要素，也是評鑑自身專業認同的核心。潘慧玲(民 93)亦指出，評鑑領域中有關方法的書籍汗牛充棟，評鑑理論的論著卻極其罕見，評鑑不該只是

應用的方法論。

台灣及中國大陸均有學者嘗試去建立教育評鑑的理論，例如：李聰明（民 61）、侯光文（民 85）、陳玉琨（民 87）及劉本固（民 89）等人。歸納這些學者的觀點，多數從教育評鑑和哲學、心理學、教育學、管理學等理論的關係進行分析（如侯光文，民 85；陳玉琨，民 87；劉本固，民 89）。這類分析雖強調了教育評鑑具有哲學、心理學、教育學、管理學等理論基礎，但深入細究，會發現此類分析的基調，在於認為教育評鑑是借用哲學、心理學、教育學、管理學等理論。「借用觀」的教育評鑑理論，雖在分析教育評鑑的理論基礎上有所助益，但對於建立教育評鑑成為一獨立的學科造型上，則明顯的有所不足與欠缺。

美國學者 Shadish, Cook 與 Leviton(1991)的著作「方案理論評鑑的基礎：實踐的理論(Foundations of Program Evaluation: Theories of Practice.)」曾對評鑑理論進行深入的後設分析。他們嘗試以評價(valuing)、知識建構(knowledge construction)、知識使用(knowledge use)、社會方案運作(social programming)、評鑑實務(evaluation practice)等五大範疇，為評鑑理論提供了完整的分析架構。之後，Shadish(1998)根據前述的五個面向，描繪了如圖 1 的理論分析架構圖：

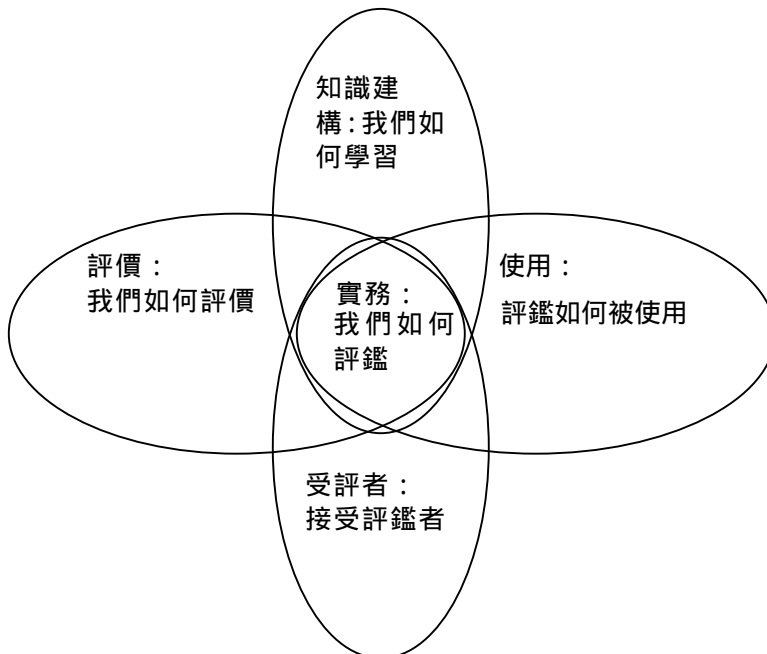


圖 1 評鑑理論的五個範疇

資料來源： "Evaluation Theory Is Who We Are," by W. R. Shadish, 1998, *American Journal of Evaluation*, 19 (1), p.2.

Shadish, Cook 與 Leviton(1991)和 Shadish(1998)認為有關評鑑理論之討論，可參照圖 1 的五大範疇來進行，即：我們如何評價、我們如何建構知識、評鑑如何被使用、接受評鑑的對象、以及實際如何進行評鑑等。另外，每個範疇尚包含更進一步的專用術語(terminology)，簡述如下(Shadish,1998)：

- (一) 評價部分：區辨描述評價、規範評價與後設評價。
- (二) 使用部分：說明工具性使用、概念性使用與說服性使用，以及長程使用、短程使用。
- (三) 社會方案部分：闡明漸進式變革與激進式變革；方案變革、計畫變革與要素變革；新方案變革與現存方案變革等。
- (四) 知識建構部分：區分了確定性的不同層次，以及知識的不同類別，例如：描述性、因果性、說明性等知識。
- (五) 評鑑實務部分：辨別了所問問題的種類，例如：有關需求下執行、委託者、效果、影響、成本等問題；有關促進使用的作法；有關詢問誰的問題；有關評鑑者的角色，以及關於所採用的評鑑方法等。

上述 Shadish 等學者所發展的評鑑理論分析架構，對於建構教育評鑑的理論，提供了更完整的思維角度，因而更能拓展人們對於評鑑的視野，故獲得國內學者的推介。潘慧玲（民 93）曾在此基礎上，針對評鑑的本質與評價、知識論及方法論、評鑑的用途及其與決策的關連、方案運作與方案理論等面向，析論評鑑理論的發展，並提出了以下的問題，值得進一步探討：

- (一) 在評鑑活動中，我們需要判斷事物的價值，如何作判斷，判斷的準則為何？
- (二) 評鑑所獲得的知識是誰的知識？誰來決定知識的真實、實用性或有效性？什麼是知識的判準？
- (三) 評鑑結果如何運用？可為決策者直接採用或只是發揮啟蒙之功能？評鑑結果與決策是否為線性關係？決策過程中涉及許多利益團體的權力角逐，這些政治因素如何影響決策？
- (四) 方案如何運作？其如何有助於社會問題之解決？方案的改變如何帶動社會的改變？是漸進式的改變或激進式的改變？

歸納而言，現代教育評鑑的發展，雖已超過半個世紀，但大多數評鑑學者關注的焦點在於評鑑方法、模式、技術、標準、步驟等面向，對於型塑評鑑成為一門專業的獨特理論部分，在過去甚少受到關注。為促使教育評鑑成為一項專業，學者們對於教育評鑑理論這塊園地的耕耘，乃屬當務之急。

肆、教育評鑑的歷史回顧

一、教育評鑑的起源

民國六、七年代，台灣學者談起教育評鑑起源，都認為教育評鑑源自於古代的中國。例如：「我國古代帝王對於太學的定期巡視與考試，乃至選拔人才所採用的鄉舉里選與科舉制度，均富有一部分評鑑的意義。」（張植珊，民 68）；或認為科舉制度是最早的系統性評鑑制度之雛形（林萬義，民 76；黃炳煌，民 72；楊文雄，民 70）。甚至，包括許多外國學者在內，談起教育評鑑的起源，大多數也認為教育評鑑的起源，是源自於古代的中國。如：「評鑑個人或方案的概念早在公元二千年前已甚明顯，中國在當時已實施文官考試制度。」（Stufflebeam & Shinkfield, 1985）；Merwin 在 1969 年指出：「約在西元前 1115 年，中國已明確有了考試的內容與方法。」（引自盧增緒，民 84）；「早在西元前 2200 年，中國就有每三年考查百官一次的辦法。」（Dubois, 1970；Popham, 1993）；「評鑑個人表現的作法，很明顯的在西元前 2000 年，中國官方就用於考查公務員的公眾服務能力」（Worthen & Sanders, 1987）。

綜觀上述，各學者的說法並不一致，尤其在時間推估上差異甚大，有的往前推至西元前 2200 年，有的推估至西元前 1115 年，或更晚近的隋朝（約在西元 600 年左右）；其共通點大致為：1. 評鑑的觀點在古代中國就已存在；2. 古代中國的考查百官、科舉制度、文官考試、鄉舉里選等，可以說是現代教育評鑑的起源。

然而，盧增緒（民 84）並不同意上述缺乏考證的說法，特別指出：

1. 國外文獻對於我國古代的評鑑觀念，大多數語焉不詳，且在事證的舉例上有內容上的疏漏、錯誤，或時間上的矛盾。
2. 國內學者引用欠缺考證的國外文獻，造成以訛傳訛、人云亦云的現象，例如：將「大比」誤解為科舉制度中之鄉試，以及將科舉制度當成是最早的正式評鑑等錯誤見解。
3. 我國最早的正式評鑑應指周代鄉遂之治中的邦比之法，即所謂周禮中所記載的「三年大比」制度，確實具有現代評鑑的意義。
4. 最早的正式評鑑活動雖始於中國，但正式的教育評鑑觀念卻產生自國外，主要為美國、英國及歐陸先進國家。

不過，許多大陸學者還是認為教育評鑑源於中國科舉制度，例如：侯光文（民 85）：「育士與選士緊密相接，形成了一整套包括教育、考試、選才在內的嚴密完整的制度，這是中國傳統文化的一大特色，並成為現代教育評價活動的源頭。」；劉本

固(民 89):「中國是考試的故鄉,歷朝歷代都有考選制度,通過考試是來選拔賢才。教育評價,就其歷史發展來說,源於我國的隋朝。」

其次,常被引用作為教育評鑑起源的記載,是禮記學記的一段文字:「比年入學,中年考校:一年視離經辨志,三年視敬業樂群,五年視博習親師,七年視論學取友,為之『小成』。九年知類通達,強立而不反,謂之『大成』。」藉以強調評鑑並非一種新觀念、舶來品,評鑑在我國教育活動中的重要地位,以及中國早期學校已明確規定評鑑的內容、標準及時間(湯志民,民 925;劉本固,民 89)。至於周朝邦比之制和禮記學記的記載孰先孰後,是另一個值得有心者探究的問題。

綜上所述,可以瞭解正式評鑑的觀念和作法,在中國古代周朝的邦比之法即已存在,而禮記學記中考校制度,則是我國學校實施教育評鑑的最早記載,但許多學者仍認為科舉制度是教育評鑑的起源,論及現代教育評鑑的觀念和實施,則是發源自近百年的美國、英國和歐陸等諸先進國家。

二、美國教育評鑑的發展歷史

關於美國教育評鑑的發展歷史,有幾種不同的看法。Worthen 和 Sanders(1987:12)採取較簡單的劃分方式,將美國教育評鑑發展的歷史分成三個時期:(一)早期歷史至 1920 年;(二)1920-1965 年;(三)1965 年迄 1987 年。有的學者以重要的人物、事件為劃分的參照點,例如:Stufflebeam 和 Shinkfield (1985),以「教育評鑑之父」泰勒(Tyler)的早期貢獻和美國重要教育法案為參照點,將教育評鑑的發展歷程分為以下五個時期:(一)前 Tyler 時期(Pre-Tyler Period):包括 1934 年之前的發展;(二)泰勒時期(Tylerian Age):自 1930 年迄 1945 年間;(三)純真時代(Age of Innocence):1946 年迄 1957 年;(四)現實時代(Age of Realism):自 1958 年迄 1972 年;(五)專業地位時代(Age of Professionalism):從 1973 年迄今。

有的學者從教育評鑑之前的心理測驗發展開始,說明教育評鑑發展的歷史,例如:Madaus & Stufflebeam(2000)將教育評鑑的歷史發展,自 1972 年開始,到 2001 年為止,分為七個時期:

- (一)改革時期(The Age of Reform,1792-1900):以 1792 年為教育評鑑改革時期之始,主因於該年為 William Parish 首度以標準化的數據記載測驗的結果。
- (二)效率與測驗時期(The Age of Efficiency and Testing,1900-1930):二十世紀初 F. Taylor 的科學管理理念影響教育行政理論的發展,強調系統化、標準化與效率。
- (三)泰勒時期(The Tylerian Age,1930-1945):此時期受教育評鑑之父 R.W. Tyler 思

想的影響，關注教育測驗與評鑑的發展。

(四) 純真時期(The Age of Innocence,1946-1957)或稱為「無知時期」(Age of Ignorance)。

(五) 發展時期(The Age of Development,1958-1972)：強調教育方案的評鑑與改良式多因素評鑑模式的發展是此時期的最大特徵。

(六) 專業化時期(The Age of Professionalization,1973-1983)：在此之前，評鑑人員面臨身分認同危機，未能確認自己應歸諸何種專業社群人員；此外，亦無任何教育評鑑的專業組織與專業雜誌，可讓評鑑人員得以相互切磋，交換心得。

(七) 擴充與整合時期(The Age of Expansion and Integration,1983-2001)：隨著經濟成長，教育評鑑的發展也不斷擴展與愈越整合。

而 Guba 和 Lincoln(1989)則從教育評鑑理念的典範或觀念角度切入，提出「代」(generations)的發展歷程，國內學者亦多人引用(如秦夢群，民 89；潘慧玲，民 93；鄭新輝，民 91)。Guba 和 Lincoln(1989)們將教育評鑑的發展自 1910 年代迄今分成四代，分別為：

(一) 第一代評鑑(1910-1930)：視評鑑為測驗。

(二) 第二代評鑑(1930-1967)：視評鑑為客觀適時的陳述。

(三) 第三代評鑑(1967-1987)：視評鑑為判斷。

(四) 第四代評鑑(1981 以後)：視評鑑為協商與溝通的歷程。

綜合上述，有關美國教育評鑑的發展歷史，言人人殊，學者們對教育評鑑的發展歷史有不同的觀點，主因在於對教育評鑑歷史人物或事件著眼點的差異。最常為學者們所引用者為 Madaus & Stufflebeam(2000)的七個發展時期說，以及 Guba 和 Lincoln(1989)的四代評鑑發展階段說。

三、現代我國教育評鑑的發展

分析我國教育評鑑的發展歷程，可以遠溯自先秦時代，亦可僅論述自 1949 年國民政府遷台後的發展。前者如侯光文(民 85)的看法，其認為我國教育評鑑與選士制度密切關係，將中國教育評鑑活動分成四個階段：(一) 古典教育評鑑萌生期：西元 606 年以前，包括先秦至魏晉南北朝時期的選士測評活動；(二) 科舉時期：606 年至 1905 年，包括隋唐以來至清末我國古典教育評鑑歷程；(三) 近代教育評鑑活動期：1905 年至 1949 年，教育評鑑與教育測量結合，呈現多種體制的多元格局；(四) 現代教育評鑑發展期：1949 年迄今。為使本文聚焦，僅說明 1949 年國民政府遷台後，我國現代教育評鑑的發展歷史。

我國現代教育評鑑的發展歷史，可以下列三個階段來加以區隔和說明：

- (一) 萌芽初創期：自 1949 迄 1974 年，此時期國內有關教育評鑑的觀念和作法，尚處於模糊、摸索、起步的階段。論及我國教育評鑑的實施，一般學者多認為是始自民國六十四年，由教育部正式開始實施的大專院校評鑑，但盧增緒（民 84）曾提起一段為人淡忘的教育評鑑歷史。民國五十二年，教育部與聯合國兒童教育基金會，簽署了國民教育發展五年計畫實施方案，在方案中已設計對該計畫的完整評鑑，包括成立評鑑委員會，並於民國五十六年發表「國民教育發展五年計畫」期中評鑑報告，至民國五十九年又公布期終報告，在當時此教育評鑑在亞洲尚屬首創；盧氏並認為該次教育評鑑具有：評鑑與計畫密切結合，具形成性與總結性評鑑之作用，以及掌握經費、強調績效責任等特點。然而，令人遺憾的是這段教育評鑑歷史的後續影響相當微弱（盧增緒，民 84）。
- (二) 成長擴展期：自 1975 迄 1999 年，此時期的學校教育評鑑工作逐漸蓬勃發展，在學術研究上亦有若干成果。首先，從教育評鑑實務而言，我國學校教育評鑑發展的重要事件，是民國六十四年，教育部為提昇大學教育水準，所進行的制度性大學教育評鑑工作。張德銳（民 89）曾將此時期大專院校教育評鑑工作分為三階段：第一階段自民國六十四年至七十一年止，以全面評鑑為主；第二階段從民國七十二年 to 民國八十年，採重點學門評鑑方式；第三階段為民國八十一年後委託學術機構試辦之學門評鑑。教育部自民國八十五年開始試辦的綜合高中，且於民國八十九年至九十二年委託國立台灣師範大學教育研究中心，辦理了四屆的綜合高中評鑑（潘慧玲、徐昊杲、黃馨慧、張志偉，民 92）。國民教育評鑑方面，由於國民教育屬地方政府權限，國小教育評鑑在民國六十六年由台灣省教育廳指定彰化縣教育局試辦，台北市、高雄市分別陸續於六十八、六十九學年度試辦；國中教育評鑑亦由教育部在民國六十八年三月指定基隆市、桃園縣先試辦，同年十二月並頒佈「國民中學評鑑暫行實施要點」，決定於六十八學年度全面實施各國中之評鑑（張德銳，民 89）。其次，就教育評鑑學術研究方面而言，1992 年國立新竹師範學院成立教學與學校評鑑研究中心，是國內第一個成立的教育學術研究中心；此時期國內的學者、學會開始從事教育評鑑專書的譯著，較為人知者如黃光雄等人（民 78）編譯的「教育評鑑的模式」，係取自 Stufflebeam 和 Shinkfield 1985 年的著作「Systematic evaluation」，伍振鷺（民 82）、中國教育學會（民 84）各主編名為「教育評鑑」的專書，以及陳漢強（民 86）主編「大學評鑑」一書等，均

使此時期的教育評鑑研究更為快速便利；論文方面若以博碩士學位論文的角
度來觀察，自 1970 年代的 2 篇，到 1980 年代的 14 篇，至 1990 年代則有 47
篇，此時期學術研究論文的量逐年擴增之趨勢相當明顯（潘慧玲，民 93）。

（三）專業導入前期：2000 年迄今，此時期特徵在於提倡教育評鑑應朝向專業化發
展的目標，並有導入評鑑邁向專業化的若干成果，如評鑑專業組織的成立、
教育評鑑指標的建立、教育評鑑的實施層面擴大、學術研究成果擴增等。
以教育評鑑研究組織而言，九十二學年度臺北市立師範學院設立教育行政與
評鑑研究所碩士班，九十三學年度國立新竹師範學院教育研究所碩士班設立
行政與評鑑組（蘇錦麗，民 93：26）；在學術專業學會方面，2000 年台灣教
育政策與評鑑學會成立，2003 年財團法人台灣評鑑協會通過核准成立；學術
研究活動方面，有關教育評鑑的博碩士學位論文，在 2000-2004 年間共有 129
篇之多，量的擴增十分驚人（潘慧玲，民 93：19），2004 年五月由國立台灣
師範大學主辦了「教育評鑑回顧與展望學術研討會」，均有助於教育評鑑的專
業向前邁進。然而，國內教育評鑑雖已有專業化之倡議與若干作為，但距離
評鑑專業化仍尚有一大段距離，因此只能說還處於導入專業的前期階段而
已。換言之，我國教育評鑑專業化較之以往雖有進步，但進展的速度仍然相
當遲緩。

伍、教育評鑑發展的挑戰

我國教育評鑑發展至今，雖有四十餘年的歷史，但回顧過去教育評鑑的歷史，
卻有相當多問題亟待解決，許多困境亟待突破。國內學者對此多有論述，或從鉅觀
的層次加以分析者，如盧增緒（民 84）認為國人欠缺正確的評鑑觀念、輕率移植和
過度借用美國教育評鑑的觀念和作法、教育評鑑被窄化為校務評鑑、教育評鑑人員
的反省與自覺薄弱、第四代評鑑與理論導向(theory-driven)的評鑑未受重視等；又
如，潘慧玲（民 93）在反思國內評鑑發展的經驗時，指出國內評鑑觀念過於窄化、
評鑑專業化有待努力、學術風貌不夠多元等缺失；或從微觀層次加以析論者，如吳
清山（民 91）對校務評鑑的實施所招致的批評，提出以下的檢討：學校人員缺乏評
鑑觀念、排拒與抗拒評鑑，校務評鑑內容設計不易、評鑑人員專業素養不足、評鑑
時間不足、經費不敷支應、評鑑結果公信力不足、評鑑報告束之高閣，以及校務評
鑑缺乏後設評鑑等。前述的批判與省思，換個角度想，正是我國教育評鑑發展上有
待努力的部分，茲將我國教育評鑑發展上的挑戰，說明如下：

一、教育評鑑本土化有待努力

大約十年前就有學者倡議教育評鑑的本土化(盧增緒, 民 84)。不容否認, 我國教育評鑑無論在評鑑理論、評鑑方法、評鑑標準與實務作為等各方面, 都可說是美國教育評鑑的翻版。移植、借用美國教育評鑑發展的成果, 雖省事簡便, 但也因為過度抄襲、套用, 而產生許多調適不良的後果, 並與我國教育評鑑本土化漸行漸遠, 喪失發展我國教育評鑑主體性的可能, 這對我國教育評鑑而言, 毋寧是一大損失。

本土化(localization)的訴求, 在於反映出我國教育學術研究學者的主體自覺與文化認同, 在體察中西各異的思維模式和文化視野中, 透過對話、理解和詮釋, 達成創造性的視野交融, 讓多元觀點啟迪教育理論的創新與進展(林秀珍, 民 88)。本土化的立場乃「立足台灣、放眼世界」, 本土化並非國際化的對立名詞, 亦非清朝末年之鎖國閉關、閉門造車, 自閉自絕於國際評鑑學術發展潮流之外, 沾沾自喜於創造出自以為是、妄自尊大的「台灣教育評鑑理論與實務」。本土化的重點在於本著開放的態度, 容納吸取多元觀點, 逐步建立我國教育學術研究的特色(吳清山, 民 88)。換言之, 教育評鑑的本土化在於深入瞭解本國特有之文化、社會脈絡, 並善於採擷諸先進國家教育評鑑的研究成果, 因應本國國情與教育實況的需求, 重新創造、研發出能達致檢視教育實況、提升教育品質、協助教育決策等目的之「台灣本土的教育評鑑」, 可惜國內此方面的努力仍嫌不夠。

二、教育評鑑專業化仍待提升

教育評鑑是一項專業性的工作, 教育評鑑專業化為學者們所關注。Worthen (1994)曾提出發展成熟的專業必備的九項條件, 以檢核評鑑是否達成專業的標準; 稍後, Worthen, Sanders 和 Fitzpatrick(1997)增修為十項, 這十項檢核的標準為:(一)具有對於評鑑專家的需求;(二)具備評鑑所需的專門知識與技能;(三)設有培育評鑑人員的正式學程;(四)提供評鑑人員穩定的職業生涯機會;(五)評鑑功能制度化(institutionalization);(六)具有評鑑人員認證程序;(七)成立評鑑人員的專業學會;(八)專業學會發展出決定會員入會資格的規準;(九)評鑑學會對於評鑑人員的培育學程具有影響力;(十)發展出引導評鑑實施的標準。Worthen 等人依照上述十項規準, 去衡量美國評鑑專業化的程度, 發現除了上述的(六)、(八)、(九)三項外, 其餘各項均已達到, 不過嚴格地說, 美國的評鑑事業尚不能稱為是完全成熟發展的專業。對照國內的教育評鑑專業化情況, 若以上述 Worthen 等人(1997)所提出的十項評鑑專業化

標準，來衡量國內的教育評鑑專業化，幾乎尚未達成任何一項標準。

三、教育評鑑的後設評鑑亟待提倡

評鑑既然是一價值判斷的歷程，評鑑過程和結果實際上無可避免的會受到評鑑方法、工具和對象，以及評鑑人員背景、專業、經驗或偏見的影響。因此，Scriven(1969)首度提出「後設評鑑」(meta-evaluation)的用法，意指對任何評鑑、評鑑系統或評鑑服務本身的評鑑，目的在避免不正確及偏誤的報告產生誤導；Stufflebeam(2000)將後設評鑑定義為：「描述、獲取和應用有關評價評鑑的實用性、可行性、適當性及正確性資料的過程，用以修正評鑑及提出評鑑的優缺點」；簡言之，後設評鑑即「評鑑的評鑑」，目的在對評鑑程序或結果，進行後續的評鑑成效評估，據以擬定具體建議，作為控制評鑑品質與改進評鑑規劃設計的參考(Martin,1982)。Nilsson 和 Hogben 1983 年指出後設評鑑的對象，不僅包括評鑑之特定研究，同時包括對評鑑本身的功能和實施的層面加以評鑑(引自 Worthen & Sanders,1987:370)。郭昭佑(民 89)認為改進學校層級評鑑的關鍵策略中，應有後設評鑑的回饋，目的在使評鑑制度的規劃、設計和實施，能與時俱進，避免重蹈覆轍，促使評鑑趨於完善，學校則因評鑑而受益、成長。至於，後設評鑑者可以是原評鑑者，亦可以是評鑑結果使用人，或其他有能力的評鑑者(胡斯淳，民 92)。

後設評鑑的意義與價值雖如上所述，但觀照我國教育評鑑的現實場景中，後設評鑑並未受到應有之重視，而顯得相當欠缺。例如，郭昭佑(民 89)分析八十七至八十八學年度國民教育階段的學校層級評鑑實況，指出所有的評鑑幾乎均無後設評鑑。因而，之後有許多有關教育評鑑的研究，包括學前幼稚園、國民中小學、高級中學、專科學校等各級學校評鑑，以及特殊教育、校務評鑑、校長評鑑系統等各式評鑑研究中，各自提出的研究結果，均在在強調與倡議應妥善規劃、實施後設評鑑(林劭仁，民 89；邱錦興，民 92；胡斯淳，民 92；許韡穎，民 90；曾淑惠，民 85；游家政，民 83；鄭新輝，民 90；穆慧儀，民 91；賴志峰，民 86；蘇慧雯，民 91)。後設評鑑非僅止於學術研究上的倡導而已，已然成為教育評鑑必備的要素，位居教育評鑑主導地位的政府行政當局，實應加緊腳步、積極有效率地將後設評鑑納入教育評鑑之實施過程中。

四、教育評鑑倫理的重視有待倡導

評鑑倫理和研究倫理一樣重要，評鑑倫理的議題提醒我們：評鑑必須符合道德上的要求。評鑑倫理的內涵可以從下列三方面加以說明(曾淑惠，民 91)：

- (一) 評鑑倫理在於規範使用合理、合法的方法蒐集資料，並對所呈現的資料負責。
- (二) 參與評鑑必須確認制度的重要性在個人好惡之上，個人的價值必須建構於倫理上。
- (三) 評鑑倫理的責任由評鑑關係人所共同負擔，包括：評鑑人員、評鑑贊助者、財物提供者、參與者及閱聽者等評鑑關係人。

引申之，評鑑倫理的範疇，包括評鑑委託人與評鑑人員間的倫理問題，評鑑人員的倫理問題，評鑑過程中的倫理問題，以及評鑑報告的倫理問題(曾淑惠，民 91)。許多國外評鑑專業組織，例如：美國全國師資認可審議會(National Council for the Accreditation of Teacher Education, 2002)、加拿大評鑑學會(Canadian Evaluation Society)、澳洲評鑑學會(Australasian Evaluation Society)等，均各訂有倫理信條，以期在評鑑的各個階段和各個層面上，詳細、明確、嚴謹的規範評鑑的實施。

對照我國的現況，可以發現評鑑小組多為臨時任務編組，評鑑委員多為大學院校教授、中小學資深校長、主任，這些人員所具備之學術地位或道德聲望，固然令人尊敬，但是否具備評鑑專業能力，難免令人生疑；加以台灣地小人稠，教育圈內人際互動頻繁，評鑑委員與受評鑑者在過去或具師生、同事之情誼，應注意與落實所謂迴避原則，以避免影響評鑑的專業性、可信度、公正性。另外，我國評鑑專業組織的成立時日尚短，似乎尚未意識到評鑑倫理的重要性，實有必要針對評鑑倫理的問題進行思考，並著手研擬評鑑倫理信條。

五、教育評鑑的種類過多有待整合

國內教育評鑑的種類、名目實在繁多，以致令接受評鑑者聞評色變、未評先拒。郭昭佑(民 89)的分析發現：八十七至八十八學年度國民教育階段的學校層級評鑑，中央或全省性的相關評鑑約 15 項，各縣市政府自行辦理者在 5-7 項；蘇秀花(民 91)的研究亦指出，教育評鑑的規劃由行政機關主導，教育部主辦的有 10 項，台北市政府教育局自行辦理的評鑑有 9 項，評鑑種類過多，且以單項評鑑為主。

教育評鑑的種類過多尚衍生出其他問題，例如：校務評鑑偏重外部評鑑，多數著重書面審查，評鑑標準項目過多；評鑑多屬強制性，多為自評和訪評，追蹤評鑑不足，後設評鑑闕如；受評對象事先說明不足，受限於訪評時間急迫與經費不足，評鑑委員無法深入瞭解；受評者防衛、應付心態濃厚，評鑑實質意義低、徒具形式等。因而，多數研究建議將校務評鑑的評鑑標準、項目加以整合，以達致評鑑的真正目的，並減輕學校基層教育者的工作負擔與心理壓力(林文榮，民 93；洪梓祥，民 92；莊忠儒，民 93；劉智云，民 93；潘俊程，民 92；蘇秀花，民 91)。

面對種類繁多的教育評鑑，令人喜少憂多。喜的是國內教育評鑑發展在量的方面明顯擴張，然而更令人擔憂的是如此多的教育評鑑，並未達成評鑑的實質意義，反而產生形式化主義的教育評鑑，對我國教育評鑑的長期發展，絕對是弊多於利。面對此一挑戰，實有必要將目前種類過多的教育評鑑加以整合。

陸、教育評鑑實施的有效策略

我國教育部於民國六十四年正式實施大學教育評鑑，迄今將近三十年之久，雖然多多少少產生了一些功能，然提升教育品質的功能仍屬有限。為了強化教育評鑑的效果，茲提出下列實施策略，以供參考。

一、設置「教育評鑑專責機構」，統籌教育評鑑事宜

我國實施教育評鑑種類繁多，就學校層級而言，包括幼稚園、國民小學、國民中學、高級中學、職業學校、專科學校等；就人員而言，包括校長評鑑、教師評鑑；就方案而言，包括體育評鑑、資訊教育評鑑、兩性教育評鑑、教訓輔三合一評鑑、交通安全教育評鑑、特殊教育評鑑、經費使用評鑑，這些評鑑多數為教育行政機關委託辦理，在時間、評鑑人員和實施方式等方面，都缺乏長期和完整的規劃，故實施效果相當有限。未來為落實教育評鑑功能，宜設置財團法人性質的教育評鑑專責機構，統籌教育評鑑事宜，才能達到教育評鑑效果。

二、整合各種教育評鑑類別，減輕受評者評鑑壓力

國內中小學校務評鑑種類繁多，若將各種方案評鑑的單項評鑑包括在內，林林總總高達十多項，造成受評學校備感壓力和心理負擔，影響到評鑑功能的發揮。因此，中央教育行政機關和地方教育行政機關宜相互協商，共同討論，將部分單項評鑑整合在校務評鑑內容之中，一方面可避免學校接受一連串的評鑑，影響到校務的正常運作，一方面亦可減少評鑑人力和時間不必要的浪費，以提升評鑑的效能。所以，整合當前各種教育評鑑類別，將力量集中在整體的校務評鑑，使學校不會受到太多的校務評鑑干擾，深信可收到較佳的評鑑效果。

三、訂定明確評鑑倫理規範，增進評鑑結果公信力

教育評鑑是一種專業性的工作，評鑑人員需要專業的評鑑知能，也需要遵守一定的評鑑倫理規範，才能符應評鑑專業化的需求。國內正式辦理教育評鑑，已將近

三十年，可惜皆未能建立評鑑的專業倫理信條，來規範或約束評鑑人員的行為，使評鑑人員不致於做出傷害受評學校的各種行為。今後為增加評鑑的公信力與客觀性，評鑑人員倫理準則的訂定，實有其必要性。理想上，教育評鑑人員倫理準則宜由專業團體訂定之，要求其會員遵守之；然而國內教育評鑑，除大學評鑑部分委託專業團體辦理外，其餘中小學評鑑，多數是由教育行政機關邀集學者專家、校長代表、教師代表和家長代表組成評鑑小組。當今權宜之計，仍宜由教育行政機關訂定評鑑倫理準則，以供評鑑人員參考，較為適切。

四、強化評鑑人員專業教育，提升教育評鑑品質

評鑑人員的專業素養，深深影響到教育評鑑的實施成效。因此，為了提升教育評鑑的效果，評鑑人員的專業教育實屬重要。而評鑑人員的專業教育，一部分來自於職前養成教育，一部分有賴於在職進修教育。由於國內之評鑑人員大都是臨時邀集而成，其素質參差不齊，水準不一，而且亦無評鑑人員的養成教育，所以評鑑效果有所受限。政府未來可以考慮邀請各大學的教育政策與行政研究所，或教育行政與評鑑研究所，規劃評鑑人員的專業學程，負責培訓教育評鑑人員；此外，政府亦可考慮辦理評鑑人員證照制度，凡是具有證照者，才有資格參與評鑑工作，以提升教育評鑑品質。

五、持續研發適切評鑑工具，提供教育評鑑參考

俗語說：「工欲善其事，必先利其器。」教育評鑑要展現其效能，具有良好信度與效度的評量工具，實屬關鍵所在。因為評鑑人員從事評鑑工作，都是以評鑑工作所列出的規準和項目作為依據，萬一工具有瑕疵，則評鑑結果必大打折扣。因此，設計一套能符應教育現場需求，而且有效評量出評鑑所要達到的目標的工具，就顯得格外的重要。所以，教育行政機關應編列一筆教育評鑑工具研發經費，成立專案小組或委託學者專家組成研究小組，有系統和長期性地從事評鑑工具的研發工作，才能發展出適切的評鑑工具。

六、實施評鑑後之追蹤評鑑，落實教育評鑑功能

教育評鑑工作的實施，不能流於「為評鑑而評鑑」，應該朝向「為改進而評鑑」，所以評鑑報告完成，只是完成評鑑工作的一個段落，事後的追蹤評鑑，更是重要。任何教育評鑑報告，除了列出受評單位的主要成果或績效之外，而且也會提出一些改進建議，供受評單位參考。這些建議相當寶貴，教育行政機關應要求受評單位依

報告提出改進時間表，並於一定期間內組成訪評小組到受評單位評鑑，以了解各校改進的實際情形。唯有做好評鑑後的追蹤評鑑，教育評鑑的功能才會彰顯。

七、辦理後設教育評鑑，持續改進教育評鑑缺失

教育評鑑告一段落之後，教育行政機關宜組成一個專案研究小組，針對此次的教育評鑑的實施情形（包括評鑑的目的、工具、程序、時間、學校反應）進行總檢討，以作為未來實施教育務評鑑的參考，讓教育評鑑能夠做得一次比一次好，最後受益者就是學校和學生，這也是教育評鑑之本質所在。所以，一次完整的教育評鑑，不僅是評鑑前的準備和評鑑的實施，更重要的是評鑑後的檢討和改進，這種教育評鑑的動態循環過程，才能確保未來繼續辦理評鑑的品質。所以做好後設教育評鑑工作，應列入每年例行性教育評鑑業務的一部分，以建立更有品質的教育評鑑機制。

八、致力教育評鑑本土化，建構適合國情評鑑模式

國內教育評鑑過於依賴外來評鑑理論，始終無法生根。因此，倡導教育評鑑本土化，建立教育評鑑主體性，實有其必要性。所以，全面喚醒教育學者對學術研究的文化自覺、文化自信，並培育質佳量足的研究人才，惕勵自勉於提升專業研究能力，以厚植教育評鑑本土化根基，朝向擺脫依賴他國學術研究成果的附庸、邊緣地位的目標而努力；此外，也要更廣泛、深入地引進、介紹與評析先進國家教育評鑑的重要文獻，以掌握國際最新進之發展潮流、思想理念，並有系統地針對我國教育制度與情境脈絡之特殊，所產生之獨特教育評鑑問題，進行深入之研究；同時也要有組織地整理、分析我國教育評鑑的研究文獻，建構出本土特有的教育評鑑理論，因應我國特有之教育文化脈絡，發展出合適、獨特、實用、創新的教育評鑑方法、模式、工具或標準，將是未來努力課題。

柒、結語

教育評鑑理論的建構，經過學者們的努力，已經慢慢建立其學科的獨特性，逐漸朝向專業的脚步邁進，讓教育評鑑成為一門有系統的知識體系，教育評鑑學日趨成熟。

教育評鑑的起源，實難論斷，部分學者們認為禮記學記中的考校制度，是我國學校實施教育評鑑的最早記載，但許多學者仍認為科舉制度是教育評鑑的起源，論

及現代教育評鑑的觀念和實施，則是發源自近百年在美國、英國和歐陸等諸先進國家。我國正式實施教育評鑑，始自民國六十四年的大學評鑑，迄今約有三十年光景，遠遠落後於歐美國家，因而一些教育評鑑的做法，常常須借助於歐美教育評鑑的理論與模式。

我國教育評鑑經過多年的努力，多多少少達到提升教育品質和改進教育缺失之效，然而因受到環境、觀念和制度諸多的限制，當今已遭遇了一些挑戰，包括教育評鑑本土化有待努力、教育評鑑專業化仍待提升、教育評鑑的後設評鑑亟待提倡、教育評鑑倫理的重視有待倡導、教育評鑑的種類過多有待整合，仍有待努力。

因此，本文特別提出未來我國實施教育評鑑的策略，包括下列八項：1.設置「教育評鑑專責機構」，統籌教育評鑑事宜；2.整合各種教育評鑑類別，減輕受評者評鑑壓力；3.訂定明確評鑑倫理規範，增進評鑑結果公信力；4.強化評鑑人員專業教育，提升教育評鑑品質；5.持續研發適切評鑑工具，提供教育評鑑參考；6.實施評鑑後之追蹤評鑑，落實教育評鑑功能；7.辦理後設教育評鑑，持續改進教育評鑑缺失；8.致力教育評鑑本土化，建構適合國情評鑑模式。

教育評鑑是確保教育品質的重要手段，隨著社會大眾對於教育品質的呼聲高漲之際，教育評鑑日益受到重視，將是大勢所趨。教育評鑑是一段漫長的歷程，其成效非一朝一夕即可達成，有賴教育行政機關、學術界、各級學校和家長群策群力，發揮集體智慧，長期努力，方能竟其功。

參考文獻

- 王秀玲和林新發（民 93）。臺灣小學教育改革政策：現況、內涵與評析。載於張明輝主編：**教育政策與教育革新**（頁 84-114）。台北市：心理。
- 吳明清（民 82）。教育評鑑在教育決策的應用——一個概念架構的說明。載於伍振鷺主編：**教育評鑑**（頁 133-144）。台北市：南宏。
- 吳芝儀、李奉儒譯（民 84）（Patton, M. Q. 原著）。**質的評鑑與研究**。台北縣：桂冠。
- 吳清山（民 88）。**教育改革與教育發展**。台北市：心理。
- 吳清山（民 91）。校務評鑑的實施挑戰與因應策略。**教師天地**，117，6-14。
- 吳清山、林天祐（民 88）。教育名詞解釋：教育評鑑。**教育資料與研究**，29，66。
- 李緒武（民 82）。教育評鑑的意義與發展。載於伍振鷺主編：**教育評鑑**（頁 1-12）。台北市：南宏。
- 李聰明（民 61）。**教育評價的理論與方法**。台北市：幼獅。

- 林天祐(民 93)。校務評鑑專業化的探討。載於張明輝主編：**教育政策與教育革新** (頁 320-338)。台北市：心理。
- 林天祐、蔡菁芝(民 90)。教育評鑑的理念分析。**教育研究**，91，36-44。
- 林文榮(民 93)。**台南縣國民小學校務評鑑實施現況調查及整合之研究**。國立中正大學教育研究所碩士論文，未出版。
- 林秀珍(民 88)。教育理論本土化的省思。**教育研究集刊**，42，1-15。
- 林劭仁(民 89)。**我國高級中學後設評鑑指標之研究**。國立政治大學教育研究所博士論文，未出版。
- 林萬義(民 76)。教育評鑑理論之研究。**研習資訊**，34，17-28。
- 邱錦興(民 92)。**臺北市國民小學校務評鑑之後設評鑑研究**。臺北市立師範學院國民教育研究所碩士論文，未出版。
- 侯光文(民 85)。**教育評價概論**。河北：河北教育出版社。
- 洪梓祥(民 92)。**台中縣國民小學實施校務評鑑之研究**。臺中師範學院國民教育研究所碩士論文，未出版。
- 胡斯淳(民 92)。**國民小學身心障礙特殊教育評鑑後設評鑑知覺研究**。屏東師範學院特殊教育學系碩士論文，未出版。
- 秦夢群(民 89)。**教育行政 - 實務部分**。台北市：五南。
- 張春興(民 78)。**張氏心理學辭典**。台北市：東華。
- 張植珊(民 68)。**教育評鑑**。臺北市：教育部教育計畫小組。
- 張德銳(民 89)。台灣學校教育評鑑制度。載於林天祐主編：**台灣教育探源**(頁 103-105)。台北市：國立教育資料館。
- 莊忠儒(民 93)。**臺北縣國民小學校務評鑑實施狀況調查研究**。國立台北師範學院教育政策與管理研究所碩士論文，未出版。
- 許韡穎(民 90)。**特殊教育後設評鑑指標之建構**。國立高雄師範大學特殊教育學系碩士論文，未出版。
- 郭昭佑(民 89)。**學校本位評鑑**。台北市：五南。
- 陳玉琨(民 87)。大學教育評估的理論基礎和未來研究的趨勢。載於胡悅倫(主編)：**海峽兩岸大學教育評鑑之研究**(頁 191-215)。台北市：師大書苑。
- 陳玉琨(民 93)。**教育評鑑學**。台北市：五南。
- 陳漢強主編(民 86)。**大學評鑑**。台北市：五南。
- 陶西平主編(民 87)。**教育評價辭典**。北京市：北京師範大學。
- 曾淑惠(民 85)。**我國專科學校後設評鑑之研究**。國立台灣師範大學工業教育研究

- 所博士論文，未出版。
- 曾淑惠（民 91）。**教育方案評鑑**。台北市：師大書苑。
- 游家政（民 83）。**國民小學後設評鑑標準之研究**。國立台灣師範大學教育研究所博士論文，未出版。
- 湯志民（民 92）。北市國中校務評鑑實施之評析。載於國立教育資料館主編：**現代教育論壇（八）**，台北市：國立教育資料館。
- 黃光雄編譯（民 78）。**教育評鑑的模式**。台北市：師大書苑。
- 黃炳煌（民 72）。**教育與訓練**。台北市：文景。
- 楊文雄（民 70）。**教育評鑑之理論與實際**。南投縣：台灣省教育廳。
- 劉本固（民 89）。**教育評價的理論與實踐**。浙江：浙江教育出版社。
- 劉智云（民 93）。**大都會區國民小學校務評鑑之研究**。國立臺東大學教育研究所碩士論文，未出版。
- 潘俊程（民 92）。**臺中縣國民小學校務評鑑之研究**。國立臺中師範學院國民教育研究所碩士論文，未出版。
- 潘慧玲（民 93）。邁向下一代的教育評鑑：回顧與前瞻。載於國立台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論文集**（頁 11-23），台北市：國立台灣師範大學。
- 潘慧玲、徐昊昊、黃馨慧、張志偉（民 92）。第四屆綜合高中評鑑之實施。**教育研究資訊**，11（3），133-156。
- 鄭新輝（民 90）。**國民中小學校長評鑑系統之研究**。國立政治大學教育研究所博士論文，未出版。
- 鄭新輝（民 91）。國民中小學校長評鑑系統建構的基本理念。**初等教育學報**，15，183-236。
- 盧增緒（民 84）。論教育評鑑觀念之形成。載於中國教育學會（主編）：**教育評鑑**（頁 3-59）。台北市：師大書苑。
- 穆慧儀（民 91）。**國民中學後設評鑑之研究——以臺北市為例**。國立台北師範學院國民教育研究所碩士論文，未出版。
- 賴志峰（民 86）。**台北市幼稚園後設評鑑之研究**。國立政治大學教育研究所碩士論文，未出版。
- 謝文全（民 92）。**教育行政學**。台北市：高等教育。
- 蘇秀花（民 91）。**台北市國民小學校務評鑑實施成效及整合之研究**。台北市立師範學院國民教育研究所碩士論文，未出版。

蘇錦麗 (民 93)。 **展望我國教育評鑑專業之發展**。輯於國立台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論文集 (頁 24-27)。台北市：國立台灣師範大學。

蘇慧雯 (民 91)。 **臺北市幼稚園後設評鑑之研究**。臺北市立師範學院國民教育研究所碩士論文，未出版。

Alkin, F. (1990). *Debates on evaluation*. Newbury Park, CA:Sage.

Boulmetis, J. & Dutwin, P.(2000). *The ABCs of evaluation: Timeless techniques for program and project managers*. San Francisco: Jossey-Bass.

Chelimsky, E. (1985). *Program evaluation: patterns and direction*. Washington, D.C.: American society for public administration.

Douglah, M.(1998).*Program development and evaluation-developing a concept of extension program evaluation*. Retrieved Aug. 31, 2004, from <http://www.uwex.edu/ces/pdande/evaluate.html>

Dubois, P. H. (1970) .*A history of psychological testing*. London: Allyn & Bacon.

Guba, E., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA.: Sage.

House, E. R. (1993). *Professional evaluation*. Newbury Park, CA.: Sage.

Joint Committee on Standard for Educational Evaluation.(1994). *Program evaluation standard*.(2nd ed.). Thousand Oaks, CA.: Sage.

Martin, P. H. (1982).Mata-analysis, meta-evaluation and secondary analysis. (ERIC Document Reproduction Service, No.228280)

McLauaglin, M.(1990).The rank change agent study revisited: Marco perspectives and mirco realities. *Educational Researcher*,19(9), 11-16.

National Council for the Accreditation of Teacher Education (2002) . *Handbook for accreditation visits*. Washington, DC: Author.

Nevo, D. (1995). *School-based evaluation: A dialogue for school improvement*. New York: Pergamon.

Popham, W. J.(1993). *Educational evaluation*.(3rd ed.). London: Allyn and Bacon.

Scriven, M. (1967).The methodology of evaluation. In R. E. Stake(Ed.). *Curriculum evaluation*. American education research association monograph series on evaluation, 1. Chicago: Rand McNally.

Scriven, M. (1969). *Evaluation skills*. Washington, D. C.: American Educational Research Association.

Scriven, M. (1994). Evaluation as a discipline. *Studies in Educational Evaluation*, 20, 147-166.

- Scriven, M. (2003). Evaluation theory and metatheory. In T. Kellaghan, D. L. Stufflebeam & L. A. Wingate(ed.) *International handbook of educational evaluation (part one:perspectives)*(pp15-31).London: Kluwer Academic.
- Shadish,W.R.(1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19 (1), 1-19.
- Shadish, W. R., Cook, T. D., & Leviton, L. C.(1991). *Foundations of program evaluation: theories of practice*. Newbury Park, CA.: Sage.
- Stufflebeam, D. L.(1973). An introduction to the PDK book: Educational evaluation and decision-making. In B. R. Worthen & J. R. Sanders(Eds.). *Educational Evaluation: theory and practice*. CA:Wadsworth.
- Stufflebeam, D. L. (2000). Foundational models for 21st century program evaluation. In D. L. Stufflebeam, A. J. Shinkfield, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed.) (pp. 33-83). Boston: Kluwer Academic.
- Stufflebeam, D. L. (2003). The CIPP model for evaluation. In T. Kellaghan & D. L. Stufflebeam(Eds.). *International handbook of educational evaluation*.(pp. 31-62). Dordrecht: Kluwer Academic.
- Stufflebeam, D. L. & Shinkfield, A. J.(1985). *Systematic evaluation*. Boston: Kluwer-Nijhoff.
- Tenbrink, T. D. (1974). *Evaluation: A practical guide for teachers*. New York: McGrew-Hill.
- Worthen, B. R.(1994). Is evaluation a mature profession that warrants the preparation of evaluation professionals? *New Directions for Program Evaluation*,62,3-15.
- Worthen, B. R. & Sanders, J. R. (1987). *Educational Evaluation: Alternative approaches and practical guidelines*. New York: Longman.
- Worthen, B. R., Sanders, J. R. & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practical guidelines* (2nd ed.). New York: Longman.

The Conception and Development of Educational Evaluation

Ching-shan Wu & Hsiang-Li Wang

ABSTRACT

Planning, implementing and evaluating are three key components for education sustainable development. In the past decade, we have been devoted to educational reforms. Educational evaluation plays an important part in understanding the performance of educational reforms and assuring the educational quality. The purpose of this article is to discuss the basic conception and recent development of educational evaluation. It divides into five parts. We begin to analyze the important conception of educational evaluation. Secondly, we illustrate the characteristics of discipline and theoretical construction of educational evaluation. Next, we review the history of educational evaluation and present the educational evaluation development in USA and Taiwan . We then to analyze the challenges of Taiwan's educational evaluation, including the efforts of localization, the promotion of profession, advocacy of meta-evaluation and ethics of evaluation, as well as the integration of various evaluations. Finally, we address eight strategies for improving our educational evaluation effectiveness.

Key Words: Evaluation, Educational Evaluation, The History of Educational Evaluation