

教育評鑑的後設評鑑

游家政、曾祥榕

【作者簡介】

游家政，台灣省桃園縣人。國立台灣師範大學教育研究所博士。現任國立台灣海洋大學師資培育中心及教育研究所教授。

曾祥榕，台灣省宜蘭縣人。國立花蓮師範學院國民教育研究所博士班研究生。現任宜蘭縣頭城國小教師。

隨時質問評鑑任一部分的優點和價值，從不會覺得太早或太遲。因為你若不質問它的品質，無疑地，別人也會如此做 (Brinkerhoff et al., 1983, 205)

摘要

本文旨在探討後設評鑑的概念、教育評鑑的後設評鑑標準與原則、教育評鑑的後設評鑑方案的規劃步驟。根據相關文獻的探討結果，本文提出以下的結論：第一，後設評鑑為一種「較高層級的評鑑」或「次層評鑑」，以「原級評鑑」為受評對象，其目的在判斷原級評鑑活動表現的價值或優缺點；第二，教育評鑑的後設評鑑標準是教育評鑑專業化的表徵，目前評鑑界已發展出評鑑標準可供使用，必要時亦可經由系統的研究程序自行訂定適用的標準；第三，後設評鑑也是一種評鑑，具有敏感的社會政治性特質，為確保教育評鑑的後設評鑑能夠順利地進行和獲得精確的結果，可參照本文所列出 16 項規劃步驟。此外，本文提出以下四點建議：(1)教育評鑑單位應依據後設評鑑標準來設計與實施教育評鑑方案；(2)任何教育評鑑方案應將「後設評鑑」列為方案的必要步驟之一，在設計和實施中、實施後進行後設評鑑；(3)後設評鑑的結果應列入教育評鑑報告中，以供評鑑結果的使用者參考；(4)教育界和評鑑界攜手合作，共同發展我國的教育評鑑後設評鑑標準。

關鍵詞彙：後設評鑑、評鑑標準、評鑑方案

Key Words: metaevaluation, evaluation standards, evaluation program

前言

台灣近年來由於社會急遽變遷，連帶地促發了接二連三的各級學校教育改革，進而引起各界對教育評鑑的迫切需求。例如台灣加入 WTO 之後，大學便面臨國際競爭的壓力，教育部除了對大學進行學門評鑑（例如管理學門、化學學門、醫護學門、師資培育學程等）外（教育部高等教育司，1993；蘇錦麗，1995），並鼓勵大學辦理自我評鑑，藉以改進教學和研究品質，吸引國內外優秀師資和學生。台北市、高雄市和台灣省大多數縣市教育局，也定期或不定期對其所屬學校進行校務評鑑，意圖經由評鑑來了解辦學績效，甚至做為校長續聘與否的重要參照標準。

尤其當大學自主、以及九年一貫的國民教育課程綱要（以下簡稱「課程綱要」）實施以後，教育專業權下放到大學和學校，如何確保教育的品質，則有賴更精良的內部人員評鑑或專業取向的教育評鑑，特別是學校或教師的自我評鑑（Simons, 1988）即認為只有經由自我評鑑的過程，教師才能持續改進學生的教育經驗。歐用生（1996，125-136）也指出只有教師不斷的自我研究與改進，專業的地位才得以獲得各界的認同。因此我們可以預見，教育評鑑無疑地將是下一波教育改革的重點。

就國內教育評鑑的研究而言，至今的發展尚未非常完備（黃光雄，2003；潘慧玲，2003；蘇錦麗，2003）。觀諸目前國內教育評鑑的運作，大都由教育行政機關負責規劃與實施，而學校只是處於被動的位置，消極地應付上級的評鑑。這種遵循由上而下的評鑑模式，不論在評鑑設計上，或就實際的運作情形，都存在一些問題，例如：評鑑概念被窄化為學生的升學率或校長的辦學績效，評鑑方法重技術而輕忽理論，評鑑內容偏重背景與輸入而忽略過程和結果，評鑑過程流於形式化的靜態資料查閱與檢核，評鑑結果的運用漸趨政治化，以及評鑑專業人員不足等（黃政傑等，1992，37；游家政，1995；游進年，2003；盧增緒，1995）。

以上的問題也非全然是我國教育評鑑所獨有，即使是教育評鑑較為發達的美國，亦曾出現許多病症（游家政，1995；PDK National Study Committee on Evaluation, 1972）：(1)逃避評鑑，認為評鑑是一種痛苦的過程，除非絕對必要，否則儘量避免接受評鑑，因此地方學區很少將評鑑列為工作要項；(2)焦慮感，評鑑者與受評鑑者視評鑑為審判，因而心生焦慮或恐懼；(3)對評鑑無反應，學校沒有對評鑑做出認真的反應，校內亦缺乏負責評鑑的系統；(4)懷疑評鑑的價值，對評鑑採取防衛的態度，不信任評鑑的結果；(5)缺乏實施評鑑的指引；(6)提出不當的建議；(7)忽略無顯著的差異，評鑑者對無顯著差異結果的解釋，忽略了其他可能的潛在因素；以及

(8)缺乏評鑑所需的要素，評鑑的實施缺乏合適的理論、或可用的概念、設計、工具、訓練有素的評鑑人員等。

為了改進教育評鑑的缺失，預防教育評鑑結果的不當利用，確保教育評鑑的品質，除了加強教育人員的評鑑素養，以及積極培育教育評鑑專業人員外，更有賴「後設評鑑」(meta-evaluation)的實施，即教育評鑑本身也必須被評鑑。自1960年代起，評鑑學者即開始投入「後設評鑑」(meta-evaluation)或「評鑑的評鑑」(evaluation of evaluations)的研究。例如：Scriven (1972)從評鑑方法論的觀點，探討後設評鑑的概念與方法；Stufflebeam (1974)區分「形成性後設評鑑」與「總結性後設評鑑」的不同，並提出34項判斷評鑑優缺點的標準，以及64項的後設評鑑實施原則；Cook 和 Gruder (1978)研究後設評鑑的模式，提出七種途徑。

另外一些研究則強調評鑑利用的改進，對影響評鑑結果利用的因素進行分析 (Alkin & Daillak, 1979; Leviton & Hughes, 1981; Patton, 1986; Weiss, 1972, 1981)，或探討評鑑未利用、誤用或濫用的現象與成因 (Agarwala-Rogers, 1977; Cook, et al., 1981)。

美國「西密西根大學的評鑑中心」(Western Michigan University's Evaluation Center)更於1975年結合了12個評鑑學會組成「教育評鑑標準聯合委員會」(Joint Committee on Standards for Educational Evaluation, 以下簡稱「聯合委員會」)¹，發展教育評鑑標準和人事評鑑標準。美國「評鑑研究學會」(Evaluation Research Society, ERS)也效法跟進，結合各類方案評鑑的專家學者，發展出方案評鑑標準。這些標準除了提供評鑑者和利害關係人判斷評鑑品質的依據外，也可做為規劃、設計與實施評鑑的準則 (Rossi, 1982)。

國內一些學校評鑑的研究也提出加強後設評鑑的建議 (教育部高等教育司, 1993, 31; 黃政傑等, 1993)，前台灣省教育廳所設計的「國小訓輔工作評鑑」也將

¹ 「教育評鑑標準聯合委員會」(Joint Committee on Standards for Educational Evaluation)及1994年發行《方案評鑑標準》一書的會員，包括：美國學校行政學會 (American Association of School Administrators)、美國教育研究學會 (American Educational Research Association)、美國評鑑學會 (American Evaluation Association)、美國教師聯盟 (American Federation of Teachers)、美國心理學會 (American Psychological Association)、諮商評估學會 (Association for Assessment in Counseling)、視導與課程發展學會 (Association for Supervision and Curriculum Development)、加拿大教育研究學會 (Canadian Society for the Study of Education)、州立學校最高主管委員會 (Council of Chief State School Officers)、高中後教育認可委員會 (Council on Postsecondary Accreditation)、全國小學校長學會 (National Association of Elementary School Principles)、全國中等學校校長學會 (National Association of Secondary School Principles)、全國教育測量委員會 (National Council on Measurement in Education)、全國教育學會 (National Education Association) 及全國學校董事會學會 (National School Boards Association)。

後設評鑑列為正式的必要步驟（國立新竹師範學院，1991，6）。不過，後設評鑑如何進行，以及根據什麼標準來判斷評鑑品質，仍有待研究。筆者以關鍵字查詢國內以「後設評鑑」為題的博、碩士論文，僅有博士論文 3 篇，碩士論文 5 篇²；至於學術期刊部分也不多，可見後設評鑑的議題仍未受到教育學術界的重視。因此，後設評鑑觀念的澄清、理論與方法的建立，對我國教育評鑑研究和實務的改進和發展應該是相當迫切。

據此，本文採文獻分析的方法，並參酌Stufflebeam、Madaus 和 Kellaghan(2000) 主編《評鑑模式：從教育和人類服務評鑑的觀點（第二版）》（*Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, 2nd ed.）的第 24、25 章的架構³，首先剖析後設評鑑的意涵，其次論述教育評鑑的後設評鑑標準與原則，第三分析教育評鑑的後設評鑑方案的規劃步驟，最後提出結論和建議，以供國內教育評鑑界和教育界的參考。

後設評鑑的意涵

從評鑑專業的角度而言，評鑑本身必須被評鑑，這種形式的評鑑活動即被稱為「後設評鑑」（Stufflebeam, 2000a）。Brinkerhoff 等人（1983, 205）在其所著的《方案評鑑訓練手冊》上寫道：「隨時質問評鑑任一部分的優點和價值，從不會覺得太早或太遲。因為你若不質問它的品質，無疑地，別人也會如此做」。換句話說，不論評鑑者、評鑑的當事者或被評鑑者為了解和改進評鑑的品質，其必要步驟之一就是後設評鑑。以下就後設評鑑的定義、目的與功能、範圍分析於后。

一、後設評鑑的定義

從後設評鑑的歷史發展可知，後設評鑑概念是基於了解和改進教育評鑑品質的需求而形成的。早在 1940 年，Orata 曾提出「評鑑的評鑑」的概念。俟後，在 1969 年 Scriven 提出「後設評鑑」一詞，並將其定義為「評鑑的評鑑」：

後設評鑑就是第二級評鑑（second-order evaluation），意指評鑑的評鑑。

² 國家圖書館全國博碩士論文資訊網（<http://datas.ncl.edu.tw/theabs/00/index.html>, 2004/08/29）。

³ 第 24 章 評鑑專業標準與原則 和第 25 章 後設評鑑方法論 的作者即為《評鑑模式：從教育和人類服務評鑑的觀點（第二版）》一書主編之一的 Daniel L. Stufflebeam。

就理論而言，後設評鑑涉及評鑑角色的方法論評估；就實務而言，後設評鑑關注明確具體的評鑑表現的評鑑（Scriven, 1972, 84）。

典型的後設評鑑，即由另一個評鑑者去評鑑一項已計劃好的或已完成的評鑑。換言之，後設評鑑將原來的評鑑者置於受評者（the valuee）的位置，對該評鑑者及其活動表現進行價值判斷（Scriven, 1991）。

Cook（1974）則以「次級評鑑」（secondary evaluation）一詞代替「後設評鑑」，將原來的評鑑相對地稱為「原級評鑑」（primary evaluation）。Cook 的定義如下：

次級評鑑即：（1）試圖經由檢核研究資料及其解釋和啟示，重新評估特定評鑑研究與潛在政策的相關性；（2）試圖利用基礎研究或評鑑的現有資料，評估當前措施或變通方案的主要假定，及其對未來政策的實證性支持程度（Cook, 1974, 159）。

Stufflebeam（1974, 69-70）則採用 Scriven 的術語，認為後設評鑑是「較高層級評鑑」（higher-order evaluation），或「評鑑的評鑑」。基本上，後設評鑑是一種評鑑的形式，其概念必須和評鑑概念相互契合：

1. 評鑑是優缺點的評估。因此，後設評鑑意指評估原級評鑑實施的優缺點。
2. 評鑑為決策和績效服務。因此，後設評鑑必須在原級評鑑實施前提供資訊，支持評鑑者做成實施評鑑的決策；以及在原級評鑑實施後提供資訊，協助評鑑者了解其評鑑工作的品質。
3. 評鑑應評估方案的目標、設計、實施和結果。後設評鑑應評估原級評鑑目標的重要性、設計的合適性、實施的妥當性、結果的品質與重要性。
4. 評鑑必須為所有受到方案影響的人服務。因此，後設評鑑應當為評鑑者及對其評鑑工作感興趣的人服務。
5. 評鑑必須由內部人員和外部人員實施。通常（但並不是常常），內部人員實施形成性評鑑，以獲得決策的資訊；外部人員實施總結性評鑑，以了解績效。因此，評鑑者必須對進形中的評鑑實施形成性評鑑，以及對已完成的評鑑活動之整體優缺點實施外部的判斷。
6. 評鑑包括描述問題、獲取所需的資訊、以及利用資訊做決策和判斷績效等程序。因此，後設評鑑者必須進行此三個步驟：首先，必須描述特定的後設評鑑問題；其次，必須蒐集、組織、分析所需的資訊；最後，必須將所獲得的

資訊應用到合適的決策和績效責任上。

7.評鑑必須在技術上是合適的、有用的和符合成本利益的。因此，後設評鑑必須滿足這些相同的規準。

據此，Stufflebeam (1978, 9-10; 2000b) 進一步將後設評鑑定義為：經由記述、獲取及運用，有關評鑑活動的效用性、可行性、適當性及正確性等描述性資訊和判斷性資訊的過程，據以引導評鑑，並公開報導其優點和缺失。

Cook 和 Gruder (1978) 則將後設評鑑限定在對總結性評鑑進行評鑑。Straw 和 Cook (1990) 承續前述的觀點，認為後設評鑑係泛指對一項評鑑或一組評鑑的技術品質及其結果進行評鑑的各種活動。

從以上的定義可知，評鑑學者普遍將「後設評鑑」視為「評鑑的評鑑」，即對「原級評鑑」進行「較高層級評鑑」或「次級評鑑」的「歷程」。首先，就「原級評鑑」而言，係指後設評鑑的對象為一個或一組「原級評鑑」的活動表現。這些活動表現包括評鑑的計畫、設計、目的、角色、方法、結果和報告，涵蓋技術和理論兩個層面。

其次，就「較高層級評鑑」或「次層評鑑」而言，係指後設評鑑是對原級評鑑所進行的第二層、第三層的評鑑。此一界定可以從「後設」(meta)一詞的意義窺知。在詞典上，「後設」的定義至少包含(Gove, 1979, 1418):「後來發生」(occurring later)或「繼先前活動之後」(in succession)或指「位於其後的」(situated behind)或「事後的」(posterior)、超越(beyond)或批評(critique)等意義。因此，「後設」一詞是相對於「原級」而言，指出「後設評鑑」的發生時機和立場：(1)後設評鑑是在原級評鑑的某些階段性活動之後進行，即對評鑑的規劃、設計、實施和總結等各項活動實施形成性後設評鑑；(2)後設評鑑是在原級評鑑完成之後進行，即對評鑑報告的品質或評鑑結果的影響實施總結性的後設評鑑；(3)後設評鑑是從較高層次的立場，對原級評鑑進行批評。

最後，就「歷程」而言，係指後設評鑑是一種系統地描述、獲取和使用敘述性和判斷性資訊的過程。具體地說，後設評鑑是一種價值判斷的過程，其步驟至少包括判斷標準的訂定、原級評鑑表現資訊的蒐集、原級評鑑表現品質或優缺點的判斷、以及後設評鑑結果的運用等。

二、後設評鑑的目的與功能

由前述後設評鑑的定義——評鑑的評鑑，可知原級評鑑是後設評鑑所要評鑑的「對象」。後設評鑑的目的即在於判斷原級評鑑的價值；也就是將原級評鑑置於受評

者的位置，對該評鑑者及其活動的表現進行價值判斷。不過，此乃是就後設評鑑的整體目的而言，不因後設評鑑的對象而有變化。在整體目的之下，依據後設評鑑在特定情境的運作，則可分為許多不同的功能。

評鑑的功能，基本上可區分為「形成性評鑑」與「總結性評鑑」兩種 (Scriven, 1967, 43)。大多數的評鑑學者都贊同這樣看法，並將後設評鑑區分為「形成性後設評鑑」與「總結性後設評鑑」。Stufflebeam (1974, 72-73) 即認為，後設評鑑的任務有二：其一，為「做決定」(decision making) 服務，經常是由內部人員所實施的「形成性後設評鑑」。在評鑑活動之前，對評鑑的設計和實施，提供適時的建議，藉以確保評鑑結果在技術上合乎妥當、有用和成本/效能的標準。其二，為「績效責任」(accountability) 服務，係指「總結性後設評鑑」，大多由外部人員對已完成的評鑑工作，提出優缺點的判斷，藉以確認評鑑者的績效和評鑑結果的可靠性。

後設評鑑除了具有品質控制的功能，尚具有統整各種評鑑研究的功能。Cook (1974, 160-174) 從「再評鑑」(to reevaluate) 原級評鑑報告資料的途徑，認為實施後設評鑑的主要理由有八項：(1)減低評鑑結果的不明確性；(2)獲得另一種觀點和評鑑獨立的精神；(3)突破學術本位的限制，建立科際整合的評鑑研究；(4)提供多層面的評鑑訓練功能；(5)整體評估各種原始評鑑的效果；(6)評鑑知識的累積和多元的資料利用；(7)對傳統評鑑知識的質疑；(8)增進評鑑專業的能力。這八項理由係基於政策與評鑑研究的關係而提出的。對決策者而言，經由資料的多元分析與利用，可以協助決策者確認方案評鑑的結論，評估方案評鑑的整體績效，獲得新的評鑑結論或新知識，做為選擇變通方案的參考。對評鑑研究而言，後設評鑑可以協助評鑑者獲得科際整合的學術訓練，增進評鑑的專業能力，擴展評鑑的研究領域。

歸結而言，後設評鑑的目的在判斷原級評鑑的價值或優缺點，而後設評鑑所具有的功能有三：第一，為「做決定」服務的形成性後設評鑑，旨在協助評鑑者改進評鑑的實施或確保評鑑的正確進行，提高評鑑的可行性；第二，為「績效責任」服務的總結性後設評鑑，旨在判斷已完成的評鑑活動之優缺點，確保評鑑品質、影響和利用；第三，為「評鑑研究」服務的啟發性後設評鑑，旨在突破學術本位的限制，建立資料多元使用和科際整合的評鑑研究途徑，藉以發現新的評鑑觀點與知識。

三、後設評鑑的對象與內容

後設評鑑究竟要評鑑什麼？其範圍和內容每因後設評鑑功能的差異而有不同的設計。如前述 Stufflebeam 將後設評鑑的功能有二：其一，為「評鑑決策」服務的形成性後設評鑑；其二，為「評鑑績效」服務的總結性後設評鑑。依據這個觀點，

Stufflebeam 將後設評鑑設計區分為形成性和總結性兩個取向。如表一所示，設計(1)至設計(4)屬於形成性後設評鑑，分別評估原級評鑑的目的、設計、過程和結果，以做成評鑑的決策；設計(5)則為總結性後設評鑑，評估原級評鑑整體工作的績效。

表一 後設評鑑的功能與對象

| 後設評鑑的功能 | 後設評鑑的對象 | | | |
|-----------------------|---------------------------|---------------------|---------------------|---------------------------|
| | 評鑑目的 | 評鑑設計 | 評鑑過程 | 評鑑結果 |
| 為「評鑑決策」服務的 形成性後設評鑑 | 設計(1) 評鑑目的之形成性評估 | 設計(2) 評鑑設計之形成性評估 | 設計(3) 評鑑實施之形成性評估 | 設計(4) 評鑑結果的品質與利用之形成性評估 |
| 為「評鑑績效服務」的 總結性後設評鑑 | 設計(5) 整體評鑑工作的優缺點之總結性評估 | | | |

資料來源：修改自 Stufflebeam, 1974, 78。

(一) 形成性後設評鑑

為了做成評鑑的決策，形成性後設評鑑必須在原級評鑑進行前和進行中實施，適時提出有關原級評鑑應如何設計和實施的建議，藉以確保原級評鑑能夠獲得正確的、有用的和符合成本／效益的結果。由表一可知，Stufflebeam (1974, 70-74) 依據評鑑對象的不同，將形成性後設評鑑分為評鑑目的、評鑑設計、評鑑實施與評鑑結果等四種設計。

1. 評鑑目的形成性評估

評鑑目的之形成性後設評鑑，旨在確認各種變通的評鑑目的，以及決定其優先順序。此一設計的資訊需求如下：(1)確認原級評鑑的當事人、可能的評鑑目的之範圍、以及評鑑目的排序規準；(2)對每項變通的評鑑目的進行邏輯分析，包括何人是這项目的之服務對象？什麼問題將被回答？為何這位當事人想要知道那個問題？什麼行動可能指引這項評鑑目的之達成？(3) 決定評鑑目的之優先順序，由評鑑小組和當事人的代表，依據選定的標準（例如明確、範圍、重要性和可行性），或採取系統的德懷術方法，進行目的之排序；(4) 提出選擇評鑑目的之建議，做為原級評鑑者及其委託人進行評鑑的基礎。評鑑目的形成性後設評鑑，可以由原級評鑑者來實施，亦可以委託後設評鑑者來進行。

2. 評鑑設計形成性評估

評鑑設計之形成性後設評鑑，旨在確認各種變通的評鑑設計，並做成優先順序的決定。由於評鑑設計的內容通常包括取樣、工具、處理和資料分析等要素，因此決定各種變通評鑑設計優先順序的依據，以技術、效用、成本／效益等層面的標準為主。此一設計的資訊需求如下：(1)確認各種變通的評鑑設計和排序的標準；(2)應用選定的標準，評估各種變通設計的優點；(3)提出選擇評鑑設計的建議。評鑑設計形成性後設評鑑的實施者，和第一種設計一樣，可以由原級評鑑者或委託後設評鑑者擔任。

3. 評鑑實施形成性評估

評鑑實施之形成性後設評鑑，旨在指導既定的評鑑設計之實施。此一設計所涉及的資料蒐集和分析，包括定期的評鑑設計的檢查和評鑑過程的追蹤，旨在判斷設計是否適當地運作，以及此一設計被實現的程度。此一設計提供兩種回饋資訊：其一，評鑑實際過程的紀錄，這些資訊將有助於解釋評鑑的結果；其二，關於評鑑問題的確證和改進評鑑活動的建議，這類回饋對於評鑑過程的管理是重要的。評鑑實施之形成性後設評鑑，可由評鑑行政人員要求原級評鑑者定期做口頭或書面的進度報告，或者聘請外部的審查人員定期檢核評鑑設計的實施情形。

4. 評鑑結果形成性評估

評鑑結果之形成性後設評鑑，旨在加強評鑑結果的品質和利用。此一設計的資訊需求包括：(1)評鑑目的之說明；(2)與評鑑目的有關的後設評鑑標準，例如技術上的適切性、效用性和成本／效益等的說明；(3)原級評鑑結果的預定使用者的確認。這些資訊是評估評鑑品質和影響的基礎，包括：評鑑報告的效度、信度和客觀性的評定，原級評鑑的開支紀錄，預定當事人使用評鑑報告的紀錄，以及這些當事人對評鑑報告效用的評定。這些評鑑成效的資訊，可以從評鑑管理者或外部的審查人員等處取得。由評鑑結果之形成性後設評鑑所獲得的回饋，包括評鑑工作的品質、影響和成本／效益的定期報告，可用來評定評鑑結果的成效，以及提出改進評鑑工作的建議。

(二) 總結性後設評鑑

為了判斷評鑑績效，必須在原級完成後進行總結性後設評鑑，評鑑的內容以整體評鑑工作為範圍，判斷評鑑目的、設計、實施和結果等的優缺點。此一設計的資訊需求如下：

1. 評鑑決定者的意圖，包括：預定服務的當事人有哪些？獲致哪些評鑑結果？選擇什麼設計達成這些目的？如何進行此設計？造成哪些特定的影響？

- 2.分析有關評鑑目的的資訊，包括：預定的當事人想從評鑑得到什麼？是否還有其他的合法當事人？這些人想得到什麼資訊？評鑑者是否考慮過各種變通目的？這些變通目的是什麼？為何未使用這些變通目的？
- 3.分析有關評鑑設計的資訊，包括：此評鑑設計在技術的適切、效用、效率和可行等標準的評定如何？是否考慮過其他的設計？選定的設計是否優於其他可能的選擇？
- 4.分析有關評鑑實施的資訊，包括：是否貫徹實行？如果沒有，遭遇哪些困難？評鑑者是否有效解決有關概念、社會政治、法律、技術、管理和倫理等方面的問題？以及實施的成本如何？
- 5.分析有關評鑑結果方面的資訊，包括：評鑑目的是否達成？獲得哪些資訊？這些資訊的品質如何？評鑑結果是否被使用？被誰使用？是否正確地使用？

總結性後設評鑑是以整體評鑑工作為範圍，判斷評鑑各項工作的優點和缺點。最後的評鑑結果必須彙總為後設評鑑報告，提供原級評鑑的當事人參考改進。由於總結性的評鑑經常被用來做成擴大、繼續或終止方案的依據，其爭議性特別明顯。因此，為了提高總結性後設評鑑可靠性和客觀性，評鑑實施者通常以外部人員為主，藉以說服原級評鑑的相關當事人。

後設評鑑的設計雖然可分為形成性和總結性兩種取向，不過形成性和總結性兩種評鑑是相對的而非絕對的概念（黃政傑，1987，59）。亦即，形成性後設評鑑可視為階段的總結性後設評鑑，而過去的總結性後設評鑑亦可做為未來教育或學校評鑑改進的形成性後設評鑑。因此，形成性和總結性後設評鑑的設計，必須依據其功能或作用，而不是依據其蒐集的資料和方法。

四、後設評鑑對台灣教育評鑑的啟示

根據上述的分析，後設評鑑對評鑑領域的重要性，就如同審計制度對於會計領域的關係一樣。如果只靠會計師的財務資訊，而缺少獨立的審計人員的詳細檢查，這個社會將會處於高度的危險之中。換言之，後設評鑑的存在是符合全民利益的（Stufflebeam, 2000b）。

同樣地，如果學校或教育體系只有教育評鑑，家長、學生、教育人員、企業和雇主、以及其他利害相關者又不信任教育評鑑的結果，那麼社會也會陷入危險之中。因此，後設評鑑需要詳細檢視教育評鑑系統，包括學校或校務評鑑、課程評鑑、教學評鑑和學習評鑑、以及教育人員評鑑（如校長評鑑、教師評鑑）等。基於以上理由，教育評鑑的後設評鑑也是符合全民利益的。就消極面而言，後設評鑑能確保教

育評鑑方案的品質，避免利害關係者受到不當教育評鑑的傷害或迫害；就積極而言，後設評鑑能發揮正面的功能，引導教育評鑑邁向專業化、學術科際化，促進教育評鑑理論和實務的發展。

檢視當前台灣教育評鑑的後設評鑑研究，根據「國家圖書館全國博碩士論文資訊網」（見註②）所蒐集的碩、博士論文（民國 45 至 93 年），以「評鑑」為題目的論文計有 377 篇（碩士 351 篇、博士 26 篇），其中包含教育評鑑研究 94 篇（分別為教育評鑑 16 篇、校務或學校評鑑 14 篇、課程評鑑和教科書評鑑 12 篇、教學和教師評鑑 35 篇、校長評鑑 10 篇、學習評鑑和學習評量 7 篇）。至於教育評鑑的後設評鑑研究則有如鳳毛麟爪，僅有 8 篇（碩士 5 篇、博士 3 篇）。相對於整體評鑑研究和教育評鑑研究，後設評鑑研究仍有待教育界投入更多的關注。

教育評鑑的後設評鑑原則和專業標準

大部分的專業和其他許多公共服務領域的成員，都必須遵守該項專業的表現和服務標準或規範，其目的在於確保消費者和社會免於受到不當服務措施的傷害。教育評鑑也不例外，在 1980 年代早期，兩個訂定評鑑標準的方案出現，並且延續到現在。其一為成立於 1975 年的「教育評鑑標準聯合委員會」（The Joint Committee on Standards for Educational Evaluation，以下簡稱「聯合委員會」），其二為成立於 1976 年的「評鑑研究協會」（Evaluation Research Society, ERS），她與「評鑑網路」（Evaluation Network, E NET）於 1986 年合併為「美國評鑑學會」（American Evaluation Association, AEA）。本節旨在剖析教育評鑑的後設評鑑標準，主要參照 Stufflebeam（2000a）的「評鑑專業標準和原則」一章和游家政（1994）《國民小學後設評鑑標準》博士論文、以及其他相關論文撰寫而成。

一、「評鑑研究協會」的評鑑標準和「美國評鑑學會」的評鑑原則

在 1977 年 7 月，「評鑑研究協會」理事長成立一個「七人委員會」，其所有成員都是評鑑專家，由 Scarvia B. Anderson 擔任主席，負責研發評鑑標準，並於 1982 年 9 月發表「方案評鑑標準」一文，提出 55 項標準，分為下列六大類（游家政，1994；ERS Standards Committee, 1982; Stufflebeam, 2000a）。

（一）規劃和協商（formulation and negotiation）

「規劃和協商」有 12 項標準，建議評鑑者在進行評鑑之前，必須儘可能地

向委託人或客戶說明，並以書面方式詳述評鑑的事項，包括如何完成此項工作、誰去做、誰受到服務、如何避免受到利益衝突的影響、如何保護參與者和人員對象、評鑑的經費預算，以及評鑑的限制等。

（二）結構和設計（structure and design）

「結構和設計」有 6 項標準，指出評鑑計畫的結構和設計需能描述系統的、合理的探究程序，且要考慮相關的脈絡情境。這些標準要求在設計評鑑時，能夠針對研究方案的價值，提出具辯明性的結論。評鑑計畫應該明確地呈現基本研究設計並證明其正當性、取樣程序、資料蒐集工具，以及和評鑑有關的方案人員、其他參與者之間必要合作關係的安排等。

（三）資料蒐集和準備（data collection and preparation）

「資料蒐集和準備」有 12 項標準，指出資料蒐集的過程必須事先加以規劃，包括選訓資料蒐集員；保護資料來源和人員對象的權利；監督、控制及紀錄資料的蒐集；控制偏差；評估程序和工具的信、效度；降低研究過程對研究方案的妨礙與干擾；以及控制資料取得的管道等。

（四）資料分析和解釋（data analysis and interpretation）

「資料分析和解釋」有 9 項標準，說明在評鑑設計和實際蒐集資料的限制下，如何調節資料的分析和解釋。這些標準要求評鑑者必須將分析程序和評鑑目的相配合；描述特定分析程序的使用並證明其正當性；運用合適的分析單位；探究量的結果在實際上和統計上的意義；藉由參照評鑑設計及消除似是而非的對立說明以支持因果解釋；以及明確區分客觀發現、意見、判斷及臆測之間的差異。

（五）溝通和公布（communication and disclosure）

「溝通和公布」有 10 項標準，強調評鑑者在整個評鑑過程當中，必須使用有效的溝通方式。這些特定的要求包括：決定發布結果的權限；配合公開資料的政策與過程以組織資料；清楚地、完整地、公正地且正確地呈現結果；說明不同結果的相對重要性；清楚說明評鑑潛在的假設與限制；做好說明評鑑過程的準備；以及依據事先的公布合約將適當結果通知有知的權利之閱聽大眾。

（六）結果的使用（use of results）

最後一類「結果的使用」有 6 項標準，強調評鑑者在評鑑的任何階段中，都必須小心地關照所有潛在使用者的訊息需求。故評鑑者應該在做出適當決定之前就要提出報告；對於結果的誤解與誤用，應該儘可能事先處理；指出評

鑑過程可能產生的副作用；精確地區分評鑑發現和建議；謹慎地提出建議；以及小心地區分他們的評鑑角色和他們可能扮演的倡議角色。

在1986年，「評鑑研究協會」和「評鑑網路」合併成為「美國評鑑學會」之後，其領導人員即決定對前述的評鑑標準，補充最新的評鑑原則，以供評鑑人員參考，其內容分為以下五大項：(1)系統的探究——評鑑者對受評對象實施系統的且以資料為根據的探究，旨在責成評鑑者在質的和量的探究上，必須達到最高的可用技術標準；(2)能力——評鑑者對利害關係人提供勝任的表現，旨在要求評鑑者必須發展、並適當地發揮他們的專業；(3)公正/誠實——評鑑者必須確保整個評鑑過程的誠實和公正，旨在確保評鑑是合乎倫理的；(4)尊重他人——評鑑者必須尊重與他們互動的人員，包括回應者、方案參與者、委託人及其他利害關係人的安全、自尊及自我價值等，旨在要求評鑑者對參與評鑑的所有團體，都應展現適當的體貼；(5)對一般大眾福利的責任——評鑑者應清楚的說明及考慮與一般大眾福利有關的利益和價值之差異性，旨在要求評鑑者不能短視，相反地要關心評鑑與社會的關聯性。基本上，「美國評鑑學會」訂出的評鑑原則與原先「評鑑研究協會」的標準是一致的，只是說明的呈現變得比較短而已。

二、「聯合委員會」的評鑑標準

「聯合委員會」在1975年成立於「西密西根大學評鑑中心」(Western Michigan University Evaluation Center)，其創始的17個成員是由12個專業組織所任命的(見註①)。這些組織及其指派的成員都是來自不同的專業領域，包括學校認可、諮詢和輔導、課程、教育行政、教育測量、教育研究、教育管理、方案評鑑、心理學、統計學及教學。「聯合委員會」所制定的評鑑標準有二：其一為1981年出版《教育方案、計畫及教材的評鑑標準》(*Standards for Evaluations of Educational Programs, Projects, and Materials*)一書，並於1994年修訂後改名為《方案評鑑標準》(*The Program Evaluation Standards*)；其二為1988年出版的《人員評鑑標準》(*The Personnel Evaluation Standards*)。到了1990年代末期，則開始從事學生評鑑標準的訂定。目前，「聯合委員會」已經被「美國國家標準局」(American National Standards Institute, ANSI)認可為全美唯一具有公信力、且得以制定教育評鑑標準的團體(Stufflebeam, 2000a)。

(一) 教育方案評鑑標準

「聯合委員會」將教育方案評鑑標準分成效用性(utility)、可行性

(feasibility) 適切性 (propriety) 及精確性 (accuracy) 等四大屬性，計有 30 項標準。表二是《方案評鑑標準》的摘要，「聯合委員會」建議評鑑者和委託人都要遵守《方案評鑑標準》，俾使其評鑑工作能夠滿足健全的評鑑應具備之四大屬性。評鑑專業人員更應該要研讀《方案評鑑標準》的所有內容，如此才能對每個標準內化，並明智地應用在評鑑的每個階段 (Joint Committee on Standards for Educational Evaluation, 1994)。

表二 方案評鑑標準的摘要

效用性 效用性標準旨在確保評鑑必須能夠為特定使用者提供所需要的資訊。

- U1 利害關係人的確認：確認參與評鑑或受評鑑影響的人，並敘述他們的需求。
- U2 評鑑者的可靠性：評鑑者必須值得信賴並具執行評鑑的能力，以確保評鑑結果的可信度和接受度。
- U3 資訊的範圍與選擇：廣泛蒐集和選擇資訊，對方案提出適切的問題，並回應委託人或其他特定利害關係人的需求和利益。
- U4 價值的確認：謹慎地描述結果解釋的觀點、程序和理論基礎，使價值判斷的基礎得以清晰明白。
- U5 報告的清晰性：評鑑報告必須清楚的描述被評鑑的方案，包括它的背景脈絡、目的、程序、以及評鑑結果，使呈現的基本資訊易於為人了解。
- U6 報告的及時性與傳播：重要的期中發現和評鑑報告必須傳達給預定的使用者，以供適時地被使用。
- U7 評鑑的影響：評鑑的規劃、執行和報導方式，應能鼓勵利害關係人堅持完成，以升昇評鑑結果的使用性。

可行性 可行性標準旨在確保評鑑必須是真實的、審慎的、圓融的、樸實的評鑑。

- F1 務實的程序：評鑑的程序必須是實際可行的，且在蒐集所需資訊時，應將干擾減至最小。
- F2 政治上的可行：評鑑的計畫與實施應該預先考慮不同利益團體的期望，以獲得他們的合作，並預防或反制任何團體簡化評鑑過程、扭曲或誤用評鑑結果等不當意圖。
- F3 成本效益：評鑑應該有效率，並提出具價值性的資訊，以證明所耗費的資源是正當的。

適切性 適切性標準旨在確保評鑑的實施是合法的、合倫理的，並保障所有的評鑑參與者和受評鑑結果影響者的福利。

- P1 服務導向：評鑑的設計應能幫助組織提出參與者的各種需求，並提供有效的服務。
- P2 正式的義務：正式的評鑑團體對評鑑之義務（做什麼、如何做、誰去做、何時做），必須獲得同意並形成書面約定，同時善盡合約上的責任、或正式地重新協商。
- P3 受評者的權利：評鑑的設計及執行必須尊重和保護受評者的權利和利益。
- P4 人性的互動：在評鑑的過程中，評鑑者應尊重所有互動者的尊嚴和價值，並避免威脅或傷害參與者。
- P5 完整和公平的評鑑：評鑑必須完整、公正地檢驗和記錄受評方案的優、缺點，使優點才得以言之有據，而問題部分也能被提出來。
- P6 結果的報導：正式的評鑑團體應確保受評鑑影響者和其他具合法權利之相關人員，可以取得完整的評鑑結果及其限制。
- P7 利益的衝突：利益衝突的處理必須開誠佈公，才不會危及評鑑的過程和結果。
- P8 會計的責任：評鑑者對資源的分配與開支，應能反映良好的績效責任程序、以及審慎的和倫理上的責任感，並使支出具績效性和恰當性。

精確性 精確性標準旨在確保評鑑對於決定受評方案優、缺點的特徵，能夠提出足夠的、有效的資訊。

- A1 方案文件：清楚、精確地描述受評方案，並形成書面文件以確認該方案。
- A2 情境分析：詳細檢視方案所在的情境，並確認出可能影響方案的因素。
- A3 目的與程序的描述：監控並詳細敘述評鑑的目的和程序，使其獲得確認和評估。
- A4 具有辯護力的資訊來源：詳盡地描述方案評鑑所使用的資訊來源，並經評估資訊的適切性。
- A5 有效的資訊：選用或發展資訊的蒐集程序，然後再實施，以確保所獲致的解釋，對預定的用途是有效的。
- A6 可靠的資訊：選用或發展資訊的蒐集程序，然後再付諸實施，如此才能確保所獲致的資訊，對預定的用途是可靠的。
- A7 系統化的資訊：系統地檢視評鑑的資訊蒐集、處理與報導，並修正任何的錯誤。

- A8 量化資訊的分析：適當地、系統地分析評鑑的量化資訊，有效地回答評鑑問題。
- A9 質性資訊的分析：適當地、系統地分析評鑑的質性資訊，有效地回答評鑑問題。
- A10 理由充分的結論：明確地充分呈現評鑑所獲致的結論，如利害關係人才能加以評估。
- A11 公平的報導：評鑑的報導步驟應該避免受到個人感情或任何群體偏見的扭曲，評鑑報告才可以公平地反映出評鑑結果。
- A12 後設評鑑：評鑑本身應該參照上述標準和其他適切的標準，進行形成性和總結性評鑑，使實施過程獲得適當的指引；在完成之後，利害關係人也能仔細地檢查其優點和缺點。

資料來源：Joint Committee on Standards for Educational Evaluation, 1994.

雖然，上述每一項標準的建立都經過系統化的過程，其內容包括：標準的說明、標準需求的解釋、理論基礎、實施的指引、陷阱的預測和防止、對過度熱心使用的警告，以及範例等。不過，「聯合委員會」認為這些標準都是發展中（developing）的文件，因此他們鼓勵每一位標準的使用者，針對標準的應用提供批評和建議。此外，「聯合委員會」也定期進行標準的複審與改善工作，以符合「美國國家標準局」（American National Standards Institute, ANSI）認可該委員會的需求。

（二）人事評鑑標準

《人事評鑑標準》旨在為教育人員和董事會成員，提供一個廣泛共識性的一般原則，據以發展和評估優良的、受尊重的、和可接受的人事評鑑系統，以及執行這些原則的有用建議（Joint Committee on Standards for Educational Evaluation, 1988）。全書內容包含了 21 項標準，其結構亦分為適切性、效用性、可行性及精確性四個基本概念。這些標準反映一個事實——人事評鑑的重要性，以及確保人事評鑑者能提供良好的、合倫理的專業表現。此書的使用者包括有學區、社區學院、四年制學院、大學、專業發展組織、以及其他教育機構。

人事評鑑的情境是不同於方案評鑑，因為人事評鑑無論做得多糟糕，基本上是無法避免的。因此，「聯合委員會」主張人事評鑑者應該優先處理適切性標準，主要理由是：適切性標準關注服務方針的議題，特別強調人員評鑑的基本目的在為學生和社會提供有效的、安全的、倫理的服務。《人事評

鑑標準》即在幫助確認學生受到很好的服務，讓服務有持續性的改善，以及讓有害的措施能夠立即被揭露並提出來。但《人事評鑑標準》也特別聲明，人事評鑑措施應該是建設性的，而且免於受到不必要的威脅或有不道德的特質。在積極的氣氛下，人事評鑑也可以幫助規劃一個良好的專業發展經驗，並且協助每一個專業者評估和加強她/他的表現。

三、國民小學學校評鑑的後設評鑑標準

游家政(1994)參照「評鑑研究學會」和「聯合委員會」方案評鑑標準的架構與內容、以及其他有關教育評鑑與後設評鑑標準的相關文獻，研擬「國民小學評鑑的後設評鑑標準」初稿，經德懷術(Delphi technique)的過程，以13位學者專家(包括教育學者、教育行政專家、國小校長、民間文教與企業人士)為樣本，經由三次連續密集的問卷，蒐集不同的意見和建立共識，建構出一套後設評鑑標準，作為判斷國民小學學校評鑑品質的參考。

此套標準係依據評鑑的流程和工作項目，分為規劃、設計、實施、資料分析、報告與結果利用等六個層面(表三)：

- 1.「規劃」層面的標準：係指在做成實施學校評鑑的決定之後，首先必須對即將進行的評鑑活動，研擬一份初步的計畫。其主要的工作要項包括評鑑目的之訂定、評鑑對象的選擇、規劃人員的成員與能力、評鑑經費的編列等。
- 2.「設計」層面的標準：係指在學校評鑑正式實施前，依據事先的規劃，界定評鑑重點、選用或編製方法與工具、訂定評鑑規準、安排評鑑程序和成立評鑑組織。
- 3.「實施」層面的標準：係指學校評鑑資料蒐集活動的進行，包括評鑑前的協調與溝通、學校自評與訪問評鑑的實施。
- 4.「資料分析」層面的標準：係指在學校評鑑資料蒐集活動完成後，所進行的資料分析與解釋。
- 5.「報告」層面的標準：報告層面係指學校評鑑結果的呈現，包括格式的安排、內容的撰寫、報告的傳播與公布。
- 6.「結果利用」層面的標準：係指行政機關、學校的結果利用、評鑑機構的評鑑管理與檢討。

表三 國民小學評鑑的後設評鑑標準摘要

1. 「規劃」層面的標準

1.1 評鑑目的

- 1.1.1 能促使學校切合教育的目標。
- 1.1.2 有助於了解學校現況、問題與困難。
- 1.1.3 有助於引導學校決策與發展。
- 1.1.4 有助於改進學校教育品質。

1.2 評鑑對象

- 1.2.1 能依據評鑑目的的需要決定全面評鑑或抽樣評鑑。
- 1.2.2 每所學校都有被評鑑的機會。

1.3 規劃人員

- 1.3.1 具備學校評鑑的規劃能力與熱誠。
- 1.3.2 能邀請專家或相關人員（如評鑑專家、行政人員、學校實務人員、家長等）參與規劃。
- 1.3.3 完成評鑑展開前的準備工作。

1.4 評鑑經費

- 1.4.1 能編列切合評鑑所需的經費預算。
- 1.4.2 能善用經費預算的額度。

2. 「設計」層面的標準

2.1 評鑑重點

- 2.1.1 能切合學校評鑑的目的。
- 2.1.2 能包含教育過程和結果的面向。
- 2.1.3 能審視學校自我改進的努力與成效。

2.2 方法與工具

- 2.2.1 能依據評鑑目的選用合適的評鑑模式。
- 2.2.2 能採取文件分析、調查、觀察、訪談等有效的評鑑方法。
- 2.2.3 能編製切合資料性質的評鑑表格。
- 2.2.4 評鑑表格實用可行。

2.3 評鑑規準

- 2.3.1 能明確說明各項規準的具體內涵。
- 2.3.2 能切合評鑑的重點。
- 2.3.3 能具有共同一致的評鑑標準。

2.4 評鑑程序

- 2.4.1 評鑑程序具體且符合邏輯順序。
 - 2.4.2 評鑑時間的規劃能達成評鑑目的。
-

2.5 評鑑組織

- 2.5.1 學校自評委員會能由校內各部門及家長代表組成。
- 2.5.2 訪評小組能由具學校評鑑能力的專家（含同僚）組成。
- 2.5.3 訪評小組人數能依據學校規模和評鑑範圍彈性調整。
- 2.5.4 訪評小組的公正性能獲得各校信任。

3. 「實施」層面的標準

3.1 協調與溝通

- 3.1.1 能向受評學校及評鑑委員說明評鑑目的、過程和判斷標準。
- 3.1.2 能明確指出學校必須準備的基本評鑑資料。
- 3.1.3 能明確說明評鑑結果的用途。
- 3.1.4 能激發學校接受評鑑的意願。

3.2 學校自評

- 3.2.1 能列入學校行事曆工作要項。
- 3.2.2 能激勵全體教師、家長代表共同參與。
- 3.2.3 能運用團體討論、問題解決與其他有效方法確實地實施。

3.3 訪問評鑑

- 3.3.1 能減少干擾學校的正常作息。
- 3.3.2 能充分利用訪評時間。
- 3.3.3 能切實了解學校背景和自評資料。
- 3.3.4 能觀察學校真實的情況。
- 3.3.5 能運用適切的方法交叉檢核有關資料。
- 3.3.6 能激勵學校有關人員切實檢討。
- 3.3.7 能按照規劃程序實施評鑑。
- 3.3.8 能接受受評學校申辯或說明。
- 3.3.9 學校及其成員的尊嚴能受到適度的尊重。
- 3.3.10 評鑑委員能彼此充分溝通評鑑結果並達成共識。

4. 「資料分析」層面的標準

4.1 方法

- 4.1.1 能解答評鑑的問題。
- 4.1.2 能切合資料的性質。
- 4.1.3 能說明資料分析的要項。
- 4.1.4 能描述資料分析的步驟。

4.2 解釋

- 4.2.1 能適度考慮學校的情境。
 - 4.2.2 能鑑別受評學校的優點和缺點。
-

4.2.3 能統合運用量與質的資料來解釋。

4.2.4 能忠實反應資料分析的結果。

5. 「報告」層面的標準

5.1 格式

5.1.1 能包含評鑑的發現和建議。

5.1.2 能適合行政機關和學校等相關使用者。

5.2 內容

5.2.1 敘述清晰且具有可讀性。

5.2.2 能兼含正面和負面的發現。

5.2.3 能指出解決問題的責任歸屬。

5.2.4 能提出具有可行性的建議。

5.2.5 能說明評鑑的限制。

5.2.6 能反映評鑑委員一致同意的意見。

5.2.7 能公正客觀報導評鑑結果。

5.3 傳播與公布

5.3.1 能適時提供給行政機關、學校和相關人員。

5.3.2 能審慎處理涉及學校權益的資料。

5.3.3 能保密涉及個人隱私的資料。

5.3.4 能適時向社會大眾公開評鑑結果。

6. 「結果利用」層面的標準

6.1 行政機關

6.1.1 能據以訂定教育改革的優先順序。

6.1.2 能據以提供學校必要的資源或協助。

6.2 學校

6.2.1 能據以訂定學校改革方案。

6.2.2 能據以改進校務的運作。

6.3 管理

6.3.1 能預防評鑑結果的誤解或誤用。

6.3.2 能安排必要的後續評鑑。

6.3.3 能追蹤評鑑對學校教育改革的影响。

6.4 檢討

6.4.1 能檢討評鑑目的之適切性。

6.4.2 能檢討評鑑目的之達成程度。

6.4.3 能檢討評鑑各項措施的適切性。

6.4.4 能將檢討結果提供未來評鑑的參考。

以上六個層面的評鑑標準共計76項，分別為：規劃層面有11項，設計層面有16項，實施層面有17項，資料分析層面有8項，報告層面有13項與結果利用層面有11項。這些標準是從使用的便利性和實用性來考量，儘量力求周詳與完備。但國小後設評鑑標準的使用對象，除了後設評鑑者為當然的使用者外，凡是國小學校評鑑的「利害關係人」都是此一標準的使用者，包括國小學校評鑑的贊助者、委託者、規劃者、實施者、受評者、以及直接或間接受到國小評鑑影響的人員。這些使用者對於後設評鑑標準的需求不盡相同，因此使用者可自行加以選用或精簡。

四、比較與評析

本節簡要敘述了美國「聯合委員會」和「評鑑研究學會」、「美國評鑑學會」所發展出的評鑑標準，以及游家政(1994)經由德懷術所建構的「國民小學評鑑標準」。比較三者之間，不難發現明顯差異和相似的地方。

首先，就制定的機構而言，「聯合委員會」是一個頗具規模的專業組織，她的17個成員係由12個專業組織所指派的，全體會員人數總計在200萬人以上。特別是1994年發行的《方案評鑑標準》，還經「美國國家標準局」認可。「評鑑研究學會」/「美國評鑑學會」則是單一專業組織，其成員數分別為1,000和2,000多人左右。至於游家政(1994)的「國民小學評鑑標準」則屬個人研究，由研究者和13位專家學者所組成之德懷術小組共同建構而成。

其次，就標準的內容而言，在組織結構上有明顯的不同。「聯合委員會」根據評鑑標準的性質，將標準分為效用性、可行性、適切性、精確性四類，計有30項標準。「評鑑研究學會」/「美國評鑑學會」和游家政(1994)則根據評鑑方案的要素，將標準分為規劃與協商(規劃)、結構與設計(設計)、資料蒐集與準備(實施)、資料分析和解釋(資料分析)、溝通與公布(報告)、以及結果的使用(結果利用)等六個層面，分別有76和55項標準。不過，就實質內涵而言，三者則大同小異，皆可視為一般的指引。

第三，在應用方面，就功能而言，「聯合委員會」的評鑑標準表面上似乎以總結性後設評鑑為主，另二者的評鑑標準則較有利於形成性後設評鑑的使用。不過，就實際而言，三者均可適用在總結性或形成性後設評鑑，端賴使用者的習慣和觀念。如就應用對象而言，「評鑑研究學會」/「美國評鑑學會」的標準和原則的適用範圍，跨越廣泛的學科和服務領域；「聯合委員會」的評鑑標準則定位在教育領域，至於游家政(1994)的評鑑標準更侷限在國民小學學校評鑑。

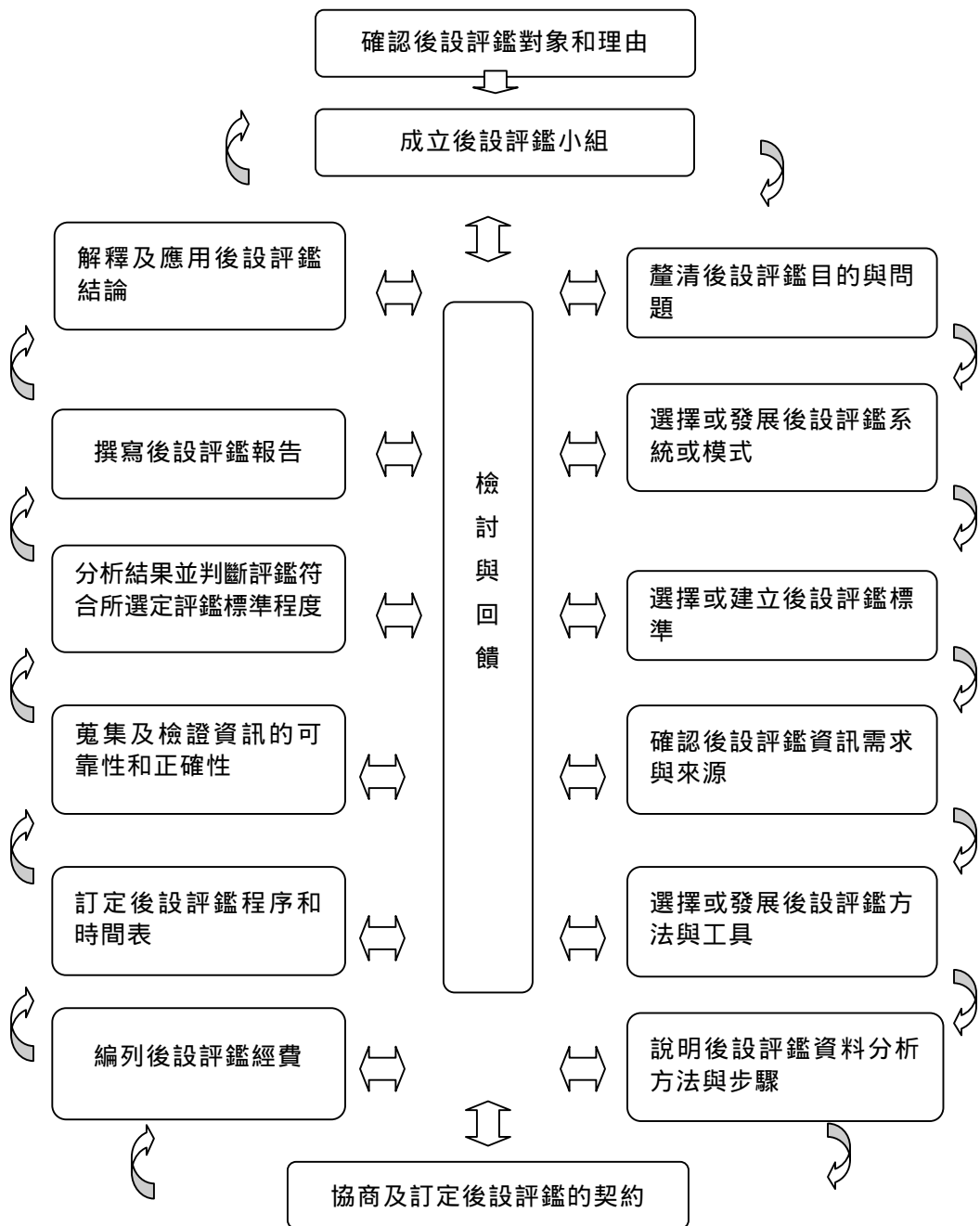
經由上述的比較，雖然各有特色，但「評鑑研究學會」/「美國評鑑學會」的標

準和原則之重點也都涵蓋在「聯合委員會」的標準裡面。再者，教育領域以外的評鑑者亦發現，「聯合委員會」詳細的標準也可以支持「評鑑研究學會」/「美國評鑑學會」的標準和原則。因此，Stufflebeam (2000b) 特別呼籲美國這兩個評鑑專業團體，除了視實際需要繼續修訂和更新他們的標準和原則之外，現在應該是攜手合作的時刻，共同遵守高的標準和原則，並朝向專業評鑑的方向邁進。

教育評鑑的後設評鑑方案之規劃

評鑑規劃對任何的評鑑而言，它都扮演關鍵性的角色，決定爾後評鑑實施的成敗。因為，不良的評鑑規劃，必然無法獲致令人滿意的評鑑結果；而良好的評鑑規劃，則能提供界定清楚的評鑑任務，確保正確的評鑑方向、評鑑過程的完整性、評鑑概念的一致性，以及獲得應有的評鑑績效 (Sanders & Nafziger, 1976, 5)。後設評鑑亦是如此，在正式實施之前，後設評鑑者必須進行規劃的工作。

筆者分析蘇錦麗 (1995)、黃政傑、李隆盛、游家政等 (1994, 99)、Brinkerhoff 等人 (1983)、Worthen 和 Sanders (1987a, 1987b)、Stufflebeam (2000b) 等觀點，發現後設評鑑與原級評鑑的規劃程序具有共通性，其步驟可歸納以下十六項原則性的步驟 (圖一)。



圖一 後設評鑑方案規劃的程序與步驟

一、確認後設評鑑對象和理由

後設評鑑者首先必須確認後設評鑑的對象，其目的在協助後設評鑑者及其委託者了解後設評鑑的範圍和內容。透過後設評鑑對象所提供的資料，相關當事人能夠很快了解後設評鑑對象的全貌。例如以學校評鑑為後設評鑑的對象時，其委託者必須提供學校評鑑的緣起、目的、內容、方法、程序、結果和限制，以及學校的基本背景等資料。

後設評鑑的理由旨在說明為何要實施此一評鑑的原因。例如：評鑑技術不當，未能判斷出方案的優缺點；或者評鑑的績效不彰，無法達成評鑑的目的。由於評鑑是一個充滿社會政治性的過程，在實施過程必須得到相關當事人的支持與合作。因此，後設評鑑者應提出充分的理由來說服當事人。

為了增強後設評鑑理由的說服性，評鑑者可以利用需求評估的方法。需求評估是描述、獲取和應用資訊的歷程，藉以決定某些事務為達成特定目的所必須的或有用的（Stufflebeam & Shinkfield, 1985）。後設評鑑需求評估的目的即在決定有哪些後設評鑑需求存在，以及如何提出這些後設評鑑需求，做為規劃和設計後設評鑑的基礎。換言之，需求評估具有為後設評鑑鋪路的作用，因為需求評估的結果能夠協助解決下列課題（Barak & Breier, 1990, 68）：(1)確認後設評鑑的需求；(2)建立後設評鑑的目的；(3)確認後設評鑑的參與人員，以及這些人員的參與程度；(4)選擇可能達成後設評鑑需求和目的的一般途徑；(5)確認後設評鑑所需的人力和其他資源；(6)建立實施後設評鑑的指導原則，例如截止時間和限制。

需求評估為後設評鑑方案規劃的基礎工作，藉以確保後設評鑑的必要性和正當性。不過，需求評估並非強制性的，只要上述六個要素已經確認，就不必再經過正式的需求評估。

二、成立後設評鑑小組

後設評鑑將由何人實施，在評鑑設計裡也要加以描述。後設評鑑小組的成員必須具有相當的能力，且受到利害關係人的信賴。由於後設評鑑是要判斷原級評鑑的品質，因此後設評鑑者至少必須具有和原級評鑑者一樣的洞察能力。

Worthen & Sanders (1987b, 378) 認為後設評鑑可以由下列三種人來擔任：(1)由原級評鑑者實施後設評鑑，即原級評鑑者利用評鑑標準判斷其評鑑品質，這種方式通常用在形成性後設評鑑。不過，為了避免個人的偏見，最好能夠諮詢外部評鑑人員，藉以提高其客觀性。(2)由評鑑消費者實施後設評鑑，即評鑑贊助者、委託者

或其他的利益相關者自行判斷評鑑計畫或報告的正確性。這種方式能否成功，端賴消費者的評鑑技術能力而定。(3)由具有勝任能力的評鑑人員擔任，即由評鑑專家來判斷評鑑的品質，這種方式最具公信力。

Barak 和 Breier(1990, 70)則提出三種典型的後設評鑑方式：(1)專家判斷，由外部的評鑑專家擔任後設評鑑人員；(2)自評，由學校的教職員擔任後設評鑑人員；(3)結合專家判斷和自評，由外部專家和學校教職員共同擔任後設評鑑人員。

Brinkerhoff 等人(1983, 208-209)分析不同組合的後設評鑑人員後建議：(1)外部的後設評鑑人員的可靠性通常高於內部的後設評鑑人員。(2)一組的後設評鑑人員所關照的技術範圍大於個別的後設評鑑人員。(3)如果必須在評鑑專家和內容專家(例如課程專家)做二選一的選擇，則評鑑專家比內容專家更具重要性。

由上述分析，後設評鑑人員可以由外部專家、原來的評鑑者、評鑑消費者等個別進行，或由前述兩種以上的人員組合。各類人員所占比例的多寡，則視後設評鑑的功能而定。在形成性後設評鑑，以內部評鑑人員為主，其餘的為輔；在總結性後設評鑑，則以外部評鑑人員為主，其餘人員為輔。例如：美國海軍陸戰隊後設評鑑小組在選擇成員時，最重要的考慮為其曾有參與軍事人員評鑑的經驗，以及具有不同種類的後設評鑑專業。新紐約市學區基本教育方案的後設評鑑，就包括了教育研究、方案評鑑、教育政策、學區運作、以及女性和少數族群等人員的觀點，並儘可能加入學校、班級層級的成員。大致而言，後設評鑑的領導者需要規劃出後設評鑑的類型，並且和適當的利害關係人互動，藉以建立合格的且為人接受的後設評鑑小組。當然，有些後設評鑑的委託者，只能僱用一位後設評鑑者，此時就必須找到最值得信賴和最有力量的後設評鑑者(Stufflebeam, 2000b)。

三、釐清後設評鑑的目的與問題

後設評鑑目的指出實施此一評鑑將獲得什麼結果，以及評鑑結果將提供給哪些讀者或利害關係人使用。例如：Barak 和 Breier(1990, 65)認為對大學校院的學程評鑑進行後設評鑑，其主要目的有：(1)獲得回饋，促使評鑑更有效率；(2)重新獲得或加強評鑑的任務；(3)確認改進評鑑可靠性的方法；(4)排除評鑑目的、方法和報告的不確定性；(5)評估評鑑對學程的影響；(6)評估評鑑對學校其他決策過程的影響；(7)決定評鑑的成本是否合理，以及實施成本效益分析的可能性；(8)修正評鑑過程的疏忽所造成的錯誤。後設評鑑的結果除了可以直接提供學校和評鑑者改進評鑑品質的依據外，對於學程規劃人員、教師和學生等也會因學程評鑑的改進而受益。1995年，「為美國而教」(THEACH FOR AMERICAN, TFA)組織對其評鑑試用教師「表

現評估系統」(Performance Assessment System, PAS)實施後設評鑑,目的就是要確認「表現評估系統」的設計和執行過程,是否能夠公平、確實地正確的評鑑初任教師(Stufflebeam, 2000b)。

為達成後設評鑑的目的,後設評鑑者必須研訂後設評鑑問題,因為後設評鑑問題是構成受設評鑑活動的基本要素,後設評鑑所要回答的問題將影響資訊蒐集的種類、資訊的型態和蒐集方法,進而決定資訊分析的選擇。換言之,後設評鑑問題是整體資訊處理的核心因素(Brinkerhoff et al., 1983, 27)。有了後設評鑑目的之後,後設評鑑者就可逐一列出必須解答的問題。

Stufflebeam (2000b)建議在選擇後設評鑑的待答問題上,可以考慮兩個評鑑的基本原則:(1)符合閱讀評鑑報告讀者所需的評鑑資訊的程度(價值),(2)符合健全評鑑要求的程度(優點)。因此,後設評鑑者必須慎重地與委託者及其他利害關係人商議,以確認後設評鑑能夠提出他們認為重要的問題。例如「全國評估管理董事會」(National Assessment Governing Board)的後設評鑑,即試圖為「全國教育進步評估」訂定成就水準,分為「基本的、熟練的、精進的」三個等級。由後設評鑑者與「全國評量管理董事會」共同界定二十餘項的特定後設評鑑問題。以下是兩個具體問題的例子:「1.1 為了要滿足並訂定『全國教育進步評估』成就水準的可信度和權威性,『全國評估管理董事會』的會員要如何適當且充分地代表『全國教育進步評估』的相關人員,以及有效的與利害關係人保持接觸?」和「2.2 為了要確保合適的、具有代表性的標準設立群體,能夠持續地和有效地訂出優異的『全國評估』成就水準,『全國評估管理董事會』訂定成就水準的政策架構和細節是否非常清楚,並且符合相關評量技術的發展現況?」

四、選擇後設評鑑系統或模式

針對後設評鑑的系統,Cook 和 Gruder (1978, 481)依據評鑑實施的方式,歸納出七種後設評鑑的模式(見表四):(1)對一個評鑑報告進行評論,即對一組評鑑資料給予事後的批評;(2)對特定方案的文獻進行評論,即對有關某一個方案的多組資料,進行事後的批評;(3)對一個評鑑或方案進行實證性再評鑑,意指運用某個方案的原始資料,評估該方案評鑑所得結論的效度;(4)對同一個方案的多組資料進行實證性再評鑑,意指運用某個方案的多組原始資料,評估方案評鑑所得結論的效度;(5)諮詢式後設評鑑,由一個或一個以上持續追蹤原級評鑑但卻不參與執行的諮詢者,對進行中的評鑑提供回饋,藉以判斷和改進該項評鑑;(6)同步進行原始資料的二級分析,意指當原級評鑑者仍在分析資料之際,由另外的人士同時進行資料的分

析；(7)多組的獨立評鑑，即由一個以上的評鑑者，對同一個方案，分別設計和進行方案影響評鑑。

表四 後設評鑑的模式

| | | | |
|---------------|-------|--------------|---------------------|
| 在原級評鑑 之後進行 | 不處理資料 | 單一資料組 | 對一個評鑑報告進行評語式評論 |
| | | 多元資料組 | 對一個特定方案的文獻進行評論 |
| | 處理資料 | 單一資料組 | 對一個評鑑或方案進行實證性再評鑑 |
| | | 多元資料組 | 對同一個方案的多組資料進行實證性再評鑑 |
| 與原級評鑑 同步進行 | 不處理資料 | 單一或多元 資料組 | 諮詢式後設評鑑 |
| | | 單一資料組 | 同步進行原始資料的二級分析 |
| | 處理資料 | 多元資料組 | 多組的獨立評鑑 |

資料來源：Cook & Gruder, 1978, 481

前四種模式是在原級評鑑完成後實施，後三種則與原級評鑑同時進行。同步進行的模式較為複雜，且成本較大；而在原級評鑑之後實施的模式，不但成本較少、較具實用性，而且是大多數評鑑者所能掌握的。在評鑑設計裡，評鑑者無論從現有的模式中選擇，或視需要自行發展，最重要的前提是：所選擇或發展的評鑑模式必須能夠達成評鑑目的。

五、選擇或建立後設評鑑標準

為了要決定原級評鑑的價值或優缺點，後設評鑑者需要一套標準，做為判斷的基礎。如上節所述，目前評鑑界已發展了數套評鑑標準，可供後設評鑑者和評鑑界自行選用。Brinkerhoff 等人（1983, 215-217）認為，選用評鑑標準必須考慮其內容和使用（表五）：前者包括標準制定者的權威性、標準的適用範圍、支持文件和更新策略；後者包括標準使用的角色、提升、追蹤和加強、合法使用和追蹤等。

表五 評鑑標準的選用依據

1. 標準的內容

(1) 權威性

- A. 該套標準的制定基礎為何？
- B. 該套標準的權威層級為何？
- C. 該套標準的可靠性和一致性經過哪些方面的檢證？
- D. 該套標準的專業判斷範圍是否已經確定？

(2) 範圍

- A. 該套標準的相關應用範圍是否明確描述？
- B. 該套標準的限制和不適用的領域是否已確認？
- C. 該套標準的修定條件為何？

(3) 支持的文件

有哪些支持該套標準的文件？

(4) 更新的策略

該套標準的更新或修訂策略為何？

2. 標準的使用

(1) 標準的角色

- A. 該套標準如何用在自我約束？
- B. 該套標準如何用在教育當事人？
- C. 該套標準如何用在教育和訓練專業人員？

(2) 標準的提升、追蹤和加強

- A. 該套標準如何用來允許和賦予繼續從事評鑑工作的權力？
- B. 該套標準如何用來建立專業人員的倫理行為？
- C. 該套標準如何用來界定優良和不合標準的評鑑？
- D. 該套標準如何用來發展品質控制評鑑？
- E. 在該專業領域，如何處罰不合標準的表現？

(3) 標準的合法使用

在合法的系統裡，不合標準的表現如何形成和接受制裁？

(4) 追蹤或控制標準的使用

如何追蹤或控制公共機構、一般的當事人和委員使用該套標準？

「聯合委員會」即針對《方案評鑑標準》的適用項目和使用步驟，提出具體的建議。「聯合委員會」認為它最適用在評鑑過程的 10 項工作，包括：決定是否要評鑑、界定評鑑問題、設計評鑑、蒐集資訊、分析資訊、作出評鑑報告、編列評鑑預算、訂定評鑑合約、管理評鑑、聘用評鑑者等。如表六所示，表的左邊是 30 項標準，上方則是 10 項評鑑工作，上、下欄交會處若有打上「X」，則代表「聯合委員會」認定這個標準特別適用在該項評鑑工作上。據此，評鑑者能夠很快地確認出和評鑑工作最有相關的那些標準。

此外，「聯合委員會」也針對標準的使用，提出五個步驟：(1)熟悉《方案評鑑標準》，(2)澄清方案評鑑的目的，(3)澄清方案評鑑的情境，(4)依據目的和情境應用每個標準，(5)決定如何使用這些結果。

當然，後設評鑑者也可以根據評鑑方案的需要，自行發展評鑑標準。Worthen & Sanders (1987b, 211-225)認為評鑑標準的建立，可區為三個層面：首先，擴散面向 (divergent phase)，儘可能從各種不同的來源蒐集資料，包括：確認利害相關者的問題與關心，整合各種不同評鑑途徑或模式，應用教育或評鑑文獻所提出的顯著議題，利用研究文獻所提出的知識，利用專業團體所發展的工具或標準，諮詢學者專家，利用評鑑專家的專業判斷等，加以綜合歸納出完整的標準清單；其次，聚斂層面 (convergent phase)，從完整的標準清單中選出最重要的、最迫切的、具有可行性的項目；最後，在評鑑的過程中還要保持適度的彈性，允許增加新的標準。

國內所發展的後設評鑑標準，大多採文獻分析、訪談、德懷術等方法，例如游家政 (1994) 即先經由文獻分析與歸納，再透過德懷術的程序，發展出「國民小學後設評鑑標準」；林劭仁 (2001) 亦採類似程序，發展「我國高級中學後設評鑑指標系統」；許韓穎 (2002) 則兼採半結構式訪談和德懷術，建構「特殊教育後設評鑑指標」。

表六 30 項標準在 10 項評鑑工作表現的相關重要性分析

| | 決 要 定 評 是 鑑 否 | 界 鑑 定 問 評 題 | 設 評 計 鑑 | 收 資 集 訊 | 分 資 析 訊 | 作 鑑 出 報 評 告 | 編 鑑 制 預 評 算 | 訂 鑑 定 合 評 約 | 管 評 理 鑑 | 聘 鑑 用 人 評 員 |
|--------------|---------------------------------|----------------------------|------------------|------------------|------------------|----------------------------|----------------------------|----------------------------|------------------|----------------------------|
| U1 利害關係人的確認 | X | X | X | | | X | | X | X | |
| U2 評鑑者的可靠性 | X | | | X | | | | X | X | X |
| U3 資訊的範圍與選擇 | | | X | X | | X | X | X | | |
| U4 價值的確認 | | | X | X | X | X | | | | |
| U5 報告的清晰性 | | | | | | X | | | | |
| U6 報告的及時性與傳播 | | | | | | X | | X | X | |
| U7 評鑑的影響 | X | | | | | X | | | | |
| F1 務實的程序 | | | X | X | | | | | X | |
| F2 政治上的可行 | X | | | X | | | | X | X | X |
| F3 成本效益 | X | | | | X | | X | | X | |
| P1 服務導向 | X | X | | | | X | | X | X | |
| P2 正式的義務 | X | | X | X | | | X | X | X | |
| P3 受評者的權利 | | | | X | | | | | | |
| P4 人性的互動 | | | | X | | X | | X | X | |
| P5 完整和公平的評鑑 | | | X | X | | X | | | | |
| P6 結果的報導 | | | | | | X | | X | X | |
| P7 利益的衝突 | X | | | | | | | X | X | X |
| P8 會計的責任 | | | | | | | X | X | | |
| A1 方案文件 | X | X | X | X | X | X | X | X | | |
| A2 情境分析 | X | X | | X | X | X | | | X | |
| A3 目的與程序的描述 | | X | X | X | | | | | | |
| A4 有辯護力的資訊來源 | | | X | X | | X | X | X | | |
| A5 有效的資訊 | | | X | X | | X | | | | |
| A6 可靠的資訊 | | | X | X | | | | | X | |
| A7 系統化的資訊 | | | | X | | | | | | |
| A8 量性資訊的分析 | | | X | | X | | | | | |
| A9 質性資訊的分析 | | | X | | X | | | | | |
| A10 理由充分的結論 | | | X | | X | X | | | | |
| A11 公平的報告 | | | X | | | X | | | | X |
| A12 後設評鑑 | X | X | X | X | X | X | X | X | X | X |

資料來源：Joint Committee on Standards for Educational Evaluation, 1994.

六、協商及訂定後設評鑑的契約

就如同任何一種評鑑一樣，後設評鑑也應該奠基於有效的合約備忘錄或正式的契約之上。因此，事先的契約是影響後設評鑑成功與失敗的因素；若沒有這樣的契約，後設評鑑的過程就會出現誤解、爭議、對評鑑結果作妥協、攻擊、以及委託人放棄合作和撤回經費等（Stufflebeam, 2000b）。

「聯合員會」和「美國評鑑學會」的評鑑標準皆建議後設評鑑者和他們的委託人應該協商，並簽定評鑑同意書，藉以幫助委託者和後設評鑑者互相了解雙方的特殊期望與責任，釐清及預防雙方之間的誤解，並為未來可能發生的評鑑爭議提供解決的基礎（參見前節的標準）。後設評鑑書的同意書應該要敘述清楚，並預留必要的彈性調整空間，允許在評鑑過程當中，經由雙方同意後做適當的修正。

七、確認後設評鑑資訊需求和來源

通常後設評鑑的資訊需求包括原級評鑑的目的、設計、過程和結果（Stufflebeam, 1974），或規劃與協商、結構與設計、資料蒐集與準備、資料分析與解釋、〔報告的〕溝通與公布、結果的利用（ERS Standards Committee, 1982）。到底要蒐集原級評鑑的全部或部分的資訊，則視後設評鑑所要回答的問題而定。因此，後設評鑑者必須根據評鑑問題來指出資訊需求及其來源，也就是從何處去蒐集回答評鑑問題所需的資訊。所蒐集的資訊或證據必須與問題相關，而且還有足夠的資訊來支持待答的問題。

以Stufflebeam所主持的「開放學習的澳洲（Open Learning Australia, OLA）」遠距教育方案評鑑的後設評鑑為例，由於從美國到澳洲需要一筆為數可觀的旅費，委託者基於控制後設評鑑費用的需求，遂以郵寄方式將適當的資訊寄到美國。這些資訊包括了大量的文件，例如信件、計畫書、預算書、契約、資料的蒐集方式、期刊和報紙的文章、實地訪談筆記、報告，以及對報告的回應等。這些文件大部分集中於「開放學習的澳洲」方案的本質、評鑑的背景、評鑑的計畫和程序、評鑑的過程、評鑑結果、方案的宣傳，以及後設評鑑的指導原則等（Stufflebeam, 2000b）。

八、選擇或發展後設評鑑方法與工具

根據前述的資訊需求和來源，評鑑者必須設計資訊蒐集的方法和工具。評鑑常用的方法有專家判斷法、觀察法和調查法（黃政傑，1987）。專家判斷法是邀請專家針對評鑑的項目予以審查，依據其所擁有的知識和專長，提供意見和判斷。觀察法

包括無論量化觀察和實地觀察，是由評鑑研究者到實際的評鑑情境，針對所要評鑑的現象，加以觀察記錄。調查法是藉由受調查者自我報告來蒐集資料，採取的方式有問卷調查、訪談和座談等。

以上各種方法都有其特色、適用對象和實施工具，因此在選擇或發展評鑑方法與工具時，必須考慮資訊的性質。在人力、經費、設備等資源和時間許可的情況下，最好採取多種方法，從多元的管道蒐集質與量的資訊。

九、說明後設評鑑資料分析方法與步驟

資料分析旨在處理所蒐集的資訊，提出評鑑的發現或結論。在規劃評鑑的階段，評鑑者必須說明資料分析的方法與步驟。評鑑研究的資料分析方法，有兩個基本的派典：其一，量的分析，如敘述統計、相關分析等；其二，質的分析，如現象詮釋、內容分析、文件或紀錄分析等。分析的步驟如下（Brinkerhoff et al., 1983, 127-129）：(1)檢閱評鑑目的或問題；(2)依據資料性質選擇量或質的分析方法；(3)說明基本的課題、趨勢、關係和問題；(4)評估可用的證據。

十、編列後設評鑑經費

評鑑是花錢的事業。在講究績效的時代，評鑑的每項費用都必須考慮成本和效益之間的平衡。因此，在評鑑計畫裡必須將後設評鑑所需的費用一一條列，以期將經費做最有效率的使用。

典型的評鑑預算，通常包含的項目有（Alkin & Stecher, 1983, 123-124）：(1)專業人事費；(2)助理人事費；(3)外部人員諮詢費；(4)材料、設備和郵電費；(5)資料處理費；(6)設備租用費；(7)差旅費。以上七項費用，評鑑者必須視實際需要酌予增減。

十一、訂定後設評鑑程序和時間表

在評鑑研究裡，經常會碰到太早評鑑或評鑑時間不夠的問題。因此，擬訂一份符合實際的評鑑程序和時間表，對任何型態的評鑑，都是同樣的重要（Gay, 1985, 34）。後設評鑑亦然，其程序和時間表的訂定，首先必須考慮的是進行後設評鑑的時機，例如，在評鑑過程中進行？或在評鑑完成之後才實施？其次要考慮的是資料蒐集所需的時間，即這個後設評鑑所需的資訊，需要花多少時間來蒐集？第三，要考慮資料蒐集的先後順序，即在蒐集某項資料之前必須先進行哪些資料的蒐集？最後，根據上述三方面的考慮，訂定後設評鑑的起訖時間，以及各項後設評鑑活動的程序和進度。

十二、蒐集及檢證資訊的可靠性和正確性

在協商契約後，後設評鑑者必須依據後設評鑑程序和時間表進行資料的蒐集，後設評鑑所需的資訊，可以從多方面來進行蒐集，例如：(1) 相關的當事人，如原級評鑑的委託者、評鑑者、受評學校的校長、教師、職員、學生和家長等的意見；(2) 文件，如原始資料、評鑑報告、評鑑表格、追蹤報告或相關的評論文獻等；(3) 委託者與受評者在評鑑實施後的表現或事件。

若現有資訊欠適當、不充分或不可信，或者現有資訊之間有相互衝突，在技術上無法回答後設評鑑的問題，則必須進一步的蒐集額外資料，並應用「三角檢定」去檢證其可靠性和正確性。例如前述「開放學習的澳洲」遠距教育方案評鑑的後設評鑑，雖然委託者提供的資訊頗為充實，但仍不足以產生委託者需要的後設評鑑報告，於是 Stufflebeam 便前往澳洲去訪談「開放學習的澳洲」方案的領導者、教職員、學生，以及一些傳統高等教育方案的領導者與教職員等見面。這些額外的訊息獲致以下的結論：「開放學習的澳洲」的需求假定是有問題的，服務的品質是參差不齊的，評鑑的結果是前後不一致的 (Stufflebeam, 2000b)。由此可見，藉由實地訪問澳洲以獲得額外的資訊，對後設評鑑報告的正確性是相當重要的。不過，若蒐集太多的資訊，不是增加系統的負擔，就是使得評鑑的成本提高或不切實際。因此，資訊蒐集的主要課題在以最低的成本，去蒐集足夠的正確資訊 (Brinkerhoff et al., 1983, 75-76)。

十三、分析結果並判斷評鑑符合所選定評鑑標準程度

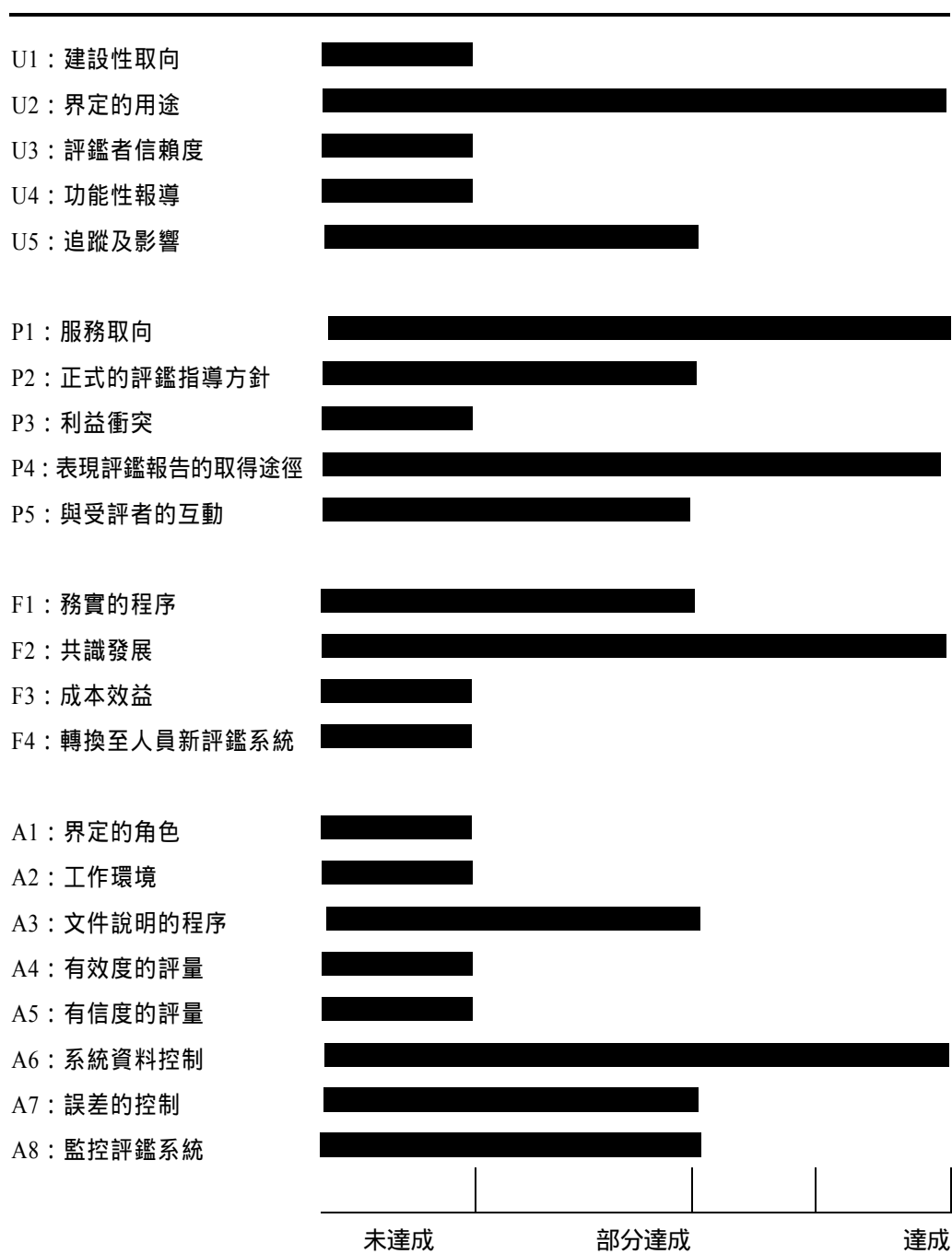
評鑑標準選定或發展完成之後，評鑑者接著需制定評鑑指標和評鑑的方式。常用的評鑑方式有：(1) 採「符合」(或「達成」)和「不符合」(或「未達成」)的二分法；(2) 採評等法，如五點量表；(3) 採加權法，對不同的標準分別給予不同比重的加權計分。

例如美國「陸戰隊人員評鑑系統」的後設評鑑，即依據「聯合委員會」所訂《人事評鑑標準》的 21 項標準，來判斷其人員評鑑系統的整體效用性、可行性、適切性及精確性。如表七所列，依據這個表格，所有的相關證據都被用來鑑定人事評鑑系統對應相關標準的優、缺點，且在每項標準的符合情形會被標示為「未達成」、「部分達成」或「達成」。整體判斷結果如圖二所示。根據表七的項目，圖二的判斷結果可以摘述並繪製成矩陣式的表八。美國海軍陸戰隊基於這樣的分析，遂決定更換其人事評鑑系統，以更符合評鑑標準的要求 (Stufflebeam, 2000b)。

表七 美國「陸戰隊人事評鑑系統」達成人事評鑑標準要求程度的判斷方式

| 標準分類 | 達成標準要求的程度 | | |
|-------|----------------|---|----------------------------|
| | 未達成 | 部分達成 | 達成 |
| 效 用 性 | 1.三個以上的標準未達成。 | 2.五個標準中至少有三個標準達成或是部分達成，至少一個標準未達成。 或 3.達成的標準少於四個，所有五個標準不是達成，就是部分達成，且沒有標準未達成。 | 4.五個標準中至少有四個達成，且無任何標準未達成。 |
| 可 行 性 | 5.三或四個標準未達成。 | 6.至少有二個標準達成或部分達成，以及至少一個標準未達成。 或 7.少於二個標準是達成的，且沒有標準未達成。 | 8.四個標準中至少有二個達成，且無任何標準未達成。 |
| 適 切 性 | 9.三個或以上標準未達成。 | 10.五個標準中至少三個達成或部分達成的，且一或二個標準未達成。 或 11.少於四個標準是達成或部分達成，且沒有標準未達成。 | 12.五個標準中至少有四個達成，且無任何標準未達成。 |
| 精 確 性 | 13.四個或以上標準未達成。 | 14.八個標準中至少五個達成或部分達成，且至少一個標準未達成。 或 15.少於五個標準是達成的，至少有四個標準達成或部分達成，且沒有標準是未達成的。 | 16.八個標準中至少有五個達成，且無任何標準未達成。 |

資料來源：Stufflebeam, 2000b.



圖二 美國「陸戰隊人事評鑑系統」達成人事評鑑標準要求的結果統計圖

資料來源：Stufflebeam, 2000b.

表八 美國「陸戰隊人事評鑑系統」達成人事評鑑標準的結果摘要

| 標準的分類 | 結果 | 表九的標題號碼 |
|-------|-------|---------|
| 效用性 | 未 達 成 | 1 |
| 適切性 | 部分達成 | 10 |
| 可行性 | 部分達成 | 6 |
| 精確性 | 未 達 成 | 13 |

資料來源：Stufflebeam, 2000b.

十四、撰寫後設評鑑報告

依據後設評鑑資訊的分析，後設評鑑者開始準備提出報告，但在某些情況下將會有好幾份的報告。例如海軍陸戰隊的後設評鑑就提出了一系列的報告，包括現行系統的評鑑、工商界或其他軍事人員的評鑑系統、三套競爭性新系統的評鑑、新系統運作和測試的實地測試計畫等。對每一個報告來說，都有一個執行摘要、一份完整的報告、一個相關資料的附錄及一套解說的投影片影本。這些報告都會以口頭和書面方式提出。所有的報告會先以草案的形式提出，經與委託者討論後才完成定稿。

評鑑報告的部分，必須要適合促進結果的利用 (Stufflebeam & Shinkfield, 1985, 182)。因此，有關後設評鑑報告的描述，必須包括的項目有：(1)後設評鑑報告的架構；(2)後設評鑑結果的用途；(3)後設評鑑者可能的偏見；(4)後設評鑑的限制。

十五、協助委託者及其他利害關係人解釋及應用後設評鑑結論

後設評鑑可以是、且經常必須是一種共同合作的工作，尤其當以幫助一個組織評鑑及改革其評鑑系統為目的時，更是如此。若目的是要保護大眾免於受到特定事物評鑑的誤導，評鑑者就必須保持適當的距離，以維護其獨立的觀點。儘管如此，後設評鑑人員仍然必須和閱讀評鑑報告的讀者適度地溝通評鑑報告，以確保他們對結果能有所貢獻、信心、利益、理解及利用。因此，交出後設評鑑報告，並不代表後設評鑑人員已經成功地完成後設評鑑的工作。更重要的，就是提升並協助委託者適切地使用後設評鑑的結果；因此，在後設評鑑開始之前，就應該為此奠定基礎 (Stufflebeam, 2000b)。

十六、檢討與回饋

誠如前述，後設評鑑基本上就是一種評鑑，因此「後設評鑑」也可以、甚至應

該是被評鑑的對象。據此而言，以上十五項要素或步驟也必須經過「再評鑑」或「再檢討」的過程，以求達到盡善盡美，合乎精確性、可行性、適切性和倫理性等評鑑標準。亦即，在後設評鑑方案的規劃與實施過程中，由規劃者或實施者進行「形成性評鑑」，藉以診斷各項步驟的問題，並加以改進；在後設評鑑方案實施完畢後，則進行「總結性評鑑」，可邀請評鑑委託者或贊助者、後設評鑑人員、以及外部的諮詢人員，進行後設評鑑方案績效的討論和批評，以供後續或其它後設評鑑方案的參考。

以上各步驟的順序並非固定不變的，而是可彈性調整或同步進行。不過，有邏輯先後順序要求者除外，例如「列出目的與問題」必須先於「選擇模式選擇」或「指出資訊需求和來源」。

結語

台灣自實施各類各級學校評鑑以來，至今已累積不少教育評鑑報告，但對於評鑑的績效與品質，卻很少進行系統的後設評鑑。尤其當評鑑結果涉及教育資源與學校利益分配之際，評鑑的可靠性與公正性，更是受到學校及其他利害關係人的質疑。例如黃政傑、李隆盛、游家政（1994）「國民小學教育評鑑之研究」，邱錦興（2004）「台北市國民小學校務評鑑之後設評鑑研究」、穆慧儀（2003）「國民中學後設評鑑之研究——以臺北市為例」、蘇慧雯（2003）「臺北市幼稚園後設評鑑之研究」等，皆對評鑑目的、評鑑項目和標準、實施程序與步驟、評鑑者的專業素養、評鑑機構與被評者的溝通、評鑑結果的運用、以及學校本位評鑑制度的建立等提出改進建議。因此，研發客觀的後設評鑑標準和設計良好的後設評鑑方案，將有助於改進與提升教育評鑑的品質，亦可做為社會各界判斷評鑑公信力的參考。本文根據前述相關文獻探討的結果，提出以下的結論：

第一，後設評鑑為一種「較高層級的評鑑」或「次層評鑑」，以「原級評鑑」為受評的對象，其目的在判斷原級評鑑活動表現的價值或優缺點。就其功能而言，後設評鑑可分為「形成性後設評鑑」和「總結性後設評鑑」兩種。前者是在原級評鑑過程中進行，大多由內部人員實施，旨在對評鑑的過程提供適當的建議，藉以確保評鑑結果的品質。後者是在原級評鑑活動結束後進行，大多由外部人員實施，旨在對已完成的評鑑工作提出優缺點的判斷，藉以確認評鑑活動的績效和評鑑結果的可靠性。至於後設評鑑的範圍，則包括評鑑的規劃、設計、實施、資料分析、報告格式與公布、結果利用等，涵蓋原級評鑑的活動過程和結果的表現。

第二，教育評鑑的後設評鑑標準是教育評鑑專業化的表徵，目前評鑑界已發展

出評鑑標準可供使用，例如「聯合委員會」的《方案評鑑標準》、「評鑑研究學會」／「美國評鑑學會」的評鑑標準與原則等可供使用，必要時亦可經由系統的研究程序自行訂定適用的標準。

第三，後設評鑑也是一種評鑑，具有敏感的社會-政治性，為確保教育評鑑的後設評鑑能夠順利地進行和獲得精確的結果，在事前必須訂定周詳的計畫與實施步驟。本文列出 16 項要素：確認後設評鑑對象和理由、成立後設評鑑小組、釐清後設評鑑目的與問題、選擇或發展後設評鑑系統或模式、選擇或建立後設評鑑標準、確認後設評鑑資訊需求與來源、選擇或發展後設評鑑方法與工具、說明後設評鑑資料分析方法與步驟、協商及訂定後設評鑑的契約、編列後設評鑑經費、訂定後設評鑑程序和時間表、蒐集及檢證資訊的可靠性和正確性、分析結果並判斷評鑑符合所選定評鑑標準程度、撰寫後設評鑑報告、解釋及應用後設評鑑結論、檢討與回饋等。這些工作要項的順序並非固定不變，而是可視情境需要或資源多寡來彈性調整或同步實施，但必須符合一定的邏輯先後順序。此外，後設評鑑方案規劃的活動並不是採直線進行，而是在各項工作要項完成後即進行形成性的評估，根據評估結果進行修正後，再繼續下一項工作要項。因此，後設評鑑方案的規劃是一個藉由不斷的回饋與修正的循環方式。

目前國內各級各類學校評鑑，在評鑑完成後，或許有舉行評鑑檢討會，但大多未實施正式的、系統的後設評鑑。因此，對於學校評鑑的效用性、可行性、適切性與精確性，沒有確切的資料可查，以至學校評鑑一再重蹈以往的覆轍，未見有具體的改善措施，學校評鑑的品質無法提升。本文提出以下四點建議：(1)教育評鑑單位應依據後設評鑑標準來設計與實施教育評鑑方案；(2)任何教育評鑑方案應將「後設評鑑」列為方案的必要步驟之一，在設計和實施中、實施後進行後設評鑑；(3)後設評鑑的結果應列入教育評鑑報告中，以供評鑑結果的使用者、後續評鑑或未來評鑑的參考；(4)教育界和評鑑界攜手合作，共同發展我國的教育評鑑後設評鑑標準。

參考文獻

- 林劭仁 (2001), **我國高級中學後設評鑑指標系統之研究**。國立政治大學教育研究所博士論文 (未出版)。
- 邱錦興 (2004), **台北市國民小學校務評鑑之後設評鑑研究**。台北市立師範學院國民教育研究所碩士論文 (未出版)。
- 國立新竹師範學院 (1992): **台灣省國民小學訓輔工作評鑑手冊**。新竹：編者。

- 教育部高等教育司 (1993): **大學校院教育評鑑**。台北: 編者。
- 許韓穎 (2002), **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)。
- 游家政 (1994), **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)。
- 游家政 (1995), **促進國小學校評鑑品質的提升 後設評鑑方案及其標準**。八十四學年度師範學院教育學術論文 (民 84 年 11 月 3 日)。
- 游進年 (2003), 中小學教育評鑑之評析 - 以校務評鑑為例。 **教育研究**, 112 期, 頁 47-61。
- 黃光雄 (2003), 談學校課程評鑑。 **教育研究**, 112, 頁 15-21。
- 黃政傑 (1987), **課程評鑑**。台北: 師大書苑。
- 黃政傑、李隆盛、游家政等 (1994): **國民小學教育評鑑之研究**。行政院國科會專題研究計畫成果報告 (計畫編號: NSC 83-0301-H-003-011)。
- 黃政傑等 (1993), **發展與改進高級中學中程計畫實施成效評鑑**。國立台灣師範大學教育研究中心專題研究成果報告 (編號: 0056)。
- 歐用生 (1996), **教師專業成長**。台北: 師大書苑。
- 潘慧玲 (2003), 教育評鑑之概念釐清與展望。 **教育研究**, 112, 頁 22-30。
- 盧增緒 (1995), 論教育評鑑觀念之形成。載於中國教育學會編, **教育評鑑**。台北: 師大書苑。
- 穆慧儀 (2003), **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)。
- 蘇慧雯 (2003), **臺北市幼稚園後設評鑑之研究**。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 (1995), **大學學門評鑑試辦計畫成效之評估**。台北: 師大書苑。
- 蘇錦麗 (2003), 談教育評鑑之專業性。 **教育研究**, 112, 頁 31-36。
- Agarwala-Rogers, R. (1977). Why is evaluation research not utilized? *Evaluation Studies Review Annual*, 2, 327-333.
- Alkin, M. C., & Stecher, B. (1983). A study of evaluation costs. From M. C. Alkin & L. C. Solmon, (Eds.). *The costs of evaluation* (pp.119-132). CA: Sage.
- Barak, R. J., & Breier, B. E. (1990). *Successful program review: A practical guide to evaluating program in academic settings*. CA: Jossey-Bass.
- Brinkerhoff, R. O. et al. (1983). *Program evaluation: A practitioner's guide for trainers*

- and educators*. MA: Kluwer-Nijhoff.
- Cook, T. D. & Gruder, C. (1978). Metaevaluation research. *Evaluation Quarterly*, 2(1), 5-51.
- Cook, T. D. (1974). The potential and limitations of secondary evaluation. In M. W. Apple, M. J. Subkoviak, & H. S. Lufner (Eds.). *Educational evaluation: Analysis and responsibility* (pp.155-235). Berkeley, CA: McCutchan Pub.
- Cook, T. D., et al. (1981). The misutilization of evaluation research: Some pitfalls of definition. *Evaluation Studies Review Annual*, 6, 727-748.
- ERS Standards Committee. (1982). Evaluation Research Society standards for program evaluation. In P. H. Rossi (Ed.). *Standards for Evaluation Practice* (pp.7-20). Newbury Park, CA: Sage.
- Gay, L. R. (1985). *Educational evaluation and measurement: Competencies for analysis and application* (2nd ed.). Columbus: Charles E. Merrill Publishing Co.
- Gove, P. B. et al. (1979). *Webster's Third New International Dictionary*. MA: G. & C. Merriam Company.
- Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. NY: McGraw-Hall.
- Joint Committee on Standards for Educational Evaluation. (1988). *The personnel evaluation standards: How to assess systems for evaluating educators*. CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed). CA: Sage.
- Leviton, L. D., & Hughes, E. F. (1981). Research on utilization of evaluation: A review and syntheses. *Evaluation Review*, 5, 525-548.
- Orata, P. T. (1940). Evaluating evaluations. *Journal of Educational Research*, 33, 641.
- Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd ed.). CA: Sage.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). CA: Sage.
- PDK National Study Committee on Evaluation. (1972). *Educational evaluation and decision making* (3rd printing). IL: F. E. Peacock.
- Rossi, P. H. (ed.). (1982). Standards for evaluation practice. San Francisco, CA: Jossey-Bass.
- Sanders, J. R., & Nafziger, D. H. (1976). *A basis for determining the adequacy of evaluation designs*. MI: Western Michigan University.
- Scriven, M. (1972). An introduction to meta-evaluation. In P. A. Taylor, & D. M. Cowley, (Eds.). *Readings in curriculum evaluation* (pp.84-86). Dubuque, Iowa: W. C. Brown Co.
- Scriven, M.(1967). *The methodology of evaluation*. In R. W. Tyler, R. M. Gagne, M.

- Scriven (Ed.), *Perspectives of curriculum evaluation* (pp.39-83). Chicago, Ill: Rand McNally.
- Scriven, M.(1991). *Evaluation thesaurus*. Newbury Park, CA: Sage
- Simons, H. (1988). Teacher professionalism and the national curriculum. In D. Lawton & C. Chitty (Eds.) *The national curriculum* (pp.78-90). London: Institute of Education, University of London.
- Straw, R. B., & Cook, T. D. (1990). Meta-evaluation. In H. J. Walberg, & G. D. Haertel (Eds.). *International encyclopedia of educational evaluation* (pp.58-60). NY: Pergamon Press.
- Stufflebeam, D. L. (1974). *Toward a technology for Evaluating evaluation*. EDRS ED NO090319.
- Stufflebeam, D. L. (1978). Meta evaluation: An overview. *Evaluation and the Health Professions*, 1, 17-43.
- Stufflebeam, D. L. (1987). Professional standards for assuring the quality of educational program and personnel evaluations. *International Journal of Educational Research*, 11(1), 125-143.
- Stufflebeam, D. L. (2000a). Professional standards and principals for evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan. (Eds.) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (chapter 24). Boston: Kluwer Academic Publishers.
- Stufflebeam, D. L. (2000b). Methodology of metaevaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan. (Eds.) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (chapter 25). Boston: Kluwer Academic Publishers.
- Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic evaluation*. MA: Kluwer-nijhoff Publishing.
- Stufflebeam, D. L.; Madaus, G. F., & Kellaghan, T. (Eds.). (2000). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (2nd ed.). Boston: Kluwer Academic Publishers.
- Weiss, C. H. (1972). *Evaluation research: Methods for assessing program effectiveness*. NJ: Prentice-Hall.
- Weiss, C. H. (1981). Measuring the use of evaluation. In J. A. Ciarlo (Ed.). *Utilizing evaluation: Concepts and measurement techniques* (pp.17-33). CA: Sage.
- Worthen, B. R., & Sanders, J. R. (1987a). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.
- Worthen, B. R., & Sanders, J. R. (1987b). *Educational Evaluation: evaluation contracts, and technical assistance*. MA: Kluwer-Nijhoff Publishing.

Metaevaluation of Educational Evaluations

Chia-Cheng Yu & Hsiang-Jung Tseng

ABSTRACT

The purposes of this article are to analyze the concepts of metaevaluation, the professional standards and principles for metaevaluation, and the design for implementation. Based on literature review, four conclusions are brought about. Firstly, metaevaluation is a type of higher-order evaluation or secondary evaluation, which reevaluates primary evaluations, educational evaluations, to judge their values and merits. Secondly, the establishment of standards for metaevaluation of educational evaluation is the symbol of professional evaluation. Since the evaluation communities have already developed standards and principles for educators, evaluators and other stakeholders to follow, but evaluators also can develop their own by systematic research methods and procedures when necessary. Thirdly, as one of various evaluations, the process of metaevaluation is extremely complicated and social-political and would be designed in detail before being implemented. The 16 steps presented in this article could be the useful guidelines, if educational evaluators try to make a reasonable plan for metaevaluation. Besides, four suggestions are proposed as follows: (1) institutes of educational evaluation would follow the metaevaluation standards to design and implement educational evaluations; (2) metaevaluation should be embedded as necessary parts to any educational evaluation, and be implemented during its beginning, process and the end; (3) the outcomes of metaevaluation should be included in educational evaluation reports for their users; and (4) it is right time for both communities, education and evaluation, to collaborate developing metaevaluation standards.

Key Words: metaevaluation, evaluation standards, evaluation program