

國中小校長評鑑系統的 後設評鑑標準

鄭新輝*

摘 要

校長評鑑是國內近年來教育行政領域頗受關注的主題。目前各縣市國中小校長評鑑系統的發展，尚處於萌發階段，因此有必要藉助於一套後設評鑑標準，作為規劃實施校長評鑑系統或檢視現有評鑑系統的規準。本研究之目的主要在探討由美國「教育評鑑標準聯合委員會」1988 年所發表的「教育人員評鑑標準」，以作為國中小校長評鑑系統後設評鑑標準的參考。其所提供的四大類 21 項標準加上擬新增的 1 項，合計 22 項標準，均提供明確的使用原理、具體實施指引與常犯錯誤，可說是目前最為完備的教育人員評鑑系統的後設評鑑標準。不僅可用來作為規劃設計中小學校長評鑑系統的規準，亦可以之進行現有校長評鑑系統實施

*鄭新輝：國立台南師範學院初等教育學系助理教授

歷程的後設評鑑。雖然該標準是在美國文化背景下形成，但在理念與做法上仍有可供參考之處。該標準曾被用來檢視英美兩國中小學校長評鑑系統，而經由國內的實證研究發現，22 項後設評鑑標準在規劃實施國中小校長評鑑系統的歷程中均相當重要，尤其是效益性與正當性標準，惟可行性與正確性標準亦是受訪者所關注的問題。

關鍵字：校長評鑑、後設評鑑、評鑑標準

國中小校長評鑑系統的 後設評鑑標準

鄭新輝*

壹、前言

校長評鑑是國內近年來教育行政領域頗受關注的主題。尤其自國內國中小校長的任用制度，在國民教育法及其施行細則於民國八十八年修正通過後，已明確規定各縣市對任期屆滿的現職校長應辦理「辦學績效評鑑」，以作為校長遴選的依據。但自實施以來，各縣市大部分均尚未建立完整的績效評鑑系統（鄭新輝，民 91），不僅缺乏對現職校長進行完善的績效評鑑，以作為校長遴選的依據，更未能提供適合校長持續性專業支持的評鑑系統，使得校長遴選僅憑簡單的檢核資料或訪問意見，即進行遴選，不僅容易流於主觀亦可能形成政治問題。而缺乏形成性評鑑的體系，也容易讓校長有算總帳的壓力與對評鑑的疑慮。因此，近年來

*鄭新輝：國立台南師範學院初等教育學系任助理教授

國內陸續有學者開始關注並努力建構完善的校長評鑑系統或方案 (江文雄等, 民 88 ; 吳德業, 民 90 ; 吳錦森, 民 89 ; 侯世昌等, 民 89 ; 倪靜貴, 民 90 ; 翁福元, 民 90 ; 張德銳, 民 88), 希望能藉由兼具形成性與總結性目的的評鑑系統, 讓校長的辦學與專業發展獲得更好的協助, 同時亦能發揮績效責任的功能。

美國在 1980 年代受績效責任運動影響, 對教育人員素質提昇與評鑑的要求增加, 更促成對教育人員評鑑系統發展與實施的重視。相對的, 對當時教育人員評鑑品質不良的問題, 包括評鑑系統過於封閉、主觀、不可靠、存有偏見、不公平、缺乏邏輯性與過度簡化等缺失, 均受到嚴厲的批評(Darling-Hammond, Wise, & Pease, 1983 ; Scriven, 1983)。許多關心教育人員評鑑的團體, 普遍認為有必要進行「後設評鑑」(meta-evaluation), 並建立一套衡鑑與改善教育人員評鑑系統品質的後設評鑑「標準」(standards), 而校長評鑑系統の後設評鑑, 亦成為關注的對象。

目前有關教育人員評鑑的後設評鑑標準, 主要以美國「教育評鑑標準聯合委員會」(Joint Committee on Standards for Educational Evaluation, 1988)(以下簡稱聯合委員會, Joint Committee) 在一九八八年所發展的「教育人員評鑑標準」(personnel evaluation standards)(以下簡稱後設評鑑標準) 最為完整。在國外亦有許多針對教師、校長或學區教育局長的評鑑系統進行後設評鑑的研究 (Candoli, Cullen &

Stufflebeam, 1997; Cullen, 1997; Ellett, et al., 1996; Glasman & Martens, 1993; Shinkfield & Stufflebeam, 1995), 均以該後設評鑑標準作為檢證的指標。此外，為提供後設評鑑規劃實施的程序，Stufflebeam (2000)更進一步提出完整的後設評鑑方法論。Stufflebeam (1995:163-164) 認為後設評鑑標準可視為一種實用性理論 (working theory)，用以指導政策規劃者規劃實施一項符合正當性 (propriety)、效益性 (utility)、可行性 (feasibility) 與正確性 (accuracy) 標準的教育人員評鑑系統。該四類價值標準屬於基本的哲學原理，共有 21 項次標準，若加上 Scriven (1997)與 Stufflebeam & Pullin (1998)近來提議增列的 1 項，合計 22 項次標準，則提供後設評鑑的主要概念；而每項次標準所提供的指引 (guidelines) 與常犯錯誤 (common errors)，更提供建構完整教育人員評鑑系統所應為與應避免的實用性假設，以及規劃實施的實用性策略。由於國內國中小校長評鑑系統的規劃與實施，尚處於萌發階段，因此有必要藉助於一套後設評鑑標準，作為規劃實施校長評鑑系統或檢視現有評鑑系統的規準。

本研究之目的主要在深入探討教育人員評鑑標準，以提供作為國中小校長評鑑系統的後設評鑑標準之參考，以下將分七部分加以探討：(一) 後設評鑑標準的基本概念；(二) 建立後設評鑑標準的基本假定；(三) 後設評鑑標準的內容；

(四) 後設評鑑標準的應用歷程 ; (五) 後設評鑑標準在國外的實證研究 ; (六) 後設評鑑標準在國內的實證研究 ; (七) 結語。

貳、後設評鑑標準的基本概念

一、後設評鑑的基本概念

後設評鑑的概念，源自 1940 年由 Orata 所提出「對評鑑的評鑑」(Evaluating evaluation) (Cook & Gruder, 1978)。後來 Scriven 於 1969 年最先使用“Meta-evaluation”一詞，作為對某項教育成果計畫評鑑的評鑑，當時 Scriven (1969) 將後設評鑑定義為：「對一項評鑑、評鑑系統或評鑑設計的評鑑」，亦即是對評鑑的評鑑。隨後 Scriven (1972) 更將後設評鑑視為「第二層級的評鑑」(second-order evaluation)。而 Cook (1974)亦提出「次級評鑑」(secondary evaluation)的概念，用以檢視「原級評鑑」(primary evaluation)的資料。Stufflebeam (2000:96)更進一步認為後設評鑑是指：「為能引導評鑑的實施並提供其優缺點，而對某一評鑑系統的正當性、效益性、可行性與正確性，進行審慎的描述、獲取及應用敘述性與判斷性資訊的歷程」。亦即後設評鑑的目的，主要在引導並確保評鑑系統能符合評鑑的正當性、效益性、可行性與正確性標準。

二、後設評鑑標準的定義

聯合委員會 (Joint Committee, 1988:7-8) 將「教育人員評鑑」定義為：「依某一專業角色與某些特定的機構目標，對個人的表現或資格進行有系統的評鑑」。而將後設評鑑「標準」定義為：「從事專業評鑑者所一致同意，作為價值測量或確保評鑑品質的共同性原則」。因此，教育人員後設評鑑標準即是：「一群專業評鑑者為確保教育人員評鑑品質，所普遍同意用以協助規劃、設計與執行評鑑的一組共同性原則與指引」。而聯合委員會所發展的共同性標準則包含：正當性、效益性、可行性與正確性標準。

三、後設評鑑的使用時機與功能

一般人對後設評鑑使用時機與功能的認知，常會以為是在評鑑結束後，對評鑑所進行的總結性評鑑，但事實並不然。完整的後設評鑑亦如同其他評鑑一般，仍具有形成性的功能，因此在校長評鑑的實施歷程中，亦可實施形成性的後設評鑑。Stufflebeam(1974 ; 1981; 2000)將後設評鑑依其實施的時機，分成兩類：(一) 形成性後設評鑑 (formative meta-evaluations)：是在評鑑規劃與實施歷程中，用以協助評鑑人員做好評鑑規劃與實施歷程的一種評鑑；(二) 總結性後設評鑑 (summative meta-evaluations)：則在評鑑完成後進行，用以協助評鑑的關注者

(audiences)瞭解評鑑的優缺點，依評鑑標準評斷其績效(merit)與價值(worth)，並作必要決定的一種評鑑。Worthen & Sanders (1987)認為形成性後設評鑑能在評鑑尚未造成無法彌補的錯誤之前獲得改善；而總結性後設評鑑則可增強最後評鑑結果的可信賴性。

此外，後設評鑑同時具有符合專業性與公共利益的雙重功能，在專業性方面，評鑑人員需以後設評鑑協助其改善評鑑的實施；在公共利益方面，消費者需以後設評鑑協助其選擇正確的產品、方案、與服務的評鑑資訊，同時亦可避免接受不正確的評鑑結果。對專業的教育評鑑人員言，後設評鑑可提供其所需的回饋，以確保評鑑的品質並提供改進的方向。對評鑑報告的使用者言，後設評鑑可提供判斷該評鑑報告結果是否可信與可應用 (Stufflebeam, 2000 : 96)。至於教育人員後設評鑑的功能，主要在提供規劃完善的教育人員評鑑系統、檢視現行評鑑系統的缺失、提供政策決定者發展與評估健全的評鑑程序之原則與實用性建議，並滿足利害關係人對評鑑的需求。

四、後設評鑑標準的需要性

提供優良的教育品質服務是所有學校與教育機構的主要責任，而教育人員的素質則是能否提供良好教育品質服務的關鍵要素。因此，教育機構必須藉由評鑑

的手段去遴選、聘用與發展有資格的教育人員，並有效的管理與促進其工作效能。然而在教育體系中，對教育人員的品質仍有許多無法滿意的現象，同時又缺乏良好的教育人員評鑑系統，或者即使有之，其評鑑系統亦不完整，實施亦未能落實其精神。而校長評鑑系統，也同樣存在這些問題。因此，為解決所存在的這些問題，許多關心教育人員素質的團體均認為有必要繼續強化教育人員評鑑系統的數量與品質。然而更重要的問題是如何發展，以及所發展出來的評鑑系統，是否能適當而有效的改善教育人員評鑑品質與發揮應有的評鑑效能。因此就有必要建立一套評鑑標準，用以判斷教育人員評鑑系統。雖然「聯合委員會」已經於 1981 年建立一套方案評鑑標準，但並不適用於教育人員評鑑。其原因在於方案評鑑是對「事」，而教育人員評鑑是對「人」，較容易引起受評者的焦慮、爭論與對簿公堂。由於二者的評鑑對象不同，評鑑標準的內涵與重點自應有別。嗣後，隨著教育人員評鑑問題的顯現，因此乃於 1988 年另行制訂「教育人員評鑑標準」。

五、後設評鑑標準的應用對象與層面

教育人員後設評鑑標準的應用對象，主要為學校與教育機構人員，並適用於所有從事專業工作者。事實上，學校或教育機構為提昇教育人員素質，需建立一套健全的評鑑系統，以作為受理專業訓練、頒發能力證明、界定角色、遴選申請

者、監督與提供表現回饋、輔導專業發展、決定薪給、對任期、晉升與終止聘任等的決定參考。因此，教育人員後設評鑑標準的應用層面亦可涵蓋整個教育人員的生涯，包括：儲備階段(preparation stage)、授證階段(licensing stage)、聘用階段(practice stage)與專業化階段(professionalization stage)；每個階段又可區分為進入(entry)、參與(participation)及結束活動(exit activities)；而每項活動均有其必要進行的評鑑與作決定。有關教育人員在每個生涯階段的不同發展活動，所需要進行的評鑑與作決定可整理如表 1 所示(Joint Committee, 1988; Shinkfield & Stufflebean, 1995:49; Stufflebean & Nevo, 1993:27-28)。

表 1 教育人員儲備、授證、聘用與專業化生涯階段的各種評鑑與作決定類型

在每個生涯階段的活動	教 育 人 員 的 生 涯 階 段							
	準備階段		授證階段		聘用階段		專業化階段	
	評鑑	作決定	評鑑	作決定	評鑑	作決定	評鑑	作決定
進入活動	供需評鑑	排序與補助訓練方案	檢視文憑	認可進入授證歷程	教育人員聘用需求評鑑	工作界定與研究	檢視教育人員與機構需求	繼續教育機會
	招募方案評鑑	方案重設計			招募方案評鑑	方案重設計	需求與成就評估	認可休假研究與特殊獎助
	申請者評鑑	學生選擇			申請者評鑑	教育人員選擇	資格證照評估	參與全國證照方案
參與活動	錄用評鑑	規劃學生方案	試用期間的導入階段	提供州的證書	工作需求與能力比較	人力分配	錄用評鑑	設計個別教育方案
	學生熟練課程要求的評鑑	分級、諮商、治療	授證測驗	部分資格證明	表現回顧	試用結束、晉級、任期、績效、薪級、專業成長、受獎	能力檢視	全國性的檢定

	評鑑進步情形	諮商、修正學生方案、終止			工作負擔調查	委屈管理		
結束活動	學生達成等級要求的期末評鑑	畢業	在指定期間內檢視成功的教學	永久性或長期證書	資源、人員需求與年資	人力減少	參與繼續教育的成就	未來休假資格
	結束的晤談	方案檢視與改善			表現回顧	終止或處罰	能力與性向的檢視	新的人力分配
	追蹤調查	方案檢視與改善						

資料來源：Stufflebean & Nevo (1993:27-28)

表中所列的各種評鑑，對提昇教育人員的素質與效能均有助益。對校長而言，自生涯準備以至結束，所有可能發生的評鑑與生涯決定，對確保學校能聘用具合適資格與有效能的校長，均有其重要性。因此，校長在生涯發展的不同階段，均需經由適當的評鑑系統，以協助學校與教育機構作適當的決定與判斷。但校長評鑑除考慮其功能與價值之外，對每一生涯階段所實施的評鑑，能否符合專業性的後設評鑑標準則更為重要。

參、建立後設評鑑標準的基本假定

教育人員後設評鑑標準建立的基本理念，主要是以五項重要基本假定作為引導 (Joint Committee, 1988:8-9)：

一、教育人員評鑑的基本目的或其他教育活動，均必須是為了能對學生與社

會提供有效的服務。因此教育人員評鑑必須能激勵與引導教育人員，更有效地服務所有學生，並提昇教育的理論與實際表現。

二、教育人員評鑑必須具建設性且避免威脅或打擊士氣。因為唯有受評者對評鑑指標與歷程感到信任、覺得有清楚的溝通與公平的對待、且認為評鑑有助於個人的成功，受評者才更有可能盡心於工作並樂於接受評鑑。

三、教育人員評鑑是規劃健全專業發展經驗的核心。因為經由評鑑可以指出個人的學習需求、興趣與責任，以作為規劃專業發展課程或訓練的依據。

四、工作描述與教育方式會隨著教育哲學觀、地區與機構的需求與政策目標不同而有差異。因此對所謂好的教學、行政作為與研究內涵，均有不一致的看法。此外，對達成教育目標所需的條件、資格與表現亦未有一致性的要求，因此隨著環境的差異，各種不同的方法均有其適用情境。所以聯合委員會對評鑑標準的應用，採取多元取向的觀點，認為評鑑標準應依所界定的教育目標、方法與角色定義作適當的應用。

五、教育人員評鑑的複雜性與重要性會隨著環境不同而有差異，有的評鑑標準在某些環境中可能很重要的，但在其他環境可能不適合甚至有反效果。因此對評鑑標準的應用就有必要花較多的時間進行審慎的規劃，尤其對評鑑結果所採取

的行動可能具有潛在的爭議時。

由這五項基本假定中，可發現幾項建立教育人員評鑑的後設評鑑標準應有的重要觀念：(一)強調教育人員評鑑本身只是一種手段，讓學生與社會受到更優質的教育服務才是最終目的。(二)重視建設性的實施歷程與受評者的接受度。(三)強調評鑑的發展性目的。(四)主張依不同情境與目的用途，彈性採用適應性的評鑑標準。(五)重視評鑑環境的複雜性，因此需審慎的應用評鑑標準。

肆、後設評鑑標準的內容

表 2 每一項後設評鑑標準在不同評鑑用途上的重要性

後設評鑑標準		訓練	授證	角色界定	遴選	專業發展	績效責任	薪級	任期決定	晉升	停聘
正當性標準	1.服務取向	高	高	高	高	高	中	中	高	中	高
	2.正式評鑑指引	高	高	中	高	高	高	高	高	高	高
	3.利害衝突	高	高	低	高	中	高	高	高	高	高
	4.管制評鑑報告	高	中	低	高	中	高	中	高	高	高
	5.與受評者互動	高	中	中	高	高	高	高	高	高	高
效益性標準	1.建設性取向	高	低	高	中	高	高	高	中	中	中
	2.明確的用途	高	高	低	中	高	高	中	高	高	高
	3.評鑑者的信譽	高	高	高	高	高	高	高	高	高	高
	4.功能性的報告	高	高	高	高	高	高	中	高	高	高
	5.追蹤與影響	低	低	高	中	高	高	中	高	高	中
可行性標準	1.務實性的程序	高	中	高	高	高	中	高	中	中	中
	2.政治的可行性	中	高	高	高	中	中	高	中	中	高
	3.財政的可行性	中	中	中	高	高	高	高	高	高	高
	4.法令的可行性	高	高	高	高	高	高	高	高	高	高
正確	1.明確的角色	中	高	高	高	高	高	高	高	高	高
	2.工作環境	低	低	高	高	高	高	中	高	高	高
	3.記錄實施程序	高	高	中	高	中	高	高	高	高	高

性 標 準	4.正確的測量	高	高	中	高	高	中	高	高	高	高
	5.可靠的測量	高	高	中	高	高	中	中	高	高	高
	6.有系統的資料控制	高	高	中	高	高	高	高	高	高	高
	7.偏見的控制	高	高	中	高	高	高	高	高	高	高
	8.監控評鑑系統	高	高	高	高	高	高	高	高	高	高

資料來源：Joint Committee (1988:17)與 Stufflebeam & Pullin (1998)

教育人員後設評鑑標準的規劃，本質上即是一種價值選擇的歷程。依據聯合委員會於 1988 年所建立的教育人員評鑑標準，共有四大類 21 項標準 199 條實施指引，包括：(一) 正當性標準 (propriety standards)：有五個次標準，61 條實施指引；(二) 效益性標準 (Utility Standards)：有五個次標準，53 條實施指引；(三) 可行性標準 (Feasibility Standards)：有三個次標準，23 條實施指引；(四) 正確性標準 (accuracy standards)：有八個次標準，62 條實施指引。後來 Scriven (1997) 與 Stufflebeam & Pullin (1998) 建議在「可行性標準」中增列一項「法令上的可行性」(legal viability) 次標準，33 條實施指引。若一併予以納入，則教育人員評鑑標準將增為四大類 22 項標準 232 條實施指引。而每項標準的重要性，依教育人員評鑑的功能與用途不同而有些差異，包括：教育訓練、授證、職位分析、遴選、績效表現、專業發展、薪級、任期、晉升、停聘等 (如表 2)。此外，評鑑程序、資料蒐集的時間表、時間與資源的投入也隨用途不同而異 (Joint Committee,

1988:15-16 ; Stufflebeam & Pullin, 1998)。以下僅就四大類 22 項後設評鑑標準，分別就其：(一) 定義與內容解釋；(二) 使用該標準的主要原理兩方面，逐一加以說明 (Joint committee, 1988 ; Stufflebeam & Pullin, 1998)：

一、正當性標準 (propriety standards)

所謂「正當性標準」(Propriety Standards) 是指「評鑑的實施應具有合法性、合倫理性並適切的關心受評者與委託者的福祉」。由於學校是為服務學生而存在，因此教育人員評鑑應以關注教育人員是否有效的符應學生的需求為目的。而正當性標準即在保護受評鑑所影響的人，包括：學生、教師、輔導員、行政人員、與評鑑人員。教育人員評鑑在「正當性標準」中有五項次標準：1. 服務取向 (Service Orientation, P1)；2.正式的評鑑指引(Formal Evaluation Guidelines, P2)；3. 利害衝突(Conflict of interest, P3)；4. 管制評鑑報告的接觸(Access to Personnel Evaluation Reports, P4)；5.與受評者的互動 (Interactions with Evaluatees, P5)。

(一)服務取向

1.定義與內容解釋：是指「教育人員評鑑應能促進良好的教育、實現機構的使命及有效的職務工作表現，使教育能符合學生、社區、與社會的需求」。由於教育人員評鑑的基本目的在引導教育人員提供高品質的教育服務。因此評鑑應協

助確保教育人員了解與追求機構的目標、釐清教育人員的責任、提供預定的服務、提昇專業能力、解聘不適任與有問題的人員；而負責規劃設計、執行、結果運用及參與教育人員評鑑者，其視野必須超越教育人員的自我利益之上。當行政人員、教師、專業人員及其他受評者的受雇權利受到重視時，學生及社區對良好教育服務的權利亦必須受到維護。

2.使用原理：教育系統存在的目的在提供滿足學生與社區的教育需求，因此系統中的所有要素，包括教育人員評鑑，均應導向於完成此一目的。

(二)正式的評鑑指引

1.定義與內容解釋：是指「教育人員的評鑑指引應登載於政策的說明、協商合約與人員評鑑手冊中，讓評鑑具有一致性、公平性、並符合相關的法律與倫理規範」。正式的評鑑指引內應明確的界定評鑑目的、內容、方法與實施程序等書面資料。如果評鑑目的、內容、方法與程序能合理且明確的說明清楚，評鑑的利害關係人就能對整個評鑑歷程有充分的瞭解。

2.使用原理：無論評鑑者與受評者是誰，評鑑均須以一致、正當與合法的態度實施。正式的評鑑指引不僅能將評鑑的決策內容與程序，讓各有關單位與個人瞭解，亦能提供專業人員檢視評鑑目的的機會；同時，亦可確保機構與受評者應

有的權利，避免可能的訴訟並增進對評鑑的信任感與支持度。亦即完整的正式評鑑指引與手冊，可增加利害關係人對表現期望的瞭解、一致化判斷的標準、確保評鑑的公平性、強化對評鑑結果的尊重、提升對評鑑本身與實施單位的信任等。

(三)利害衝突

1.定義與內容解釋：是指「利害衝突應予確認且以開放、坦誠的心胸加以處理，以避免因利害衝突而影響到評鑑的過程與結果」。評鑑者與受評者的利害衝突隨時可能在評鑑過程中產生。其原因通常受評鑑者個人的利害關係所影響，尤其當評鑑人員只有一位時，最容易產生問題。利害衝突的來源有許多種：評鑑者與受評者的關係、不同的教育哲學觀、政治偏好與道德規範、受評者的聲望、地位或政治影響力等，均可能產生利害衝突，例如：評鑑者可能受組織內領導者的壓力或威脅，而產生非專業性的判斷結果。由於評鑑過程中存在許多利害衝突，因此，有必要對這些可能的衝突加以有效控制，以確保評鑑的公平性與可靠性。

2.使用原理：利害衝突可能會暗中破壞整個教育人員評鑑系統的公正性。尤其當利害衝突影響到評鑑過程與結果時，判斷的精確性將受到挑戰。評鑑結果如是受到個人私利所影響，而非受組織或專業共享的目標所影響時，教育人員評鑑所作決定即喪失其可信度。

(四)管制評鑑報告的接觸

1.定義與內容解釋：是指「對教育人員評鑑報告的接觸，應限制只具有合法身分者才准查閱，以確保資訊使用的適切性」。亦即對有權接觸評鑑報告及相關資訊者，必須予以明確規範與限制。通常只容許具合法身分或有專業上的需求者，才有查閱的權力，例如：受評者、有權以該評鑑報告為基礎作決定的上級長官以及督導該評鑑進行者等。此外，對評鑑報告的適當用途亦應予以明確界定，例如：專業發展、輔導、表揚、記錄或績效責任等用途。

2.使用原理：對評鑑報告的接觸，是一件具高度敏感性的問題。因為報告中的資訊通常會成為作決定的重要依據，例如：獎賞、晉升、留任、專業發展與專業表現的判斷等。如果將評鑑報告揭露給無權接觸者，可能會讓受評者難堪或造成傷害。反之，評鑑報告的資訊不應拒絕提供有責任對受評者作重要判斷或作決定者，否則可能會因缺乏這些資訊，而對受評者做出不正確與不公正的決定。

(五)與受評者的互動

1.定義與內容解釋：是指「評鑑應以專業、審慎、有禮的態度去描述受評者，使受評者的自尊、動機、專業聲望、表現及對教育人員評鑑的態度能獲得增強，或至少不致遭受不必要的傷害」。亦即評鑑者必須瞭解受評者個人與專業上的需

求並予以適當反應。發展評鑑者與受評者間良好的信賴與專業性關係，並遵循機構協議是相當重要的。評鑑者必須視受評者為可經由正確的評鑑回饋獲得能力的人，並且能提供受評者具體改進的專業性建議與服務。

2.使用原理：評鑑者與受評者的互動關係存在於整個評鑑過程中。若評鑑者與受評者能分享專業責任與人性尊嚴，即可減低受評者的焦慮與對評鑑的負面感受。對評鑑結果的發現更能產生建設性的用途並促成專業成長。如果評鑑能以此種態度進行，則受評者的士氣與評鑑歷程的可靠性將更能獲得增強。

二、效益性標準 (Utility Standards)

所謂「效益性標準」是指「讓評鑑具有教育上的價值性、適時性與影響力」。不僅要求教育人員評鑑應提供個人與團體有用的資訊，改善其表現；同時也要求評鑑應事先決定用途，例如：對人員的選擇、任期決定或提供專業發展方向；而實施評鑑的人員亦應具備專業與可信度。效益性標準將教育人員評鑑視為教育機構用以遴聘優秀成員並經由適時性的回饋機制，以激勵並引導員工提供更高品質服務的一項努力。在「效益性標準」中有五項次標準：(一)建設性導向(Constructive Orientation, U1)；(二)明確的用途 (Defined Uses, U2)；(三)評鑑者的信譽 (Evaluator Credibility, U3)；(四)具功能性的報告(Functional Reporting, U4)；(五)

追蹤與影響 (Follow-up and Impact, U5)。

(一)建設性導向

1.定義與內容解釋：是指「教育人員評鑑應具建設性，以便協助機構發展其人力資源，鼓勵與協助受評者提供卓越的服務」。評鑑若有助於促進學生、教育人員與組織目標的達成，即具有建設性。亦即評鑑如能協助機構正確選擇與留任優秀人才、增強好的工作表現、提供改進工作表現的方向、釐清專業發展的需求、表揚卓越表現人員、協助處理不適任人員、提升專業水準及促進合作與和諧關係時，即具有建設性。反之，如果評鑑只是表面上為符合官僚體制下的績效責任要求、作為控制與威嚇的目的或當作政治上的工具，則不具有建設性。

2.使用原理：教育人員評鑑如具建設性，則能激勵與支持教育人員與組織實現其目標與責任。亦即具建設性的評鑑，不僅能提供個人必要的回饋資訊、強化專業知能、提升工作熱忱、增進工作效率；同時亦有助於組織遴選與留任優秀人才、提升專業能力、激勵專業發展與促進合作關係。教育人員與機構一旦具備這些良好的特性，則可提供學生更好的服務，且能在穩定或變革階段，維持高效能的表現與積極的自我形象。

(二)明確的用途

1.定義與內容解釋：是指「教育人員評鑑的使用者與目標用途應予確認，使評鑑所要探討的問題能獲得適當的描述」。亦即評鑑的規劃者應先與評鑑的使用者商議並釐清評鑑目標。評鑑的使用者可能涵蓋行政人員、學校管理委員會、教職員、認可單位、學生及其他對評鑑結果能合法接觸者。至於評鑑的用途，大致上包括：教育人員的遴選、授證、專業發展需求診斷、績效責任、晉升、獎賞、決定薪級、表揚及終身聘用等。

2.使用原理：教育人員評鑑的實施歷程應受評鑑用途所引導。因此有必要先確認評鑑的使用者並明確化所需的資訊及其用途，俾讓評鑑者能正確的決定評鑑問題、所欲蒐集的資料與觀察行為、應呈現的報告與評鑑結果的處理。此外確認評鑑用途亦可避免資訊被誤用。資訊蒐集是依目的而來，因此如將資訊用在其他的目的上未必合適，且可能造成錯誤的傷害。

(三)評鑑者的信譽

1.定義與內容解釋：是指「評鑑系統應由具評鑑資格、技能與權力者負責管理與執行，而評鑑者亦應專業性地進行評鑑，使評鑑報告受到尊重與採用」。因此，教育機構必須對管理與執行教育人員評鑑者的任命、訓練、支持與監督賦予最大的關注。只有具備合適資格者，才能賦予評鑑的重責大任，而其權責亦應加

以明確化。對執行評鑑規劃者，例如：行政人員與教職員等，應訓練使其充分瞭解機構的評鑑政策與程序。對受評者的資格與能力表現進行評鑑，均應由具相關專業知能者負責。例如，當行政人員在評鑑教師的學科領域知識時，通常應輔以學科內容專家的判斷。負責評鑑者應以最專業與敏感性的能力，進行評鑑的規劃、資料蒐集與提出評鑑報告。評鑑的回饋會議應有系統的依特定議程進行。機構應監控評鑑系統的進行，並掌握評鑑者是否有能力完成其評鑑責任。

2.使用原理：評鑑的接受度絕大部分與受評者對評鑑者的權威性、專業性、敏感性、可信賴性、效率與效能表現等有很大關係。如果評鑑者未被受評者認為具有可信賴性，例如：不具備評鑑權威、不熟知受評者的職位與專業領域、不能免於利害衝突等，則受評者可能對評鑑採取不合作、抗拒或抨擊的態度，使得評鑑的成效大打折扣。換言之，可信賴的評鑑者可以強化評鑑歷程，並有助於建設性評鑑用途的實現。

(四)具功能性的報告

1.定義與內容解釋：是指「評鑑報告必須正確清楚、適時貼切，讓評鑑結果對受評者與其他關心者具有實用價值」。為能發揮評鑑的效用，評鑑報告所提供的資訊，必須能滿足使用者並讓受評者改進其工作表現。評鑑報告應直接述明評

鑑目的與適切的工作描述與標準。所使用的語言與名詞，必須能讓所有利害關係人立即瞭解，而且應明確說明評鑑資料的實用性含意。此外，評鑑報告亦應提供受評者回饋，以協助其改進。亦即評鑑報告不僅應反應工作期望與表現指標，同時亦應指出受評者的優缺點，並提供具體的改進建議。

2.使用原理：由於評鑑報告可能成為判斷受評者專業表現、地位與發展需求的重要依據，因此評鑑結果應提供清楚、有用與有關聯的資訊，同時亦應適時提供作決定與採取行動的方法。評鑑報告本身不應成為目的，而應當作達成評鑑目的的手段，協助與引導受評者及其他利害關係人採取有效的行動。模糊的用語及建議會減低評鑑的價值；正確提供優缺點的評鑑報告，可以協助減少評鑑的系統性誤差，且有助於激勵與增進受評者改進其工作表現。

(五)追蹤與影響

1.定義與內容解釋：是指「評鑑結果應加以追蹤，讓評鑑的使用者與受評者能了解評鑑結果，並採取適當的行動」。評鑑的目的不僅在做好評鑑工作，而且也要瞭解評鑑的目的與價值在於善用評鑑的發現。因此，評鑑應協助使用者瞭解評鑑且能依評鑑結果採取適當的行動。例如：視導人員應與受評者一起設計合適的行動計畫，以協助受評者改進其缺失，強化其優點。

2.使用原理：追蹤活動應該是評鑑歷程的自然結果。例如僅對受評者描述其與所規定標準有關的表現。

三、可行性標準 (Feasibility Standards)

所謂「可行性標準」是指「要求評鑑系統應盡可能的易於實施、所需的時間與資源具有效率、適當的經費需求、及其他所需考慮點均具有可行性」。教育人員評鑑在「可行性標準」中有四項次標準：(一)務實性的程序(Practical Procedures, F1);(二)政治上的可行性 (Political Viability, F2);(三)財政上的可行性 (Fiscal Viability, F3);(四)法令上的可行性 (Legal Viability, F4)。

(一)務實性的程序

1.定義與內容解釋：是指「教育人員評鑑的實施程序應予審慎規劃與執行，使評鑑能在最少干擾與成本中獲得所需的資訊」。教育人員評鑑的實施程序是一系列能產生效果的計畫行動，這些行動必須對評鑑者、受評者與其他評鑑資訊的需求者有幫助。實施程序的每一步驟均應以易懂的言語加以明確表示，以便於瞭解。此外，評鑑程序必須能發揮效率，評鑑歷程應能提供有用的回饋，且應避免妨礙機構與成員的工作，因此最好能融入組織的定期作業中。評鑑的實用性固然重要，但仍不應損及正確性。因此，評鑑程序必須能產生正確資訊，同時兼顧實

用性並符合相關的法令與評鑑政策。亦即評鑑程序必須符合資訊的需求、資源的有效性、評鑑目標與政策、以及評鑑的實用性建議。

2.使用原理：不實用的評鑑程序不僅無效率，且會損及個人的表現與組織的效能與效率；同時也會傷害機構的行政信譽與員工士氣。因此，機構必須避免讓人對評鑑程序感覺厭煩、浪費、過於複雜、有強迫性或表面化。

(二)政治上的可行性

1.定義與內容解釋：是指「教育人員評鑑系統的發展與管理，應以合作的方式進行，讓所有關心的團體均能建設性的參與制訂」。亦即有關教育人員評鑑的決策內容與實施程序等，應由教育人員、評鑑者與其他利害團體一起協商制訂，俾讓評鑑目標、內容、方法與程序形成可理解與可接受的共識。教育機構的人員不僅應忠實的執行此一政策與程序，而且必須定期檢視與適切的修正。此外，有關評鑑的各項規定，應讓受評者、使用者及其他外部團體所接受。

2.使用原理：如果教育人員評鑑的政策與程序易於瞭解，且是由合作中發展出來並被所有利害團體與官方所接受，則較能確保該方案能繼續合作與有效的執行，包括：對該方案的支持、對目標與方法的共識、能有效的實施、對評鑑結果與報告具信任感等。通常要讓受評者支持一項教育人員評鑑系統，則該系統必須

是法定、公平且能提供受評者改善缺失的正當程序，並肯定其優良表現。如果評鑑不具政治上的可行性，則誤解或分裂就有可能發生，而負責執行評鑑者也可能難以產生預期的效果。

(三)財政上的可行性

1.定義與內容解釋：是指「應提供教育人員評鑑所需的足夠時間與人力財力資源，讓評鑑計畫能有效能與有效率的實施」。教育人員評鑑需要實質上的資源，以有效的發揮其功能。評鑑系統必須訓練評鑑人員與工作人員，讓受評者、其他個人與團體參與，並依情況遴請外部專家協助。其他所需資源還包括：空間、設備、材料、設施、時間等。檢視財政上的可行性，其重點主要在於瞭解評鑑所需的人力與財力資源是否足以完成評鑑工作。

2.使用原理：提供教育人員評鑑所需經費與其他資源的正當性理由包括：可遴選優秀的新進人員、改進成員的表現、改善對學生的服務、改進組織的運作與一般福祉等。資源分配是檢視組織對教育人員評鑑是否支持的最佳明證。足夠的資源提供成功的可能性；缺乏資源會讓評鑑作不適當的妥協或過於簡略。當有限的資源妨礙評鑑的專業性與效能時，評鑑資料的運用就有可能受到限制或對組織與個人產生負面的影響。然而單靠資源的支持並無法保證評鑑方案的成功，評鑑

系統最好能整合到組織中，並將所獲得的資訊作更有效能的運用。

(四)法令上的可行性

1.定義與內容解釋：是指「教育人員評鑑須符合中央與地方的法令要求，使評鑑人員能成功的執行符合公平、效率與效能的評鑑」。評鑑者在進行教育人員評鑑以及後續運用評鑑結果時，必須顧及相關的法令並避免可能動搖評鑑結果的法律困擾。影響的法令規範，由中央以至地方，構成一套複雜的法令要求。此項標準屬通論性，其目的在提供全面性的指導原則以進行公正的評鑑，因此在實際進行評鑑時，仍須以適用該評鑑的相關法令規定為準。法令上的可行性最基本的原則在於符合正當程序（due process）的要求，包括：「程序性的正當程序」（procedural due process）與「實質性的正當程序」（substantive due process）兩方面。「程序性的正當程序」是指員工在評鑑過程及評鑑結果的處理上，均有公平表達意見與被告知的機會與權利；對重大的人事決定（例如：解雇），必須有公正無私的覆核機制（review mechanism）。一般說來只要遵循教育人員評鑑標準中的其他 21 項標準，就可以符合大部分正當程序的要求。而「實質性的正當程序」則是指不得以不公平、任意、或者不合理的行為或理由來對付受評者。因此應明確定義受評者的角色職掌與要求改善的目標，並讓員工了解績效評估的標準；此

外雇主亦必須以適當的程序，確保對員工所作的決定是正當且公平的。如果評鑑者或雇主使用工作職掌以外的資訊來為員工做出重要評斷，就是違反實質上的正當程序。此外評鑑結果如果是基於畏懼法令紛爭而留任不適任或者不負責任的教育人員，就跟違反正當程序解聘一樣有問題。教育人員評鑑與解聘過程還必須有責任感。如果不能及時解聘不適任員工，造成浪費公帑，就是違反專業倫理。此外機會均等亦應遵守正當的程序，解決的方法包括：遵循正式的評鑑指引、調整評鑑以符合工作職掌所需之任用條件、控制偏見、解決利益衝突、監測評鑑系統等標準。

2.使用原理：由經驗得知教育人員評鑑稍不留意即容易產生申訴，甚至訴訟案件。評鑑者即使遵循所有 21 項評鑑標準去蒐集與評估被評鑑者的資料，但所做出來的判斷仍可能導致法律問題。所以，評鑑者與被評鑑者均需藉助法令上的可行性標準及實施指引，來確保教育人員評鑑能符合法令規定並避免興訟。雖然遵循此一標準以及其他 21 項標準，仍不可能預防所有的爭端，但若輔以有能力的法律顧問，應該可以減少法律問題。

四、正確性標準 (accuracy standards)

所謂「正確性標準」是指「要求所獲得的資訊是具有技術上的正確性，且評

鑑的結果是由資料邏輯推論而來」。教育人員評鑑在「正確性標準」中有八項次標準：(一)明確的角色 (Defined Role, A1);(二)工作環境 (Work Environment, A2);(三)記錄實施程序 (Documentation of Procedures, A3);(四)正確的測量 (Valid Measurement, A4);(五)可靠的測量 (Reliable Measurement, A5);(六)有系統的資料控制 (Systematic Data Control, A6);(七)偏見的控制 (Bias Control, A7);(八)評鑑系統的監控 (Monitoring Evaluation Systems, A8)。

(一)明確的角色

1.定義與內容解釋：是指「受評者的角色、責任、表現目標、與所需的資格應清楚的界定，使評鑑者能判定有效的評估資料」。在整個評鑑歷程開始規劃設計前，所有的團體對規定的職位角色與表現期望應先形成共識。所列舉的表現期望可以來自明確的工作描述、實際專業活動的檢討、受評者自己對工作的要求。成員的工作與角色描述通常須符合機構的目標，並反映出最重要的核心任務。職位期望 (position expectations) 可分成三部份：(1) 職位資格 (qualifications)：是指實現職位責任與表現期望所必須的經驗、知識、技能、執照或證照。(2) 職位責任 (responsibilities)：是指持有該職位所賦予的任務。(3) 表現目標 (performance objectives)：是指受雇者被期望的表現結果。

2.使用原理：審慎發展與充分描述其職位角色、責任、表現目標與資格是明確化評量指標的必要條件。

(二)工作環境

1.定義與內容解釋：是指「受評者的工作環境應予確認、描述及記錄，使環境因素對工作表現的影響與限制能在評鑑中予以考慮」。工作環境會影響個人的實際表現，包括：(1)組織結構與運作歷程。(2)社區的特性。(3)時間、空間、教材與設備、人力等資源。(4)工作關係、班級大小、專業自主性、同事間的專業支持與貢獻程度。(5)經濟與社會變動。(6)偶發事件與天然災害等。這些因素在評鑑歷程中均應加以考慮，以便了解其對受評者的可能表現影響。在評鑑報告中，環境背景因素應予說明，以提供公平的表現評量並決定表現期望水準，或修正機構對受評者應有的支持。然而，環境背景的影響亦不應被用來作為合理化不良表現的理由。

2.使用原理：讓教育人員承擔無法掌控或具影響力的績效責任，會讓其產生不公平感與打擊服務士氣；而未能考慮環境因素對受評者表現的可能影響，也會損及評鑑歷程的正確性。由於教育人員必須在動態與複雜的環境中工作，因此教育機構必須瞭解環境的影響力，以協助受評者提升工作效能。此外，在實施教育

人員評鑑時更應檢視環境背景因素，以提供公平、正確的評鑑，並協助受評者與機構更有效的掌握環境的影響力。然而，評鑑者、受評者與機構不應以環境的影響作為合理化不合標準服務的藉口。

(三)記錄實施程序

1.定義與內容解釋：是指「實際進行的評鑑程序應予記錄，讓受評者與其他評鑑的使用者能評估與目的有關的實際進程序」。評鑑程序記錄的重點包括：評鑑步驟、格式、申訴程序、報告與記錄的時程、追蹤與正當程序等的紀錄。此一紀錄可提供受評者及其他評鑑使用者，對證據的蒐集與評鑑進行的程序，有更清楚的概念。

2.使用原理：評鑑的效能與評鑑者、受評者及其他利害關係人對評鑑程序的瞭解與接受程度有關。一個具一致性、系統性與審慎記錄評鑑方法的評鑑系統，才能確保其符合評鑑的公平性與合法性。不良的評鑑設計或實施會產生錯誤的評鑑結果。因此，為讓評鑑的設計與實施發揮效能，有必要協助所有與評鑑有關的人員了解，並邀請其對評鑑的設計與實施歷程提供意見。

(四)正確的測量

1.定義與內容解釋：是指「測量的程序應基於所描述的角色與目的而選擇或

發展，讓評鑑者對受評者的推斷能更有效與正確」。正確性是指測量到所欲測量的東西。亦即由測量工具或程序所獲得的證據，可用以支持推論的程度。正確性並非指評鑑工具或程序本身，因此一項特定的測量可能對某一目的言是正確的，但對其他目的則不具正確性。本項標準所用的「測量」一詞，是取其最廣義的定義，包含用來蒐集作為個人資格、表現或效能評鑑的資訊，所採用的任何工具或程序。在教育人員評鑑中，評鑑內容的決定應符合「明確的用途」與「明確的角色」所要求的標準。

2.使用原理：正確性是評估任何評鑑歷程中最重要問題。如果評鑑要符合其預期的目的，則所做的推論與判斷必須能經得起檢驗。用來蒐集資料的工具與程序的選擇、發展與實施，以及用來綜合資訊與推論的基礎，均必須清楚的與評鑑目的相連結，俾能作正確的判斷、推論與決定。而這些均需加以書面記錄。

(五)可靠的測量

1.定義與內容解釋：是指「測量程序的選擇、發展與實施應具可靠性，讓所獲得的資訊能提供受評者表現的一致性指標」。測量的可靠性是指對表現的評鑑提供一致性資訊的程度。任何資料蒐集的程序與工具，在測量過程中均可能存在導致不一致的誤差來源。這些不一致可能存在於：(1) 受評者對相同測驗的反應

或在不同情境中的表現；(2)一觀察者在不同情境中對受評者相同行為的詮釋；

(3)不同觀察者對相同受評者表現所做的評定。測量的一致性應包括：內部一致性 (internal consistency)：指不同的指標 (indicators) 在相同規準 (criterion) 中的一致性；觀察者的一致性 (observer agreement)：指不同的觀察者對相同行為觀察的一致性；穩定性 (stability)：在不同情境中觀察某一行為表現的一致性。

確保測量可靠性的步驟與水準應加以記錄並讓利害關係團體瞭解。

2.使用原理：教育人員評鑑工具與程序應具有符合評鑑用途的可靠性水準。

缺乏可靠性的程序也會缺乏正確性，受評鑑所影響的人，應有權瞭解強化測量可靠性的努力與所達成的水準。

(六)有系統的資料控制

1.定義與內容解釋：是指「評鑑所使用的資訊應確保其安全性，且應審慎的加以處理與保存，讓資料的保存與分析能與評鑑資料的蒐集一樣重要」。評鑑資訊在蒐集、保管與取用的每個階段，均應審慎的處理並做正確性檢核。正確性的檢核工作涵蓋：資料的記錄、資料的保管、資料檔案的取用以及資料的分析等。資料庫的完整性應確實受到安全保存。資料與文件均應予以保密，並嚴格管制資料的接觸與取用程序。

2.使用原理：有系統的資料控制可減低資料庫的誤差。雖然評鑑者對評鑑取樣與測量誤差的控制能力有限，但對減低評分、偏好、資料輸入與分析方面的誤差，有較高的主控性。如果誤差未能察覺，則會影響評鑑的正確性、扭曲資料分析結果，並產生誤導與誤判；而未具區辨力的評鑑發現，亦會損及評鑑系統的信賴性。

(七)偏見的控制

1.定義與內容解釋：是指「評鑑過程應提供預防偏見的安全措施，使受評者的資格能力或表現，獲得公平的評定」。在教育人員評鑑歷程中隨時可能潛存著偏見。如果所做的判斷來自局部或與其工作無關的資訊，則評鑑即存有偏見。無偏見的評鑑需以指標規準為基礎，且所獲得的資訊與受評者的工作具有關聯性。

2.使用原理：偏見的存在可能會整體削弱該評鑑系統的價值。偏見會扭曲資訊蒐集的歷程，且會影響所作建議、所採取的行動與所作決定的正確性。而偏見也可能導致資源的浪費與興訟。

(八)評鑑系統的監控

1.定義與內容解釋：是指「教育人員評鑑系統應定期與有系統地加以檢討，以便能夠做適當的修正」。教育人員評鑑系統包含評鑑目的、內容、方法、工具、

程序、評鑑報告與結果的運用等要素。每一組成要素與整個系統均應定期接受評鑑，或在遭受質疑時，立即進行檢視。檢視的內容包括：依原訂計畫實施的程度、達成評鑑目的的程度、以及受評者與其他關係人接受的程度等。一般說來，檢視的標準在於評鑑系統是否符合評鑑的標準。如果評鑑未達原訂目的，則系統應做適當的修正或改變原先的期望。檢視的結果應用來解決問題並提升評鑑系統的效能與公平性。如果需要做改變，則評鑑的正當性、效益性、可行性與正確性標準及其實施指引，即可應用在系統的修正中。

2.使用原理：做好教育人員評鑑是一件困難且容易出錯與發生抱怨的事，因此有必要隨時對該系統的規劃實施歷程進行檢討與修正。即使某一評鑑系統運作情形令人滿意，但仍需隨時加以檢討改進。一個規劃良好的評鑑如果缺乏評鑑者與受評者主動發現問題並賦予關注，則該系統無法成功的實施。此外，隨著個人與工作條件的改變，亦有必要對評鑑程序做定期性的檢討。而瞭解受評者抱怨、確保評鑑的公平性、避免申訴與興訟、並找出評鑑系統待改進之處，亦相當重要。

伍、後設評鑑標準的應用歷程

一、後設評鑑的實施歷程

為能有系統的進行教育人員與方案評鑑の後設評鑑，Stufflebeam (2000)將後設評鑑規劃實施歷程分成十個步驟，以作為進行後設評鑑的實施程序：(一)決定後設評鑑的利害關係人(stakeholders)並與之互動。(二)建立合格的後設評鑑團隊。(三)界定後設評鑑的問題。(四)取得判斷評鑑系統或特定評鑑的評鑑標準之共識。(五)與委託人建立後設評鑑的契約。(六)蒐集並檢視現有的相關資訊。(七)進一步蒐集所需的資訊，如：實地晤談、觀察與調查。(八)分析蒐集所得量與質的資訊，並判斷其符合所選定評鑑標準的程度。(九)準備與提出所需的評鑑報告。(十)協助委託人及其他利害關係人解釋與應用評鑑發現。

二、後設評鑑標準的應用歷程

有系統的運用教育人員評鑑標準可以提供作為建立、評鑑與改善校長評鑑系統的基礎。聯合委員會提出五個應用該標準的實施步驟：(一)熟悉所有的評鑑標準；(二)釐清評鑑系統的目的；(三)描述現行的評鑑系統；(四)應用評鑑標準；(五)決定評鑑結果的處理。Stufflebeam(1993)認為經由這五個步驟，亦可將教育人員評鑑標準應用在改善校長評鑑系統上。

(一)熟悉所有的評鑑標準：對評鑑標準的瞭解是檢視與改善校長評鑑系統品質的基礎。因此，所有參與改善評鑑歷程者，均應充分瞭解每項標準的基本使用

原理、指引、常犯錯誤與範例等，以作為順利進行下一步驟的重要基礎。

(二)釐清校長評鑑系統改善的目的：本步驟的目的在對評鑑的使用目的作充分的瞭解。因此，必須經由與評鑑有關的利害關係人直接互動來獲得瞭解，並參考評鑑標準的基本原則與實施指引。在此一歷程中，最好由評鑑指導委員會代表利害關係人，逐一檢視下列問題：1.為何要實施評鑑？ 2.有哪些工作要被評鑑？ 3.誰要使用評鑑結果？ 4.評鑑會影響何種決定或行動？ 5.評鑑的重點在資格、能力、表現、效能或特殊成就？ 6.該評鑑系統想要擁有何種影響？

(三)描述現行的校長評鑑系統：釐清校長評鑑系統的目的後，評鑑指導委員會或其他參與評鑑系統改善的團體，應：1.蒐集評鑑歷程中的有關資料文件。2.提出對該評鑑系統的書面描述。3.審慎的檢視該書面描述與證明文件，以決定是否已具備充分的資訊，能以每項評鑑標準開始進行該系統的評鑑。如果所需資訊尚未完備，則應反覆蒐集，直到可以進行下一步驟為止。

(四)判斷現行校長評鑑系統的品質：當評鑑系統的書面描述與證明文件齊備，即可開始以評鑑標準進行該系統的品質判斷。此項工作可以由評鑑指導委員會、受過訓練的評鑑小組、或個別評鑑人員進行。每項標準必須分開來作判斷，以決定其符合程度。其步驟可先檢視每項標準的內容，其次列出評鑑系統在該標

準上的優缺點，最後作出評鑑系統符合該標準的程度：由符合，部分符合，到不符合。當每項標準均完成判斷後，即可得出整個評鑑系統的輪廓，並描述其主要優缺點。對評定欠佳的標準，評鑑指導委員或判斷人員應進一步確認所發現的優缺點，並作為修正改善評鑑系統的依據。

(五)改善評鑑系統：當現行評鑑系統的優缺點經確認後，評鑑指導委員會即可提出改進評鑑系統的建議與計畫。由於聯合委員會在所提的每項評鑑標準中，均列出規劃或改善評鑑系統的詳細實施指引，因此，對評鑑系統較弱的標準，即可進一步使用這些指引來進行改善工作。但由於不可能一下子對每項改善系統的建議全部完成，所以有必要提出改善校長評鑑系統的一般性計畫，並設定改善的優先順序與時間表。該優先順序的設定，需考量該系統所處的教育與社會背景，以及本身的特殊需求、偏好、限制與可用的資源等。

陸、後設評鑑標準在國外的實證研究

有兩篇研究 (Cullen, 1997; Martens, 1991)，分別以 21 項教育人員後設評鑑標準，評估英國全國性的校長評鑑系統，以及美國部分學區的校長評鑑系統。其應用歷程與評鑑結果整理如下：

一、Cullen 的研究

Cullen (1997) 以質的研究方法，進行英國全國性中小學校長評鑑系統的後設評鑑研究。她選擇 15 位英格蘭學區內的中小學校長，進行深度訪談，並輔以文件分析法探討有關校長評鑑的資料。其次，有系統的運用 21 項教育人員評鑑標準的原理、指引與常犯錯誤，進行所得資料的綜合分析研究，以確認該評鑑系統在每項評鑑標準的優缺點。以下僅就該研究的發現，就英國中小學校長評鑑系統符合四大類 21 項教育人員評鑑標準的程度與優缺點作綜合性敘述。

(一)正當性標準方面

英國的全國性校長評鑑系統在符合正當性標準的優點方面，包括：1.提供明確的「正式評鑑指引」(P2)；2.明確管制誰有權接觸評鑑報告 (P4)；3.受評者與評鑑者有積極性互動，並提供受評者有價值的經驗 (P5)。在缺點方面，包括：1.未以服務學生需求為導向 (P1)；2.該評鑑系統雖有保護措施以減低利害衝突 (P3) 與評鑑者的不適當影響；但由於讓受評者得以控制評鑑過程，使得評鑑成為軟性評鑑，同時也缺乏判斷表現的明確標準。

(二)效益性標準方面

該評鑑系統大部分均能符合效益性標準的要求，包括：1.具建設性取向

(U1), 以協助校長強化其專業知技能 ; 2.明確的界定評鑑資訊用途與使用者 (U2), 且讓受評者參與評鑑規則的訂定 ; 3.具體的建立評鑑者信譽 (U3); 4.對評鑑報告亦建立明確、適時與有關聯性的報告程序，屬於功能性的報告 (U4); 5.在評鑑週期的第二年，對評鑑結果進行追蹤與發揮影響 (U5)。

(三)可行性標準方面

該評鑑系統在符合可行性標準的優點方面，包括：1.具政治上的可行性(F2)，雖然家長、學生與社區成員未參與評鑑的發展歷程，但許多利害關係團體均大致同意最後的評鑑計畫；2.雖然在安排評鑑活動的程序上有些困難，但該評鑑系統在執行與提供適當的評鑑資訊方面，具務實性的程序 (F1)。在缺點方面：由於缺乏評鑑的足夠時間與經費，因此較不具有財政上的可行性 (U3)。

(四)正確性標準方面

該評鑑系統在符合正確性標準的優點方面，包括：1.有良好的實施程序記錄 (A3)，讓受評者及其他人可以檢視；2.由教育標準局與教師訓練局負責評鑑系統的監督與評鑑 (A8)，並提出檢討報告；此外尚有部分符合標準的優點，包括：1.明確的界定角色 (A1)，但缺乏判斷受評者表現的明確指標；2.資料蒐集方面採取多元化的來源、觀察者與情境，以發現測量的不一致 (A5)，但仍缺乏信度的

實證性評估；3.有一套文件保管系統，可以有系統的控管資料 (A6)，但並不是很清楚；4.有超過一個以上的評鑑人員且採用多元化的證據來源，可以減低可能的偏見 (A7)，但仍缺乏判斷校長表現的明確標準。在缺點方面，包括：1.對考慮工作環境的限制上，缺乏明確的要求或指引 (A2)；2.雖然以多樣化方式來蒐集資料，以強化判斷的正確性，但測量工具與方法的效度仍有問題。

二、Martens 的研究

Martens (1991) 以晤談法探討美國地方學區在實施校長評鑑的歷程中，使用四大類 21 項教育人員評鑑標準的情形。他以 Southern California county 七個聯合學區內 13 位教育行政人員及 14 位國中小學校長為對象，分別就各學區：現有校長評鑑系統的發展歷程、現有校長評鑑系統、現有追蹤歷程與現有的評鑑訓練系統等問題加以提問。所得資料並以事後回溯分析 (Ex-Post analysis) 作內容分析，以了解其符合 21 項教育人員評鑑標準中各實施指引 (guidelines) 的程度，俾決定每項評鑑標準被運用的情形。表 3 列出各評鑑標準的運用情形。

表 3 21 項「教育人員評鑑標準」指標運用的百分比

順序	評鑑標準	代號	百分比	順序	評鑑標準	代號	百分比
1	建設性取向	U1	74.0	12	利害衝突	P3	43.4
2	務實性的程序	F1	62.4	13	正式的評鑑指引	P2	43.3

3	與受評者互動	P5	62.3	14	明確的角色	A1	37.0
4	追蹤與影響	U5	58.0	15	財政上的可行性	F3	33.0
5	評鑑人員的信譽	U3	56	16	系統的資料分析	A6	32.0
6	管制評鑑報告接觸	P4	53	17	記錄實施程序	A3	30.0
7	明確的用途	U2	51.4	18	測量的正確性	A4	16.0
8	具功能性的報告	U4	51.4	19	政治上的可行性	F2	13.0
9	服務取向	P1	50.0	20	監控評鑑系統	A8	2.0
10	工作環境	A2	48.0	21	測量的可靠性	A5	0.0
11	偏見控制	A7	46.0				

資料來源：Glasman & Martens (1993:54)

由表中可明顯看出，四大類 21 項教育人員評鑑標準被運用在校長評鑑系統的重視程度有別。其中，以「效益性標準」最受重視，其所含的五項次標準，均位於 21 項標準的前八項內，包括：建設性取向 (U1)、追蹤與影響 (U5)、評鑑人員的信譽 (U3)、明確的用途 (U2) 與具功能性的報告 (U4) 等五項，其指標運用百分比均在 50% 以上。在「正當性標準」方面，五項次標準的排序介於第三位與第十三位之間，其中與受評者互動 (P5) 位居第三，管制評鑑報告的接觸 (P4) 位居第六較受重視；在「可行性標準」方面，三項次標準的排序較分散，介於第二與第十九之間，其中的務實性的程序 (F1) 相當受重視，位居第二，惟另兩項次標準：財政 (U3) 與政治上的可行性 (U2) 則位居倒數第七與第三。在「正確性標準」方面，相對的較受忽視，八項次標準的排序均在後半段，介於第十至

第二十一位，尤其是對評鑑系統的監控（A8）與測量的可靠性（A5），分居倒數第二與第一位。Glasman & Martens (1993)認為：不同的決策要求，會有不同重視程度的評鑑標準。在實施校長評鑑系統中，效益性標準最受關注的主因在於校長評鑑的目的，主要在適時的發揮其影響力與提昇教育服務的品質。而正當性標準亦被認為相當重要，其主因在於讓校長評鑑的實施程序能符合倫理性並顧及受評者的尊嚴。在可行性標準方面，受學區實用性哲學影響，務實性的程序亦受相當重視。相對的，對正確性標準的要求，由於其技術上的困難度較高與評鑑人力的缺乏，因此在獲取正確與可靠性資訊方面較難達成。Glasman & Martens (1993)也發現，在較大的學區裡，由於受評校長與評鑑人員較多，因此需要更正式化的評鑑系統，而對評鑑標準的重視程度也較小學區高。

依據 Glasman & Martens (1993)的分析，該評鑑系統：（一）在優點方面：各學區每年均正式實施一次校長評鑑，而實施的歷程通常較強調合作的方式（如：U1, P5）、積極與公平的態度（如：A7）、一致性（如：U5, P2）與符合規定的正當程序（如：P4, U2, A1）。尤其是強調校長評鑑的建設性取向及重視評鑑人員與受評校長間的互動，以減輕校長對評鑑的心理壓力。各學區實施校長評鑑的主要目的包括：專業成長、改善經營缺失與達成學校目標等，兼顧校長評鑑的形成性

與總結性目的。而在整個校長評鑑歷程的文件中通常亦含蓋校長職責角色的工作描述、校長評鑑的決策目的、每年的目標設定程序、評鑑會議、觀察與其他資料蒐集活動、各種訊息來源的比較等說明。評鑑人員與受評者的角色亦有明確的界定。對新的受評校長亦提供定期研習課程。(二)在缺點方面：相較於教育人員評鑑標準的要求而言，各學區對校長評鑑的年度活動紀錄，仍不夠正式化與周詳，例如：缺乏正式的觀察表、在觀察與資料蒐集後，缺乏書面評論與會議、缺乏期中書面報告。此外，亦缺乏對評鑑人員實施正式的密集訓練，包括：表現評鑑技術、激勵方法、衝突管理與對相關法規的瞭解等。對時間的安排、評鑑程序的發展與修正、使用者對評鑑的回饋、對測量程序上信度與效度的研究等亦有不足。總而言之，缺乏時間、人力與經費是學區實施校長評鑑系統的最大限制。

經由上述的分析可發現，在英國全國性校長評鑑系統所強調的評鑑標準，與 Glasman & Martens (1993)在美國 Southern California county 七個學區校長評鑑系統所強調的評鑑標準仍有許多相似之處。在英美兩國，「效益性標準」最受重視，而「正確性標準」較受忽略。「可行性標準」在英國比在美國受重視，其原因或許來自兩方面：其一為校長評鑑系統實施的涵蓋層面不同；其二為哲學背景的差異。在英國，校長評鑑系統屬於全國性適用範圍，其影響層面較廣且大，政治層

面的考量因素較受重視，因此所發展出來的評鑑計畫必須更強調其可行性；反之，在美國，校長評鑑系統隨各學區的決策而有不同，各學區除重視其務實性的程序外，對政治上的可行性要求則較弱。此外，英國社會受實利主義哲學影響，在校長評鑑系統的政策規劃過程中，對可行性標準的要求自然比美國來得重視。

柒、後設評鑑標準在國內的實證研究

研究者（鄭新輝，民 91）以自編問卷，對國內各縣市 1,379 位國中小校長、主任、教師、家長、教育學者與各縣市教育行政人員（受試者）進行問卷調查，並選取 24 位各類代表（受訪者）進行訪談，以了解其對 22 項後設評鑑標準的重要性程度之看法。由研究結果發現：

一、整體而言，受試者對四大類 22 項後設評鑑標準，均普遍認為相當重要。其中「正當與效益性標準」的重要性顯著高於「正確與可性標準」；而在「正當與效益性標準」中，「效益性標準」的重要性亦顯著高於「正當性標準」，可見受試者認為校長評鑑系統的後設評鑑標準之重要性依序為：「效益性標準」、「正當性標準」、「可行性標準」與「正確性標準」。亦即規劃實施校長評鑑系統最應重

視的是：該評鑑系統能否對校長或學校教育，提供適時與有價值的影響力；其次為該評鑑系統的規劃能否符合評鑑的倫理與正當性；再其次為校長評鑑系統是否易於實施，有足夠的時間、資源與經費，且法令規範亦明確可行；而資料蒐集與評鑑結果均符合正確性的要求。至於不同性別、年齡、教育程度、職務身分、有無評鑑經驗與服務地區的受試者，對所列舉的「校長評鑑系統的後設評鑑標準」的反應，均無顯著差異。

二、若將受試者的反應，依後設評 22 項次標準的重要性程度加以排序，由表 4 可發現：排序在前十二項的次標準中，(一) 屬於「效益性標準」的五項次標準全部包含在內(佔 100%)，包括：「建設性取向標準」、「明確的用途標準」、「評鑑者的信譽標準」、「具功能性的報告標準」與「追蹤與影響標準」；(二) 屬於「正當性標準」者有三項(佔 60%)，包括：「服務取向標準」、「利害衝突的處理標準」、「與受評者互動標準」；(三) 屬於「可行性標準」者亦有兩項(佔 50%)，包括：「務實性的程序標準」與「法令上的可行性標準」；(四) 屬於「正確性標準」者則僅有兩項(佔 25%)，包括：「偏見的控制標準」與「檢視評鑑系統標準」。可見效益性與正當性仍是最受重視的標準。此一結果與 Glasman & Martens (1993)在美國 Southern California county 七個聯合學區實施校長評鑑所獲得的實證研究結果

相近似。其中僅在「檢視評鑑系統」一項，國內受試者比 Southern California county 重視，然而在考量「工作環境因素」方面，則相對較低。

表 4 國中小校長評鑑系統「後設評鑑標準」的各項次標準重要性程度排序

序號	次標準名稱	平均數	標準差	序號	次標準名稱	平均數	標準差
1	p5 與受評者互動	3.64	.53	12	u2 明確用途	3.52	.55
2	u3 評鑑者的信譽	3.59	.53	13	a1 明確的角色	3.51	.53
3	a7 偏見的控制	3.59	.52	14	a2 工作環境	3.51	.54
4	p1 服務取向	3.59	.52	15	a5 可靠性測量	3.49	.54
5	u4 具功能性的報告	3.58	.53	16	f2 政治上的可行性	3.48	.57
6	u5 追蹤與影響	3.58	.53	17	a4 正確性測量	3.47	.54
7	u1 建設性取向	3.57	.54	18	a6 有系統的資料控制	3.47	.53
8	f4 法令上的可行性	3.55	.53	19	p4 管制評鑑報告接觸	3.46	.63
9	p3 利害衝突的處理	3.54	.56	20	p2 正式的評鑑指引	3.45	.57
10	f1 務實性的程序	3.54	.52	21	f3 財政上的可行性	3.45	.55
11	a8 檢視評鑑系統	3.54	.54	22	a3 記錄實施程序	3.42	.55

三、若分析受訪者對規劃實施完善的國中小校長評鑑系統，應符合哪些標準的看法，並以 22 項後設評鑑標準為分類架構，則可發現在未提供引導的開放性問題反應中，受訪者對 22 項後設評鑑標準，僅四項無人提及，包括：管制評鑑報告的接觸、財政上的可行性、法令上的可行性與記錄實施程序等。而在關注的十八項標準中，則以「p2 正式的評鑑指引」、「p5 與受評者互動」、「u3 評鑑者的

信譽」、「u4 具功能性的報告」、「u5 追蹤與影響」、「f1 務實性的程序」、「a1 明確的角色」、「a2 工作環境」、「a4 正確性測量」、「a5 可靠性測量」、「a7 偏見的控制」等十一項被提及的次數較多。尤其是「u5 追蹤與影響」、「p5 與受評者互動」、「a4 正確性測量」、「a7 偏見的控制」、「p2 正式的評鑑指引」、「u3 評鑑者的信譽」等六項最受重視。若與問卷調查結果相比較，受訪者似乎對「正確性標準」的重視程度明顯增加；其次為「效益性標準」與「正當性標準」。此一結果亦反映出受訪者對校長評鑑系統在規劃實施歷程中，應強調「正式與有系統性的規劃設計、客觀公正且正確的資料蒐集、追蹤輔導評鑑結果的效能」等之重視。若由 Glasman & Martens (1993)在美國七個聯合學區的實證研究結果或本研究的發現，均顯示受訪者對「效益性標準」的重視程度最高，其次才是「正當性標準」；而在訪談的資料中，國內受訪者對「正確性標準」的關注程度亦不小，亦即國內受訪者對具「功能性與工具性取向」的「效益性與正確性標準」的要求，要比具「合倫理性與合法性取向」的「正當性與可行性標準」的要求要來得高。

四、若將受訪者的反應，依規劃實施校長評鑑系統的三階段歷程並配合後設評鑑標準加以歸納，則各階段應重視的後設評鑑標準主要有下列幾項：

(一) 規劃設計階段：1. 應有嚴謹客觀的評鑑規劃，包括評鑑目的、人員、

指標、方法、程序結果的處理等 (p2 正式的評鑑指引); 2. 規劃過程應公開化、多元參與，且具務實性 (f1 務實性的程序、f2 政治上的可行性); 3. 應明確化評鑑的目的或用途 (u2 明確的用途); 4. 評鑑應以提昇教育品質及校長辦學能力為主要目的 (p1 服務取向、u1 建設性取向); 5. 評鑑委員應具專業技能與態度，並遵守利益迴避原則 (u3 評鑑者的信譽、p3 利害衝突處理); 6. 評鑑指標應符應校長的角色與責任 (a1 明確的角色)。

(二) 資料蒐集整理階段：1. 應以專業、尊重的態度進行評鑑並與受評者有良好的互動 (p5 與受評者互動); 2. 資料蒐集應考慮校長所處的工作環境差異性並作個別化考量 (a2 工作環境); 3. 應正確客觀的測量與多元化的蒐集資料，以確保資料的正確性與一致性 (a4 正確性測量、a5 可靠性測量); 4. 所獲得的資料應有系統的整理 (a6 有系統的資料控制); 5. 應嚴防偏見，並公平的對待每位校長 (a7 偏見的控制)。

(三) 評鑑結果處理階段：1. 評鑑結果清楚明確，並提供具體建議 (u4 具功能性的報告); 2. 讓校長知道評鑑結果，配合評鑑目的，做適當處理與追蹤 (u5 追蹤與影響); 3. 進行後設評鑑 (a8 檢視評鑑系統)。

綜合言之，在規劃實施國民中小校長評鑑系統的歷程中，本研究所探討的四

大類 22 項後設評鑑標準均適用於國內的校長評鑑系統，而其重要性則以效益性與正當性標準最受關注。若以次標準的重要性相比較，則位居前三分之一者依序為：與受評者互動 (p5)、評鑑者的信譽 (u3)、偏見的控制 (a7)、服務取向 (p1)、具功能性的報告 (u4)、追蹤與影響 (u5)、建設性取向 (u1)。亦即在規劃實施國中小校長評鑑系統時，必須以提昇教育品質為終極目標；評鑑目的應具建設性，以協助校長不斷專業成長與改進；在評鑑過程中應慎選評鑑人員，尊重且公平的對待受評者；而評鑑結果更應提供受評者具體實用的建議，協助其採取適當的行動計畫，並落實追蹤與輔導工作。因此，若將四大類 22 項後設評鑑標準依其重要性程度重新排列，則國內國民中小校長評鑑系統在規劃實施歷程中，可遵循的各項後設標準之內涵與適用階段，可綜合整理如表 5。

表 5 四大類 22 項校長評鑑系統的後設評鑑標準、內涵與適用階段

後設評鑑標準	適用階段	後設評鑑標準的內涵
一、 效益性 標準	1.評鑑者的信譽	規劃設計
	2.功能性的報告	結果處理
	3.追蹤與影響	結果處理
	4.建設性取向	規劃設計
	5.明確的用途	規劃設計

二、 正當性 標準	1.與受評者互動	各階段	評鑑委員應以專業、尊重的態度進行評鑑，避免傷害受評校長的自尊。
	2.服務取向	規劃設計	校長評鑑的目的應能符合教育、學生、社區與社會需求。
	3.利害衝突	各階段	應公開與誠實的面對利害衝突，對與受評校長有利害關係者，應迴避擔任評鑑委員。
	4.管制評鑑報告	結果處理	評鑑報告應限制只有合法身分者才可查閱。
	5.正式評鑑指引	規劃設計	校長評鑑應有完整的規劃並編印成評鑑手冊或指引。
三、 可行性 標準	1.法令的可行性	各階段	評鑑的規劃與實施應符合相關的法令規定，使評鑑委員能更公平而有效地執行校長評鑑。
	2.務實性的程序	規劃設計	應妥善規劃評鑑流程，俾能在最少干擾與最低成本下，正確有效的獲取資訊。
	3.政治的可行性	規劃設計	校長評鑑系統的規劃與監控，應邀請專家學者、受評校長與其他有關人員共同參與合作來完成。
	4.財政的可行性	規劃設計	校長評鑑的規劃與實施應能提供足夠的時間、資源與經費。
四、 正確性 標準	1.偏見的控制	各階段	評鑑過程應嚴防偏見產生，並使受評校長獲得公平的評定。
	2.監控評鑑系統	各階段	校長評鑑系統應定期的進行檢討並做適當的修正。
	3.明確的角色	規劃設計	應明確界定受評校長的角色、責任與表現目標，讓評鑑委員能正確的蒐集評鑑資料。
	4.工作環境	資料蒐集	應明確描述受評校長所處的環境背景，並能考量這些環境因素對校長表現的可能影響。
	5.可靠的測量	資料蒐集	資料蒐集與測量程序應能確保所獲的資料具有一致性或可靠性。
	6.正確的測量	資料蒐集	資料蒐集與測量程序應依所描述的校長角色與目的進行，讓測量或資料蒐集更具正確性。
	7.有系統的資料控制	資料蒐集	校長評鑑所獲得的資訊應有系統的妥善處理與保存。
	8.記錄實施程序	各階段	應詳細記錄評鑑實際進行的程序，以檢核是否與預定程序相符。

捌、結語

校長評鑑系統的發展已有相當的歷史，在美國教育評鑑發展的歷程中，由重視學生的學習成就表現，所促成的標準化測驗運動的發展；到 1960 年代中期為能了解各種教育補助方案的執行成效，因此強化教育方案的評鑑，更進而發展出

方案評鑑標準以作為後設評鑑的基礎；到了 1980 年代由於學生的學習與方案的成效未盡理想，因此將教育評鑑的重心轉向執行教育方案與影響學生學習表現最直接的教育人員身上，並在績效責任運動的推波助瀾下，開始展開對教育人員的評鑑，先由教師再轉向教育行政人員，而校長評鑑也成為重要的教改任務。各州與各學區亦逐步發展其校長評鑑系統，藉以落實校長的績效責任，並引導其專業成長，進而提昇辦學校能與學生的學習表現。然而校長評鑑自實施以來，由於政策規劃未完備、設計未周延、執行未落實、方法與工具未合適、評鑑結果的處理未獲重視；而評鑑人員也未受必要的專業訓練；受評者也未獲得充分的尊重，使得教育人員評鑑系統的發展經常受到質疑與批判。因此，教育評鑑標準聯合委員會，才在 1988 年發展出教育人員評鑑標準，以作為教育人員後設評鑑的標準。其所重視的正當性、效益性、可行性與正當性標準，已成為檢視各種評鑑系統的重要價值哲學。而所提供的四大類 22 項標準不僅有明確的使用原理，且提供具體的指引，可說是目前最為完備的後設評鑑標準。不僅提供該標準的應用程序，近來更提出後設評鑑實施的方法論。雖然該標準係在美國文化環境下形成，但在理念與做法上仍有可供參考之處。為建立國內完善的國中小校長評鑑系統，在規劃設計上仍可以以該標準為基礎進行規劃，或以之進行現行校長評鑑系統的後設

評鑑。讓國中小校長評鑑系統能符合正當性、效益性、可行性與正確性標準，以充分發揮並確保校長評鑑系統的品質與效能。由國內實證研究中發現，22 項後設評鑑標準在校長評鑑系統規劃實施上的重要性均相當高。而受試者對校長評鑑系統的要求，如同英美兩國一般，均以效益性標準最受重視，包括評鑑者的信譽、具功能性的報告與落實追蹤輔導等，而受訪者亦對正確性標準相當關注，因此未來在校長評鑑系統的規劃設計上，應特別重視對該兩項標準的要求。

參考書目

- 江文雄等(民 88)。台北市中小學校長評鑑方案之探討。台北市政府教育局委託研究論文，台北。
- 吳德業(民 90)。苗栗縣國民小學校長評鑑制度實施現況之調查研究。國立新竹師範學院國民教育研究所學校行政碩士班碩士論文，未出版。
- 吳錦森(民 89)。國民小學校長評鑑制度之研究。國立台中師範學院國民教育研究所碩士論文，未出版。
- 侯世昌等(民 89)。台北縣國民小學校長評鑑之研究。行政院八十八年研考經費補助研究論文，台北。
- 倪靜貴(民 90)。國民中學校長評鑑制度之研究。載於國立嘉義大學國民教育研究所(主編)，中小學校長專業成長制度規劃(頁 255-275)。高雄：復文。
- 翁福元(民 90)。國民中小學校長評鑑制度之初探：以阿拉巴馬州為例。載於國立嘉義大學國民教育研究所(主編)，中小學校長專業成長制度規劃(頁 147-173)。高雄：復文。
- 張德銳(民 88)。國民中小學校長評鑑系統的初步建構。初等教育學刊，7，15-38。
- 鄭新輝(民 91)。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。

- Candoli, I. C., Cullen, K., & Stufflebeam, D. L. (1997). *Superintendent performance evaluation: Current practice and directions for improvement*. Norwell, MA: Kluwer Academic.
- Cook, T. D. (1974). The potential and limitations of secondary evaluation. In M. W. Apple, M. J. Subkoviak, & H. S. Lufner, Jr. (Eds.), *Educational evaluation: Analysis and responsibility* (pp. 155-235). Berkeley, CA: McCutchan.
- Cook, T. D., & Gruder, C. L. (1978). Metaevaluation Research. *Evaluation Quarterly*, 2(1), 5-51.
- Cullen, K. (1997). An evaluation of the United Kingdom's national system of headteacher appraisal. *Studies in Educational Evaluation*, 23(2), 103- 130.
- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.
- Ellett, C. D., Wren, C. Y., Callender, K. E. Loup, K. S., & Liu, X. (1996). Looking backwards with the Personnel Evaluation Standards: An analysis of the development and implementation of a statewide teacher assessment program. *Studies in Educational Evaluation*, 22(1), 79-113.
- Glassman, N. S., & Martens, P. A. (1993). Personnel evaluation standards: The use in principal assessment systems. *Peabody Journal of Education*, 68(2), 47-63.
- Joint Committee on Standards for Educational Evaluation (1988). *The personnel evaluation standards: how to assess systems for evaluation educators*. Newbury Park, CA: Sage.
- Martens, P. A. (1991). *A study of the nature and extent of use of standards in the principal evaluation process in selected districts*. Unpublished doctoral dissertation, University of California, Santa Barbara.
- Scriven, M. (1967). The methodology of evaluation. In *AERA monograph series on curriculum evaluation*, 1 (pp. 39-83). Chicago: Rand McNally.
- Scriven, M. (1972). An introduction to metaevaluation. In P. A. Taylor & D. M. Cowley (Eds.), *Reading in curriculum evaluation* (pp. 84-86). Dubuque, IO: Brown.
- Scriven, M. (1983). Evaluation ideologies. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation*. Boston, MA: Kluwer-Nijhoff.
- Scriven, M. (1997). Due process in adverse in adverse personnel action. *Journal of Personnel Evaluation in Education*, 11 (2), 127-137.
- Shinkfield, A. J., & Stufflebeam, D. L. (1995). *Teacher evaluation: Guide to effective practice*. Norwell, MA: Kluwer Academic.
- Stufflebeam, D. L. (1981). Metaevaluation: Concept, standards, and uses. In R. A. Beck (Ed.), *Educational evaluation methodology: The state of the art* (pp. 146-163). Baltimore, MD:

Johns Hopkins University.

Stufflebeam, D. L. (2000). The methodology of metaevaluation as reflected in metaevaluations by the Western Michigan University Evaluation Center. *Journal of Personnel Evaluation in Education*, 14(1), 95-125.

Stufflebeam, D., & Pullin, D. (1998). Achieving legal viability in personnel evaluations. *Journal of Personnel Evaluation in Education*, 11(3), 215-230.

Stufflebeam, D., & Nevo, D. (1993). Principal evaluation: New directions for improvement. *Peabody Journal of Education*, 68(2), 24-46.

Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.

Meta-evaluation Standards for School Principal Appraisal

Hsin-hui Cheng^{*}

ABSTRACT

The appraisal of school principals has been receiving attention in the field of education administration in Taiwan. At a time when every local government is developing its own evaluation system, it is most urgent to have a set of meta-evaluation standards to be referred to during the planning and implementing process. These standards are indispensable for examining current evaluation systems for that matter. It is the aim of this study to explore the feasibility of such a set of meta-evaluation standards in the Personnel Evaluation Standards issued in 1988 by Joint Committee on Standards for Education Evaluation in the United States. The 21 standards in four categories, with the proposed new standard on legal viability, provide clear rationale and guidelines to be followed, and to avoid common errors. They constitute a most comprehensive set of meta-evaluation standards for education administrator appraisal to date. These standards not only can serve as criteria for designing systems of school principal appraisal, they can be used to meta-evaluate ongoing principal appraisal systems as well. These 22 standards have been used to examine school principal appraisal systems in the United States, as well as Britain. Although they were formulated in an American cultural background, these standards are useful reference in both theory and practice in Taiwan. The empirical research in Taiwan also revealed that these 22 meta-evaluation standards should be taken more seriously when planning and implementing school principal appraisal systems. Most important of all were the utility and propriety standards. Nevertheless the feasibility and accuracy standards were also frequently mentioned among the interviewees.

^{*} Hsin-hui Cheng: Assistant Professor, Department of Elementary Education, National Tainan Teachers College.

Key words: Principal appraisal, Meta-evaluation, Evaluation standards