

第參章 文獻探討：後設評鑑

第一節 後設評鑑的要素與架構

壹．後設評鑑的意義、內容和標準：

後設評鑑的意義，一般是指對某一教育評鑑活動的評鑑，評鑑該項評鑑方案是否合理？評鑑工作是否有效？評鑑結果是否可靠？是否依照一定的評鑑理論和標準？對教育評鑑本身進行評鑑與研究的工作。教育評鑑的後設評鑑與教育評鑑兩者關係密切。但也有以下的不同；從研究對象而言；教育評鑑以受教者的變化和造成這些變化的各種因素為對象；進行研究和評鑑，而後設評鑑則是把教育評鑑本身，當作對象來加以研究和評鑑（侯光文，1999）。從目的和作用上而言；教育評鑑是通過對被評對象的價值判斷，為教育決策提供依據和服務，或進行總結性結果來分等級，或回饋矯正改進工作；然而後設評鑑則是為了規範教育評鑑活動，使其減少偏差，實現評鑑目的，發揮評鑑功能，以獲得良好評鑑效益。

後設評鑑如以內容和標準，針對評鑑方案的評鑑而言應包含：1.評鑑對象要明確；2.評鑑目的要正確；3.評鑑標準要合理，權重分配要恰當；4.評鑑工作計劃周密，安排合理；5.評鑑的方法科學可行，評鑑信息的收集、分析、處理得當。如以評鑑組織與過程而言應包含：1.評鑑組織要健全，職責要明確；2.評鑑人員能履行職責，按照方案要求，正確運用評鑑方法，且客觀公正地展開評鑑；3.評鑑須符合道德要求；4.被評者能態度正確，資料完備。如以評鑑結果的效用而言，評鑑結果要可靠有效，應包含：1.採集信息和處理過程的分析，結合必要的抽樣複評，認定評鑑有較高的信度和效度；2.評鑑結果的解釋合理，結論恰當獲被評對象認同接受；3.對被評對象具促進改善功能；4.評鑑人力、財力和時間的投入合理，社會反映良好。

除上述後設評鑑的內容和標準外，其基本要求亦包括以下幾個方面：

一、可靠性(dependability)和準確性(accuracy)：

可靠性是指評鑑的信度(reliability)，它反映了對評鑑非系統誤差控制的程度，信度高代表評鑑的結果是穩定可靠的(stability)，否則評鑑的一致性(consistency)較差，可靠程度不高。可分為評分者信度、再測信度、複本信度和內部一致性等。準確性是指評鑑的效度(validity)，他反映了對評鑑系統誤差控制的程度，效度高代表評鑑能達到評鑑目的，能正確、真實地測量出想測量的東西，評鑑效度包括：評鑑方案設計效度；又可分為內容效度(content validity)、結構效度(construct validity)；評鑑實施效度；又可分為信息採集效度和信息處理效度；結論效度；又可分為結果合成效度和效標關聯效度(criterion-related validity)。其涵蓋的範圍主要包括：評鑑目的要明確；評鑑內容和標準要合理；評鑑信息收集的方法和處理的技術要完善；多人評鑑或一人多次評鑑結果有較好的一致性，信度較高；評鑑結果能準確地反應被評對象的真實情況，有較高的效度。

二、適用性(propriety)和可行性(feasibility)：

重點包括：評鑑步驟要求明確；對所有被評對象，有較好的適用一致性；評鑑標準明確，內容具代表性；評鑑方法具體可行，評鑑者易於掌握；評鑑經費、人力、物力耗費不大。

三、效用性(utility)和精確合理性(accuracy)：

重點包括：具正確導引作用；實現評鑑目的，發揮評鑑功能；評鑑結論正確能得到被評對象的認同，對被評對象的工作具明顯的改進功能；評鑑者能顧全被評者的合法權益。

貳、後設評鑑的方法

從整體而言後設評鑑是教育評鑑工作的一部分，其採用的方法可從教育評鑑的方法中加以移植運用，如觀察法、訪談法、問卷法、訊息處理的集中趨勢、

相關程度、統計推斷、綜合評判，相對評鑑與絕對評鑑，形成性評鑑與總結性評鑑等方法。至於其所採行的主要方法有如下四種，說明如下：

一、內容分析法

內容分析法是對評鑑中各種資料內容及其效用，依照評鑑目的進行客觀深入的研究、分析和評判的方法。其實施的步驟包括：

確定後設評鑑研究對象與目的。

依據研究對象與要求，設計內容、分析向度及標準。

收集資料，作好歸類。

系統分析與評斷。

對結果的處理與分析。

二、經驗總結法

經驗是指人的經歷和體驗，有時也泛指人們在實踐中獲得的知識與技能，經驗總結法是根據後設評鑑目的，依照後設評鑑標準，對評鑑中的事實經驗，透過理性思維，揭示客觀規律的活動，其實施的步驟包括：

對教育評鑑的事實經驗，進行收集、整理、篩選、準確選出富有代表性的事實。

透過對照、比較、分類進行初步的分析、綜合、抽象和概括對評鑑事實與關係作一定的歸納，形成經驗性概念。

深入的分析 and 抽象，透過判斷和推理，尋找出本質和規律。

三、評鑑信度分析法

教育評鑑信度反應了評鑑結果的可靠程度，一般採用定量分析，可分為評分者信度、內部一致性、穩定性和等值性（林清山，1997）。

評分者信度：

主要有相關係數法和肯德爾和諧係數法。相關係數法；一般用於評鑑

兩個評分者評分的一致性，其步驟如下：

隨機抽取一定數量的被評學校。

分別由兩位評分者，依評鑑標準對被評學校作出評定，得到兩列分數。

計算兩列分數的相關係數（積差相關或等級相關）。

作出評鑑，可根據相關係數假設檢驗的方法，作出統計判斷。一般要求相關係數應在 0.09 以上。

肯德爾和諧係數法；當評分者超過兩人以上時，可用此法，其步驟如下

設有 k 個評鑑人員，對 n 個被評對象評分。

將每位評鑑人員的分數，按順序化為等級分數。

計算和諧係數
$$W = \frac{S}{\frac{1}{12}K^2(N^3 - N)}$$

其中， W 為和諧係數。

S 為每個被評對象，所獲得等級分數之和，

R_j 與所有等級分數的平均數 \bar{R} 之差的平方和，即

$$S = \sum_{j=1}^n (R_j - \bar{R})^2$$

進行顯性檢驗，有兩種情況：

當 $K \leq 20$ 且 $N \leq 7$ 時，屬於小樣本，檢驗統計量 S 與肯德爾一致性係數 S 臨界值表中的臨界值對照，若 S 值大於或等於臨界值，認為意見是一致性的，否則為不一致。

內部一致性：

評鑑指標中各指標間或測驗卷各項目間，如果有較高的正相關，就具有較好的內部一致性，或稱之為同質的，一般而言，內部一致性愈高，項目愈同質，它所代表的分數意義就愈明確；教育評鑑往往包含許多不同性質的評鑑內容，從整體而言，其內部一致性的可能不高，但對同範

疇的內部應當要有較高的內部一致性。以下簡要說明常用的克隆貝克 α 係數(Cronbach's coefficient alpha) (謝廣全，民 85)。

$$\alpha = \left(\frac{K}{k-1} \right) \left(1 - \frac{\sum Si^2}{Sx^2} \right)$$

其中 K 為項目數。

Si^2 為第 i 項目得分的方差。

Sx^2 是整個評鑑的方差。

信度要求在 0.5 以上

例一：某次評鑑包括 8 個項目，10 個被評對象，得分如表 3-1-1 該評鑑的內部一致性係數計算如下表 3-1-1：

表 3-1-1 評鑑內部一致性係數計算

被 評	1	2	3	4	5	6	7	8	總 分
A	10	12	8	11	9	8	12	15	85
B	8	10	11	10	9	9	10	13	80
C	11	7	7	12	8	6	11	14	76
D	9	8	10	7	10	7	8	10	69
E	7	11	6	6	7	10	11	8	66
F	8	8	8	8	7	7	9	7	62
G	6	8	7	9	8	8	5	10	61
H	7	9	5	6	5	6	8	7	53
I	10	4	3	7	6	4	6	6	46
J	4	5	6	3	5	4	2	5	34
Si	3.61	2.36	2.11	2.55	1.62	2.09	2.96	3.32	Sx=14.90
Si ²	13.0	5.56	4.89	6.49	2.64	4.36	8.76	11.05	Sx ² =222.01

可得 $\sum Si^2 = 56.76$

$$\alpha = \left(\frac{K}{K-1} \right) \left(1 - \frac{\sum Si^2}{Sx^2} \right) = \frac{8}{8-1} \times \left(1 - \frac{56.76}{222.01} \right) = 0.71$$

穩定性和等值性

穩定性是對同一群體用同樣的評鑑項目，在不同時間進行的評鑑，所得結果間的相關係數是否一致，目的在檢查評鑑實施產生的結果，是否穩定？

等值性是基於相同的評鑑目的，採用不同的評鑑方案針對同一群體實施評鑑所得的兩組結果，求出其間的相關係數，借以研究和判斷兩方案的異同，借由其中一方案所得的結果，可以推論另一方案應得的結果。

第二節 後設評鑑的分析

壹、評鑑信度

一、信度的概念

信度是指評鑑結果的穩定性或可靠的程度，亦即評鑑的結果是否真實客觀地反映學校實際表現（胡中鋒，李方 1999）。可從以下三方面加以說明：

信度是指實測值與真實質相差的程度

$$X = T + E$$

X：表示實測值；T：表示真實值；E：表示誤差值；其中測量誤差越小，信度就越高。

信度是指統計量與參數間的接近程度

統計量是指樣本的各种數字特徵（如樣本的平均數，標準差等），參數是指學校總體的各种數字特徵（如總體平均數，標準差），統計量越接近參數，可靠性越高，信度就越高。

信度是指兩次重複評鑑或等值評鑑間的關聯程度

對同一對象進行兩次重複評鑑或等值評鑑，計算兩次評鑑的相關係數，相關係數越高，信度就越高。

二、信度的類型

評分者信度：

不同的評分者對同一份試卷或評分量表往往給分不同，甚至有很大的差距，如作文評分，品德評分，創造力評分，評分者間的評分越一致，評分信度越高，反之則越低。

計算評分者的一致性參係數，若為二人評 n 份試卷可用斯皮爾曼等級計算（林清山，1992），若為三人以上時，則需計算肯德爾和諧係數。

重測信度：

是指同一試卷或評分表格，在不同時間內對同一組受評對象進行測試，實施兩次測試實得分數的相關係數。

其程序為：測驗 A_1 適當時距=>測驗 A_2

A_1 表示試卷（或量表）A 進行第一次測驗

A_2 表示試卷（或量表）A 進行第二次測驗

複本信度：

或稱等值性係數，當同一個測試不適合實施兩次時，則採用該測試的另一個平行測試（或稱複本），複本的內容、題數、格式、難度、平均數、標準差均應與原測試一樣。

其程序為：測量 A_1 最短時距=> 測驗 B_1

試卷 A 與 B 等值，兩次測試盡可能在最短時間內進行

內部一致性：

為了避免使用兩次測試的麻煩，根據一次測試，把一次測試內容對等分為兩部分，比較兩部分評分一致程度，來評估信度係數，稱為內部一致性，其計算方法有兩種（王文中等，民 88）。

折半信度：常以題目的奇數為一組，偶數為一組，計算兩組

的相關係數，以斯皮爾曼一布朗公式校正。其公式為：

$$\gamma_{tt} = 2r_{xy} / (1 + r_{xy})$$

r_{xy} 為兩組測試分數的相關係數

r_{tt} 為整個測試的信度係數

庫爾德—理查森公式法(Kuder-Richardson reliability)只需測試一次，後以正確得分數為基礎,或根據個人總分的平均數，和標準差，計算信度係數，其中可用

$$\gamma_{KR11} = 1 - \frac{0.8 \bar{x} (k - \bar{x})}{K_s^2}$$

\bar{x} 為個人總分的平均數；

S^2 為個人總分的方差

K 為題目數。

三、關於教育評鑑的指標的信度，可從三方面說明：

同一評鑑者（評鑑組或個人）運用同一評鑑指標，在短時間內重複評鑑同一對象，兩次（或以上）的評鑑結果，越接近則信度越高，反之，則越低，可稱為重複信度。 不同評鑑者使用同一評鑑指標在短時間內，評鑑同一對象，視其結果是否一致的程度，可稱為評鑑者信度。 同一評鑑者使用兩種（或以上）的評鑑指標，同時評鑑同一對象，所得結果與事實相符的程度，較能反應實際情況者，信度較高，可稱為事實信度。

一般而言，評鑑指標力求客觀，指標間界線清楚，評鑑標準規定明確，給分標準界線確定，有利於提高信度，此外評鑑過程的安排，評鑑方法的運用，評鑑委員的價值觀，對評鑑標準的理解程度，專業能力，評鑑工作的訓練落實程度等，皆會影響評鑑的信度。

四、提高信度的方法有

適當增加評鑑量表題目的數量，題目數量愈多，題目的代表性愈大，可提高評鑑的信度、效度，但太多的題目數也會造成無法操作，只能適當增加。評鑑量表題目難度要適當，評鑑量表難度適中，能使評鑑的信度達到最

大，也能使評鑑的區分度達到最大。

評鑑程序力求統一，包含簡報內容、參觀內容、檢視場地內容，查閱資料內容、訪談內容、座談內容、訪評委員討論內容，綜合座談內容等均應力求統一。

評鑑的時間要充分，如果時間不夠，訪評委員無法從容評鑑所有問題，也就無法真實反應受評者的實際內涵。

評分要客觀，減少評分誤差，根據參照準則的評分標準評分，最好是多位專家分別評分，平均計算或是單一位評分者多次評分，對結果誤差數大者給予複評，其次訪評委員的疲勞程度亦須加以注意。

貳、評鑑效度

一、效度的概念

效度是指評鑑結果的準確性和有效的程度，亦即評鑑是否達到預期目的（胡中鋒、李方，1999）。

從 $X=T+E$ （ X 為測量值、 T 為真實值、 E 為誤差）進一步分析實測值的方差為

$$\sigma_x^2 = \sigma_r^2 + \sigma_e^2 = \sigma_v^2 + \sigma_x^2 + \sigma_e^2$$

σ_v^2 為潛在真分數方差

σ_x^2 為系統誤差方差

σ_e^2 為隨機誤差方差

效度的計算公式為 $\gamma_{xy} = \frac{\sigma_v^2}{\sigma_x^2}$ 可知，隨機誤差方差和系統誤差方差的和越少，也就是潛在真分數方差越大，效度越高。

教育評鑑指標的效度是指評鑑預期目標與評鑑指標相關程度，評鑑指標、評鑑標準及各元素設置越準確、越具代表性，其效度越高。在實際運用上，如果指標不全，未能全面反映所要評鑑的目標或指標不當，造成誤導教育工作或指標數重分配不合理，皆會降低教育評鑑的效度，只有評鑑結果能真實、準確地

反映所欲測量的功能和特性時，才是有效的或高效度的。

在教育評鑑中，評鑑對象大都屬於精神層面，只能透過其具有可測性的外部表現（如語言，行為等）的測量，間接認識其心理活動或心理特徵，然而此一間接認識從外部表現到內心世界，一般僅具有相關關係而無一定的函數關係，並不能準確無誤的反應一位具有主觀能動性的個人心理狀態，人往往又會有意識地調節自己的外部行為，掩飾自己的內心活動，更增加了認識精神層次的難度，也降低了評鑑的效度。

二、效度的類型

內容效度：是指根據測試目的，提出所欲測量的內容和預期反應所達到的效度，也就是測量內容的代表性程度。其估計的方法有：1.邏輯分析法：根據教育學理論、教育法規、課程標準、學校工作實務，規劃學校教學與行政工作的範圍和深度，再以邏輯分析法，估計評鑑時能在多大程度上代表這些內涵，評鑑內容能否測出所要測量的特性和功能。2.評鑑內容與校務工作內容比較法；先製作評鑑的雙向細目表，分別列出評鑑內容和校務工作內容及其應佔的百分比，再根據此兩表相符合的程度，判斷內容效度的高低。

效標關聯效度：又稱經驗效度或統計效度，是以得分和效標之間的相同關係數來表示測試效度的高低。效標以顯示測試與所欲測量特性的變量，作為效度的參考尺度，例如，以各學校畢業成績為效標來檢驗學生升學考試的成績，計算兩者之間的關係數，就是升學考的效標關聯效度。

結構效度：是指一個測試能實際測量出理論上的概念或心理特性的程度（王文中等，民88），其目的在於使用心理學的概念，分析解釋分數的意義，其方法是先從某一結構理論出發，找出各項關於心理功能或行為的基本假設，據此編制測試問題，再由果溯因，檢驗測試結果與心理學的理論是否相符？分析其間的相關因素，是由累積結果的證據來說明測試結構的

效度。

三、提高效率度方法

不同的測試有不同的效度係數（邵國柱，1995）。如果原測試與獲得效標的測試間越相似，則效度係數越高。如智力測驗與標準測驗之間的相關係數應達到 0.6~0.8 才算合於要求。

影響效度高低的因素有：

評鑑的設計方面：評鑑目的、評鑑標準、評鑑的訓練都須做出合理的安排。

評鑑的實施方面：如評鑑流程、評鑑方法、評鑑委員的要求都須依照標準化的程序安排。

受評對象的心理狀態方面：受評者的態度、動機、情緒能否真實的反應學校真實的狀況。

樣本取樣方面：樣本要能代表所評的全體對象，當樣本容量增加時，測量的誤差有相互抵消的趨勢，結果較為穩定；樣本的異質性，樣本分數全距越大，效度係數越高。

提高效率度的方法有：

控制評鑑系統誤差：包含評鑑組織、職業及分工，評鑑課程訓練，定期的評鑑制度，徵詢被評對象的意見等。

精心編製量表：量表內容要確實反映評鑑目的，量表題目敘述必須清楚、簡明、難度適合有足夠的鑑別度。

要善加進行評鑑：依照評鑑手冊進行，不做超出規定的解釋，掌握時間，遵循評分標準。

擴充樣本的容量和代表性：不論是訪談對象，查閱資料對象，問卷對象，座談對象均應增加樣本對象的代表性，使隨機誤差相互抵消。

合理處理信度效度關係：信度是效度的必要條件，雖然信度高的評鑑未必效度高，但效度高的評鑑，信度一定高，信效度的關係是，效度的最大值

等於信度的平方根〈邵國柱，1995〉。

適當增加評鑑量表題數：增加量表題數可以提高信度，也可提高效率。

參、評鑑標準的難度(item difficulty)

一、難度的概念

難度是指達到評鑑標準的難易程度，一般是以被評的總成績與滿分的比值，難度是一個相對的概念，是由參予評鑑所有被評群體，整體表現所決定的。

二、難度的計算

基本公式

$$P = \bar{X} \div K$$

P 表示難度指標， \bar{X} 表示所有被評學校的平均得分，K 表示該項的滿分。例如，教務與實習範疇滿分為 100 分，所有學校在這個範疇的平均得分為 85 分，其難度為 $P = \bar{X} \div K = 85 \div 100 = 0.85$ ，難度偏易。

採用極端分組法計算難度

其步驟如下：

1. 依得分排序，找出總人數 27% 的高分組與 27% 的低分組。

$$2. P = \frac{X_H + X_L - 2nL}{2n(H - L)}$$

P 表示難度指數， X_H 表示高分組得分總和， X_L 表示低分組得分總和，n 表示總校（科）數的 27%，H 為最高得分，L 為最低得分。

例如，某評鑑標準，高分組得分總數和為 40 分，低分組得分總數和為 15 分，有 37 學校參與評鑑，此項標準最高得分為 8 分，最低得分為 2 分，則此評鑑標準的難度為：

$$P = \frac{X_H + X_L - 2nL}{2n(H - L)} = \frac{40 + 15 - 2 \times 10 \times 2}{[2 \times 10 \times (8 - 2)]} = 0.125$$

此題偏難。

三、難度對評鑑的影響

難度影響評鑑分數的分佈型態

難度過大或過小，都影響評鑑分數的分布；難度接近 0，難度越大，評鑑分數集中在低分段，其分數分布呈正偏態；難度接近 1，難度越小，評鑑分數集中在高分段，其分數分布呈負偏態（邱淵等，1992）。

難度影響評鑑分數的離散程度

難度過大或過小，造成分數集中在低分段或高分段，評鑑分數的離散程度變小，往往造成評鑑結果的實際狀況，與各校客觀存在的事實產生差異，因此，適當的難度，其分數的分佈範圍才有可能達到最大。

評鑑標準的難度影響評鑑的鑑別能力

適當的難度可使區分度達到最大。

評鑑標準的適宜難度

依計算公式得出 P 的範圍為 $0 \leq P \leq 1$ ，當 $P=0$ ，難度最大(所有學校均得 0 分)，當 $P=1$ 時，難度最小(所有學校均得滿分)，依常態分配，大多數評鑑標準的難度在 0.3~0.7 之間，使整體評鑑標準平均難度為 0.5 左右 (0.45~0.55 之間)，只有適當的難度才能區分出各學校不同程度的最大效果，也才能使得分呈常態分配，而對全部都滿分或都零分的標準應予刪去。難度指數等級如表 3-2-1：

表 3-2-1 難度指數等級

等級類別	難度指數範圍
易	$0.7 < P < 1$
中	$0.3 < P \leq 0.7$
難	$0 < P \leq 0.3$

肆、評鑑結果的鑑別度〈item discrimination〉

一、鑑別度的概念

鑑別度是指評鑑結果對受評鑑學校實際表現的鑑別程度，具有良好鑑別度

的評鑑，實際表現優良者應得高分，反之得低分。鑑別度又稱區分度(胡中鋒、李方，1999)。他是篩選評鑑指標的重要依據，以 D 表示；當 D>0 時，表示正鑑別，意指實際表現優良的學校，得到高分，實際表現不佳之學校，得到低分；當 D<0 時，表示負鑑別，意指實際表現優良的學校得到低分，實際表現不佳的學校反得到高分；當 D=0 時，表示零鑑別，意指實際表現與得分之間沒有多大的關係。鑑別度的高低直接影響到評鑑效果的信度和效度。鑑別度與評鑑信度的關係如表 3-2-2、3-2-3：

表 3-2-2 區分度與評鑑信度的關係

度鑑別	信度
0.1225	0.00
0.16	0.42
0.20	0.63
0.30	0.84
0.40	0.915
0.50	0.949

1962 R.L.艾伯

表 3-2-3 鑑別度的最大值與難度的關係

難度(P)	1.00	0.90	0.70	0.50	0.30	0.10	0.00
區分度的最大值	0	0.2	0.6	1.00	0.60	0.20	0.00

由上表可知，難度適中可使鑑別度達到最大值。

二、鑑別度的計算

極端分組法：高分組和低分組各取總人數的27%，依公式

$$D = \frac{X_H - X_L}{n(H - L)},$$

D 表示區分度， X_H 表示高分組得分總數， X_L 表示低分組得分總數，n 表示總校（科）數的 27%，H 表示最高得分，L 表示最低得分。

使用內部一致性係數計算鑑別度：可採用點二列相關、兩列相關、 ϕ 相關係數等方法計算相同係數表示鑑別度。

三、提高鑑別度的方法

使評鑑標準的難度適中。

著重評鑑輸出成果，使質優學校能得高分，質低學校得低分，使分數儘量分佈在整個分數量尺上。

四、評鑑結果的鑑別度

鑑別度是越高越好，一般可參照下表為標準。

表 3-2-4 採用極端分組法計算鑑別度的評鑑標準

鑑別度等級	鑑別度	評 鑑
1	0.4 以上	非常良好
2	0.30~0.39	良好，如能改進更好
3	0.20~0.29	尚可，但通常須要修改
4	0.19 以下	差，必須淘汰或改進以提高鑑別度後方可使用

伍、評鑑失當的幾種表現

評鑑因未能實現最佳，未能達到滿意程度，必然出現失當或不當之處（馬俊峰，1994）茲討論如下：

一、評鑑失準

評鑑失準，即是失之準確或失之正確，也就是說出現了錯誤，即評鑑結果與實際情況出現了錯誤，包含評鑑標準有誤、評鑑方法有誤、收集資料有誤、判斷有誤、計算有誤等等。有以下情況：

把有價值的當成沒有價值的，或反之亦然。

把小價值當成大價值，或反之亦然。

把重要價值當成非重要價值，或反之亦然。

把不可能或不太可能實現的價值當作可能或很可能實現的價值。

二、評鑑失全

評鑑失全是片面性的一種錯誤，不準是指整體上的錯誤，不全是指基本上正確，但未顧及全面，以偏概全，整體上是錯誤，但未及全面。有以下情況：

只注意到有利的一面而忽略不利的一面，反之亦然。

把握了一地區、一方面的利害，卻忽略了其他地區和其他方面的利害。

把握了一時、近期的價值對長遠的價值瞭解不足。

把握了表層的價值而對深層的價值理解不夠。

三、評鑑失時

評鑑未能即時回饋，即時效性。

評鑑內容方面的「時差」，評鑑者原先對某一事件的價值，隨著時間的推移，這事件及其影響已經發生了一定的變化，而評鑑者對此注意不夠，仍堅持先前的評鑑結論，於是出現了評鑑失誤。換言之，人們遇到與以前曾經經歷的相同事件，由於不同時空條件有不同的價值，仍沿用以前的評價和經驗，以致於原來確切的評鑑變不太確切。

四、評鑑失宜

有以下情況

本可兩全其美卻只顧了一邊失了一邊或兩邊未能全美，評鑑本可雙方受益，但最後只能一方受益，或是只顧集體利益而犧牲個人利益，或是只注意局部利益而失掉整體利益，或是只注意物質價值而失去精神價值等等。利未得全，害未避夠，兩利相權取其大，兩害相權取其輕，面臨評鑑標準的不盡理想，時間的緊促，各種利益的排列順序未十分清楚，最大效益只獲得了較小效益，最小傷害變成了最大的犧牲。

第三節 後設評鑑的設計與實例

壹、設計

教育評鑑標準聯合委員會之方案評鑑標準美國教育評鑑標準聯合委員會於 1981 年針對教育方案、計畫和教材的評鑑而訂定「教育方案、計畫和教材評鑑標準」(Standards for Evaluations of Educational Programs, Projects, and Materials)，共有三十項標準，分為效用性、可行性、適切性和精確性等四大類，1994 年加以修訂並更名為「方案評鑑標準」(The Program Evaluation Standards) (The Joint Committee on Standards for Educational Evaluation, 1981; 1994)。這套標準是由美加地區主要的教育專業組織有系統的發展而成，並由美國國家標準協會(American National Standards Institute, ANSI)所認可，以作為判斷教育評鑑的美國國家標準(Stufflebeam, 1994)。從標準編排的方式看來，第一版係以 A, B, C, D 來對四大項標準編號，第二版則是以四大項標準英文字母的大寫(U, F, P, A)來編號。以下列出上述第二版教育評鑑標準聯合委員會(1994)「方案評鑑標準」：

效用性準(Utility Standards):

效用性標準是要確保評鑑可提供特定當事人所需的實用資訊。這些標準如下：

U1 利害關係人的確認

參予評鑑或受評鑑影響的聽取人必須加以確認，以充分考慮他們的需求。

U2 評鑑者的可靠性

進行評鑑的人員必須是值得信任的，且有能力勝任評鑑工作 以利評鑑發現達到最大的可信度和可接受性。

U3 資訊的範圍與選擇

所收集的資訊，在範圍和選擇上必須注重有關受評對象的適切問題，並反映特定當事人的需求與興趣。

U4 評價的解釋

評鑑報告必須描述評鑑發現的觀點、程序和緣由，必須詳細描述，俾使價值判斷的基礎清楚明白。

U5 評價的清晰性

評鑑報告必須描述受評對象及其背景、評鑑目的、程序與發現，俾使聽取人能輕易了解評鑑內容、緣故、發現、結論和建議。

U6 報告的及時性和傳播

評鑑發現和報告必須傳播給想知道的使用者，俾使他們可得以及時使用和運用這些發現。

U7 評鑑的影響

評鑑的規劃與實施必須使當事人有跡可循。

可行性標準 (Feasibility Standards):

可行性標準是要確保評鑑是務實的、謹慎的、圓融的和簡約的。這些標準如下：

F1 務實的程序

評鑑的程序必須符合實際，俾使半途中斷的可能性減至最小，並能獲得所需的資訊。

F2 政治上的可行性

評鑑的規畫與實施必須考量各種利益團體不同立場的期待，俾便得到他們的合作，以及避免或消除他們對評鑑運作的抵制或對評鑑結果的偏見或誤用。

F3 成本效益

評鑑應該是有效率的，且產生相當價值的資訊，使資源的花費是有用的。

適切性標準(Propriety Standards)：

適切性標準是要確保評鑑是合法、合倫理、以及符合評鑑涉入者和受評鑑結果影響者的福祉。這些標準如下：

P1.服務的導向

評鑑必須設計為幫助組織陳述或有效的服務全部參與者的需求。

P2.正式的協議

評鑑的參與人員的義務（工作內容、方法、人員、時間）必須達成書面協議，俾使這些人在協議的所有條件下履行義務或正式重新協商。

P3.受評者的權利

評鑑的設計與實施必須尊重和保護受評者的權益與福祉。

P4.人際的互動

評鑑應尊重人性的尊嚴和價值，在他們與其他人之評鑑的互動時，不致於使參與者受到威脅或傷害。

P5.完全和公平的評估

評鑑必須完全公平地評估和紀錄被評鑑方案的優點和缺點，俾使優點被延續而問題被陳述。

P6.發現的公佈

評鑑的正式參與者應確定評鑑的發現與限制的完整資料，提供給易受評鑑影響的人，而其他人具有了解結果的合法權利。

P7.利益的衝突

利益衝突必須公開和誠懇的處理，俾使這種衝突不致危及評鑑的程序與結果。

P8.會計的職責

評鑑者對資源的分配與開銷必須反應出良好的績效程序，而且是謹慎和合乎道德責任。

4.精確性標準(Accuracy Standards)：

精確性標準是要確保評鑑在技術上顯示和傳達受評對象特點的適當資訊，已決定它的價值或優點。這些標準包括：

A1 方案說明

被評的方案應詳細的正確的描述說明，使方案得以清晰地被確認。

A2 背景分析

方案存在的背景必須詳細評估，俾使確認其受評對象的可能影響。

A3 目的與程序的描述

評鑑之目的與程序必須詳細追蹤與描述，俾使它們得以被確認與評估。

A4 有辯護力的資訊來源

資訊來源必須詳盡描述，俾使資訊精確性得已被評估。

A5 有效的評鑑

收集資訊的工具和程序必須經過選擇或發展，俾使在施行後能確保其解釋合理，對特定用途是有效的。

A6 可靠得評鑑

收集資訊的工具和程序必須經過選擇或發展，俾使在施行後能確保其所獲得的資訊，對特定的用途具有充分的可靠性。

A7 系統的資料

評鑑資料的收集、處理和報導必須加以檢討或改進，俾使評鑑結果沒有缺失。

A8 量化資訊的分析

評鑑的量化資料必須適切和系統的分析，以確保有關的解釋能獲得支持。

A9 質化資訊的分析

評鑑的質化資料必須適切和系統的分析，以確保有關的解釋能獲得支持。

A10 證明正當的結論

評鑑所獲得的結論必須明確地驗證，以便聽取人得以進行評估。

A11 無偏差的報導

評鑑程序必須提供安全保護，以免評鑑發現和報告受到評鑑人員個人感覺和偏見所扭曲。

A12 後設評鑑

貳、實例比較

我國國內五個學校層級後設評鑑研究之比較如表 3-3-1：

表 3-3-1 國內五個學校層級後設評鑑研究之比較

作者 研究名稱 比較項目	蘇錦麗 (民84)	曾淑惠 (民85)	林邵仁 (89)	游家政 (民83)	賴志峰 (民86)
研究學校層級	大學	專科學校	高級中學	國民小學	幼稚園
研究方法	1.文獻分析法 2.問卷調查法 3.觀察與訪談法 4.文件分析法	1.觀察法 2.訪問法 3.問卷調查法 4.文件分析法 5.專家座談法	1.文獻分析法 2.得懷術	1.文獻分析法 2.得懷術 3.專家座談法	1.文獻分析法 2.檔案文件分析法 3.問卷調查法
後設評鑑標準之主要架構	CIPP 評鑑模式 分背景評鑑 投入評鑑 歷程評鑑 產出評鑑 兼以適當性 周延性 效能性 回應性檢視	CIPP 評鑑模式 分背景評鑑 投入評鑑 歷程評鑑 產出評鑑	CIPP 評鑑模式 分背景評鑑 投入評鑑 歷程評鑑 產出評鑑	評鑑研究會(ERS) 之〈方案評鑑標準〉 (standards for program evaluation)	教育評鑑標準聯合委員會之〈方案評鑑標準〉 (the program evaluation standards)
後設評鑑標準之內容	4 個評鑑類別 12 個評鑑重點 85 個評鑑問題	4 個評鑑類別 12 個評鑑項目 85 項評鑑標準	5 個一級指標 63 個二級指標	6 個評鑑層面 77 項評鑑標準	4 個評鑑類別 30 項評鑑標準
研究屬性	質，量並呈	質，量並呈	質，量並呈	質，量並呈	質，量並呈