

教育研究集刊

第五十一輯第二期 2005年6月 頁31-65

後設評鑑研究：以 2001 年大學校院 實施自我評鑑計畫成果報告書為例

黃曙東 蘇錦麗

摘要

本研究旨在建構一套具有信度與效度的後設評鑑檢核表，以分析 2001 年 9 所實施自我評鑑計畫大學校院的成果報告書，判斷其優缺點，進而推論 34 所母群體大學的整體報告書品質，並瞭解 34 所大學實施自我評鑑計畫之現況。

本研究採用文件分析及專家檢核方法。首先，分析七年內國內外 22 篇後設評鑑研究報告。其次，依據美國線上評鑑資源圖書館（OERL）所建構的品質結論報告準則，並經專家審核與研究者試評，修正完成方案評鑑報告檢核表，包括執行摘要、受評方案描述、評鑑概述、評鑑設計、分析程序及結果與建議共六大項 27 個品質準則項目、三等量表，以及每項目之量的判斷依據；再者，由兩位研究者據之檢核，以判斷 9 所大學校院成果報告書的優缺點，進而推論 34 所大學整體報告書的品質。最後，提出結論與建議。

關鍵詞：大學評鑑、後設評鑑、評鑑報告、檢核表、內部評鑑／自我評鑑

黃曙東，東南技術學院工業管理學系講師

蘇錦麗，國立新竹師範學院初等教育學系教授

電子郵件為：sthuang@mail.tnit.edu.tw; Sujl@tpts5.seed.net.tw

投稿日期：2004 年 7 月 12 日；修正日期：2004 年 10 月 1 日；採用日期：2005 年 1 月 7 日

Bulletin of Educational Research
June, 2005, Vol. 51 No. 2 pp. 31-65

A Metaevaluation Study of Self-evaluation for Taiwan's Colleges and Universities

Shu-Tung Huang Jin-Li Su

Abstract

In 2001, 34 of Taiwan's colleges and universities participated in a Self-Evaluation Project (SEP), and this study set out to analyze the effectiveness of the SEP. More specifically, it set out to understand the quality (or accuracy) of the self-evaluations reported by these institutions and to determine certain problematic aspects of the current implementation of the SEP. The research methods used include content analysis and experts' reviews. The authors asked 3 experts to revise the Quality Criteria for Reports checklist developed by the On-Line Evaluation Resource Library and, in order to construct proper validity and reliability for this checklist, randomly selected 9 colleges and universities as samples. This metaevaluation study indicated both the strengths and weaknesses of the system of self-evaluation used by these colleges and universities. In this paper the authors report these findings and also give some suggestions for a more effective self-evaluation process in the future.

Keywords: university evaluation, metaevaluation, evaluation report, checklist, internal evaluation/self-evaluation

Shu-Tung Huang, Lecturer, Department of Industrial Management, Tungnan Institute of Technology

Jin-Li Su, Professor, Department of Elementary Education, National Hsinchu Teachers College
E-mail: sthuang@mail.tnit.edu.tw; Sujl@tpts5.seed.net.tw

Manuscript received: July 12, 2004; Modified: Oct. 1, 2004; Accepted: Jan. 7, 2005

壹、緒 論

一、研究動機

大學評鑑基於實施評鑑驅動力 (motive) 的不同, 可分為內部評鑑 (internal evaluation) 與外部評鑑 (external evaluation) 兩種; 所謂內部評鑑, 係大學基於自主、自治的精神, 自動自發的原則, 發揮自我決定與自我批判的能力, 由大學本身或其指定團體所規劃和執行的評鑑活動, 目的在於藉由建構一套自我品質管制機制以持續改進大學教育品質, 進而證明其教育作為符合社會期望目標, 因而被一般學者認為是有效提昇機構教育品質之主要關鍵; 至於外部評鑑則常基於外在認證或認可的要求下, 由大學之外的團體或小組來執行評鑑活動, 通常以符合績效責任要求為其主要目的 (蘇錦麗, 1997: 21)。

我國大學評鑑工作自 1975 年實施以來, 均由教育部全權負責辦理, 此係屬「外部評鑑」。自 2001 年始, 教育部為有效提高教育品質, 特編列相關預算, 並訂出「2001 年大學校院實施自我評鑑計畫補助申請要點」, 藉由經費補助以鼓勵各大學建立自我評鑑機制, 此係屬「內部評鑑」, 亦是我國首次由教育部補助大學校院實施自我評鑑¹計畫。根據教育部高教司 (2002) 資料, 教育部共審訂 34 校予以補助, 各校補助金額自 80 萬至 20 萬不等, 總計共補助 1890 餘萬元。

根據研究者詢問教育部業務單位相關人員的結果, 教育部並無在此計畫之始, 對 34 所學校召開相關的評鑑說明會、工作坊。在計畫實施中, 亦無委託學術單位提供學校諮詢或輔導。在計畫結束後, 各校除向教育部辦理經費報銷外, 亦需繳交一份成果報告書, 並將其放置於高教司網站上, 以供社會大眾查閱。除若干繼續申請 2002 年自評經費補助學校外, 教育部並未對其他學校之報告內容進行審核, 且亦未召開檢討會或委託學術單位來進行整體自評實施成效評估。事實上, 基於各校經費補助之績效責任及主管機構督導責任之要求, 教育部實有必要瞭解

¹ 即內部評鑑, 在本研究中, 兩者實隱含等同意義, 故在本文以下章節中, 兩者皆交互使用。

各校自我評鑑計畫之實施情形與整體成效，此係後設評鑑（metaevaluation），即「評鑑的評鑑」之概念。

一般而言，後設評鑑之對象、範圍、種類及方法頗為廣泛（Scriven, 1991; Stufflebeam, 2000b），以本計畫為例，其探討範圍可大至由教育部遴派一訪視小組至各校瞭解其實施現況，並評估計畫之整體成效，亦可小至審核各校經費是否依照原先預算之編列，以及其運用情況。在衡諸時間、資源之有限性，以及資料取得性之條件下，本研究擬自教育部網站下載各校繳交教育部之成果報告書，針對其內容與格式，加以分析，藉此瞭解各校實施自我評鑑計畫之若干現況與成效。此外，本研究欲採用何種工具，以檢核評鑑報告書？又如何建構一套具有信效度的工具？以目前歐美先進國家針對教育方案與學校校務之後設評鑑而言，有四類較為專家學者採用之評鑑標準或檢核表，包括由美國評鑑協會（American Evaluation Association, AEA, 1994）、線上評鑑資料圖書館（On-line Evaluation Resources Library, OERL, 2001）、Stufflebeam（1999）及 Scriven（2004）所分別發展者。其意義、目的、內容及應用為何？何者可以修正後適用於本研究？或者本研究必須另行建構之？上述這些議題都值得深入探討，爰為本研究之動機。

二、研究目的、方法及步驟

基於上述動機，本研究旨在建構一套具有信度與效度的後設評鑑檢核表，並依據檢核表的準則以立意抽樣方式，分析 2001 年 9 所實施自我評鑑計畫大學校院的成果報告書，判斷報告書的品質，進而推論 34 所母群體大學的整體報告書品質，並瞭解 34 所大學實施自我評鑑計畫之現況。研究結果除供教育行政主管機關未來擬定大學相關評鑑制度時之參考外，亦能作為相關後續研究之參考。

為達成本研究目的，本研究採用文件分析及專家檢核方法。首先透過文獻探討，瞭解後設評鑑與檢核表發展之理論與實務，並運用文件分析法分析近七年國內外 22 篇後設評鑑研究報告。其次綜合文獻探討與文件分析所得，並依據 OERL 所建構的品質結論報告準則，經專家審核與研究者試評結果，修正完成方案評鑑報告檢核表，包括執行摘要、受評方案描述、評鑑概述、評鑑設計、分析程序及結果與建議共六大項 27 個品質準則項目、三等量表，以及每項目之量的判斷依

據；其後由本研究兩位研究者依據所建構檢核表的準則進行 9 所大學校院成果報告書的檢核，以判斷其優缺點，進而推論整體報告書的品質。最後提出結論，並自教育部、大學校院及後續研究等三方面提出建議。

貳、文獻探討與文件分析

本節針對後設評鑑之意義與目的、後設評鑑的實施種類與途徑、檢核表之意義及種類等，進行文獻探討，以及針對國內外相關後設評鑑的研究進行文件分析。

一、後設評鑑之意義與目的

Scriven (2001) 指出，在目前 20 餘種深具效用性之評鑑種類中，學科內評鑑 (intradisciplinary evaluation) 與後設評鑑是格外重要的兩項評鑑領域。其中，後設評鑑係跨越學科領域的，是維持評鑑真實性，且顯示評鑑是可以自我參照 (self-referent) 的。Scriven (1967) 也是首位提出以「後設評鑑」概念來評量教育成果評鑑制度，並將其定義為：「一套評量評鑑、評鑑制度或評鑑工具的流程」，簡言之即「評鑑的評鑑」(Stufflebeam, 2001b)。之後，Stufflebeam (1974) 針對美國俄亥俄州立大學評鑑中心於 1963 到 1973 年實施的評鑑，所遭遇如概念、社經、契約／合法性、技能、管理、倫理道德、效用性考慮等議題，提出一套後設評鑑邏輯架構，用以瞭解評鑑過程之優缺點。

Chelimsky (1987) 在評鑑公共政策方案時，將後設評鑑定義為一種重新分析一個以上評鑑之方法，並用以決定在公共政策或方案評鑑中，評鑑人員與利害關係人可以學習到什麼，並認為後設評鑑之目的在檢驗原級評鑑 (primary evaluation) 設計、資料蒐集及分析方法之有效性；之後，教育評鑑標準聯合委員會 (The Joint Committee on Standards for Educational Evaluation, JCSEE)，在訂定方案評鑑的效用性 (utility)、可行性 (feasibility)、適切性 (propriety) 及精確性 (accuracy) 四類標準時，亦將後設評鑑列入精確性標準要項之一，並界定後設評鑑之定義為：「評鑑本身應該依據其他三類標準和其他適切的標準，進行形成性與總結性的評鑑，使實施過程獲得適當的指引，利害關係人也能仔細地檢視評鑑的

優缺點」(Joint Committee, 1994: 185)。

最後，Stufflebeam (2001b) 指出後設評鑑的操作性定義為：「針對評鑑的效用性、可行性、適切性及精確性、評鑑制度化本質、評鑑人員能力行爲 (competent conduct)、廉潔／誠實 (integrity/honesty)、尊重他人，以及社會責任 (social responsibility)，進行敘述性與判斷性資訊之描述、獲取及運用的過程，目的在正確引導評鑑實施並報告評鑑的優缺點。」換言之，後設評鑑的運用可以有效改善評鑑實施 (Scriven, 1967, 2001; Stufflebeam, 1974, 2001b)，也是決定評鑑是否符合專業化之主要關鍵 (Worthen, Sanders, & Fitzpatrick, 1997)。

二、後設評鑑實施種類與途徑

Schwandt 與 Halpern (1988) 的研究指出，後設評鑑有兩大研究途徑：分析原級評鑑的方法論是否適當以及檢視評鑑結論報告是否良好與正確；其中良好與正確的結論報告應有以下特質：

1. 中肯性：資料、證據應與評鑑目的有直接相關；
2. 信度：具備值得信賴之證據，且可以作為價值判斷之依據；
3. 充份性：在合理誤差內，價值判斷應有充足的證據；
4. 代表性：樣本的資料蒐集可以合理推論到受考驗的母群體；
5. 時效性：應有效衡量資料蒐集之成本與時間的限制。

Scriven (2004) 則指出可透過下列途徑進行後設評鑑之資訊蒐集，以作為檢驗原級評鑑之優缺點、限制及運用：

1. 運用關鍵評鑑檢核表 (Key Evaluation Checklist, KEC)，來檢核評鑑過程；
2. 使用後設評鑑專門檢核表 (Special Metaevaluation Checklist)；
3. 重複實施評鑑並比較其結果產出；
4. 以不同評鑑方法論實施評鑑，並比較其結果產出；
5. 對照 AEA 方案評鑑標準。

其中與本研究相關之檢核表的主要內涵，將在下節敘述。

三、後設評鑑檢核表之意義及種類

評鑑檢核表是快速地蒐集有關評鑑實施資料的一種方法 (Talisman, 1999；引自曾淑惠, 2002：200)，同時在評鑑使用上亦是一種有用的方法 (Stufflebeam, 2001a)。而後設評鑑檢核表則是一種可以快速、容易、精準地蒐集資訊，引導評鑑程序，並對照評鑑專業標準，最後整體判斷原級評鑑之優缺點及價值的有效方法。

茲針對被廣泛引用且與本研究相關的四類檢核表，分述其內容如下：

(一)Stufflebeam (1999) 設計之方案後設評鑑檢核表 (Program Evaluations Metaevaluation Checklist)

本檢核表乃依據方案評鑑標準 (Program Evaluation Standards) 所建構 (參見表 1)，並依四類標準內分項檢核要點之數目 (6 或 10 條)，區分為長、短條文兩種版本。以長條文版為例，每個檢核要點均為 1 分，每一分項總分為 10 分，若評定 9-10 分係表示傑出，7-8 分係十分良好，5-6 分為良好，3-4 分為普通，0-2 分則為不佳。在分項項目中以 P1 (服務導向)、A5 (有效資訊)、A10 (證明為正當的結論) 及 A11 (公平客觀的報導) 格外重要，若這些項目評定為不佳者，該評鑑係屬失敗之評鑑實施。

(二)線上評鑑資源圖書館 (OERL, 2001) 發展之評鑑資源品質準則 (Quality Criteria for Evaluation Resources) 檢核表

OERL 在國家科學基金會 (National Science Foundation, NSF) 的經費贊助下，根據教育評鑑標準聯合委員會 (JCSEE) 發展的標準，設計了一套評鑑資源之品質準則檢核表，包括計畫、工具及結論報告等三方面的準則，每方面的準則又包括若干項目，如表 2。

(三)關鍵評鑑檢核表 (Scriven, 2004)

本檢核表係 Scriven 於 1991 年開始發展，並經多次修正後定稿，內容包括四大部分：

1. 序言：包括執行摘要、前言、評鑑方法論等。
2. 評鑑基礎：包括背景與脈絡、描述與定義、消費者、資源、價值等。
3. 評鑑分項：包括過程、結果、成本、可概括性、比較性等。

表 1 Stufflebeam 方案後設評鑑檢核表

效用性標準	U1 利害關係人之確認。U2 評鑑人員的可靠性。 U3 資訊範疇與選擇。U4 評鑑價值之確認。U5 報告清晰性。 U6 報告的時效性與公布。U7 評鑑影響。
可行性標準	F1 務實的程序。F2 考量政治效應。F3 成本有效性。
適切性標準	P1 服務導向。P2 正式契約。P3 受評對象的權利。P4 人際互動。 P5 完整且公平之評量。P6 評鑑發現之公開。P7 利益之衝突。 P8 財務之職責。
精確性標準	A1 方案的文件說明。A2 背景分析。A3 目的與程序的描述。 A4 有辯證能力的資訊來源。A5 有效資訊。A6 可靠資訊。 A7 系統化之資訊。A8 量化資訊的分析。A9 質性資訊的分析。 A10 證明為正當的結論。A11 公平客觀的報導。A12 後設評鑑。

資料來源：作者整理。

表 2 OERL 評鑑資源之品質準則檢核表內容

品質計畫準則	1.方案描述：受評方案特徵，方案參與者，閱讀評鑑報告者及其他利害關係人，方案背景。 2.評鑑概述：評鑑意旨／目的，評鑑問題，評鑑人員的可靠性，利害關係人之參與。 3.評鑑設計：評鑑方法論之取向，資訊來源與取樣，評鑑工具，資訊蒐集程序與時程，後設評鑑。 4.分析程序：量化分析，質性分析。
品質工具準則	1.設計：符合規劃意旨，有原則的評量設計，項目建構與工具發展，品質保證。 2.工具品質：效度，信度，評量的誤差。 3.工具運用：工具資訊之準備，報導測驗工具建構之實際狀況，工具可使用性。
品質結論報告準則	1.執行摘要。 2.方案描述：受評方案特徵，方案參與者，閱讀評鑑報告者及其他利害關係人，方案背景。 3.評鑑概述：評鑑意旨／目的，評鑑問題，評鑑人員的可靠性，利害關係人之參與。 4.評鑑設計：評鑑方法論之取向，資訊來源與取樣，評鑑工具，資訊蒐集程序與時程，後設評鑑。 5.分析程序：量化分析，質性分析。 6.結果與建議：解釋與結論，建議事項，利害關係人的檢核與利用。

資料來源：作者整理。

4. 結論：如總體重要性、建議與解釋、責任及正當性、報告與支援、後設評鑑等。

(四)Stufflebeam (1999) 發展之評鑑人員評鑑指導原則之檢核表草案 (Guiding Principles Checklist for Evaluating Evaluators Draft)

本草案 (參見表 3) 係依據 AEA 評鑑人員指導原則 (Guiding Principles for Evaluators) 而設計, 旨在協助評鑑人員能運用此原則來規劃與引導評鑑, 以及評量評鑑成效。

表 3 Stufflebeam 評鑑人員評鑑指導原則之檢核表草案

原級評鑑之特性	包括原級評鑑之名稱、委託人、經費提供者、經費成本、時間流程、評鑑人員、利害關係團體、其他評鑑對象、意圖、評鑑關鍵問題、主要方法論、優缺點, 以及關鍵報告等。
相關後設評鑑之關鍵項目	包括後設評鑑名稱、委託者、經費提供者、其他後設評鑑報告閱讀者、後設評鑑目的、主要服務/報告提供、評鑑關鍵問題、主要方法論、期程、契約名稱、後設評鑑者與原級評鑑之關係, 以及後設評鑑者與原級評鑑對象之關係。
評斷原級評鑑之主要資料來源	資料包括評鑑契約、提案報告、期中與期末報告、特殊附錄表件、報紙期刊、會議資料、回應文件、新聞稿、文件摘要, 及出版物等; 並可透過實地訪談及電話訪問來蒐集資訊。
五大標準:	
有系統的資料蒐集	<ol style="list-style-type: none"> (1) 為了蒐集精確可靠資訊, 應具備量化資料分析之技術標準。 (2) 為了蒐集精準可靠資訊, 應具備質性資料分析之技術標準。 (3) 能保證評鑑委託人瞭解評鑑問題之限制與優點。 (4) 能保證各利害關係人瞭解評鑑問題之限制與優點。 (5) 保證評鑑委託人瞭解評鑑方法之限制與優點。 (6) 保證各利害關係人瞭解評鑑方法之限制與優點。 (7) 提供足夠詳細之評鑑方法與步驟以供瞭解、批判及解釋評鑑之結果。 (8) 在結論報告中, 清楚說明評鑑過程與結果之各種限制。 (9) 充份討論各項足以影響評鑑結果解釋之評鑑價值、假設、理論、方法, 以及分析與結論。

表 3 Stufflebeam 評鑑人員評鑑指導原則之檢核表草案 (續)

能力	<ul style="list-style-type: none">(1)展現具備足以建立及維護評鑑可信度之教育理念、人格特質及豐富經驗。(2)展現足以確保評鑑品質之專業能力與技巧。(3)對各評鑑參與人與受評單位應清楚說明任何可能之重大評鑑限制。(4)具備評鑑工作所需之特殊專案能力與資格。(5)透過課程、研討會、工作坊、自我學習，以及與有評鑑工作經驗者一起工作學習等，評鑑人員能不斷提昇自我評鑑能力。
廉潔與誠實	<ul style="list-style-type: none">(1)能反覆不斷同評鑑委託人與利害關係人討論並確定所有評鑑活動之重要內容。(2)能真誠與委託人和利害關係人協商評鑑問題與限制、評鑑方法、結果限制、成本、責任區分以及如何適切使用結果等，並紀錄協商過程與協定內容。(3)事先提醒委託人各種可能產生評鑑資料與結論誤導或誤判之可能。(4)與委託人共同防止資訊偏差及結果誤解之情形。(5)能順利完成評鑑之過程並協助消除各項評鑑之障礙。(6)若有影響評鑑工作不能廉潔與誠實的進行時，應舉證並採取可行之導正行動。(7)在改變評鑑計畫前，應與委託人協商並加以紀錄。(8)適當告知委託人及重要利害關係人任何可能影響評鑑範疇與結果之重大改變。(9)告知並記錄對核准作評鑑計畫重大改變之理由以及其可能產生之正向及負向影響。(10)告知委託人及評鑑參與人任何在評鑑實施與結果所獲得之各種利益。(11)記錄各項保證利益衝突能有效管控之步驟。(12)結論報告能鑑定各種利益衝突之方法、程度及管控，使該衝突無法危及評鑑工作之進行。(13)能公開、真誠的報導評鑑程序、資料、結果及資料限制等。(14)協助評鑑報告之讀者能對評鑑報告進行正確解釋與運用，並避免誤解與誤用。(15)在合理限制下，嘗試防止並修正任何他人重大評鑑工作之誤用。(16)在評鑑報告中，應公布申請評鑑團體之名稱及其各項財務資源。

表 3 Stufflebeam 評鑑人員評鑑指導原則之檢核表草案 (續)

尊重他人	<ol style="list-style-type: none">(1)防止或降低任何違反良善專案評鑑人員之倫理規範與標準。(2)對違反專案評鑑人員倫理規範，能採取安全保護之行動，如將評鑑計畫交由公正第三者檢查評定。(3)在保障人權議題上，應符合任何聯邦、州之法規。(4)尊重評鑑參與者各種非正式協商所同意之精神與規定，如告知可能提供資訊者是否願意參與資訊提供或退出評鑑之權力。(5)清楚且真誠的告知參與者相關保密與不具名之範圍與限制。(6)預先仔細評量並決定評鑑潛在利益已大於不利時，再執行評鑑工作。(7)在執行評鑑工作過程中，應充份尊重利益關係人之廉潔與自我價值。(8)避免引起不必要傷害；報導負向結論應公正、持平且真誠。(9)採取任何擴大評鑑利益及避免不必要傷害之步驟。(10)提昇評鑑之社會公平性價值，如提供評鑑貢獻者資訊回饋與其他適當之利益。(11)能充份展現尊重對評鑑參與者相關文化、宗教、性別、殘障、年齡、人種等議題之看法。
社會與公眾福祉之責任	<ol style="list-style-type: none">(1)評鑑計畫與結論報告中，應包括並展現不同利益關係人與評鑑方案本身之利益與遠景。(2)針對評鑑計畫與結論報告中，若省略重要評鑑價值與遠景，應提出辯駁理由。(3)針對評鑑計畫與結論報告中，應充份考慮評鑑方案之各種假設、暗示及負向影響。(4)提供給有知權利之評鑑報告閱讀者，有關評鑑資料使用之途徑。(5)清楚、簡單且精確地報告評鑑過程與結果。(6)向不同有知權利的評鑑報告閱讀者，提供各種不同適切的結論報導。(7)有效及快速分送評鑑報告給有知權利的評鑑報告閱讀者。(8)告知各利害關係人，有關結論報告提供給評鑑報告閱讀者及其使用途徑之情形。(9)對評鑑結果與發現應真實呈現。(10)如有對委託人或最重要利害關係團體行使差別對待時，應公平、清楚且負責任強調該利害關係團體之利益與需要。(11)能有效指出，並強調若有利害衝突發生時所引起無法真誠報告之困難。(12)支持社會公眾權利，向評鑑結果學習並從中獲益。(13)分析並揭露社會大眾福祉及利益關係人之利益。

資料來源：作者整理。

四、文件分析國內外相關後設評鑑之研究

研究者自教育研究資訊中心 (ERIC)、國際論文摘要 (Dissertation Abstracts International)、網際網路 (Yahoo, Google, Metacrawler 等)、專業評鑑期刊、國際性評鑑研討會，以及各國評鑑組織之網站等，輸入搜尋名詞，包括：後設評鑑與後設評鑑報告 (metaevaluation report)，文獻搜尋時間在 2004 年 1 月至 3 月，共搜尋到 42 篇近 7 年來有關後設評鑑之實證研究報告，其中 15 篇符合後設評鑑之意義與做法。此外連同我國各級各類校務或學門後設評鑑之 7 篇報告，共計 22 篇；由於篇幅限制，本研究僅分析 22 篇報告內容之研究者與出版時間、後設評鑑對象／名稱、後設評鑑功能、後設評鑑種類、實施途徑／取向，以及資料蒐集工具／方法等六方面；其中表內之後設評鑑功能、種類係指發動後設評鑑時間點及發動者之身分，實施途徑／取向意指依據何類標準或指標來進行原級評鑑品質之判斷，至於資料蒐集工具／方法乃代表用以評斷該原級評鑑品質、優缺點之資料來源。分析結果如表 4，並將分析結論摘錄於第五章結論與建議之章節中。

表 4 國內外後設評鑑實證研究分析表

研究者與出版時間	後設評鑑對象／名稱	後設評鑑功能	後設評鑑種類	實施途徑／取向	資料蒐集工具／方法
Stufflebeam (2000b)	為美國而教 (The Teach for America) 中教師表現評估系統 (Performance Assessment System)	總結性評鑑	內部	依據 AEA 人員評鑑系統 (Personal Evaluation Systems)	建構一套完整實施流程；量化與質性
Stake 與 Davis (1999)	伊利諾大學教學研究與課程評鑑中心 (Center for Instructional Research and Curriculum Evaluation, CIRCE) 的讀者加強寫作 (Reader Focused Writing, RFW) 之方案評鑑	形成性與總結性評鑑	內部	依據七項重大及九項次要之評鑑關注議題	質性 (文件分析、訪談，專家會議，召開後設評鑑法庭會議)

表 4 國內外後設評鑑實證研究分析表（續）

Grasso (1999)	CIRCE 的 RFW 方案評鑑	總結性評鑑	外部	依據 AEA 方案評鑑標準及評鑑人員指導原則	評鑑報告之內容分析
McKinley (1999)	WMU 密西根公立學校學術 (Michigan Public School Academy, PSA) 方案評鑑	總結性評鑑	內部	依據 AEA 方案評鑑標準	評鑑報告內容分析、檢核表檢核
Sanders (1999)	Pierre et al. 多元孩童發展方案 (Comprehensive Child Development Program, CCDP) 評鑑	總結性評鑑	外部	依據 AEA 方案評鑑標準及評鑑人員指導原則	評鑑報告之內容分析
Scriven (2000b)	Esbensen & Osgood 的甘氏兒童發展遲緩教育與訓練 (Gang Resistance Education and Training, GREAT) 方案評鑑	總結性評鑑	外部	依據 KEC 檢核表	量化/質性
Bjornkilde 與 Bason (2000)	1993 年至 1997 年丹麥大學評鑑	總結性評鑑	外部	依據評鑑重大議題	量化 (問卷調查)、質性 (訪談、個案研究)
Gunnarsson 與 Rup precht-Hjort (2000)	21 個瑞典基金方案評鑑	總結性評鑑	外部	依據評鑑重大議題	量化 (成本調查/技術輔助分析、問卷調查) 質性 (深度訪談、外部專家判斷)
Swedish International Development Co-Operation Agency, Sida (2001)	NGO 在高棉推動民主與人權方案 (Diakonia & Forum Syd's) 之評鑑	形成性與總結性評鑑	外部	針對重大評鑑議題及利害關係人參與情形	質性 (訪談、資料內容分析)

表 4 國內外後設評鑑實證研究分析表（續）

Bustelo (2002)	1995-1999 年西班牙兩性平等方案評鑑中共 11 個評鑑過程	總結性評鑑	外部	依據評鑑品質管控目的以及認識評鑑過程為目的等二大途徑	專家判斷
Migotsky 與 Stake (2001)	WMU 之進階科技教育方案 (Advanced Technological Education Program, ATE)	總結性評鑑	外部	依據 14 項重大議題	量化 (成本調查 / 技術輔助分析、問卷調查) 質性 (訪談、專家判斷)
Hartmann 與 Loizides (2001)	WMU 之 ATE 方案評鑑中網路調查系統部分	形成性評鑑	內部	依據 AEA 方案評鑑標準及特殊網路調查之專業標準	檢核表專家判斷檢核
Stufflebeam (2000b)	評鑑模式 (Evaluation Model) 書中 22 類評鑑模式	總結性評鑑	外部	針對評鑑主要目的、服務對象、問題來源、問題特質、評鑑方法論、優點等議題	內容分析、專家判斷
Scott-Little、Hamann 與 Jurs (2002)	1996-2001 年美國各區域之學後方案 (After School Program) 評鑑	總結性評鑑	外部	依據 AEA 方案評鑑標準	質性 (評鑑報告之內容分析) 量化 (後設分析)
Stufflebeam、Gullickson、與 Wingate (2002)	夏威夷岡薩約多基金會之 (Ke Aka Ho Ona) 方案評鑑	形成性與總結性評鑑	內部	依據 AEA 方案評鑑之後設評鑑檢核表	檢核表檢核
黃政傑、李隆盛、游家政 (1994)	81、82 學年度臺北市國小總務工作評鑑	總結性評鑑	外部	依據美國評鑑研究學會之方案評鑑標準，並修正為評鑑規劃、設計、實施、資料分析、結果利用等六層面	專家座談，實地評鑑觀察

表 4 國內外後設評鑑實證研究分析表（續）

蘇錦麗 (1995)	82 學年度「中國電機工程學會」、「中華民國管理科學學會」、「中國機械工程學會」之大學學門評鑑	總結性評鑑	外部	依教育評鑑標準聯合委員會之教育評鑑標準	量化（問卷調查） 質性（觀察、訪談、內容分析）
曾淑惠 (1996)	82 學年度至 84 學年度專科學校評鑑	形成性與總結性評鑑	外部	依據 CIPP 模式，再發展出 85 項評鑑標準	量化（問卷調查） 質性（觀察、訪談、文件分析）
賴志峰 (1997)	82 學年度至 84 四學年度臺北市公私立幼稚園評鑑	總結性評鑑	外部	依據 AEA 方案評鑑標準，再行部分修正	量化（問卷調查） 質性（文件分析）
胡悅倫與陳漢強 (1998)	86 學年度大學綜合評鑑	總結性評鑑	外部	參考 AEA 方案評鑑標準，修正發展 30 個題項	量化（問卷調查）
蘇錦麗、楊瑩、王偉中、呂鴻德、詹惠雪 (2000)	86 學年度大學綜合評鑑試辦計畫	總結性評鑑	外部	依據黃政傑等人所發展之評鑑標準	質性（觀察、訪談、文件分析） 量化（問卷調查）
蘇慧雯 (2002)	90、91 年度臺北市公、私立幼稚園評鑑	總結性評鑑	外部	參考 AEA 方案評鑑標準，修正發展「臺北市公私立幼稚園評鑑問卷調查表」	量化（問卷調查）

資料來源：作者整理。

參、實證研究之設計與實施

本節旨在說明本研究的設計與實施，包括樣本與取樣、檢核表工具建構、以及實施程序與資料分析三小節。

一、樣本與取樣

2001 年實施自我評鑑計畫之 34 所大學校院的成果報告書為本研究之母群體，在綜合 34 所大學所在區域、公私立別、學校性質，以及評鑑對象等因素之平衡性後，由於時間限制，本研究決定採取立意取樣，共抽樣 9 所大學（如表 5）進行專家檢核。樣本占母群體比例為 29.01%，同時在取樣時已注意樣本間之可能差異性變化，故檢核結果應可推論母群體整體報告書的品質。

表 5 九所抽樣大學之分布

公私立別	學校性質	所在區域	自我評鑑對象／範疇
公立：5 所 私立：4 所	綜合大學：7 所 單科大學：2 所	北區（新竹以北）：5 所 中區（苗栗至雲林）：1 所 南區（嘉義至屏東）：2 所 東區與離島：1 所	校務綜合評鑑：2 所 學門評鑑：4 所 方案評鑑：1 所 校務綜合評鑑與學門評鑑： 2 所

資料來源：作者整理。

二、檢核表工具之建構

如同上述，本研究在探討四類被歐美國家廣泛引用的檢核表後，發現 OERL（2001）所發展品質評鑑結論報告檢核表最能符合本研究目的之達成，惟衡諸中外情境與本研究性質差異等因素後，研究者仍先針對檢核表進行若干修正。此外，為建立研究工具之信效度，研究者首先將修正後之檢核表送請三位教育行政與評鑑專家學者審核，再根據其意見再次修正。接著，兩位研究者挑選兩所大學之成果報告書加以試評，再根據試評結果充份討論後修正檢核表，最後確定檢核表在內容與格式方面皆屬完整、可行後，始完成定稿如表 6。

值得注意的是，本檢核表除採用美國評鑑專業社群通用之三等量表（符合、部分符合、不符合，以及不適用或無法判斷）外，亦增列「量的判斷依據」一項，意即針對每一品質準則項目的三等量表分別所需達成的程度，加以規劃。例如：

在「執行摘要」此欄位中，3等（符合）係指有完整摘要，2等（部分符合）有粗略摘要，1等（不符合）則無摘要。此設計除有利於檢核者的判斷外，亦有利於提昇檢核者之間的一致性。

表 6 方案評鑑報告檢核表

方案評鑑報告名稱：						
檢核人姓名（簽名）：						
檢核日期：		年	月	日		
品質準則項目		量表與量的判斷依據		符合 (3)	部分符合 (2)	不符合 (1) 不適用或 (NA) 無法判斷
1.執行摘要	1-1 執行摘要應提供方案評鑑報告中的基本資訊，包括清楚概述評鑑意旨／目的，評鑑方案之目的、實施及衝擊，以及由評鑑結果引申出之結論與建議。					3：有完整摘要 2：有粗略摘要 1：無摘要
2.受評方案描述	2-1 受評方案之特徵 應該清楚描繪下列受評方案之特徵： *方案之目的（外顯與內隱）與目標； *為達成目的所設計之主要活動內容； *方案之所在地與執行場所； *方案之期程； *為完成方案所使用之資源； *短程與長程之預期結果。					3：達成 4 項以上 2：達成 2-3 項 1：達成 1 項以下
		若執行場所超過一處以上，評鑑中應詳述每一評鑑場所的地點，以及場所間可能產生之預期差異。				

表 6 方案評鑑報告檢核表（續）

	2-2 受評方案之參與者、閱讀評鑑報告者及其他利害關係人	應確認不同的利害關係團體（此處係指教育部，以及學校受評單位、評鑑相關組織、參與人員），描述他們與方案的關係，以及表達他們對方案重要性的不同觀點。					3：確認 2 項（教育部及學校） 2：確認 1 項（教育部或學校） 1：皆未確認
	2-3 受評方案之背景	應詳細描述評鑑方案的背景因素，俾使利害關係人瞭解背景資訊對方案的實施與結果所產生的衝擊，以及協助閱讀評鑑報告者解釋評鑑。					3：詳細描述 2：大約描述 1：未描述
3.評鑑概述	3-1 評鑑意旨／目的	評鑑意旨／目的應該： *自利害關係人的目的及其對評鑑結果的有意用途方面，加以說明。 *被充份詳細地描述，以協助利害關係人從評鑑結果中推斷重大意義。					3：達成 2 項 2：達成 1 項 1：皆未達成
	3-2 評鑑問題	應對照下列準則，證明評鑑問題的正當性： *問題的解答對哪些利害關係人是有用的，而且如何產生用途？ *問題的解答將如何提供新資訊？ （若由於時間、資源的限制，或資料蒐集工具之不足，使得評鑑問題無法獲得處理，惟評鑑報告仍應加以描述評鑑問題。）					3：達成 2 項 2：達成 1 項 1：未達成
	3-3 評鑑人員的可靠性	評鑑人員之專業資格應該被具體指明，俾便建立評鑑結果的可信度。					3：列出訪評人員名單，其職稱、資格條件頗佳 2：僅指出訪評人員人數 1：僅指出採用訪評方式 NA：未採用訪評方式

表 6 方案評鑑報告檢核表（續）

	3-4 利害關係人之參與	針對評鑑的計畫、資料蒐集與分析，以及結果解釋，評鑑報告皆應詳細說明如何將利害關係人之立場及觀點納入考慮，蓋利害關係人一方面能協助評鑑人員瞭解方案的目的與目標、建議資料來源、檢閱發現；另一方面，藉由積極參與，利害關係人比較可能見到結果的可信度、有用性及恰當性，也比較不可能刪減評鑑的活動或隱藏結果的適當用途。					3：將全部利害關係人之立場及觀點納入考慮 2：將部分利害關係人之立場及觀點納入考慮 1：皆未納入考慮
4.評鑑設計	4-1 評鑑方法論之取向	4-1-1 評鑑報告應敘述所使用的評鑑方法論之取向（量化方法、質性方法或綜合兩者；自評、訪評或綜合兩者），以及在時間與成本的限制下，評鑑取向如何產生有用資訊，以協助達成評鑑意旨／目的或回答評鑑問題。					3：敘述方法論取向，且方法論取向合宜 2：敘述方法論取向 1：未敘述
		4-1-2 評鑑取向必須建立在良好的方法論架構與「最佳實務」的文獻分析基礎上。					3：達成 2 項 2：達成 1 項 1：皆未達成
		4-1-3 評鑑方法論的取向應注重方案的輸入部分及檢驗量化產出的評量結果，以及質性與其他混合方法的運用，俾便捕捉一複雜方案之顯著的、未預期的影響、衝擊及結果。換言之，評鑑人員應測量多重結果，並自多重來源蒐集資料（三角測定法），俾便確證評鑑之發現，並同時提供多重觀點。					3：注重輸入或產出結果，以及運用量化或質性方法 2：注重輸入或產出結果，或者運用量化或質性方法 1：皆未符合

表 6 方案評鑑報告檢核表（續）

		4-1-4 描述重要的評鑑設計（包括：取樣、研究方法、資料分析）之限制。					3：描述 2 項（取樣、研究方法或資料分析）以上 2：描述 1 項（取樣、研究方法或資料分析） 1：皆未描述
	4-2 資訊來源 與取樣	4-2-1 方案評鑑設計中，應充份詳細敘述資訊來源（如：問卷調查分析表、訪談結果、法令規章），以證明足夠的資訊可達成評鑑意旨／目的。					3：充份描述 2：部分描述 1：未描述
		4-2-2 應描述提供資訊的各種團體（如：行政主管、教師、家長、學生）。若採抽樣方式，應描述每一種樣本的選取準則、選取方法、樣本大小、是否包括比較組或控制組、是否分派參與者到實驗組與比較組，以及如何分派？並應指明樣本具全部母群體代表性的程度。					3：全部描述 2：部分描述 1：未描述
	4-3 評鑑工具	評鑑報告應描述不同工具（如：評鑑準則／標準／指標）之目的、內涵、如何使用，以及信、效度之建立。					3：至少針對 1 個工具，做大部分描述 2：至少針對 1 個工具，做小部分描述 1：皆未描述
	4-4 資料蒐集 程序與時 程	資料蒐集（如：自評或訪評）時，應描述： *在何時及如何蒐集適當種類的資料。 *自樣本及其他目標來源取得主要資料所需採取的步驟。 *資料如何達到效度的準則。					3：描述 3 項以上 2：描述 1-2 項 1：皆未描述

表 6 方案評鑑報告檢核表（續）

		<p>*如何藉由系統化的訓練資料蒐集者、持續資料蒐集及計分步驟，以具有信度。</p> <p>*資料蒐集程序如何限定方案參與者在時間上與精力上的投入。</p>				
	4-5 後設評鑑	<p>應該定期審查評鑑意旨／目的與程序，尤其是長期縱貫性的評鑑，俾便判斷評鑑設計、工具及程序是否適宜地掌握方案的實施、影響及成果。</p>				<p>3：大部分達成</p> <p>2：小部分達成</p> <p>1：未達成</p>
5.分析 程序	5-1 量化分析	<p>*量化分析的程序對評鑑問題的提出及資訊特色的分析是適當的。例如：從量化分析引申出結論時，實施上的顯著性與複製性，以及統計上的顯著性等問題皆應被考慮。</p> <p>*應報導實際提供資料者的數目，如回收率等。</p> <p>*應描述量化資料分析之潛在限制，及其對結論與解釋之可能影響。</p>				<p>3：達成 2 項以上</p> <p>2：達成 1 項</p> <p>1：皆未達成</p> <p>NA：未採用量化分析</p>
	5-2 質性分析	<p>*質性分析（如：自評或訪評分析結果）的程序對評鑑問題的提出及資訊特色的分析是適當的。例如：當評鑑進行時，需藉由蒐集一種以上來源的證據，以及獨立證實所得之推論，以確認由質性分析而得的發現是精確的。</p> <p>*應描述質性資料分析之潛在限制，及其對結論與解釋之可能影響。</p>				<p>3：達成 1 項以上</p> <p>2：1 項中，部分達成</p> <p>1：皆未達成</p> <p>NA：未採用質性分析</p>

表 6 方案評鑑報告檢核表（續）

6.結果 與建 議	6-1 解釋與結 論	6-1-1 在評鑑報告中，應以平衡、無偏差的方法，完整、公平地註記方案預期與非預期的優點與缺點，俾便加強優點及提出問題所在。倘若因時間或成本之限制而無法獲得適切的資料，報告中應記錄此種遺漏，並估計其對方案衝擊力與效力的整體判斷，可能造成的影響。					3：完整、公平地註記優點與缺點 2：呈現優點與缺點 1：未呈現
		6-1-2 假若評鑑執行及蒐集資料之地點超過一處以上，評鑑報告中應先分別比較、對照每個評鑑地點之結果，最後仍應提出一概括整體方案的結果。					3：全部提出 2：部分提出 1：未提出 NA：蒐集資料之地點只有一處
		6-1-3 結論中應結合方案目的與評鑑問題，做廣泛說明，並加以報導發現。					3：大部分達成 2：小部分達成 1：未達成
		6-1-4 在結論中，評鑑人員應開放、真誠地提出其判斷所依據的價值觀與看法，以促進閱讀評鑑報告者之瞭解。					3：大部分達成 2：小部分達成 1：未達成
		6-1-5 藉由將評鑑成果連結到其他評鑑人員所支持的方法取向與實務，俾使結論對促進評鑑社群的專業卓越，有所貢獻。					3：大部分達成 2：小部分達成 1：未達成
	6-2 建議事項	*評鑑報告應適切地對目前評鑑方案中的利害關係人，或其他從事相似目的、重點、範疇及參與團體的評鑑方案人員，提出建議事項。					3：達成 2 項 2：達成 1 項 1：皆未達成

表 6 方案評鑑報告檢核表（續）

		*建議事項應基於強有力的發現，而不是軼事性的證據，不論後者是如何地有說服力。				
6-3 利害關係人的檢閱與利用	6-3-1	在與利害關係人分享評鑑報告方面，一方面，主要利害關係人需檢閱報告初稿，且所有利害關係人及其他對結果有法定權利的人，皆應有收到報告的機會。另一方面，閱讀評鑑報告者要認為評鑑人員能清楚地、真誠地呈現其判斷與建議，且有方法能取得這些詳加描述的資料。				3：大部分達成 2：小部分達成 1：未達成
	6-3-2	評鑑報告必須以反應式的型態與格式呈現。並應對不同需要與觀點之閱讀評鑑報告者，提供不同版本之評鑑報告。				3：有敘明 1：無敘明（註）
註：基本上，實施自我評鑑計畫之 34 所大學校院皆須繳交教育部一份成果報告，教育部並將其置放於相關網站上，本研究所採用之 34 所大學校院之成果報告即自網站下載而得。惟除此成果報告外，34 所大學校院是否另備有不同版本的評鑑報告，提供給校內受評單位或相關單位，或社會大眾？此問題除非各校在其成果報告中，有所敘明，否則本研究實無法得知。故本項目量的判斷依據僅包括兩等第：「3：有敘明」及「1：無敘明」。						

資料來源：作者整理。

三、實施程序與資料分析

在完成上述取樣與檢核表定稿之建構後，兩位研究者隨即進行專家檢核。兩位研究者皆有教育評鑑之研究與實務經驗，其中一位研究者更多次主持大學相關評鑑之規劃及後設評鑑之研究，對評估原級評鑑有一定程度之理論基礎與實務經驗。

本研究實施程序之具體做法為：首先由兩位研究者詳閱 9 所樣本大學校院的成果報告書後，再對照各校達成檢核表上品質準則項目「量的判斷依據」的程度，各自逐項分別進行判斷、評分。量表等第包括：符合、部分符合、不符合，以及不適用或無法判斷；若評定為「不適用或無法判斷」時，不予計分，若評定為「符合」、「部分符合」、或「不符合」時，則分別給予 3、2 或 1 分。

接著，為求兩位研究者評分之一致性，本研究採用評分者複評信度考驗 (interrater reliability test)，進行 Spearman 等級相關分析之考驗；以電腦軟體 SPSS10.0 版計算相關 rho 值，兩位研究者在 9 校 27 細項每一題比對之評分相關分別為：.804、.853、.824、.869、.812、.790、.761、.797 及 .836，數值頗高，此表示在信心水準 95% 下，兩位研究者之評定分數頗具一致性與可靠性。

接下來，為求 9 所樣本大學校院的成果報告書在每一檢核題項之整體平均分數，意即其成果報告書達成品質專業化之程度，故分別計算兩位研究者在 9 校所檢核之各細項平均分數結果，如表 7：

表 7 研究者檢核九所大學評分結果

題項 檢核者	2			3				4									
	1			3-1	3-2	3-3	3-4	4-1				4-2		4-3	4-4	4-5	
	2-1	2-2	2-3					4-1-1	4-1-2	4-1-3	4-1-4	4-2-1	4-2-2				
第一位	1.89	2.34	2.56	2	1.89	1.11	2.50	1.44	2.56	2	2.33	1.11	2.22	1.89	1.56	1.78	1.44
第二位	2	2.78	2.56	1.67	1.78	1.33	2.50	1.56	2.56	1.89	2.44	1.33	2.11	1.89	2	1.67	1.56
平均分數	1.94	2.61	2.56	1.83	1.83	1.22	2.50	1.50	2.56	1.94	2.39	1.22	2.17	1.89	1.78	1.72	1.50
		2.34			1.75				1.91								

表 7 研究者檢核九所大學評分結果（續）

題項 檢核者	5		6							
	5-1	5-2	6-1					6-2	6-3	
			6-1-1	6-1-2	6-1-3	6-1-4	6-1-5		6-3-1	6-3-2
第一位	2	1.63	2.89	Na	2.22	1.11	1	1.78	1.33	1.22
第二位	2.33	1.63	2.89	Na	1.89	1	1.11	1.89	1.44	1.22
平均分數	2.17	1.63	2.89	Na	2.06	1.06	1.06	1.83	1.39	1.22
	1.90		1.66							

最後，進行成果報告書檢核結果之有效推論與詮釋。由於篇幅限制，無法說明 9 所大學的成果報告書各細項結果之個別意義，僅解釋在六大項目達成該品質準則程度的整體意涵。以量化結果而言，若平均分數在 2.50 分之上，表示該項目屬於「符合」等第；在 1.50 至 2.49 分之間，表示「部分符合」；在 1.49 分之下，則表示「不符合」。除此量化分數表示檢核結果外，亦針對報告書在該品質準則項目所表現的優缺點，以文字加以敘述，期能呈現質量兼具的檢核結果。

肆、結果分析

本節旨在說明 2001 年 9 所實施自我評鑑計畫大學校院的成果報告書的檢核結果，進而推論 34 所母群體大學的整體報告書品質，並瞭解 34 所大學實施自我評鑑計畫之若干現況。如同上述，檢核結果係自品質準則六項目，包括：執行摘要、受評方案描述、評鑑概述、評鑑設計、分析程序，以及結果與建議等，分別說明 9 所大學校院的成果報告書之檢核結果，包括表示達成該項目程度之量化分數，以及在該項目所表現的優缺點之質性敘述。

一、執行摘要：部分符合（1.94 分）

品質準則項目內容：評鑑成果報告應清楚概述此次自我評鑑之意旨／目的、評鑑方案之目的、實施及衝擊，評鑑結論與建議。

優點：各校皆能清楚定義該校自我評鑑之目的係以自我改進為主，績效責任為輔，同時評鑑結果運用亦能回應評鑑意旨，且清楚說明結論報告之用途。

缺點：大部分學校雖能有一執行摘要，惟未能清楚報導評鑑必要資訊，包括評鑑緣起、法源依據、評鑑組織與資源、評鑑問題、評鑑指標建構內容與分析、評鑑方法論、評鑑活動設計、結論與建議；此外，大部分評鑑報告也未能符合簡明、直接、以利害關係人為之主之語言來報導。

二、受評方案描述：部分符合（2.34 分）

品質準則項目內容：應清楚描繪受評方案之特徵（如方案目的與目標、活動內容與場所、期程與資源）、確認不同利害關係團體（教育部，以及學校受評單位、評鑑相關組織、參與人員）、描述他們與方案的關係及其對方案重要性的不同觀點，以及描述對方案的實施與結果會產生衝擊的受評方案之背景因素。

優點：大部分學校都能描繪自我評鑑之內、外在動機，對於評鑑活動內容與期程、使用資源、評鑑組織等議題，亦有所介紹。

缺點：絕大部分學校皆未能清楚描繪主要利害關係人之特質與需求，也未能提供協助各主要利害關係人參與評鑑規劃與執行之服務與學習的機會，同時在描繪方案背景與脈絡時，未能運用多層次、不同來源之資料來說明，致使評鑑報告閱讀者未能清楚瞭解該評鑑範疇與內容，以及評鑑最後產出之影響與整體效益。

三、評鑑概述：部分符合（1.75 分）

品質準則項目內容：應自利害關係人的目的及用途方面，加以說明評鑑意旨／目的；應描述評鑑問題，若有需要，尚須以產生用途及提供新資訊之準則，證明評鑑問題的正當性；評鑑人員專業資格應被具體指明；最後，評鑑報告皆應詳細說明如何將利害關係人之立場及觀點納入考慮。

優點：各校結論報告中，大都有介紹評鑑人員之姓名、職稱，以顯現其學術聲望及評鑑結果的可信度，以及對受評單位之參與情形，亦有所著墨。

缺點：絕大部分學校在評鑑規劃前，都未能詳細評估各受評單位之需求與可能的問題，且未能預先考慮各利害關係人之觀點。此外，針對利害關係人的涉入，大部分學校雖有針對受評單位辦理說明會，然該會性質僅屬準備受評工作之表格填寫說明而已，未能針對受評單位如何透過授權、民主對話、積極參與等途徑，將參與者觀點納入評鑑推動來考量。且亦未能針對如何透過解讀與回應結論報告，來尋求自我改進途徑，以達到參與評鑑及組織專業成長的目標。

四、評鑑設計：部分符合（1.91分）

品質準則項目內容：應敘述評鑑方法論之取向（量化、質性或綜合兩者，自評、訪評或綜合兩者）及其建立的基礎；能確證評鑑發現的多重資訊來源及多重觀點；提供資訊的各種團體；重要的評鑑設計（包括：取樣、研究方法、資料分析）之限制；不同工具（如：評鑑準則／標準／指標）之目的、內涵、如何使用及信、效度之建立；務實、可靠的資料蒐集（如：自評或訪評）程序與時程，以及自評工作品質管控的後設評鑑。

優點：各校皆能針對所使用評鑑方法論的取向，或詳細或粗略地加以描述；絕大部分的學校皆參考其他大學或教育部以往所建構的（例如：86學年度大學綜合評鑑試辦計畫或中程校務發展計畫）評鑑指標內容，且實際運作過程皆未產生重大問題。

缺點：各校皆使用一般的方法論取向，惟大都未能針對評鑑理論與「最佳實務」進行分析，也未對受評單位之人事物地等特質進行評估，以決定何者為符合評鑑意旨之方法論，以及受評單位內誰可以提供何種具信、效度資訊之種類與範圍；此外，也未能從受評單位瞭解什麼是最重要的評鑑需求，以便形成資訊來源與評鑑關鍵問題，再透過關鍵問題，來決定評鑑方法論。至於各校評鑑指標建構方式雖大都參考他校或教育部所建構者酌予修正，但除未能因應受評單位之需求與關鍵問題外，也未能針對所使用的評鑑工具，先行試評後始予定稿，亦未能依據評鑑計畫所規定之程序來施用。最後，在各校結論報告中，亦沒有發現有內部

或外部發動之後設評鑑，以判斷原級自我評鑑實施之情形與影響。

五、分析程序：部分符合（1.90 分）

品質準則項目內容：應詳述量化或質性分析的程序對評鑑問題的提出及資訊特色的分析是適當的，並描述量化或質性資料分析之潛在限制，及其對結論與解釋之可能影響。

優點：各校都採用質性、量化、或綜合兩者之分析方法；同時大部分學校亦採用校內相關單位先行自我評鑑與外部專家蒞校訪評兩種方式來蒐集、分析資料，以判斷受評單位的優缺點。

缺點：各校皆未對資訊本質與評鑑問題，先行透過探索性分析，以瞭解其性質與使用之合宜性；其次，也未能在評鑑計畫等書面資料中，針對資料蒐集之每一程序作詳細說明與定義，以確保資訊蒐集之正確性及符合評鑑意旨；最後，也未能報導量化或質性分析程序之任何限制，及其可能對結論與解釋之影響。

六、結果與建議：部分符合（1.66 分）

品質準則項目內容：在評鑑報告中，應以平衡、無偏差的方法，完整、公平地註記方案的優缺點、發現，以及提出結合方案目的與評鑑問題的結論；應對方案中的利害關係人或其他方案人員，提出建議事項；並應對不同需要與觀點之閱讀評鑑報告者，提供不同版本之評鑑報告。

優點：各校皆詳細記錄各評鑑委員之書面建議事項，亦有部分學校針對建議事項的回應與後續追蹤輔導，予以說明。

缺點：各校皆未能先行建立一套評鑑報告公布約定，該協議可以清楚界定誰是評鑑報告主要閱讀者、各種報告（期中、期末、報告草案、正式報告）公布流程與步驟、協助利害關係人解讀或運用結果報告之服務種類、如何保障有意或無意誤用報告之措施、出版符合不同需要之結論版本、程序，以及如何針對結果報告回應或追蹤輔導等。此外，各校結論報告大都未能依據評鑑發現作適當反應，且部分學校將各訪評委員的意見與建議，逐字摘錄，並未透過評鑑結論報告會議，先行整合達成共識，再討論並聚焦如何針對整體重點問題作回應與追蹤輔導，以

彰顯自我評鑑之價值性；同時亦未對評鑑方案整體優缺點以及對受評單位或學校產出與影響，進行評估並報導。

伍、結論與建議

本研究旨在評估 2001 年實施自我評鑑計畫大學校院的成果報告書的品質，並瞭解其實施自我評鑑計畫之若干現況。首先依據文獻理論，建構一套具有信度與效度的檢核表，其次依據檢核表的準則，抽樣檢核 9 所大學校院的成果報告書，判斷其優缺點，進而推論整體報告書的品質，並瞭解計畫實施之若干現況。最後研擬具體建議，俾政府有關單位與後續研究之參考。

一、結論

綜合前文之分析結果，本研究歸納以下三方面結論。

(一) 自後設評鑑報告之文件分析得知，歐美後設評鑑研究在下列若干做法，可供我國後設評鑑潛在運用者加以借鏡。

1. 就後設評鑑對象而言：我國後設評鑑研究對象大都僅限於方案評鑑或校務評鑑之實施；然歐美後設評鑑對象範圍則較為廣泛，包括評鑑整體實施、評鑑系統、評鑑模式、評鑑實施計畫、評鑑目的、評鑑活動與過程設計、評鑑方法論、評鑑資料蒐集工具與方法、評鑑報告，以及評鑑人員行為與倫理規範等。

2. 就後設評鑑實施途徑而言：我國檢核原級評鑑研究之實施途徑大都依據方案評鑑標準，或自行建構之焦點議題。反觀歐美研究取向，尚包括使用專業後設評鑑檢核表；再者，所參照之專業標準，除方案評鑑四大標準外，亦包括 AEA 評鑑人員之五大指導原則等。

3. 就後設評鑑發動者而論：後設評鑑因發動者不同可區分為內部與外部後設評鑑。我國後設評鑑皆屬外部後設評鑑，意即由獨立第三者，於評鑑結束後，基於評估原級評鑑符合專業化標準之程度，所發動之總結性評鑑。然歐美後設評鑑研究除包括外部後設評鑑，亦有不少內部後設評鑑之研究，即由原級評鑑團體主動委託第三者或由原評鑑規劃人員，於評鑑實施結束前，為管控原級評鑑之品質

而進行以自我改進為目的之形成性評鑑。此外，歐美部分著名學者在其所負責之原級評鑑結論報告中，亦會有一個章節報導其內部後設評鑑之結果；或者在期刊發表的評鑑結論報告中，會有歡迎獨立第三者針對其原級評鑑實施或結論報告內容，進行外部後設評鑑評論或研究之敘述。

4. 自後設評鑑的資料蒐集與分析方法觀之：我國與歐美後設評鑑之資料蒐集與分析方法大抵類似，皆依據原級評鑑之性質決定資料蒐集及分析方法；若原級評鑑係屬參與式或建構主義者評鑑模式等質性評鑑，則後設評鑑資料之蒐集與分析方法則偏重質性資料分析之判斷；倘若原級評鑑之資料蒐集與分析為質性與量化方法綜合使用時，其資料蒐集與分析方法則包括：問卷調查、成本分析及專業技術輔助分析等量化方法，以及實地觀察、訪談、相關資料（例如：正式報告、期中報告和各類型會議等資料）之文件分析等質性方法。值得注意的是，部分歐美後設評鑑研究為了確保其評鑑方法論之嚴謹性，於後設評鑑報告正式公布前，會主動提供後設評鑑的草案報告，由原級評鑑團體等先作批判與回應，或者召開後設評鑑法庭會議，供原級評鑑團體辯護後，再行出版。

(二) 後設評鑑檢核表是一種可以快速、容易、精準地蒐集資訊，引導評鑑程序，並對照評鑑專業標準，最後整體判斷原級評鑑之優缺點及價值的有效方法。

1. 後設評鑑檢核表應能呈現該評鑑範疇領域的專業知能，經由建構過程除可充份瞭解該範疇應然面之標準，亦可因實然面之限制而給予調整與回饋修正。

2. 本後設評鑑檢核表建構流程為：瞭解歐美各國評鑑檢核表之種類與用途，分析各種檢核表內涵，界定檢核對象之性質，初步決定採用何種檢核表，深入分析該檢核表內涵、定義與限制，修正並完成檢核表初稿，再由專家學者審查檢核表初稿，修正檢核表，實施試評，並依據試評結果修正檢核表後完成定稿；本建構流程頗為嚴謹、可行。

3. 評鑑係一跨學門的社會科學研究，各校自評計畫成果報告的撰寫格式與內容，都應嚴格遵守評鑑學理與原則，本研究所建構之檢核表，應可供各大學建置自我評鑑機制、實施評鑑，以及撰寫自我評鑑報告時，引導並管控評鑑是否符合效用性、可行性、適切性及精確性之專業化標準。

(三) 運用檢核表檢核各大學自評計畫成果報告書之結果，發現除少數學校，各大學成果報告書皆未能充份展現出一份具有專業性、系統性之研究報告；此外，教育部亦宜更積極地推動高等教育評鑑朝向制度化與專業化方向發展。

1. 整體各大學自評計畫成果報告書的量化檢核結果，在檢核表品質準則全部項目中（如：執行摘要、受評方案描述、評鑑概述、評鑑設計、分析程序，以及結果與建議）皆屬「部分符合」程度，分數自 2.34（受評方案描述）至 1.66（結果與建議）不等。換言之，除少數學校，各大學自評報告書之內容呈現與解釋，如自我評鑑目的、意旨之擬定，評鑑方法論之定義，評鑑規劃（評鑑設計、活動與期程、資料分析等），結論報告之撰寫與運用等方面仍不甚理想，有許多值得再改善之空間。

2. 教育部在「2001 年大學校院實施自我評鑑計畫」中所扮演的角色，應不僅只限於事前經費補助，事後「將各校資料上網以供參考」。此外評鑑專業社群涉入情形亦不多見（僅有訪評人員蒞校訪評）。大部分學校在缺乏評鑑規劃人才情形下，教育部應於實施計畫開始時召開相關的評鑑說明會、工作坊；在計畫實施中，可委託學術單位提供學校諮詢或輔導；計畫結束後再對成果報告書之內容進行審核、召開檢討會或委託學術單位進行實施成效評估，以提供最佳實務典範、經驗交流等強化評鑑知能之提昇，如此才能有效協助各大學積極建構自我品質管制之機制。

二、建議

根據結論，本研究自教育部、大學校院及後續研究三方面提出以下若干建議。

(一) 教育部

1. 建構一套完善資訊服務平臺：內部評鑑能夠有效提昇大學教育品質，故政府應建構一個結合經費補助、輔導、品質管控，及經驗交流等功能之服務平臺；同時亦應積極引導尚未建置自我評鑑機制之大學，儘速導入內部品質管控機制，或協助已建置自我管控機制之大學，分享彼此評鑑經驗。此外，亦可藉由短期研習方式培養各校評鑑專業人才，精進評鑑專業知能，使其能更有效達成各校品質

管制之預期目標。

2.整合評鑑社群資源：政府應竭盡所能整合專業評鑑社群之可用資源，並加速推動大學評鑑中心之成立，如此將可提供各大學推動內部評鑑時，更有效輔導諮詢之服務。

3.輔導高等教育評鑑系所或課程之成立：基於評鑑專業化發展有賴於評鑑基礎研究之推動，故教育部應積極輔導各大學之教育、人力資源、企業管理、行政管理、心理學等系所，增設評鑑專業課程，或輔導籌設評鑑專業系所，以落實「評鑑為一專業學科」及評鑑專業化的目標。

4.建構一套評鑑品質管控之機制：對大學申請經費推動自評工作，應建立一套追蹤輔導或後設評鑑等管控機制，並應於計畫結束後召開檢討會議，透過經驗交流、推展各校最佳實務典範，俾提供各校人員評鑑專業成長之機會。

(二) 大學校院

1.積極培育校內評鑑專業人才：目前各校推動內部品質管控專業化最大問題，乃缺乏相關評鑑規劃人才，故各校應鼓勵或輔導對評鑑或相關教育品質管理有興趣之教師，積極涉入或參與各項評鑑專業發展之實務與研究，從經驗中學習，並在校內建構評鑑專業知識分享體系。

2.建構一套自我品質管控機制：基於強化競爭並發展各校特色之精神，各校應積極建構一套符合自我需求之常駐教育品質管控機制，包括：領導者強力支持，建置實施自我評鑑與品質保證之法源依據，常設性質之品質保證組織，鼓勵全員參與，建構正向評鑑內在動機，形塑優質評鑑文化，以及提供評鑑專業知識與必要之資源和條件等。

(三) 後續研究

1.加強大學後設評鑑之研究，以維持原級評鑑之品質，並引導評鑑實施更能有效符合專業化標準，如此評鑑利害關係人才能緊密檢視評鑑之優缺點與價值，並決定如何回應評鑑之結論與後續追蹤。

2.他山之石可以攻錯，應持續探討歐美各國高等教育評鑑之制度與做法，同時亦應持續研究先進國家高等教育內部評鑑與政府外部品質管控機制之整合做法與配套措施。

3. 方案評鑑檢核表可以快速蒐集評鑑實施之資料，各種類型用途與目的之檢核表，亦值得再深入研究、發展及運用；此外，本研究檢核表係採用三等量表，若運用五點量表是否較適應國情且有較佳之鑑別度，此亦值得未來後續研究加以嘗試。

4. 後續研究可針對本後設評鑑研究，再進行評鑑。

參考文獻

- 胡悅倫、陳漢強（1998）。1997 學年度大學綜合評鑑之後設評鑑研究。載於胡悅倫（主編），*海峽兩岸大學教育評鑑之研究*（頁 3-105）。臺北市：師大書苑。
- 教育部高教司（2002）。34 校自我評鑑計畫獲本部補助。*高教簡訊*，130，10。
- 曾淑惠（1996）。我國專科學校後設評鑑之研究。國立臺灣師範大學工業教育研究所博士論文，未出版，臺北市。
- 曾淑惠（2002）。教育方案評鑑。臺北市：師大書苑。
- 黃政傑、李隆盛、游家政（1994）。國民小學教育評鑑之研究。行政院國家科學委員會專題研究報告。臺北市：國立臺灣師範大學教育研究中心。
- 賴志峰（1997）。臺北市幼稚園後設評鑑之研究。國立政治大學教育研究所碩士論文，未出版，臺北市。
- 蘇慧雯（2002）。臺北市幼稚園後設評鑑之研究。臺北市立師範學院國民教育研究所碩士論文，未出版，臺北市。
- 蘇錦麗（1995）。大學學門評鑑試辦計畫成效評估之研究。臺北市：師大書苑。
- 蘇錦麗（1997）。高等教育評鑑：理論與實際。臺北市：五南。
- 蘇錦麗、楊瑩、王偉中、呂鴻德、詹惠雪（2000）。大學後設評鑑研究。臺北市：揚智。
- American Evaluation Association (1994). Guiding principles for evaluators, Retrieved February 24, 2004, from <http://www.wmich.edu/evalctr/checklists/guiding-principles.pdf>
- Bjorkkilde, T., & Bason, C. (2000). *Metaevaluation of the evaluations of higher education in Denmark*. Paper presented at the 4th EES conference, Lausanne. Retrieved February 25, 2004, from http://www.europeanevaluation.org/pdf/5-1_bjorkkilde.pdf
- Bustelo, M. (2002). *Metaevaluation as a tool for the improvement and development of the evaluation function in public administrations*. Retrieved February 24, 2004, from <http://www.europeanevaluation.org/docs/BusteloSevillaEES.pdf>

- Chelmsky, E. (1987). The politics of program evaluation. In D. S. Cordray, H. S. Bloom, & R. J. Light (Eds.), *Evaluation practice in review* (pp. 5-22). San Francisco: Jossey-Bass.
- Grasso, P. G. (1999). Meta-evaluation of an evaluation of reader focused writing for the veterans benefits administration. *American Journal of Evaluation*, 20(2), 355-370.
- Gunnarsson, V., & Rupprecht-Hjort, M. (2000). *Metaevaluation of the structural funds in Sweden – What lessons can be learnt and how to utilize this knowledge in the future*. Retrieved February 23, 2004, from http://www.europeanevaluation.org/pdf/6-1_gunnarsson.pdf
- Hartmann, D., & Loizides, G. (2001). *Metaevaluation of the web-based ATE survey evaluation system*. Retrieved February 24, 2004, from <http://www.wmich.edu/~evalctr/ate/webbasedmetafinal.pdf#search=hartmann%26loizides>
- Joint Committee on Standard for Educational Evaluation (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage.
- McKinley, K. H. (1999). *Metaevaluation report of the evaluation of the Michigan public school academy initiative*. Retrieved February 25, 2004, from <http://www.wmich.edu/evalctr/charter/reports/metaeval.html>
- Migotsky C., & Stake R. (2001). *An evaluation of an evaluation: CIRCE's metaevaluation of the site visits and issue papers of the ATE program evaluation*. Retrieved February 25, 2004, from <http://www.wmich.edu/~evalctr/ate/sitevisitmetafinal.pdf>
- On-line Evaluation Resources Library (2001). Quality criteria for evaluation resources. Retrieved February 27, 2004, from <http://oerl.sri.com/crit.html>
- Sanders, J. R. (1999). Metaevaluation of “the effectiveness of comprehensive, case management interventions: Evidence from the national evaluation of the comprehensive child development program”. *American Journal of Evaluation*, 20(3), 577-582.
- Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, 23(4), 387-419.
- Schwandt, T. A., & Halpern, E. S. (1988). *Linking auditing and metaevaluation*. Newbury Park, CA: Sage.
- Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Perspectives of Curriculum evaluation*. Chicago: Rand McNally.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Beverly Hill, CA: Sage.

- Scriven, M. (2000a). *The logic and methodology of checklists*. Retrieved March 2, 2004, from <http://www.wmich.edu/evalctr/chechlists/logic.htm>
- Scriven, M. (2000b). *Metaevaluation of gang resistance education and training program*. Retrieved February 27, 2004, from <http://eval.cgu.edu/lectures/evtypes/metaeval/greatex.htm>
- Scriven, M. (2001). Evaluation: Future tense. *American Journal of Evaluation*, 22(3), 301-307.
- Scriven, M. (2004). Key evaluation checklist. Retrieved February 27, 2004, from <http://www.wmich.edu/evalctr/checklists/kec.htm>
- Stake, R., & Davis R. (1999). Summary of evaluation of reader focused writing for the veterans benefits administration. *American Journal of Evaluation*, 20(2), 323-343.
- Stufflebeam, D. L. (1974). *Metaevaluation*. Kalamazoo, MI: Western Michigan University Evaluation Center.
- Stufflebeam, D. L. (1999). Program evaluation models metaevaluation checklist. Retrieved February 27, 2004, from <http://www.wmich.edu/evalctr/checklist/eval-model-metaeval.html>
- Stufflebeam, D. L. (2000a). The CIPP model for evaluation. In D. L. Stufflebeam, G. F. Madaus & T. Kellaghan (Eds), *Evaluation models: Viewpoint on educational and human services evaluation*. (2nd ed.) (pp. 279-317). Boston: Kluwer.
- Stufflebeam, D. L. (2000b). The methodology of metaevaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds), *Evaluation models: Viewpoint on educational and human services evaluation*. (2nd ed.) (pp. 458-471). Boston: Kluwer.
- Stufflebeam, D. L. (2001a). Evaluation checklists: Practical tools for guiding and judging evaluations. *American Journal of Evaluation*, 22(1), 71-79.
- Stufflebeam, D. L. (2001b). The metaevaluation imperative. *American Journal of Evaluation*, 22(2), 183-209.
- Stufflebeam, D., Gullickson, A., & Wingate L. (2002). *The spirit of Consuelo an evaluation of KE AKA HO 'ONA*. Retrieved February 27, 2004, from <http://www.wmich.edu/evalctr/pubs/consuelo/>
- Swedish International Development Co-Operation Agency (2001). *Report on a Meta-evaluation of the evaluation of Diakonia and Forum Syd's Cooperation with Cambodian NGOs*. Retrieved February. 27, 2004, from <http://web.forumsyd.se/arkiv/faktabanken/evalu/001e2431-000f>
- Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practical guidelines*. (2nd ed.). White plains, NY: Longman.