

邁向下一代的教育評鑑：回顧與前瞻

潘慧玲

壹、前言

評鑑概念的起源甚早，在國外的文獻中，似均認為最早的評鑑開始於古代的中國。例如，Dubois (1965, 1970) 提到公元前二千二百年前的中國就有每三年考察百官一次的辦法；Popham (1993)、Stufflebeam與Shinkfield (1985) 等人亦提及中國早在四千年前就已舉行公職人員考試。面對上述的說法，盧增緒(1995)認為在時間上缺乏信史可據，因之，在經查考後，他斷定Dubois論著中提到的「每三年考察百官」，所指的應是周代鄉遂之治中的邦比之法，而非許多文獻以訛傳訛所指的科舉制度。評鑑概念雖有如此之早的起源，但落諸國內教育領域的實施，則為民國五十年代之事。¹

民國五十二年教育部接受聯合國兒童教育基金會之資助，簽署了國民教育發展五年計畫實施方案，在方案中規劃有評鑑計畫，其後於民國五十三至五十八年進行了完整的教育評鑑，期終報告於民國五十九年完成（盧增緒，1995）。如果該項計畫被視為評鑑之濫觴，則教育評鑑的實施至今可說已有將近四十年的歷史，只是其間並未持續，僅間歇性地進行一些局部與試辦性工作，而第一個全面性推動的評鑑——大專校院評鑑，一直要到民國六十四年才出現，不過之後仍未見評鑑的積極推動。直至近幾年，評鑑之風開始盛行。受到這波教育改革的影响，為講求績效責任，高等教育及後期中等教育的評鑑接連實施；而民國84年《國民教育法》的修訂，將校長辦學績效納入遴選考量，使得北高兩市及地方縣市政府紛紛辦理國民中小學的校務評鑑或校長評鑑。此外，中小學教師專業評鑑也是這兩年不斷討論的議題，《教師法》的修訂將為其建立法源的基礎。

見到評鑑終能在教育領域中逐步發展，固然令人欣喜，但如果未被正用，則評鑑之功不僅無法彰顯，徒留的也恐僅是誤用的惡果！因之，在提筆撰寫這樣的一篇文章時，有著焦慮與期盼的複雜心情。焦慮的是發展有近四十年光景的評鑑，雖有許多論述的生產，但學術的風貌尚不夠多元，而民國六〇年大學評鑑的實施牽動後來實務的開展，多年來的教育評鑑一直著重於外部評鑑，且以美國「認可制」(accreditation) 為取法之宗，²形成評鑑模式的單一化。加上近幾年勃興的評鑑，推動過於頻繁，造成了實務工作者的心存抗拒。故而與國外相較，國內不論是實務的推動或是學術的耕耘，實在還有好長的一段路要走！如何能夠立足

¹ 此處所指的評鑑不含涉測驗評量。

² 國內所實施的教育評鑑，主要取法自專家意見導向評鑑模式中的認可制，只是國內執行時並未完全採擷認可制之所有作法，例如評鑑標準的定期公布與認可機構作認可決定，國內即未採行。簡言之，國內參採之作法是將評鑑分自評與訪評兩階段，且評鑑結果以專家到校半天至兩天的訪評意見為依據。

於過去，借鏡他國的知識與經驗，以開展國內的評鑑領域，成為內心的殷盼。

為讓「過去」成為「未來」發展的立足平台，批判性地回顧十分重要，唯國內這類文章並不多見，盧增緒（1995）是其中一擲地有聲者。該文以犀利筆觸指陳多年來國內教育評鑑之失，諸如正確的評鑑觀念未能形成、教育評鑑被窄化為校務評鑑、學者引介國外論著未能窺其全貌、第四代評鑑與理論導向（theory-driven）評鑑³未受重視等。時隔九年的今天，昔日之問題是否均已改善，看來並不全然，教育評鑑領域尚待釐清的概念與開拓的議題還有許多。本文主旨即在梳理過去學者們在評鑑領域的耕耘成果，而這樣的成果無可諱言的是西方的，尤其是美國的！如此做的目的在於呈現評鑑領域的多元學術風貌，擴展過去國人所忽略的面向，並點出評鑑並非僅是技術性的操作，其中蘊含著知識論、方法論、政治、倫理等議題的思考與探究。

教育評鑑之內涵，如果加以細分，可依評鑑客體（object）之不同，分為方案評鑑（program evaluation）、校務評鑑、課程評鑑、教材評鑑、人員評鑑等。上述中的課程評鑑，如評鑑重點放在課程方案上，係為一種方案評鑑；如評鑑重點放在各個發展階段上，則為課程發展評鑑。雖然教育評鑑之內涵囊括不同的類別，但在美國，一提及教育評鑑，指涉的通常是方案評鑑，而在英國，教育評鑑則等同於課程評鑑。有鑒於國內除了人員評鑑外，其餘所執行的評鑑，如校務評鑑、課程方案評鑑、教材評鑑等，使用的模式均不出方案評鑑涉及的範疇，故為求聚焦，本文將以方案評鑑為主要範圍進行討論。文章首先探討自1960年代以降近四十年發展歷史的評鑑領域，在有關評鑑的評價、評鑑的知識建構、評鑑使用、評鑑實務及方案運作的論述上，有何建樹；其次，就國內經驗作一反思，分就評鑑組織、指標、人員、方式、結果處理、追蹤評鑑等作檢討；最後則提出未來之展望，探討下一代教育評鑑的可能圖像。「下一代的教育評鑑」一詞轉借了Guba與Lincoln（1989）「第四代評鑑」的語彙，不過，本文的意圖不在循著Guba與Lincoln的思考理路，為第一代視評鑑為「測驗」、第二代視評鑑為「陳述客觀事實」、第三代視評鑑為「判斷」以及第四代視評鑑為「協商與溝通」，繼續找尋第五代評鑑的出路，而是希望透過往昔經驗的探索，勾勒出未來發展的趨勢，以作為國人努力的方向。

貳、回顧教育評鑑領域耕耘的成果

為系統性地回顧評鑑領域至今所耕耘的成果，以下分就評鑑的專業化、評鑑模式的建立、評鑑理論分析架構的形塑以及評鑑理論的發展等四部份進行分析。

一、評鑑的專業化

³理論導向評鑑強調透過方案理論（program theory）的建立，瞭解方案如何產生預期成效，如此評鑑不僅止於評估方案的成效，尚能瞭解方案之所以有效或無效，中間歷程的闡釋因素，此主張將於後面作較為詳盡之介紹。

雖然評鑑最早始於中國，然評鑑領域的走向專業，則是在西方的場域中進行。有關美國的評鑑發展歷史，Stufflebeam 及其同僚補充之前 Stufflebeam 在 1985 年與 Shinkfield 所分的五期，成為七個時期：(一) 改革時期 (Age of Reform)，自 1792 年至 1900 年；(二) 效率與測驗時期 (Age of Efficiency and Testing) 時期，包括 1900 至 1930 年的發展；(三) 泰勒時期 (Tylerian Age)，從 1930 年至 1945 年；(四) 純真時期 (Age of Innocence)，從 1946 年至 1957 年；(五) 發展時期 (Age of Development)，從 1958 年至 1972 年；(六) 專業化時期 (Age of Professionalization)，自 1973 年至 1983 年；以及 (七) 擴充與統整時期 (Age of Expansion and Integration)，自 1983 至 2001 年 (黃光雄，1989；Stufflebeam & Shinkfield, 1985；Stufflebeam, Madaus, & Kellaghan, 2000)。在這幾個時期裡，評鑑較為成形與發展可說是從第五時期開始。當 1957 年蘇聯的人造衛星打上太空，美國受到極大的震撼，重新檢討自己的教育施為。接續著 1960 年代，詹森總統推展的大社會理想，教育被視為落實社會平等的重要手段。因而，隨著 1965 年通過的《初等與中等教育法案》，聯邦政府灑下大筆經費推動不同的教育方案，而評鑑也因教育方案的成效需要檢視而大為風行。當時的盛況，可從評鑑者難求，學校教師都得減少授課時數去作評鑑而可見一斑 (Worthen & Sanders, 1987)。1960 年代的聯邦政府為美國打造了教育評鑑發展的基礎，到了 1970 年代，教育評鑑乃逐漸走上專業之途，許多的專業學會、期刊與專書紛紛於 1973 年後出現。

相較於美國，英國的評鑑與課程發展是聯結在一起的，教育評鑑可以說是等同於課程評鑑。英國初期的課程評鑑籠罩於泰勒目標模式的思考架構，第一本課程評鑑領域中的出版品----Wiseman 與 Pidgeon 於 1972 年所著的 *Curriculum Evaluation* 即為一例。然而約在同時，Lawrence Stenhouse 提出對於目標模式的疑義，掀起 1970 年代英國課程評鑑另類取徑 (approach) 的風潮 (Hopkins, 1989)。1980 年代的教育評鑑開始分為兩大類，一是著重績效的評鑑，此可如學校層級的評鑑，以及學童學科表現的評量；另一則是在專案補助 (categorical funding) 風潮下所作的課程評鑑。

綜觀教育評鑑的發展歷史，可知自 1960 年代末期奠基後，1970 年代即開始朝向專業化的方向發展，專業學會紛紛成立，如美國的 American Evaluation Association (1985 年由 Evaluation Network 及 Evaluation Research Society 合併)、澳洲的 Australasian Evaluation Society、加拿大的 Canadian Evaluation Society、中美洲的 Central-American Evaluation Association 以及歐洲的 European Evaluation Society。上述的五個學會曾於 1995 年假加拿大溫哥華舉辦聯合國國際評鑑會議，會中有來自 66 個國家，超過 1600 位的評鑑者共聚一堂回顧過去並展望未來，會議結果更集結成 *Evaluation for the 21st Century* 一書 (Chelimsky & Shadish, 1997)。

對於評鑑是否能夠成為專業，Worthen、Sanders 與 Fitzpatrick (1997) 曾提出十項的檢核規準：(一) 具有對於評鑑專家的需求；(二) 評鑑具有所需之獨特知識與技能、(三) 設有評鑑人員培育之正式學程；(四) 提供穩定的評鑑生涯機

會；(五)評鑑功能建制化 (institutionalization)；(六)具有評鑑人員認證程序；(七)設有評鑑人員的專業學會；(八)專業學會中訂有決定會員資格的規準；(九)評鑑學會對於評鑑人員培育學程具影響力；(十)訂有評鑑實施的標準。依照以上的十項規準衡度評鑑專業化的程度，可發現具有良好發展歷史的美國，除了上述的(六)、(八)、(九)規準外，均已達成。

二、評鑑模式的建立

在過去的三十年中，評鑑理論五花八門，有近六十種正式、非正式的評鑑模式或取徑 (approach) 被提出 (Worthen & Sanders, 1987)。近年來，新的評鑑模式仍不斷地創發，例如 Empowerment Evaluation (Fetterman, 1994, 1997a, 2000; Fetterman, Kaftarian, & Wandersman, 1996)、Cluster Evaluation (Sanders, 1997)。

面對不同的評鑑模式，在作介紹時，學者各有不同的分類方式。其中 Popham (1993) 將評鑑模式分為五類：目標達成模式、強調輸入的判斷模式、強調結果的判斷模式、輔助作決定模式、自然模式。與上述分類有些雷同的是 Worthen 與 Sanders (1987)、Worthen、Sanders 與 Fitzpatrick (1997) 的六個分類：目標導向模式、經營導向模式、消費者導向模式、專家意見導向模式、抗詰導向模式、參與者導向模式。至於 Stufflebeam (2000) 以評鑑的意圖性將 20 種模式劃分為三大類：問題/方法導向模式、改進/績效導向模式、社會議程導向模式。深究這些不同的評鑑模式，可發現其哲學基礎不同，反映出客觀與主觀的知識論立場；而其方法論背景的差異，則反映出量化與質化的評鑑方式。

在上述的分類中，以 Worthen、Sanders 與 Fitzpatrick (1997) 之區分性較高，故以下以其分類說明學者們所建立之評鑑模式。⁴

在目標導向模式中，評鑑所著重的是確定目標以及決定方案目標所達成的程度，此模式源自 Ralph W. Tyler 於 1930 年代所作的八年研究。Tyler 曾提出六個步驟進行評鑑：(一)將目的或目標加以分類；(二)用行為術語定義目標；(三)尋找能顯示目標達成的情境；(四)發展或選擇測量的技術；(五)蒐集表現資料；(六)將表現的資料以行為表達的目標做比較。

經營導向模式強調資訊的蒐集，以滿足決策者作決定所需的資訊需求，此模式最為著名的當屬 Stufflebeam (1971, Stufflebeam & Shinkfield, 1985) 的 CIPP (context, input, process, product) 模式，其鑒於目標模式的不周全性，漸次發展出包括四階段的評鑑模式：情境、輸入、過程與結果評鑑，此四類評鑑並非需要一氣呵成全部執行，而是可配合所需而選擇採用。

消費者導向模式旨在發展產品的評鑑資訊，以便消費者從中選擇所需產品或服務，Scriven 所發展的產品檢核表可說是此模式之典型代表 (Stufflebeam & Shinkfield, 1985; Worthen, Sanders, & Fitzpatrick, 1997)。

專家意見導向模式所仰賴的是專家對於評鑑對象進行專業判斷，這類模式最常見到的是認可 (accreditation) 制度。Kells (1983: 9-10) 曾對美國的認可制

⁴以下評鑑模式之說明整理自 2002 年拙著《方案評鑑的緣起與概念》，載於《教師天地》，117, 26-31。

下一定義：「係一『自願的過程』，透過『非官方的學術團體』，採用『同僚評鑑』(peer evaluation)，以檢視被認可之機構，是否已達成『自我研究』(self-study)中所定的『目標』，並符合『評鑑的標準』。」從這樣的定義中，可以瞭解美國的認可制是一自願、自我管制的活動，通常由民間的專業組織而非政府承擔起教育品質維持與確保的工作，而認可工作的進行，除請外部的專業人員進行專家評鑑，亦側重機構內部的自我評鑑。現今認可制包含幾個重要要素：(一)公佈的標準、(二)機構的自我研究、(三)外部評鑑小組、(四)實地訪評、(五)實地訪評小組的報告，通常包含建議、(六)由一些知名人士組成小組審查報告、(七)最後的報告以及認可機構所做的認可決定 (Scriven, 1984, 引自 Worthen, Sanders, & Fitzpatrick, 1997)。美國許多的專業團體，如全國性的「高等教育認可審議會」(The Council on Higher Education Accreditation, CHEA)、 「全國師資培育認可審議會」(National Council for the Accreditation of Teacher Education, NCATE)，以及區域性的「中部各州學院與學校協會」(Middle States Association of Colleges and Schools)等均定期進行專業認可的工作。國內雖未實施正式的認可制度，唯在各級學校的校務評鑑概採此類評鑑模式。

抗詰導向模式係以計畫性的方式呈現贊成與反對雙方評鑑者的觀點。與許多評鑑模式試圖減少個人偏見的情形不同的是，此類模式知覺到偏見是無可避免的，因此力求偏見的平衡。Owen (1973) 與 Wolf (1975) 是此類模式的代表人物，他們修正法庭模式，由正反雙方評鑑者蒐集相關資料，於公聽會中交叉辯論。在這類模式中，處理抗詰程序的規則十分具有彈性，並以自由評鑑證據的方式代替繁複的證據規則，最後的評鑑結果則由仲裁者加以評斷 (引自潘慧玲譯，1989)。

最後一類的參與者導向模式，源自質性研究的傳統。自 1967 年開始有學者對於教育評鑑被機械式取徑所主導的情形加以回應，故於 1970、80 年代開始興起另一種評鑑取徑，強調以整體方式瞭解評鑑情境，偏好自然探究的方法，側重方案參與者的聲音，並回應利害關係人之需求。Stake (1978) 的回應式評鑑 (Responsive Evaluation)、Guba 與 Lincoln(1981)的自然式評鑑 (Naturalistic Evaluation) 均屬此類模式。

三、評鑑理論分析架構的形塑

評鑑是什麼？評鑑之異於其他領域之特性或本質是什麼？評鑑要具有其獨特性，當有什麼樣的知識基礎？易言之，評鑑不該只是應用的方法論，它當具備有別於其他領域的獨特知識。這些問題是有些學者在評鑑經過二十多年的發展後，卻成為「方法」獨霸，所致力的探究。從 1960 年代起勃興的評鑑領域，評鑑方法的書籍汗牛充棟，評鑑理論的論著卻罕見 (如：Cronbach et al., 1980; Shadish, Cook, & Leviton, 1991)，其中 Shadish、Cook 與 Leviton (1991) 的力作可說是一本對於評鑑理論所做的後設分析，他們企圖以五個面向----評價 (valuing)、知識建構 (knowledge construction)、知識使用 (knowledge use)、

社會方案運作 (social programming)、評鑑實務 (evaluation practice)，作為勾劃學者所提相關論述的依據，實則提供了一個相當完整的評鑑理論分析架構。

Shadish 在其 1997 年美國評鑑學會年會的理事長專題講演中，提出在理論的論述中，尚缺乏後設理論的語言類別與架構，以作為論辯之基礎，因之，他進一步精緻化其與 Cook、Leviton 於 1991 年所提的五個面向之用語——所有評鑑理論之論辯均可區分為我們如何評價、我們如何建構知識、評鑑如何被使用、評鑑客體如何運作與改變、在實際限制下實務如何作最妥當地執行等主題 (詳圖 1)。在上述五個語言類別中，又設計了細部的語彙。例如在評價部分，區分了描述評價、規範評價與後設評價。在知識建構部分，區分了確定性的不同層次；知識的不同類別 (如：描述性、因果性、說明性知識)。在使用部分，區分了工具性使用、概念性使用與說服性使用；長程使用與短程使用。在社會方案部分，區分了漸進變革式與激進式變革；方案變革、計畫變革與要素變革；新方案變革與現存方案變革。在評鑑實務部分，區分了所問問題的種類 (如：有關需求、執行、委託者、效果、影響、成本等的問題)；有關促進使用的作法；有關詢問誰的問題；有關評鑑者的角色；以及有關採用的評鑑方法等 (Shadish, 1998: 11)。

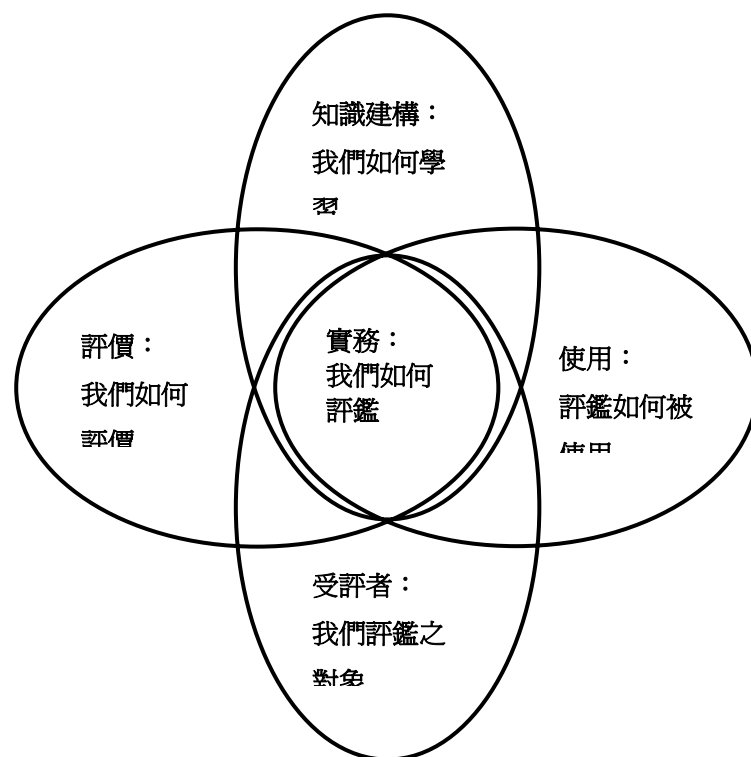


圖 1 評鑑理論的五個主題

資料來源：Shadish (1998: 2)。

對於 Shadish 及其同僚 (Shadish, 1998; Shadish, Cook, & Leviton, 1991) 所發展的評鑑理論分析架構，我認為甚具完整性，主要是因為這樣的架構，開拓了吾

人對於評鑑認識的視野。評鑑理論並非僅限於評鑑執行活動（即評鑑實務層面）的討論，其所涉及的非僅是評鑑模式及方法的選擇；相反地，評鑑理論牽涉了評鑑的整體活動，故而評鑑理論除包括模式、方法外，尚含科學哲學、公共政策、評價理論、使用理論等的論述。舉例而言，我們在評鑑活動中所探究的問題可有以下數項：

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

四、評鑑理論的發展

上述 Shadish、Cook 與 Leviton (1991) 及 Shadish (1998) 所提的分析架構雖極具分析的參考性，不過依據自己關懷的面向，以及議題所曾引發的爭論性，以下分就評鑑的本質與評價、知識論、方法論、評鑑的用途及其與決策的關連、方案運作與方案理論等五方面作探討。過去學者們在這五方面主張的差異，就其內裡，可發現主要係因研究典範的不同所致使。在社會科學中，典範的更迭引發學術大論戰，評鑑領域未能置身事外，亦呈現類似之景象，尤其後現代主義思潮的注入，使得許多原本認為理所當然的知識、真理與價值，產生了動搖！

（一）評鑑的本質及評價

評鑑之成其為評鑑，而非其他領域，尤其是評鑑之異於研究，究竟有何獨特性？這是此處想要探究評鑑本質的用意。翻開相關文獻，可看到對於評鑑意義的不同詮釋：有人界定為對於既定目的是否達成的評定；有人界定為有系統地蒐集與提供資訊，以便作決策；有人界定為判斷事物的價值；有人則將其等同於專業判斷。以上的不同界定，以「對於事物價值的評估是評鑑」最廣為接受 (Hopkins, 1989)。不過，對於這樣的定義，亦有學者提出異議，認為評鑑不僅在評估價值，亦在改進方案 (Cronbach et al., 1980)。這樣的主張與英國 Stenhouse 所提倡的教師為課程發展者與評鑑者之想法相符應。

面對不同的釋義，或許我們該加以釐析，這些釋義的基準點為何？Scriven (1967) 甚早在其 "The Methodology of Evaluation" 一文中，即指出評鑑的目標 (goal) 只有一個，那就是決定評鑑客體的價值，但角色 (roles) 可以多重，此概念類於後來 Worthen 及其同僚所區分的評鑑目的 (purpose) 與用途 (uses) (Worthen, Sanders, 1987; Worthen, Sanders, & Fitzpatrick, 1997)，他們同樣主張

評鑑目的是形成對於被評鑑者的價值判斷，可是評鑑用途，如以教育領域而言，可包括以下數項（Worthen, Sanders, & Fitzpatrick, 1997: 10）：

1. 促進教師彰權益能（empowerment），使其在學校經費分配上有較多的發言權。
2. 判斷某一學習領域的課程品質。
3. 認可達到最低認可標準的學校。
4. 決定學校反暴力方案的價值。
5. 滿足外部贊助機構對於取得學校方案效能報告的需求。

由上面的分析，可知評鑑的目的（目標）與用途（角色）是須加區分的，只是與其說判斷價值是評鑑的目的（目標），不如說是評鑑的本質；評鑑之異於其他活動，當在其具有價值判斷（value judgement）之本質。然而，如何作價值判斷？價值判斷之規準為何？Beauchamp（1982）曾提出評價（valuing）理論有三類：

1. 後設理論：對於評價本質與評價證成所做的研究；
2. 規範理論：擁護特定價值優先性的理論；
3. 描述理論：不特別擁護某一價值為最好，僅在描述價值的理論。

Scriven（1980）是評鑑界論及後設理論者，他甚且提出評價邏輯的四部曲：1. 選擇價值規準、2. 設定達到規準的表現標準、3. 測量在每項規準上的表現，並與標準作比較、4. 綜合結果以形成一價值陳述。除了後設理論的主張，Scriven對於價值的看法，受到 John Rawl 正義理論的影響，他的需求論隱含了評價的規範理論。與 Scriven 類似主張的是 House（1980）。House 在分析各種價值後，認為評鑑應遵循 John Rawl 的正義理論，他曾道：

評鑑究其本質是一政治活動，……所涉及的不只是觀念的陳述，而是分配的社會機制……評鑑不應只是真實的，也應是正義的。（House, 1980: 121）

相對於規範理論的是描述理論，雖然許多學者在其評鑑中並不刻意地談論價值，但傾向利害關係人導向評鑑的主張者，多是考量利害關係人的需求與看法，評價不在使用預設之價值，而在利害關係人的價值。Stake（1975）曾清楚表明說評鑑者應當描述價值，因為我們並無一個絕對正確的預設價值，在一個多元社會中，評鑑者不應將自己的倫理觀強加在所評鑑的方案上，評鑑者對於價值的判斷應訴諸於利害關係人。。

（二）知識論及方法論

何種評鑑知識為真、為有效的，身為評鑑先驅人物的 Donald T. Campbell，曾提出內在效度是檢證良好研究的規準，倡導實驗研究方法（Campbell & Stanley,

1963)，並致力於因果知識的追尋。為使研究結果類推於其他群體與情境，他更提出了外在效度的概念。對於判斷知識價值的權柄掌握在評鑑者手上的看法，隨著社會科學界典範的轉移，要重視參與者聲音的籲求愈來愈強烈。利害關係人導向、參與導向或是協同（collaborative）導向的評鑑，顛覆了評鑑者是一價值中立者與客觀超然者的想法。知識是否有唯一的判準？誰該來決定知識真實、實用性或有效性？評鑑中所呈現的是誰的知識？成為新一波評鑑者所關切的。

除了知識論，有關質量方法的論爭曾是評鑑場域中的大災難。就評鑑發展歷史觀之，1960年代所盛行的是量化評鑑方法，然好景不常，1970、80年代，以量化為主導的評鑑領域，就如其他社會科學一般受到猛烈的攻擊（經濟學除外）。質性方法力求於評鑑領域中取得正當性地位的力量，可說來自兩方面。第一個是長期以來，尤其是社會學領域，有關質量方法的論戰。社會學家在1970年代進入評鑑領域時，發覺質性方法不受重視，便將質量的論辯帶進，此類學者可如 M. Q. Patton。另一股力量則來自具量化方法訓練背景的學者，尤其是教育領域，他們拒斥量化方法的不適當知識論立場，而傾向質性方法的使用，這類學者可如 E. G. Guba 與 Y. S. Lincoln、R. E. Stake 等（Cook, 1997）。質量方法在評鑑領域的論戰歷時十多年，雖然具有智性的啟發，但學者們咸認就實用的立場言，質量方法均是評鑑應當使用的方法。評鑑戰場的停歇是許多學者的期盼，Cook（1997: 36）即曾言：「讓我們隱喻式地草擬一和平文件，並朝向更具生產性的評鑑戰場邁進」。只是這樣的論爭直至今日雖有歇息的景象，卻仍餘波蕩漾。

（三）評鑑的用途及其與決策的關連

當談到評鑑的用途時，過去學者專注的焦點在於評鑑結果如何被使用，其影響決策程度如何？代表較為早期評鑑思想的 Scriven 與 Campbell 咸認評鑑結果可以用來改進社會（Shadish, Cook, & Leviton, 1991），所以賦予評鑑工具性用途的意圖十分明顯，然政治運作力常介入決策歷程，Weiss（1973）是最早提出此種現象的觀察者。對於評鑑結果是否可以做為決策的依據，Weiss 持保留態度，她提出七種社會科學研究與決策關連的模式，⁵而認為啟蒙模式是社會科學研究最常用以進入政策領域的方式（Weiss, 1979）。

除了對於評鑑結果使用的關注，最近幾年開始跳脫上述拘限的思考，注意到評鑑在歷程中亦可發揮其用，例如：評鑑可培育成員能力（Fetterman, 1994）、可促進學校的發展（Hopkins, 1989）、可帶動組織的改進（Sanders, 2002）等。為了釐清評鑑的用途，可在不同的時間點發生，我將傳統上關注評鑑結果所發揮的用途，稱為「終結性用途」，而將近幾年側重歷程中評鑑所發揮的用途，稱為「過程性用途」。⁶以下分述幾個過程性用途的論述。

Fetterman 在其彰權益能評鑑的論述中，定義彰權益能評鑑為：「使用評鑑的

⁵ Carol Weiss 將社會科學研究在政策領域的使用情形分為七種模式：線性模式、問題解決模式、互動模式、政治模式、戰術模式、啟蒙模式與「研究為社會中智識事業的一部份」模式（詳潘慧玲，1999）。

⁶ 「過程性用途」一詞在 Patton（1997a, b）著作中曾提及。

概念、技術與發現，涵育改進與自我決定」(Fetterman, Kaftarian, & Wandersman, 1996: 5)，並說明彰權益能評鑑的五個面向：1. 訓練參與者執行他們的評鑑，如：能力培育；2. 評鑑者扮演促進者和教練的角色，而非評判者；3. 評鑑者為喪失權能者辯護，及/或支持喪失權能者為他們自己辯護；4. 闡明 (illumination)；5. 參與者的解放。由此觀之，此取徑倡導評鑑的主權在參與者手上、評鑑係由參與者執行、評鑑者僅為促進者或教練，故評鑑的重要功能便在教導參與者相關技能以實施評鑑。

Hopkins (1989) 在其 *Evaluation for School Development* 一書中，特別釐清「學校改進的評鑑」(evaluation of school improvement)、「為了學校改進作評鑑」(evaluation for school improvement)、與「評鑑即學校改進」(evaluation as school improvement) 等三個詞彙，並希冀評鑑除用來作為評鑑學校改進之情形、促進學校作改進，更可過渡到評鑑本身即是學校改進的一環。為了闡明這樣的想法，他發展了一個評鑑應用於學校發展的概念架構。此架構描繪了學校內部環境與外部環境。在內部環境中，以學校發展計畫為中心，這項計畫以學校脈絡之需求為出發，回應外部環境中全國性與地方性教育改革的要求，並將教師與學校發展的各项革新措施含納其中。評鑑在整體架構中的角色是作為學校發展計畫的回饋機制。

從組織的觀點看，評鑑的發展朝向建立能力 (building capacity) 與制度化 (institutionalizing)，Sanders (2002) 認為尚不足夠。因為建立能力在組織中多是外加的任務，且常間歇性地執行，故能帶動組織發展的成效是片段、有限的；而評鑑制度化的努力雖使評鑑被組織所肯認，但制度化並不必然導致評鑑的被使用，或是全組織均能珍視評鑑。故而，Sanders 提出主流化 (mainstreaming) 的概念，讓評鑑成為組織中日常運作的一部份，如此方有帶動組織發展的可能。

(四) 方案運作及方案理論

評鑑是對評鑑客體展開探究與評估的活動，一般人在作如此解讀時，常將焦點置於「如何作評鑑」上，而忽略了評鑑客體 (如：方案)。事實上，方案的執行，本身就是一種介入，有其預期的介入成效。因之，方案的運作是否能帶來改變，所帶來的改變是點滴式的，還是大幅度的？方案可以為社會解決問題嗎？這些是過去學者們談論方案運作所關心的重點。唯除此外，此處還想引入另一個重點，那就是探究方案運作之能產生效果，其間歷經何種運作機制？Weiss (1998)、Chen (1990)、Chen 與 Rossi (1992) 等人試圖以方案理論揭開過去忽視的「黑箱」，為介入的方案為何產生成效，尋找有力的理論依據；扭轉評鑑過度「方法導向」的現象，使之趨向「理論導向」。

在論及方案運作是否帶來改變時，Cronbach、Rossi 等人知覺到社會方案受到政治之影響，逐步的改變有助於短程社會的變革 (Berk & Rossi, 1977; Cronbach, 1982)；激進式的改變則具長程之效果。只是較為可行與較為實際的是方案點滴式的變革。這與 Scriven (1972) 認為評鑑是為方案找出解決問題的方法，而執

行這樣的方法，便有助於社會改變的論調有所不同。Cronbach 與 Rossi 的論點顯得較為務實，更能留意方案運作所受到的政治環境影響。

對於方案理論的重視，Weiss 是一位早期的呼籲者。方案理論導向的評鑑主要有三個目的：1. 回答有關方案的因果問題；2. 說明方案如何運作；3. 提供改進的建議 (Smith, 1994)。因此，方案理論導向的評鑑可說是一種介入研究的形式，其中在經過介入的處理後，實徵性地檢證其因果機制。這樣的作法，有助於我們較為確切地掌握方案介入之所以會產生效果的原因，唯其狹隘地關注因果連結的類推，也受到其他學者的挑戰 (如：Patton, 1989)。

參、反思國內教育評鑑發展的經驗

教育評鑑在國內發展時間的長短類於美國，但不管在評鑑學術的多樣態上，或是實務推動的專業性上，卻都有所不足。反思過去一路走來的足跡，分就評鑑概念的應用、評鑑的專業化以及評鑑的學術風貌等方面作討論。

一、評鑑概念過於窄化

國內推動多年的評鑑，其樣態顯得十分單一，主要是因為對於評鑑概念的認識不夠充分，窄化的概念造成評鑑實務的推動出現諸多問題。

(一) 評鑑客體：以校務評鑑為最大宗

如果從民國 64 年全面性推動的大專校院評鑑開始算起至今，進行最多的教育評鑑可說是校務評鑑。教育評鑑依評鑑客體之不同，原可有方案評鑑、課程評鑑、教材評鑑、人員評鑑等，但因大學校務評鑑的推動，民國 64 年之後，即陸續於其他教育階段實施校務評鑑，這樣的作法又因校長遴選制度的變革，校長辦學績效成為遴選的參照依據，使得各縣市中小學校務評鑑年年進行。

相較於英、美，美國對於教育政策、方案的評鑑十分頻繁，以補償教育的「提早開始」(Head Start) 方案為例，其前後就曾做過多次的評鑑；另分權化下的教育措施——特許學校，亦有數州進行其成效評鑑，如：西密希根大學評鑑中心為密希根州所做的評鑑 (Horn & Miron, 2000)。至於英國，教育評鑑與課程發展相連結，故課程評鑑作得很多，評鑑也就不會只是一種績效責任的考核，謀求改進與發展在英國脈絡裡是被強調的。

(二) 評鑑模式：主要採擷認可制的作法

學術與教育措施的殖民性格，一直是台灣的「宿疾」，目前盛行於國內的評鑑作法即採自美國的認可制。認可制是所有目前已發展之評鑑模式中的一種，且此模式在 Stufflebeam (2000) 認為 21 世紀將仍適用的 9 個模式中，所得到的評價最低。Stufflebeam 以教育評鑑標準聯合委員會 (Joint Committee on Standards for Educational Evaluation) 所發展的效用性 (utility)、可行性 (feasibility)、適當性 (propriety) 及精確性 (accuracy) 作為依據，針對 20 種評鑑模式作評估，指出

決定/績效、消費者導向、認可制、應用導向、案主中心、民主導向、建構導向、個案研究、結果監控/附加價值等 9 個模式是在 21 世紀最具希望會被持續應用的。認可制在 9 個模式中得到五個等級中的中間一級（60 分）。這樣的評價雖有 Stufflebeam 主觀好惡的成分在，唯其顯現了評鑑模式本可多元，因應評鑑目的之不同，國內評鑑模式不當僅限於認可制的採用。

（三）評鑑目的：過於強調績效的評估

校務評鑑以及這幾年也實施不少的單項評鑑（如：體育評鑑、營養午餐評鑑、兩性教育評鑑等），都是為了瞭解執行成效所做的評鑑。台北市自民國 90 年所推動的國中小校務評鑑，在規擬的文件中，雖也強調自我改進的目的，不過實際上，在評鑑結束後，所公布的評鑑結果如何成為學校改進與發展的回饋機制，並未具體見到。由於績效的評估過度被強調，使得學校實務工作者「聞評色變」，而文書檔案整理所得到的代價，並未真正回饋於學校的改進。在洪梓祥（2002）所做的台中縣國民小學實施校務評鑑的調查中，即有學校人員反應校務評鑑的內容過於偏重書面資料。

（四）評鑑方式：忽略內部評鑑的功能

內部評鑑與外部評鑑因應著評鑑目的之不同，可作彈性選擇。一般而言，內部評鑑較適用於形成性評鑑，謀求方案之改進，而外部評鑑則較適用於總結性評鑑，以瞭解方案的整體成效。只是多年來的教育評鑑因著重於績效責任的講求，故多執行外部評鑑。雖說校務評鑑分兩階段進行，第一階段由學校先自評，第二階段再由外部的學者專家進行實地訪評，唯評鑑結果係以訪評者之判斷為準，因之，內部評鑑之功能並無法完全彰顯。

自 1980 年代起，西方國家興起「學校重整」(school restructuring) 運動，學校本位管理成為其中的一項重要革新策略，而學校本位評鑑應運而生。Nevo (1995) 以以色列經驗陳述了學校本位評鑑的概念與作法，他提到學校本位評鑑分為四個階段：1. 透過工作坊進行基本訓練；2. 建立學校內部評鑑團隊；3. 評鑑建制化；4. 建立內外部評鑑對話。由於學校本位評鑑強調學校的自發性，它並非只是對外部評鑑所做的回應，故內部評鑑十分重要。尤其內部評鑑可使學校人員免除績效考核的壓力，實際面對學校問題謀求解決，對於學校的改進與發展，有其重要功能。雖說學校本位評鑑可以只採用內部評鑑，唯為增加檢視學校的角度與視野，在內部評鑑的實施步入正軌後，可結合外部評鑑的運用，這正是 Nevo 所建議的學校本位評鑑第四階段的作法。

二、評鑑專業化有待努力

教育評鑑在近幾年並未因受重視程度的增高，而步向專業化。以前述 Worthen 等人（1997）所發展的十項檢核評鑑專業化的規準來看，除了「評鑑所需的獨特知識與技能」及「評鑑實施的標準」兩項可以借用自國外而勉強過關外，

其餘八項規準均未能達到。回顧評鑑在國內的實務推動，可知在評鑑組織、指標、人員、時程、結果的處理，以及追蹤評鑑上，均有改善的空間。有鑒於校務評鑑係採自美國認可制的作法，故在作以下之檢討時，乃以美國認可制之作法為標竿。

（一）評鑑組織

美國認可制採自願方式實施，而評鑑組織為民間專業團體，例如在大學階段，認可團體分兩類：一為機構的認可；一為學門的認可。負責機構的認可，又可分全國性認可協會與區域性協會，其中區域性協會共有六個：「中部各州學院與學校協會」、「新英格蘭學院與學校協會」、「北中區學院與學校協會」、「西北區學院與學校協會」、「南區學院與學校協會」、「西區學院與學校協會」。至於學門的認可則由各學術專業認可協會負責（顏若映，1997；蘇錦麗，1997）。國內過去所推動的校務評鑑多由政府發動，委託學校承辦，並沒有常設的專業評鑑團體負責，導致評鑑的進行常因承辦單位的不同，而有差異的品質。直至去（2004）年終有「台灣評鑑學會」的成立，其接受教育部委託承辦今年度之大學評鑑，唯推動成效如何，尚待評估。

（二）評鑑指標

美國認可制中的一項標準程序是事前公布評鑑標準或指標，而其評鑑標準或指標通常是歷經多年方研議完成。以全國師資培育認可審議會（NCATE）為例，其認可標準四十多年來已歷經多次修正，每次修正約需兩年，修訂通過後 18 個月正式生效，修訂程序係由師資培育研究學者、師資培育單位專業人員、教師組織代表共同參與草擬，後經民主程序送「單位認可委員會」審核，審核通過後交「執行董事會」成員三分之二以上同意始得通過（張德銳，1997；NCATE, 2002）。反觀國內，評鑑指標之研擬時間過於倉促，有些縣市公布評鑑指標時間亦過慢，使得受評學校沒有充分時間作準備。

有關評鑑指標，過去多以行政業務作為切入點，故設計有整體校務、教務、學生事務、總務、輔導等項目。近年來，設計思維漸作改變，開始從學校效能展現之面向規擬，以台北市九十年國中校務評鑑為例，評鑑內容即分為「組織與行政」、「課程與教學」、「環境與設備」、「教師與專業」、「訓導與輔導」、「家長與社區」等六項（湯志民，2002）；國小校務評鑑則分為「行政管理」、「課程教學」、「專業發展」、「訓導輔導」、「家長參與」、「學校特色」等六大類目。而高雄市嘗試將全面品質管理的理念融入校務評鑑，於九十二學年度起以「學校評鑑」取代「校務評鑑」，國中階段評鑑內容分「校務領導」、「校務推動策略」、「資源規劃與運用」、「教學與評量」四領域；國小階段評鑑內容則分「學校經營」、「教學實施」、「學校特色」三領域。

（三）評鑑人員

為培育專業的評鑑人才，美國有許多大學均設有研究所階段的評鑑學程；而

實施認可制的專業團體，對於進行實地訪評人員，亦有其培訓方式，以 NCATE 為例，其目前儲備之評鑑委員約有四百位，其中師資培育者、教師、教育決策者/專門領域專家各佔三分之一。評鑑委員為無給職，三年一任，得連任一次，之後如想再任評鑑委員，則要接受另一訓練（張德銳，1997；NCATE, 2002: 93）。

至於國內，訪評委員均為臨時編組，中小學的評鑑，訪評委員多為大學教授、中小學資深校長，有些縣市則另邀請家長與教師會代表參加。這些委員甚或瞭解教育，但是否具備評鑑知能則未得知，而評鑑在執行時，說明會的辦理卻又鮮能發揮提昇訪評委員評鑑素養的功能。此外，評鑑人員的倫理信條十分重要，美國 NCATE 尚訂有迴避條款（如：受評學校有指導過的研究生、在過去十年內曾擔任過受評學校顧問等情形均須迴避），另為掌握評鑑人員之品質，在訪評後，受評學校需對訪評委員的表現作評鑑；訪評委員亦需對訪評同儕與訪評小組召集人的表現作評鑑（NCATE, 2002: 156）。相對照於國內，受限於地區狹小，人際互動較為頻繁，使得迴避原則不易做到。而這些均對評鑑的專業性造成影響，難怪學校實務工作者會表示評鑑結果的公正性與客觀性受到質疑，如洪梓祥（2002）對於台中縣國小人員所做的調查發現；以及有評鑑委員應由具專業素養、公正人士擔任之期望，如陳劍賢（2001）對於台東縣中小學教師調查研究之建議。

（四）評鑑時程

美國認可制一般分兩階段進行，先由學校進行自我評鑑，再由外部專家進行實地訪評。自評與訪評時間均排有充分的準備時間。以中部各州學院與學校協會為例，其自評時間通常不少於一個學年，而自評距離訪評之時間為三個學期

（Middle States Commission on Higher Education, 2000: 14）。另 NCATE 之作法是受評單位至少要在兩年前提出接受認可的申請，接下來要通過認可的第一道關卡是「先決條件的滿足」，受評單位在 NCATE 人員的指導和協助下，在訪評小組實地評鑑之前的 18 個月提出書面資料，說明自己在各項條件上的達成情形，之後便接受訪評，亦即初始的認可審查，如通過，則每 5 年接受一次持續的認可審查。在訪評的天數安排上，則有四天的長度（張德銳，1997；NCATE, 2002）。對於國內的受評單位而言，通常被給予進行自評的時間，多達兩個月，少則僅一個月；訪評時間過去曾短至半天，現則以一天為多數，少數如台北市國小校務評鑑則有兩天。綜言之，國內不管是自評或訪評時間上的安排，較之美國，均顯不足。

（五）評鑑結果的處理

對於美國認可制而言，評鑑的結果是作為是否通過認可的依據，國內教育評鑑雖採認可制的作法，但最後並無認可之動作，故如何處理評鑑結果是需謹慎思考的。歷年來的作法，有些將評鑑結果公布，有些則不公布；有些做為經費補助之依據，有些則束諸高閣，未加利用。一般而言，當評鑑要做為自我改進時，評鑑結果是不加公布的；若要進行績效評估，則需公布評鑑結果。此外，評鑑報告

出爐後，先送給學校一份，學校如有不同意見，可以進行申訴，這是美國認可制均有的設計，國內則少見。台北市自民國九十年開始辦理之國民小學校務評鑑引進了此種作法，是一良好開端。

(六) 追蹤評鑑

為持續追求進步，進行改善，追蹤評鑑是必需的，前所述 NCATE 在初始的認可通過後，要求受評單位每隔 5 年接受持續的認可，即是一種追蹤評鑑。國內往往作完評鑑就了事，尚未建立追蹤評鑑之制度。

三、學術風貌不夠多元

學術成果的發表形式包括專書、研究報告、期刊論文及學位論文等，其中的學位論文，一般係以較長時間所作的有系統探究，故雖未能代表學術領域耕耘之全貌，卻也能發揮管窺之效。為瞭解評鑑領域的研究情形，本文以「全國博碩士論文摘要」及「中華博碩士論文摘要」資料庫為搜尋範圍，以「教育評鑑」為關鍵字，結果篩選出與學校教育相關之碩博士學位論文計 192 篇。

自 1978 年至 2003 年，教育評鑑碩博士論文的生產量與年代的先後呈現了正向相關，亦即隨著時間的愈接近，論文出現的數量愈多。在 1970 年代僅出現 2 篇論文；緊接著 1980 年代出現 14 篇；1990 年代論文開始增多，共出現 47 篇；至 2000 年代，在近 4 年中，生產之篇數最多，共達 129 篇。

如進一步以評鑑客體做為碩博士論文探究的焦點，可發現研究主題共有九類，其中以教師評鑑（32.8%）與學校評鑑（24%）為最大宗；其次則為單項評鑑（20.3%）、課程評鑑（7.3%）、教材評鑑（7.3%）、校長評鑑（6.8%）；主任評鑑、學生同儕評鑑與未涉及任何評鑑客體者分別有 1 篇，各佔 0.5%（詳表 1）。其中教師評鑑之主題大多為 2000 年代之作品，易言之，除了 1980 年代的 3 篇，1990 年代的 9 篇外，其餘 51 篇（81%）均為 2000 年至 2003 年之作，由此可知，教師評鑑在這幾年的受重視，而因其為新興議題，是一項尚未推動的制度，故對於教師評鑑實施方式、指標建構與可行性的探討便熱門起來。與教師評鑑相較之下，學校評鑑自 1970 年便開始有論文之撰寫（1 篇），1980 年代零星出現 4 篇，1990 年代有 12 篇，到了 2000 年代，可能是因校務評鑑的大力推動，生產之論文增多起來，共 29 篇（63%）。

除了教師評鑑、學校評鑑外，在這波教育改革脈絡下，因應著部分課程決定權的下放與教學革新的推動，有關課程評鑑之主題開始出爐，而教材（包括教科書、網路教材、學習網站等）的評鑑，也紛紛呈現。與課程、教材評鑑在數量上相差不多的是校長評鑑。原本校長遴選制度的變革，將校長辦學績效納入遴選考量的作法當會刺激校長評鑑著作之生產，然因大多縣市以校務評鑑代替校長評鑑，使得校長評鑑論文數量遠不及教師評鑑。另值得一提的是，通常教育人員評鑑除了校長、教師外，尚含教育局長，唯截至目前為止，國內還未有關於教育局長評鑑的研究。

再者，此處所歸類的單項評鑑係指非全校性的，僅是針對某項業務或某項實施方案所做的評鑑，諸如特殊教育、輔導工作、通識教育、教學及評量改進班、戶外環境教育研習之評鑑等。事實上，國內單項評鑑之作法與學校評鑑雷同，僅評鑑範圍多寡不同而已。故就評鑑實施的相似度作進一步之歸併，學校評鑑、單項評鑑可歸為一類；校長、教師、主任評鑑可歸為一類，如此看來，本文所分析之論文，其處理的評鑑客體以學校/單項評鑑拔得頭籌（44.3%）；教育人員評鑑居次（40.1%）。

就上述逐漸有著多樣化主題之論文，並不代表其探究的概念是多元的。在這192篇的論文中，不管研究的對象為何，其內容多在討論指標的建構、實施的方式或是學校人員對於實施方式或現況的意見調查。對於國外已經開發的論題，如評鑑的政治、倫理；評鑑方法論；評鑑的使用；檢證評鑑知識的規準；新興的評鑑模式（如彰權益能評鑑）等，鮮少觸及。而即使在處理不同的評鑑客體，背後的思維多落入認可模式的窠臼，故不管是評鑑學校、單項業務/方案、課程、教材或教育人員，所設計的作法均為建構指標，然後進行自評與訪談。多年來的努力似無拓寬評鑑模式視野之企圖，對於多元模式除能加以引介外，更希冀能在實地場域中進行實做。Fetterman 在 1994 年所提出的彰權益能評鑑，雖因典範之轉移而漸能為人所接受，使得他在美國評鑑學會理事長專題演講中能以之為題，唯不可諱言的是，仍有人對於該模式抱持懷疑態度，Stufflebeam (1994) 即曾以客觀評鑑角度提出批判，而 Patton(1997b)、Scriven(1997) 亦有與 Fetterman(1997b) 往返之文章論辯。彰權益能模式就在 Fetterman 大力推動下，除在許多學校、社區場域中操作，也透過學術論辯，使其要表達的概念愈來愈清楚。學術的開展本應如此，只是國內尚須激發更多的學術活力與想像力，來進行理論與實踐的探究。

表 1 碩博士論文研究主題之分佈

項目	學校評鑑	單項評鑑	課程評鑑	教材評鑑	校長評鑑	主任評鑑	教師評鑑	學生同儕評鑑	無指涉客體	總計
篇數	46	39	14	14	13	1	63	1	1	192
百分比	24%	20.3%	7.3%	7.3%	6.8%	0.5%	32.8%	0.5%	0.5%	100%

肆、展望下一代的教育評鑑

對於評鑑未來的發展，有人在 Guba 與 Lincoln (1989) 所提出的第四代評鑑後，以行動研究作為第五代的評鑑（如：Merten, 2004）。然而，置身一個多元社會中，評鑑將更具多元性與包容性，「行動研究」似無法涵攝整個評鑑的未來發展趨勢，故下面我以四個展望點出下一代教育評鑑的形貌。

一、評鑑將更加科際整合化

評鑑的應用，除了教育、心理與健康領域之外，亦擴展至其他領域，故評鑑究為一門學術領域或一門超學術領域，學者們嘗有討論。由於評鑑並不限於某一學門的應用，故 Scriven (1991) 稱其「超學門」(transdiscipline) 領域，就像邏輯、設計、統計等工具學術領域一樣。

在評鑑學會中，成員通常並非單純僅是某一領域之人士；在美國評鑑學會中，便有教育、心理背景的成員；而在美國評鑑學會、加拿大評鑑學會、中美洲評鑑學會、澳洲評鑑學會以及歐洲評鑑學會於 1995 年所聯合舉辦之國際評鑑會議，與會的 1600 多位評鑑者便具不同之背景。

Smith 在 2001 年為學者們的評鑑回顧文撰寫緒論時，提出他的觀察：許多作者指出一些其他領域的想法注入了評鑑領域，諸如資訊科技、工程、管理科學、人力資源發展、環境科學、法律等 (Smith, 2001)。而科技的進展將可能改變評鑑資料的蒐集與分析方法；也可能改變評鑑報告的形式，電子視聽報告會愈來愈多。評鑑變成「說故事」，以虛擬實境的方式呈現，這在未來都有可能成真。故而，不論就評鑑的學術屬性、學術團體成員的背景訓練以及評鑑領域發展到其他學門知識的匯入，皆顯見其科技整合之勢愈趨強烈。

二、評鑑將愈趨專業化

要步入專業，執業人員的認證是必要條件。在美國，教師品質的掌控，係透過培育機構認可 (accreditation) 與個人認證 (certification) 兩種制度。對於機構的認可，是確保養成教育品質達到最低的認可標準；對於個人的認證，是為審視個人已具備從事該門專業的能力。評鑑要成為專業，亦須步此之途。

目前在評鑑領域發展較為成熟的美國，誠如前述，其專業化之程度在專業會員的汰選、專業學會對於評鑑人員培育學程具影響力以及評鑑人員的認證等方面，尚未能做到。Bickman (1997) 曾提出美國評鑑學會可致力於專業化的努力，目前已在探索自願的認證過程，加拿大評鑑學會事實上已開始此一作法。另則同時考量認可制度，選擇一套課程與活動以訓練有能力的評鑑者，確定每位評鑑者具有共同的受教經驗。對於認證與認可的作法，有人持抗拒態度，認為每一評鑑均是獨特的，所進行之程序亦十分多樣，不易規範出共同標準，而且即使認證也可能有不良的評鑑者存在。雖有如此的異議，美國評鑑學會仍秉持邁向專業化的原則，繼續推動相關規擬工作，故預期數年後，將有可能見到推展的成效。

三、評鑑將更具包容性

社會科學界的多元典範並存，將使評鑑愈趨包容性。過去所見的質量大戰，今後將不復見。評鑑所運用的不僅是量化方法，質性方法亦被肯認，質量方法的相互補充運用，有助於評鑑研究的充實。另評鑑將彰顯其民主之功能，不僅講求評鑑參與面的擴大，回應利害關係人的需求，對於社會或機構中不符平等、正義之情事，亦將揭示。評鑑除了績效責任及自我改進的功能，提昇參與者的權能，使其能自主、解放，都是評鑑可以發揮的功能。

四、評鑑將成為組織發展的內建機制

為了講求績效責任，教育當局常透過其深植於實證主義典範之結構性預設（structural assumptions），希望採用工具理性策略，獲致最大的教育成果，評鑑在此思維下，只會淪為管理主義下的產物。在體認今日環境中複雜的政治因素與權力關係，讓教育組織能夠發展，學校成員能夠彰權益能，需要扭轉管理導向的評鑑概念。Sanders（2002）敘述其倡議評鑑成為組織內建機制的想法，與彰權益能評鑑、參與式評鑑有相關，但不全然相同；內建機制觀的特點是組織文化珍視評鑑，組織的運作使評鑑得以持續，且持續使用評鑑增進組織效能。如果評鑑內建機制化，評鑑便成為一項工具，透過它，組織可以學習、更新，發揮效能，甚而達到卓越。評鑑要能內建機制化，它必須發展成組織中的一項核心價值。展望未來，評鑑將發揮較往昔為多的功能，幫助組織的成長與進步。

參考文獻

- 洪梓祥（2002）。台中縣國民小學實施校務評鑑之研究。台中師範學院國民教育研究所碩士論文，未出版，台中市。
- 張德銳（1997）。大學教育學程及其評鑑。載於陳漢強（主編），**大學評鑑**（頁327-368）。台北市：五南。
- 陳劍賢（2001）。國民中小學校務評鑑之研究--以台東縣為例。台東師範學院教育研究所碩士論文，未出版，台東市。
- 湯志民（2002）。台北市國民中學校務評鑑之評析。**初等教育學刊**，11，25-50。
- 黃光雄（1989）。評鑑導論。載於黃光雄（主編），**教育評鑑的模式**（頁1-52）。台北市：師大書苑。
- 潘慧玲（1999）。教育研究在教育決策中的定位與展望。**理論與政策**，13(2)，1-15。
- 潘慧玲（2002）。方案評鑑的緣起與概念。**教師天地**，117，26-31。
- 潘慧玲（譯）（1989）。Owens 與 Wolf 的抗詰式的評鑑途徑。載於黃光雄（主編），**教育評鑑的模式**（頁301-322）。台北市：師大書苑。
- 盧增緒（1995）。論教育評鑑觀念之形成。載於中國教育學會（主編），**教育評鑑**（頁3-59）。台北市：師大書苑。

- 顏若映 (1997)。大學評鑑模式 (一) --認可制度。載於陳漢強 (主編), **大學評鑑** (頁 219-256)。台北市: 五南。
- 蘇錦麗 (1997)。高等教育評鑑: 理論與實際。台北市: 五南。
- Beauchamp, T. L. (1982). *Philisophical ethics: An introduction to moral philosophy*. New York: McGraw-Hill.
- Berk, R. A. & Rossi, P. H. (1977). Doing good or worse: Evaluation reexamined. In G. V. Glass (Ed), *Evaluation studies review annual* (Vol. 2, pp. 77-89). Beverly Hills, CA: Sage.
- Bickman, L. (1997). Evaluating evaluation: Where do we go from here. *American Journal of Evaluation*, 18(1), 1-16.
- Campbell, D. T. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Chelimsky, E., & Shadish, W. R. (Eds.). (1997). *Evaluation for the 21st century*. London: Sage.
- Chen, H. T. (1990). *Theory-driven evaluation*. Newbury, CA: Sage.
- Chen, H. T., & Rossi, P. H. (1992). *Using theory to improve program and policy evaluations*. Westport, CT: Greenwood.
- Cook, T. D. (1997). Lessons learned in evaluation over the past 25 years. In E. Chelimsky, & W. R. Shadish (Eds.), *Evaluation for the 21st century* (pp. 30-52). London: Sage.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco, CA: Jossey Bass.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1980). *Toward reform of program evaluation: Aims, methods and institutional arrangements*. San Francisco, CA: Jossey Bass.
- Dubois, P. H. (1965). *A test dominated society: China 1115B.C.-1905A.D.*. Proceedings of 1964 Invitational Conference on Testing Problems (pp. 3-11), Educational Testing Service.
- Dubois, P. H. (1970). *A history of psychological testing*. London: Allyn & Bacon.
- Fetterman, D. M. (1994). Empowerment evaluation: Presidential address. *Evaluation Practice*, 15(1), 1-15.
- Fetterman, D. M. (1997a). Empowerment evaluation and accreditation in higher education. In E. Chelimsky, & W. R. Shadish (Eds.), *Evaluation for the 21st century* (pp. 381-395). London: Sage.
- Fetterman, D. M. (1997b). Empowerment evaluation: A response to Patton and Scriven. *Evaluation Practice*, 18(3), 253-266.
- Fetterman, D. M. (2000). *Empowerment evaluation*. London: Sage.

- Fetterman, D. M., Kaftarian, S. J. & Wandersman, A. (1996). *Empowerment evaluation: Knowledge and tools for self-assessment & accountability*. Thousand Oaks: Sage.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Horn, J., & Miron, G. (2000). *An evaluation of the Michigan charter school initiative: Performance, accountability, and impact*. The Evaluation Center, Western Michigan University.
- Hopkins, D. (1989). *Evaluation for school development*. Philadelphia, PA: Open University Press.
- House, E. R. (1980). *Evaluating with validity*. Beverly Hills, CA: Sage.
- Kells, H. R. (1983). *Self-study process: A guide for post-secondary institution* (2nd ed.). New York: Macmillan Publishing.
- Merten, B. V. (2004). *Five generations of evaluation: A meta-evaluation*. Retrieved May 11, 2004, from web.syr.edu/~bvmarten/eva/act.htm
- Middle States Commission on Higher Education (2000). *Designs for excellence: Handbook for institutional self-study*. Philadelphia: Author.
- National Council for the Accreditation of Teacher Education (2002). *Handbook for accreditation visits*. Washington, DC: Author.
- Nevo, D. (1995). *School-based evaluation: A dialogue for school improvement*. New York: Pergamon.
- Patton, M. Q. (1997a). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (1997b). Toward distinguishing empowerment evaluation and placing it in a larger context. *Evaluation Practice*, 18(2), 147-163.
- Patton, M. Q. (1989). A context and boundaries for a theory-driven approach to validity. *Evaluation and Program Planning*, 12(4), 375-377.
- Popham, W. J. (1993). *Educational evaluation* (3rd ed.). London: Allyn and Bacon.
- Sanders, J. R. (1997). Cluster evaluation. In E. Chelimsky, & W. R. Shadish (Eds.), *Evaluation for the 21st century* (pp. 396-404). London: Sage.
- Sanders, J. R. (2002). Presidential address: On mainstreaming evaluation. *American Journal of Evaluation*, 23(3), 1-19.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.

- Scriven, M. (1972). The methodology of evaluation. In C. H. Weiss (Ed.), *Evaluating action programs: Readings in social action and education* (pp. 123-136). Boston: Alley & Bacon.
- Scriven, M. (1980). *The logic of evaluation*. Inverness, CA: Edgepress.
- Scriven, M. (1984). Evaluation ideologies. In R. F. Connor, D. G. Altman, & C. Jackson (Eds.), *Evaluation studies review annual* (Vol. 9). Beverly Hills, CA: Sage.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed). London: Sage.
- Scriven, M. (1997). Empowerment evaluation examined. *Evaluation Practice*, 18(2), 165-175.
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1-19.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, MA: Sage.
- Stake, R. E. (1975). To evaluate an arts program. In R. E. Stake (Ed.), *Evaluating the arts in education: A responsive approach* (pp. 13-31). Columbus, OH: Merrill.
- Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher*, 7, 5-8.
- Smith, N. L. (1994). Clarifying and expanding the application of program theory-driven evaluations. *American Journal of Evaluation*, 15(1), 83-87.
- Smith, M. F. (2001). Evaluation: Peview of the future #2. *American Journal of Evaluation*, 22(3), 281-300.
- Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in Education*, 5, 19-25.
- Stufflebeam, D. L. (1994). Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. *Evaluation Practice*, 15(3), 321-338.
- Stufflebeam, D. L. (2000). Foundational models for 21st century program evaluation. In D. L. Stufflebeam, A. J. Shinkfield, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed.) (pp. 33-83). Boston: Kluwer Academic.
- Stufflebeam, D. L. & Shinkfield, A. J. (1985). *Systematic evaluation*. Boston: Kluwer-Nijhoff.
- Stufflebeam, D. L. Madaus, G. F., & Kellaghan, T. (2000). *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed.). Boston: Kluwer Academic.

- Weiss, C. (1973). Where politics and evaluation research meet. *Evaluation, 1*, 37-45.
- Weiss, C. (1979). The many meanings of research utilization. *Public Administration Review*, September-October, 426-431.
- Weiss, C. H. (1998). *Evaluation* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Worthen, B. R. & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.
- Worthen, B. R. & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.
- Worthen, B. R., Sanders, J. R. & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practical guidelines* (2nd ed.). New York: Longman.