

# Scriven 評鑑理論創新觀點之省思

邱慧芳

國立台北科技大學技術及職業教育所博士生

曾淑惠

國立台北科技大學技術及職業教育所副教授

## 摘要

本文旨在探討 Scriven 過去三十多年迄今在評鑑理論中的主張及其創新觀點，並藉由 Scriven 評鑑理論創新觀點的詮釋，提出在教育中的省思。透過文獻發現，Scriven 對評鑑有如下的主張：評鑑者需要具備多重技能、評鑑是一門學科、各領域皆有評鑑的需求等。而 Scriven 在評鑑理論中的創新觀點，包括：目標中立評鑑、消費者導向評鑑，及後續的形成性評鑑和總結性評鑑、業餘評鑑和專業評鑑，內在評鑑、收益評鑑與混合式評鑑，以及學科內評鑑、預測評鑑等，說明了評鑑應包含多元向度的層面，且涉及多種測量的等級，需使用多種方法的思維。

最後，根據 Scriven 的創新觀點提出教育中的省思與看法：(一)進行目標中立評鑑的彈性與問題，(二)形成性與總結性評鑑的謬誤，(三)對後設評鑑的迴響，(四)預測評鑑的指標與準確度，(五)消費者導向評鑑的優窳。希望藉由本文之探討，提供相關領域學者及評鑑實務工作者，在進行相關研究或制定策略時參考。

**關鍵詞：**消費者導向評鑑、形成性評鑑、總結性評鑑、預測評鑑

## 壹、前言

Michael Scriven 於 1928 年生於英國，1948 至 1950 年間在澳洲墨爾本大學就讀學士學位，曾在 Claremont 大學的行為與組織科學研究所任教，榮獲數學榮譽博士及澳洲墨爾本大學的數理邏輯哲學榮譽博士，並於 1956 年在牛津大學取得哲學博士學位。他曾任教於美國、澳洲、紐西蘭等地的數學系、哲學系、心理學系、歷史與哲學科學系，及教育學系等，其中包括在美國加州柏克萊大學任教了 12 年。Scriven 也曾任職於帕拉阿圖的行為科學研究中心、艾伯塔大學的理論心理學研究中心、普林斯頓的教育測驗中心 (ETS)、聖塔芭芭拉的民主體制研究中心，和澳洲的社會科學學院，並曾在哈佛大學的國家科學基金會擔任研究員，創立墨爾本大學社會心理研究中心，並主持牛津大學社會心理研究，此外，Scriven 也曾擔任美國教育研究協會 (American Educational Research Association) 和美國評鑑協會 (American Evaluation Association) 的前總裁

(CGU, 2008)。自 2004 年開始，Scriven 在西密西根大學擔任哲學教授並兼任評鑑中心副主任。

Scriven 曾出版四百本以上的刊物，含括批判思考、技術研究、電腦研究等範疇，最著名的就是評鑑。我們可以在 42 種以上的期刊中，看到 Scriven 的文章，除了上述領域，甚至在精神病學方面也有所涉略。Alkin (2004) 指出，這世紀以來，大約有二十種已知的領域使用評鑑，範圍包括當代平面藝術的藝術評論、調酒美學 (品酒)，餐廳檢視、從專業的製造設計評論產生的生產評鑑、人事評鑑、政策研究、績效評估 (著名的有學生、戰士、運動家)、方案評鑑、車輛的道路駕駛測驗、環境作用評鑑、投資組合評論等。Scriven 以其豐富的歷練與學養，力倡評鑑須能應用在真實情境中 (如：方案評鑑、人員評鑑、產品評鑑、專業評鑑、後設評鑑等)，而產生的結果也非常的重要 (如：改善方案評鑑、改善產品評鑑、較好的評估方法、新的評鑑工具、改善評鑑和科學的觀點)。據此，本文整理並闡述近半

個世紀來 Scriven 在評鑑理論中倡議的創新觀點以饗讀者。

## 貳、Scriven 對評鑑的主張

Scriven (2001a) 認為評鑑的真實情況比評鑑的結論還要更簡單，同時也是更複雜；簡單的是在各層次之間，複雜的是在細節的部分。在各層次之間，可以肯定的是這些論點都是在將評鑑的重要角色，呈現到較簡單的觀念本質中。複雜的細節是最難的部分，就是在評鑑理論中，評鑑是一個多功能、多角色的企業解決問題的辦法。誠如 Scriven (2001a) 認為「評鑑扮演很多角色，像演員一樣，有時候會陷入迷思到底這些角色的真實性是什麼。」在這模糊關係之中，評鑑似乎也引起了一些爭議。茲就 Scriven 對評鑑的主張闡述如下：

### 一、評鑑者需要具備多重技能

Scriven 認為在現今許多創新的理論觀點，及評鑑多元化的現象下，評鑑者確實需具備一些特殊的實證研究技能，並能整合實證方法，包括社會科學研究中的假設檢

驗，所以評鑑者的實證技巧必須超越傳統的社會科學研究。舉例來說，有的評鑑採用技巧的觀察，如：對於病人情況或舉動等的真實片段的醫學評鑑。同樣的，評鑑不是只和判斷有關，它可能關係到測量而不是建立標準，如：評鑑跳高或田徑的績效。評鑑者需要知道如何找出方案或政策的所帶來的其他作用，這對評鑑來說是非常重要的技能，但卻不是假設檢驗，評鑑者需要能發覺這些，才能做出整體評鑑結論 (Coffman, 2003)。這技能還包括決定相關技術的能力，以處理具爭議性的價值觀念和問題，及其他整合相關的評價和事實的結論的綜合技能。這樣才能確定判斷的結果是否能與相同的計劃、政策或產品的多元評價（可能是相互矛盾的）相互調和。所以評鑑者在整個評鑑的過程中，都需要以非常嚴謹的態度與方法來處理。

### 二、評鑑是一門學科

Scriven 曾指出：評鑑是一門學科，相較於其他學科，評鑑更要求其學習的表現方式 (Alkin, 2004)。因此，Scriven 認為評鑑本

身是一門學科，甚至包含跨學科（transdisciplines）的領域，且評鑑對其他學科提供服務，但它還是有自己的自主力。這些跨學科的領域，從小型（如：測量）到大型的（如：機率和統計）都有。

要呈現它的學科狀態，可以另外列出它的項目和彼此關係，並拼湊出其間的邏輯規則。Scriven（2001b）提到分類的過程乃藉由將假定和類別剝離，如上面所提的例子：測量、統計、機率和邏輯，來引導發現程序中可能已經被接受的謬誤。比方說我們發現教科書中度的方法有錯，或我們在評鑑老師的標準流程中發現錯誤，或是常將好的老師和對其教學的價值標準混為一談，這是不科學也非倫理的錯誤。所以一個企業想發展評鑑的核心邏輯，就要致力於將思考的過程分類－這就是將評鑑視為一門學科的貢獻。

誠如 Scriven（2001b）提到在一般教育或文科學位中，主要須改變的是加快這門新學科的腳步，並且突破學科之間的隔閡，其中需調和的有：（1）用新學科的思維來取

代傳統的思維。（2）須打破假設檢定模式（hypothesis-testing model）在社會科學研究典範中的堅固地位。（如：方案評鑑和其他社會科學一樣複雜，但它並不經過假設檢定的過程，也不是探索性質的工作。）（3）將之應用在一般教育的領域。經過這些努力，應該有些學校會想要開設這學科，且為他們帶來更多的歷史榮耀。許多人會哀嘆教育沒有價值，若將教導評鑑的應用技術，結合實行倫理分析的關鍵技術才能真正符合願景所需。

在專業發展之後，才會體認到要把專業當作是一門學科，像評鑑就是如此。近年來，我們開始看到在這領域中有更多的討論，我們正朝向視評鑑為一門學科邁進，但仍有很長的路要走（Coffman, 2003）。

### 三、各領域皆有評鑑的需求

Scriven 認為評鑑的角色非常廣泛，甚至包括了每天日常生活中的談話或活動。因此，我們也應著重這麼廣大的領域中與評鑑有關的部分。當我們去購物，許多人在買東西前也會經過仔細的評估；當

需要繞道而行時，我們會選擇一條替代道路去商店或去公司；在穿越馬路時，我們會評估交通狀況是否有危險性；當我們看電視或聽廣播時，我們會評估它所提供的廣告或電視內容，決定是否要轉台；而我們在考慮工作、生涯、學校、生活住所、醫生、藥物、休假、運動、減肥和生活型態等，也會使用一些不是很正式或經仔細思考的評估，或在完全外行的情況下會有較複雜的評估。

隨著評鑑的需求擴大，我們需要一些基本技能在評估這些日常生活的事物上，並透過父母、兄弟姐妹和同儕的協助，因為人類是很難離群索居的。在這過程中，我們更了解到「評鑑是實際生活中不可或缺的部分」，因此有人認為評鑑不是理性的行動的論點，只是一種績效的展現，並且認為這不是科學。或許，未來對於這些日常生活事物，我們可以期待看到更多系統化的論點，就像我們最近看到醫學上的大變革，也就是「以證據為基礎的醫學」(evidence-based medicine) (Scriven, 2001b)。

根據 Scriven 的觀點，一個決策的論定，必須經過嚴謹的、系統化的、有技巧的評鑑，並且只能解釋在限定的領域中。因此，如 Scriven (2001b) 所提未來的評鑑可能會被使用在理性的談話和政策決定中，這在資源分派、改善生活所需、評鑑的研究和技術、持續獲得財富、投資和教育等方面會有很大的影響。而比較冷門的評鑑工具、跨學科評鑑，若發展的較為完善時，對於純科學也會有頗大的影響。

### 參、Scriven 提出的評鑑理論創新名詞

1960 年代美國政府注入大量教育基金，造成教育市場擴大，每年教育產品的銷售超過六億美金，教育及相關產業的競爭增加，市場策略也變得主動且多元，加上教育觀的改變，消費意識多元化，注重消費品質，但這些市場策略卻少從消費者的最佳利益考量。基於此，開始有評鑑者大力推動消費者教育，並要求提出產品效能的客觀證據，Scriven 就在此一時代背景下力倡評鑑者須從消費者需要的

觀點，來確認真正的成果及評估成果的價值，而不問目標為何（黃光雄，2005）。Scriven 在評鑑領域先後提出許多創新名詞，正反映了他認為評鑑應包含多元向度的層面，運用多種觀點，且涉及多種測量的等級，需使用多種方法的思維。

### 一、目標中立評鑑

Scriven 認為目標本位的評鑑容易受到先入為主的具體目標影響，所以要評鑑成果，須對無論是否為意圖性的成果加以評鑑，因此產生「目標中立評鑑」(goal-free evaluation)理論的觀點。Scriven (1973)主張方案的副作用幾乎可以決定對方案評鑑的成果，因此評鑑人員不在評鑑受評主體的意圖，而是評鑑其成就。Gredler (1996)在闡述此一觀點時指出「目標中立評鑑」有兩大組成要素，一是界定效果：包括正面效果和負面效果；二是建立因果關係：首先要界定問題的成因，接著描述特徵在因果關係中的定位，最後界定事件與成果的因果連結。Fitzpatrick、Sanders 及 Worthen

(2004)在審視目標中立評鑑的主張後提及，「目標中立評鑑」有五大特徵，一是評鑑人員有意避免知道方案目標；二是不能因方案目標窄化評鑑探究的焦點；三是強調評鑑的真實成果，非方案意圖成果；四是評鑑人員與方案管理者和方案成員儘少接觸；五是考慮了非預期副作用的可能性。

「目標中立評鑑」和目標評鑑可同時進行，前者的優點是：(1)較無強制性；(2)較易適應中途目標的改變；(3)較能減少副作用；(4)較能避免社會的、知覺的、和認知的偏見；(5)較能面對專業的挑戰；(6)較能公平地考慮各種價值（王文科，1989）。而評鑑方案的所有效果應不要受到目標限制，但並非無目標，在目標中立評鑑中對方法論的主張，需視評鑑的情形採取適當的程序與方法，而評鑑的內容主要在找出方案的預期與非預期的正負效果，並釐清效果與事件的因果關係。其價值判斷的參考，需依據消費者的需求或標準來決定。

### 二、消費者導向評鑑

「消費者導向評鑑」(consumer-oriented evaluation) 重方案成果，而不重視方案過程。Scriven 研究發現以產品檢核表的觀點發展標準化的格式，發現副作用 (side-effect) 的影響很大，因此在消費者導向評鑑的相關論述中，多處論及檢核表的發展與應用。例如 Scriven (2005) 將消費者導向評鑑檢核表，分成下列種類：(1) 衣物清單 (laundry list)：強調的是不同的類別；(2) 順序性檢核表 (sequential checklist)：重視檢核內容的順序性，如航空班次檢核表；(3) 反覆性檢核表 (iterative checklist)：可反覆檢核，如可進行多次的形成性評鑑，最後進行總結性評鑑；(4) 診斷性檢核表 (diagnostic checklist)：如問題解決 (trouble-shooting) 列表；(5) 參考標準或優點檢核表 (criteria or merit checklist)：重視達成度的衡量。從 1974 年發展產品評鑑檢核表及關鍵評鑑檢核表以來，2000 年起在西密西根州立大學評鑑研究中心更接受美國國家科

學委員會的委託，發展了一系列的評鑑檢核表 (Wingate, 2002)，其重要性可見一斑。Scriven (2005) 及 Stufflebeam (2001) 也仍舊撰文主張檢核表的價值性包括：(1) 減少遺忘或疏漏；(2) 容易了解；(3) 降低月暈效應；(4) 降低羅莎效應；(5) 促進對評鑑領域的了解；(6) 減少對某評鑑工作的加權問題；(7) 發展容易；(8) 提供省思機會。而消費者導向評鑑檢核表主要貢獻則為對沒有時間或資訊進行評鑑的人員提供了有用的評鑑服務，同時對於選擇產品及方案的指標具有深入的知識。

另外，王文科 (1989) 也曾指出，消費者導向評鑑取徑的優點包括：集中焦點且保護使用產品及服務的消費者。但由於進行消費者導向評鑑時，需要高度可信賴且有能力的專家，以及足夠的資源提供專家能完整地進行研究；所以要進行一次完善且蒐集到真實可靠的資料的評鑑是有很多挑戰。當然在消費者導向評鑑亦有其限制，Scriven (2005) 就提及消費者導向評鑑運用上的限制包括：(1) 開

發成本大，須衡量其成本效益；(2) 檢核表若無法適時更新，可能造成消費者誤判；(3) 只重結果，不重歷程，是否符合消費者需求？(4) 評鑑工作者依賴評鑑檢核表，易阻礙評鑑進步。

### 三、形成性評鑑、總結性評鑑與歸因性評鑑

Scriven 認為評鑑的重要性如同「顧客服務」，並認為評鑑最重要的就是判斷價值。顧客對於方案或產品設計的目標是否達到並沒有興趣，只有對改善方案有間接的興趣；他們主要關心的是，是否達到自己的需要。因此，Scriven 將評鑑分成形成性 (formative) 評鑑與總結性 (summative) 評鑑兩種，前者有助於方案或目標的發展，後者可以用來評定方案或目標的價值 (Stufflebeam & Shinkfield, 2007)。若方案或目標有繼續發展的必要，則採形成性評鑑；若希望探究的是其價值，則採總結性評鑑；但也可能同時採用形成性評鑑與總結性評鑑。近年來，Scriven (2004) 在形成性評鑑與總結性評鑑之外，增加了第三類評鑑，即「歸

因性 (ascriptive) 評鑑」。他認為歸因性評鑑與方案或目標發展的過程沒有關聯性，但是在對方案發展進行評鑑時，在形成性或總結性評鑑之外，探詢方案運作過程或成果背後的成因與意涵，例如：歷史學家對歷史事件的評鑑；雖然歷史學家可能探討歷史事件的過程 (形成性評鑑)，或事件的結果 (總結性評鑑)，但對於分析與判斷事件背後的意義反而是他們有興趣的，因此有歸因性評鑑的產生。

### 四、後設評鑑

後設評鑑 (metaevaluation) 即「評鑑的評鑑 (the evaluation of evaluation)」，其目的在維持評鑑的真實性。Scriven (1983) 將後設評鑑定義為「由另一位評鑑者去評鑑一項已計畫或已完成的評鑑，也就是將原來的評鑑者置於受評者的位置，對該評鑑者及其評鑑活動之表現進行價值判斷」。

自後設評鑑的概念提出以來，有許學者如 Stufflebeam 及 Patton 在其對評鑑理論的主張中均強調後設評鑑的重要性。對於進



行後設評鑑時，Scriven (2008) 提醒應有的注意事項包括：(1) 在可能範圍內，由評鑑者或其他人在完成任何報告的最後草稿後進行；(2) 有可能的話，也可藉由一位外部評鑑者來評鑑（後設評鑑者）。此外，評斷後設評鑑的主要標準是：(1) 有效性；(2) 實用性（通常針對客戶、對象和利害關係人）；(3) 可信性（選擇利害關係人—尤其是出資者，及監管機構，通常也是方案的工作人員）；(4) 成本效益；(5) 倫理／合法性，包括可能產生的利害衝突和保障人權(Scriven, 2008)。Scriven(2008) 宣稱可以用於後設評鑑中的方式有：(1) 使用 KEC、方案評鑑標準等；(2) 使用特別的後設評鑑檢核表；(3) 用同樣的方式複製評鑑，並比較結果；(4) 用不同的方法來評鑑，和比較結果。在實用上，Scriven 在「主要評鑑檢核表」的最後一項規範進行後設評鑑，目的在確定其優點、限制或其他用途。

後設評鑑的議題在當前已受到國內外評鑑理論與實務研究人員相當的關注，許多學者也曾提出

補充性的概念，雖然發展至今有些論點或做法與 Scriven 最初的想法不盡雷同，但對 Scriven 所提後設評鑑的定義卻是至今未曾被檢討或推翻的。

### 五、業餘評鑑與專業評鑑

Scriven (1967) 認為由內部人員擔任評鑑者的工作稱為業餘評鑑 (amateur evaluation)，也就是自我評鑑 (self-evaluation)；而由外部人員擔任則稱為專業評鑑 (professional evaluation)。在方案發展早期，Scriven 傾向使用業餘評鑑，因為當發展者兼作方案的評鑑者時，可能因追求成功而容許模糊的目標及探索性的發展程序，而促進評鑑的創造力；但如果太早投入專業評鑑，可能會抑制靈感的創造。為避免業餘評鑑過於主觀，Scriven 同時強調專業評鑑的重要性。Scriven 建議最理想的方法是內部人員和外部人員都能納入體系之中，透過民主化的團體歷程、溝通、討論、決定，進行評鑑工作。Stufflebeam 及 Shinkfield (2007) 認為這兩種評鑑都需要高水準的

技術，無論在評鑑的理論或方法論都需要經過特別的訓練。雖然內部評鑑與外部評鑑並非創新的觀點，然而 Scriven 以業餘評鑑及專業評鑑兩個名詞來詮釋，仍屬對評鑑中專有名詞的創新。

## 六、內在評鑑、收益評鑑與混合式評鑑

Scriven (1967) 對「認可模式」加以批評，認為這樣先入為主的內在規準有所偏頗，於是將評鑑區分為內在 (intrinsic) 評鑑 (又稱本質評鑑) 與成果 (payoff) 評鑑 (收益評鑑) 兩種形式。「內在評鑑」係指教學工具本身的評鑑，並從內部特質與要件進行評鑑。

「收益評鑑」包括測驗分數、工作表現或健康狀態等，即針對實施處理後所獲得之成果進行評鑑。此外，Scriven 提出一種兼顧內在與收益評鑑的「混合式評鑑」(hybrid evaluation)，認為應整合兩種評鑑，即是於形成性評鑑中，使用內在評鑑形式來輔佐，於總結性評鑑中，則採取收益評鑑形式來判別方案實施成效，這樣的構想就是所謂的混合式評鑑。

雖然混合式評鑑的概念並非專用於內在與收益評鑑兩者的混合，然而 Scriven 提這樣的概念，基本的想法只在於兩者之間會有依使用時機及用途的區隔去混合運用而已。

## 七、學科內評鑑

Scriven (2001b) 提出學科內評鑑 (intradisciplinary evaluation)，認為評鑑是有原則的，如歷史學家評論拿破崙戰役。因為要強調的是一門學科必須有條有理，而不只是意見的陳述 (即使是用一些科學儀器來裝飾 (如占星術)，或對精心之作的品嘗 (如酒的評論等)，意即成為一門學科的過程中必須論證其符合科學的原則。評鑑在處理時 (評鑑的資料、假設、理論) 是有根據且可複製的，此評鑑的取向即是基於憑據原則。所以沒有科學不經過評鑑的過程，因為沒有評鑑就不能區辨它是不是假科學；要能區辨好壞科學，是作為一個稱職的科學家的必要條件。因此，要判斷價值，而不只是精心之作的品嘗，就要有相當具體的要素，無論是科學事業、科

學刊物，或每一個看似有理的科學方法論都一樣。

## 八、預測評鑑

Scriven (2007) 以預測評鑑 (predictive evaluation) 一詞含括先前評鑑領域中的許多觀點，所包含的內容包括：(1) 產品評鑑 (product evaluation)：產品必須具備顯著的耐久性，也就是它的價值，並評估耐久性可能帶來的危害；(2) 方案評鑑 (program evaluation)：相同的，方案必須是為了任何價值而可持續進行的（除非只做歷史評鑑）；(3) 人員評鑑 (personnel evaluation)：至少有 50% 的選拔／提拔／晉升／保留／任期評鑑（通常被稱為事前評鑑），都需要先預測未來可能維持現在的水準或是有更好的表現，尤其是危機預測。事後評鑑，其中包括評鑑過去服務或成就的獎勵，也就是要溯及既往；(4) 政策分析 (policy analysis)：大約 75% 的政策分析，會被視為是未來決策訂定的指標，因此必須提出未來可能發生什麼情況。其餘的 25% 是事後評鑑，只是做為紀錄，並不

涉及預測；(5) 建議評鑑 (proposal evaluation)：「建議評鑑」屬於 100% 的預測，除非是作為教學之用。(6) 檔案評鑑 (portfolio evaluation)：在投資案件的檔案中，要 100% 的預測：投資的時間、金錢或研究工作是否恰當。但值得注意的是，如果這些檔案，被用來作為是否聘雇人員的依據，這就變成了人員評鑑。

## 肆、對 Scriven 創新觀點的省思

### 一、進行目標中立評鑑的彈性與問題

黃光雄 (2005) 認為「目標中立評鑑」，是一種創新的技術。Stufflebeam 及 Shinkfield (2007) 亦指出「目標中立評鑑」，是一種創新的取向，這對於消費者導向評鑑有很大的助益。此外，Youker (2005) 認為「目標中立評鑑」中，評鑑的時間分配、財政和其他可用的資源、專門知識和能力都是額外要考慮的部分，以確保目標中立評鑑的準確性。

過去，傳統評鑑方式乃先計畫目標，然後測量方案效能，並評估

效能是否符合目標；而目標中立評鑑較能適應方案過程中目標改變，因此彈性較大，也較能發現方案所產生的副作用。雖然，Scriven 認為要評估消費者的需要，來判斷方案的結果，即所謂需求評估（needs assessment），同時目標中立評鑑是不受目標侷限，並非沒有判斷標準。但是，需求評估的結果，是否能真正符合真實情況，並就此判斷方案的價值與良窳，是否理想與現實間產生了間隙？在知識產生快速的時代下，誰又有足夠的視野與能力擔任此類評鑑的評鑑委員？這些都是在進行目標中立評鑑時需要進一步思考的議題。

## 二、形成性與總結性評鑑的謬誤

自 Scriven 提出形成性與總結性評鑑的概念以來，迭有學者存有觀念上的誤謬，因此，Scriven（1991）發表〈超越形成性與總結性評鑑〉專文，指出並討論十項有關且常見的誤解觀念，然而時至今日，仍有諸多學者對這兩個名詞提出討論，例如：Allen（2003）認為「形成性評鑑」的重點在改善正

在進行的工作，但是資金有限，往往是形成性評鑑的一大限制。所以在方案規劃前，事先將這部分的預算考慮進來，於是方案進行中，透過形成性評鑑，適度的評估與修正未來方案執行的內容，這才是長期解決衝突的首要之道。「總結性評鑑」方面，Conrad 在 2006 年邁向卓越高等教育評鑑國際學術研討會中提到，「總結性評鑑」的問題較多，過去他曾提到「評鑑員偏見」（rater bias）的問題，若是遇到有心人士的操縱，均可改變評鑑呈現出來的最終結果。所以，要提升教育品質，在制度上應該是由內而外、由下而上，並以品質促進替代品質保證。此外，以形成性與總結性來區分評鑑似乎有所限制，評鑑的本質似乎應更廣泛的定義。

## 三、對後設評鑑的迴響

後設評鑑應考慮評鑑之設計、方法和結論是否完整，且完整的後設評鑑應有其標準。自 Scriven 於 1969 年提出「後設評鑑」一詞後，許多後設評鑑標準產生，包括 Cook 和 Gruder（1978）、Stufflebeam（1981）、Schwandt

和 Halpern (1988)、Lynch 等人 (2003) 為後設評鑑增加的標準，及方案後設評鑑標準 (Stufflebeam, 1999)、人員後設評鑑標準 (Stufflebeam, 2000) 等。在實際執行時，應就方案的規劃、設計、實施、報告及結果利用等進行後設評鑑標準規劃與實施，並依其結果作為修正改善之依據，並能確實執行。

#### 四、預測評鑑的指標與準確度

「預測評鑑」方面，預測其實是一項非常棘手的事，也許可以將過去的表现作為日後表现的指標，但這指標不一定是預測的最佳準則，所以最好依據收集到的和結合所有可利用的指標來預測；舉例來說，假設研究者有證據證明這位候選人，優於其他任何候選人，也就是這位候選人沒有任何的犯罪記錄或剽竊證據，但這就是好的預測嗎？當然不是，我們應該要以不同的角度去找出那些關鍵要素。但是最基本的問題，就是除了過去的表现，還有什麼其他的證據可提高我們的預測精準度？因此，什麼時

候表現轉壞、什麼時候表現轉好、如何長期工作都能表現良好等，這些為將來建立高品質的工作潛力，都是我們需要關注的問題。

#### 五、消費者導向評鑑的優窳

消費者導向評鑑包含對目標中立評鑑及評鑑檢核表等概念，著眼於教育產品的消費所面臨的複雜情境，在「以民為本」、「顧客至上」的想法中，消費者導向評鑑無疑地提供一個更簡明易用的工具來保護消費者及教育服務工作者，對於評鑑工作者而言可以更有效率地處理評鑑的工作。Alkin (2004) 指出，Scriven 的觀點中最重要的一個問題，就是評鑑者應考慮的是究竟評鑑對人類的福祉有什麼樣的貢獻，甚至對我們所處的整個宇宙的福利。所以，「消費者導向評鑑」正如此觀點所述，將所欲評鑑之要項或主要指標列出，發展出評鑑檢核表，並蒐集整理評鑑工具，以利評鑑者據以進行評鑑之工作。Stufflebeam 及 Shinkfield (2007) 亦指出，消費者導向評鑑者要找尋用哪種取向的評鑑時，最重要的就是要確認消

費者的需求。因此，我們不可否認「消費者導向評鑑」著重消費者的觀點，確實有別於其他評鑑理論，且對評鑑的貢獻不容小覷。加上近年來消費意識的抬頭，「消費者導向評鑑」確實可讓消費者在選擇產品或方案時，更具有鑑別力。但是，除了成本效益、消費者誤判、是否符合消費者需求，及評鑑工作者對評鑑檢核表的依賴等問題外，但其並未針對特定對象的消費者提供更好的服務，以及可能增加的成本是否會轉嫁給消費者等，都是實施「消費者導向評鑑」待解決與釐清的問題。

## 伍、結語

評鑑是找尋知識須努力的部分，包括科學、技術、法規、倫理和其他人文主義的真實部分。而我們每一個人，在評鑑過程中都扮演許多角色，如：評鑑者、被評鑑者或評鑑的使用者等。基本上，評鑑是最簡單的形式，用行動連結知識的一種過程。評鑑試著只依據字彙的價值，去決定和描述已確定的觀點，而不是去描述優點、是否值得或其意義。當然，價值常根據脈

絡而來，但有時多少會依整體脈絡來考慮，也就是推論的邏輯，或多少會賴於超然的倫理標準，如扼殺或扭曲的罪惡感。許多事經過描述後，無論是好或壞的事情，都有可能合宜的或不適合的。但這些事情並非評鑑的一部分，只有社會背景下的結果才是。

在 Scriven 的觀點中，評鑑本身是一種嚴謹的戒律，可以和任何的科學比較其效力，並可從其脈絡中區分出核心的問題。或許有的人會修正評量以促成結果，但卻違反了專業的處理準則。Scriven 的消費者導向評鑑檢核表方面，雖有其優點、缺點與限制，但卻陸續更新以更符應使用者的需求。因此，Scriven 在評鑑領域的貢獻與執著，以及其用不一樣的角度來看評鑑，帶給後起之研究者不一樣的研究思維方式。

## 參考文獻

王文科譯（1989）。Scriven 的消費者評導向評鑑取徑。載於黃光雄編譯，*教育評鑑的模式*（頁 355-395）（D. L. Stufflebeam 和 A. J.

- Shinkfield 著)。台北：師大書苑。（原著出版於 1985 年）
- 黃光雄（2005）。*系統的評鑑：理論與實務的自我教學指引*。台北：師大書苑。
- Alkin, M. C. (2004). *Evaluation roots: tracing theorists' views and influences*. CA: Sage.
- Allen, M. A. (2003). *Michael Allen's guide to E-learning: Building interactive, fun, and effective learning programs for any company*. NY: John Wiley。
- CGU (Claremont Graduate University, 2008). *Michael Scriven*. Retrieved on October 16, 2008 from <http://www.cgu.edu/pages/4745.asp>
- Coffman, J. (2003). Reflecting on the past and future of evaluation. *The Evaluation Exchange*, 9 (4), 7.
- Cook, T. D., & Gruder, C. L. (1978). Metaevaluation research. *Evaluation Quarterly*, 2 (1), 5-51.
- Fitzpatrick, J. L., Sanders, J. R. & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3<sup>rd</sup> ed.). Boston: Pearson Education, Inc.
- Gredler, M.E. (1996). *Program Evaluation*. New Jersey: Prentice-Hall.
- Lynch, D. C., Greer, A. G., Larson, L. C., Cummings, D. M., Harriett, B. S., Dreyfus, K. S. & Clay, M. C. (2003). Descriptive metaevaluation: Case study of an interdisciplinary curriculum. *Evaluation & the Health Professions*, 26, 447-461.
- Schwandt, T. A., & Halpern, E. S. (1988). *Linking auditing and metaevaluation*. Newbury Park, CA: SAGE Publications.
- Scriven, M. (1967). The methodology of evaluation. In

- R. E. Stake (Ed.), *AERA monograph series on curriculum evaluation* (No.1). Chicago: Rand McNally.
- Scriven, M. (1973). *Goal-free evaluation*. C.A.: McCutchan Publishing Corporation.
- 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. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin & D. C. Phillips (eds.), *Evaluation and education: At quarter century* (pp.19-64). Chicago: Nsse.
- Scriven, M. (2001a). An overview of evaluation theories. *Evaluation Journal of Australasia*, 1 (2), 27-29.
- Scriven, M. (2001b). Evaluation: future tense. *American Journal of Evaluation*, 22 (3), 302-307.
- Scriven, M. (2004). Reflection. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences*. CA: Sage.
- Scriven, M. (2005). Checklists. In S. Mathison (Ed.), *Encyclopedia of evaluation*. CA: Sage.
- Scriven, M. (2007). Predictive evaluation. *Journal of Multidisciplinary Evaluation*, 4 (7), 71-73.
- Scriven, M. (2008). *Key evaluation checklists*. Retrieved on October 16, 2008 from <http://www.wmich.edu/evalctr/checklists>
- Stufflebeam, D. L. & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. Hoboken, NJ: John Wiley and Sons.
- 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. (1999). Program  
evaluations metaevaluation  
checklist. *Evaluation  
Checklists Project*. Retrieved  
on October 16, 2008 from  
[http://www.wmich.edu/  
evalctr/checklists](http://www.wmich.edu/evalctr/checklists).
- Stufflebeam, D. L. (2000b). The  
methodology metaevaluation as  
reflected in metaevaluations by  
the Western Michigan  
University Evaluation Center.  
*Journal of Personnel  
Evaluation in Education*, 14 (1),  
95-125.
- Stufflebeam, D. L. (2001).  
*Evaluation model*. San  
Francisco: Jossey-Bass.
- Wingate, L. A. (2002). *The  
evaluation checklist project:  
The inside scoop on content,  
process, policies, impact, and  
challenges*. Paper prepared at  
the American Evaluation  
Association Conference.  
Washington DC, November 6-9,  
2002.
- Youker, B. W. (2005). Ethnography  
and evaluation: Their  
relationship and three  
anthropological models of  
evaluation. *Journal of  
MultiDisciplinary Evaluation*,  
2 (3), 113-142.
- 收件日期：99年03月01日  
修改日期：99年04月23日  
接受日期：99年04月30日

## **The Introspection of Scriven's Evaluation Theories Innovative Views**

**Chiu, Hui-Fang**

PhD Program Student, Institute of Technological and Vocational Education,  
National Taipei University of Technology

**Tseng, Shu-Hui**

Associate Professor, Institute of Technological and Vocational Education,  
National Taipei University of Technology

### **Abstract**

The article aimed to explore M. Scriven's innovative views in the evaluation theories in the past three decades, by the explaining of it to offer the educational introspection. Through the literature, we can find M. Scriven has some opinions of evaluations, including: goal-free evaluation, consumer-oriented evaluation, formative evaluation and summative evaluation, amateur evaluation and professional evaluation, intrinsic evaluation and payoff evaluation and hybrid evaluation, and intradisciplinary evaluation, and predictive evaluation. Explains that evaluation should contain multi-dimensional stratification plane, involves many kinds of survey ranks and must has the thoughts of using many kinds of methods.

Finally, according to the Scriven innovation viewpoint proposed that in the educational introspection and thoughts: (1) flexibility and problems of proceeding the goal-free evaluation (2) a falsehood of formative evaluation and summative evaluation (3) the feedback of meta-evaluation (4) the index and accurate rate of predictive evaluation (5) the advantages and disadvantages of customer-oriented evaluation. Hope to provide reference resources by interpreting and deliberating M. Scriven's innovation view in the evaluation theories to those related

field scholars and practice workers when they are proceeding related researches or formulating strategies.

**Keywords:** consumer-oriented evaluation, formative/summative evaluation, predictive evaluation