

# 教育評鑑概念與模式之研究

丘慧芬

## 壹、前言

近年來，教育評鑑（Educational Evaluation）一詞因教育部在民國六十四年對大學院校實施的評鑑，普遍受到國內教育界人士的注意。而台灣省及台北市高級中學的評鑑，更顯示出教育當局欲藉教育評鑑的實行，來一方面改進現有教育設施中的缺失，進而提高教育的素質；另一方面，也藉教育評鑑的發現，來證明現行教育的實際成效，以符合「教育績效責任制」（Educational accountability）的要求。理論上雖然有少數美國學者提出中國傳統的科舉制度可稱做是教育評鑑活動的最早雛型，（註一）而實際上我們每個人隨時隨地也在對所接觸的人、事、物做著各種有意無意的贊美或批評，但是，要找出一系列客觀、科學與精確的評鑑理論和技術，或許仍有必要了解先進國家的評鑑發展，作為我們今後實施評鑑的參考與借鏡。筆者不敏，乃將美國近十幾年來有關教育評鑑基本概念演變與其模式發展做一理論分析的研究，並撰成此文以就教於方家，祈收拋磚引玉之效。

教育評鑑的原始發展過程，最早可追溯到由一八九〇年代萊斯（J.M. Rice）拼音成績的比較研究，到一九二〇年代桑代克（E.L. Thorndike）大力提倡的測驗運動，及一九三〇年代泰勒（K.W. Tyler）等人的八年研究，然而教育評鑑在美國正式受到重視却不過只有短短幾十年的歷史。黃炳煌先生指出教育評鑑運動能在一九五〇年代開始在美國正式被推進，其原因不外乎：美國社會大眾對公立學校教育的不滿，美國聯邦教育撥款之遽增，七〇年代初期教育財源枯竭及教育行政之分權化。（註二）事實上，教育績效責任制實為直接促使教育評鑑運動正式見諸施行的一大原因。教育績效責任制的觀念首次出現，是因為在一九六〇年代美國國會通過了國防教育法案（National Defense Educational Act）及初等中等教育法案（Elementary Secondary Education Acts），由於美國聯邦政府機構必須負責以鉅款補助有關教育的革新計劃，同時又須證明這些教育革新計劃確實有其成效，所以首次感到了必須注意「績效」的壓力，此種壓力更直接引起各種有關學校行政措施及教師教學是否具有效果的問題，進一步就促成了由美國聯邦政府機構以大量經費支持對教育革新計劃進行的委任式正式評鑑運動。

一九六〇年代由政治行政管理促進的關於「績效」、「成本—效益」（Cost-effectiveness）、及「品質保證」等一系

列觀念在美國地方區域並未受到重視，但七〇年代以後，各州人民開始要求地方教育提出績效證明，換句話說，地方州政府與一般大眾也開始明瞭教育評鑑的重要性了。今日，教育評鑑在我國受到普遍重視，也可說是教育行政當局爲了使教育投資得到最大的效益、教育品質達到相當的水準、教育實施獲得顯著的績效所必然採行的一項措施。

教育評鑑是一個複雜多端的歷程。爲了解評鑑的基本概念與其模式發展，筆者擬就評鑑概念演變的歷史、評鑑意義的探究、評鑑計劃的組成因素及評鑑的模式比較等幾個方面作一分析論述，並批判美國教育界目前對評鑑的有關爭議，最後結論且提出筆者個人對今後國內發展教育評鑑的幾點建議，祈學者方家不吝指正。

## 貳、教育評鑑概念的歷史演變

教育評鑑概念多年來有很大的改變，隨其概念的轉變而引起相關的一些改變問題，包括：(一)評鑑對象爲何、(二)評鑑內容爲何及(三)如何完成評鑑等。吾人必須明瞭評鑑概念的發展是因爲教育設施上的需要，而隨着評鑑概念的發展，進一步又影響了教育上的設施，兩者有着交互影響的作用。評鑑概念的演變主要是因爲評鑑活動與教育的三方面發生了交互的作用。(註三)第一種交互作用是因爲評鑑的實施，一方面影響了教育上各種學習理論，一方面又同時受到各種學習理論與教育方法的影響；其次是由於評鑑本身與其在教育上扮演的角色產生了交互作用；再者即起因於評鑑活動與評鑑在科技方法上的發展有了交互影響。以下乃就上述三方面來探討評鑑概念的演變過程：

### (一)評鑑概念與教育理論的相互關係

吾人欲明瞭課程、教育哲學、學習理論及教學方法對評鑑概念的影響，可追溯至廿世紀的初期。前言提及，萊斯在一八九四年所做拼音成績的比較研究到泰勒等人著名的八年研究，是教育評鑑運動之濫觴。到了一九五六年泰勒應邀參加美國測驗問題會議時，提出心理學界早期所熱衷的官能心理學已漸爲行爲心理學派所取代；後又指出，一九一八年到一九二五年期間，美國教育界受當時聯合論學習心理派影響往往以極端狹隘看法界定教育的目標，到了一九二五年新興起的學習理論及傳統進步主義教育哲學派的興起，強調教育目標的界定應注意到兒童行爲的整體性，杜賓(J. E. Dobbin)後來批評課程改變成就測驗時，發現不但學習理論在基本觀念上產生很大改變，即便學校組織上也有了快速的變化，(註四)因此導致了從一九三〇年代早期以來有關評鑑概念上的轉變。

史塔屈(D. Starch)在一九一六年美國教育研究學會出版的第十五輯年刊上提出一個半世紀以來皆爲人重視的概念，

即：「教育評鑑應該注意到學生在學習上的個別差異。」（註五）由此而演變出的評鑑概念反映在教育上就產生了一「同質團體」與「個別教學」等設施。（註六）

教育設施上需有評鑑提供的資料，評鑑也相對的影響教育設施的形成。德索（P.L. Dressel）認為美國四〇年代間評鑑對教育實施的影響引起不少的爭議。（註七）由於當時測業專家的目的在顯示測驗的科學價值，雖然他們竭力避免被世人指控為支配課程或教學方法，但實際上也不易完全避免對教學的影響。職是之故，德索就認為教材專家「若能重視評鑑，就可洞察學生心態，了解教材間的關聯與意義。」（註八）

## （二）評鑑在教育上扮演的角色與其本身的相互關係

美國教育界期望評鑑扮演的角色，可從以下三方面來作一說明：

### （1）在一般教育計劃上所扮演的角色

廿世紀初期，美國教育界認為評鑑扮演了協助學校解決其問題的角色，但彼時基本重點仍為測驗的運用，以期了解學生分班、學校組織、學科學習、教學方法、教學時間以及視導等問題。一九〇〇年到一九一〇年間，由於成就及心理能限測驗的發展，促進了教育行政方面的鉅幅革新，例如：評估學生的進步結果、診斷學習困難、了解特殊性向、評估教學效果等，都因有了客觀的科學方法而解決許多問題。然而這些測驗多半着重教育的輸入資料（Input），所以就有以下重視教育結果（輸出資料）的研究出現。

### （2）在教學上所扮演的角色

一九三八年司凱茲（D.E. Scates）在美國教育研究學報（Review of Educational Research）中說到：「要改進教學測驗，可依三方面進行：①更謹慎考慮測驗的目的何在，②必須測量學生整體行為的結果，以及發展適合新需要的妥善工具。」（註九）司凱茲基本上概括出實施評鑑上一個十分有意義的運動，這個運動最早由泰勒在一九三〇年代初期創始，有關泰勒的看法，可在一九三一年到一九三三年間俄亥俄（OHIO）州立大學出版一系列教育研究學報（The Educational Research Bulletin）中窺見其詳。後來這些學報被編纂成「編製成就測驗」（Constructing Achievement Tests）一書。（註十）自此，教育界也由狹隘的教育目標轉而重視兒童整體行為的教育目標。泰勒認為教學目標與評鑑目標之間有着緊密的關係，故而提供了一些目標設計的明確法則。

泰勒主張目標的設計須以學生行為為基礎，這些目標可進而做為教學及測驗的目標。他認為評鑑若只着重學生知識與技能，未免有失週全，因此提議評鑑應注意到教育上全部的學習結果，亦即不但要重視知識與層次較高心靈活動的認知結果，且須

評鑑教育上的情意學習結果。泰勒所提的原則與方法自彼時起即已形成了評鑑的重要依據，而衆所週知由泰勒在一九三三年至一九四一年負責的八年研究即爲最佳證明。

欲使評鑑在教學計劃中扮演更合理更積極的角色，導致一九三〇年代人們開始考慮教學前的評鑑活動。彼時許多建議認爲學校應在每年秋季舉行評鑑以做爲教師教學計劃的依據。同時反對只將評鑑用於教學成就的評估上。

### (3) 評鑑在學生自學上扮演的角色

辛浦森(R. H. Simpson)認爲：「學習大部份是在教室外發生的。」(註十一)因此主張若能使學生具備整套的計劃能力，即可評鑑自身的學習情形，也必能由此獲得更多的學習。自我評鑑的概念就如此產生了。

### (4) 評鑑概念演變與評鑑內容的關係

由於評鑑概念的演變與其在教育過程中所扮演的角色有關，顯然就引起了評鑑對象與評鑑內容的改變。評鑑內容的改變包括：①一般學習成就的評鑑、②整體行爲(認知、情意及環境因素等)的評鑑、③學生起點行爲的評鑑及④學習團體的評鑑。就一般學習成就方面來說，廿世紀前半葉，評鑑多半只注重與特定學習技能有關的偏狹目標，到了一九六〇年以後，評鑑就大幅度轉向評量學習的全部結果，也藉此明瞭學生是否能應用所學至實際的生活領域。接着再討論學生起點行爲的評鑑。教育上配合學生的現有能力和知識程度狀況，以開始教學活動並進行評鑑，就稱做起點行爲的評鑑。接受評鑑學生起點行爲並儘量發展其潛能的這種哲學觀念，需有二種時間不同的測驗，一方面在測量學生教學未實施前的狀況；另一方面在測量學習後的改變情形，但此期的評鑑活動多半已將重點放在團體學習的成就情形，而不在只重個體的學習結果。至於整體行爲的評鑑，必須再度提及泰勒的影響。事實上，教育目標分類手冊認知領域及情意目標手冊(註十二)二書的出版即爲一例。評鑑範圍由簡單的認知擴大至較高心智活動層次，再轉而重視情意的教育目標。此外也開始注意環境因素(例如：家庭收入、城市大小)對學習的影響了。有關評鑑究應着重個人或注意團體的學習結果，也可做一歷史的追溯探討。由注重學習者個別差異而導致診斷個人學習困難，補救教學計劃等，可說都是偏重個人學習的評鑑。近年來，許多研究已經注意到教育評鑑與教學方法、教材及課程的關係。有趣的是早在一九二〇年代出版的書籍中已經有學者注意到學校全面系統的效率，包括教學與教學成果的問題。(註十三)廿世紀初期的學者，事實上也相當重視以評鑑團體的學習成就做爲評鑑學校系統、學校計劃及教學效果的依據。毫無疑問在課程計劃下發展出的新課程諸如數學上的SMSG(School Mathematics Study Group)課程及生物學上的BSGS(Biological Science Curriculum Study)課程等就引起教師們提出必須有課程計劃的評鑑。此外又因電腦在教學上的應用，更有了必須評鑑的理由，結果就產生了使用教學機等「硬體式」(Hard Ware)系列的課程。近年來這些課程上的發展活動導致課程發展式評鑑，即司瑰文(Scriven)在一九六七年提出的「形成性評鑑」(Formative Evaluation)

），（註十四）而選擇已發展了的課程產品，亦即司瑰文所說的「總結性評鑑」（Summative Evaluation）。（註十五）美國教育研究學會（American Educational Research Association）後來也在一九六五年組成課程委員會，出版了一系列有關課程評鑑的專題論文。到了一九六六年，美國國防教育法案（National Defense Education Act）通過後，即要求在此法案第一及第三類計劃下的研究都須有評鑑報告，也就是注重課程評鑑的新趨勢。至此，評鑑就由學習個體行為的評估轉而重視教育團體及教育計劃的評鑑了。

### （三）評鑑與其本身科技發展的相關關係

評鑑本身科技發展所引起評鑑概念的轉變，大致可分為三種情形：①評鑑開始注重學生在多種學科上的成績、②開始以常模（Norm）解釋測驗分數的意義及③根據個人潛能解釋學習成就。由於這方面的交互關係主要是談及測驗工具的設計發展結果，可說偏重在測驗的領域較多，故不予詳細追究其發展過程。

以上係就評鑑概念的演變做一歷史回顧，吾人可發現評鑑一詞在早期（甚至在今日）與測驗的發展有密切的相關，值得再提的是，評鑑的產生是配合教育設施的需要，但由於評鑑活動的進行也相對的影響了教育上許多設施的形成。學習理論的演變，導致學生整體行為評鑑的產生；教育上同質團體與個別化教學的設施，又起因於評鑑概念轉變之結果，其次，因為評鑑角色的不同，產生了全盤教育效果的評鑑；進一步更由學習個體的行為評鑑概念轉成重視團體學習成就的評鑑。

總而言之，今日美國教育評鑑的重點對象在於全盤教育計劃的評鑑，舉凡學校行政、教學方法、教學效果及課程計劃等皆為其評鑑的內容。

## 參、教育評鑑意義的探討論述

在評鑑概念演變的歷史回顧中，明顯的指示出評鑑與測驗之間有著難以分割的關係，那麼「評鑑」是否就是「測驗」呢？雖然少數教育學者仍持有肯定的答案，（註十六）但是今日美國的教育評鑑界已逐漸廢棄此說，我國學者黃炳煌、李緒武及張植珊等人也分別提出「測驗重在客觀事實的獲得，評價重在事實之解釋、珍斷與價值判斷的說明，」（註十七）足見評鑑雖然常須藉重測驗的方法蒐集所需資料，本質上仍各有所重，不應視為一物。至於教育評鑑是否即是一種專業判斷（Professional judgement）也值得研究。專業判斷重在主觀的意見判斷，對形成判斷的資料已缺少應有的客觀了解。如此說來，教育評鑑的意義界定誠為不易。晚近美國教育評鑑人員對評鑑意義的界定，多以「決策導向」（Decision Oriented）為基礎，將評

鑑視為繼續不斷蒐集資料以協助制訂決策的歷程。以下先就國內外幾位學者有關評鑑意義的看法做一論述，配合筆者已見，再歸納出一個綜合的意義。

司德克(R. E. Stake)認為教育評鑑必須對所要評鑑的教育計劃(Educational Program)提出詳盡的說明與充份的判斷。(註十八)

司瑰文(M. Scriven)提出教育評鑑是根據一組價值量表蒐集並整理學習成就資料的活動，其目的在對受評者作一價值判斷。(註十九)。

波爾佛斯(M. M. Provos)指出評鑑是一個決定目標與成就之間是否有差距存在的歷程，其目的在決定是否要維持、取銷或改進受評的教育計劃。(註二十)此與泰勒早期主張評鑑就是將學生的學習成就與所列的目標作一比較的歷程(註二十一)可說有類似之處。

克朗巴克(L. J. Cronbach)提出評鑑的意義在蒐集運用資料以便制訂有關教育設施的決策。(註二十二)

史達夫賓(D. L. Stufflebeam)認為教育評鑑乃是為了判斷各種有關決策的改變途徑時說明、蒐集並提供有用資料的一種歷程。(註二十三)此與艾爾金(M. C. Alkin)所提出評鑑就是要確定決策制訂的範圍並選擇合宜資料加以分析的歷程(註二十四)也有相同之點。

我國學者李聰明先生認為「教育評價是利用所有可行的評價技術評量教育所期的一切效果。」(註二十五)陳瑞龍先生又主張「教育評量就是以教育價值為基準，依照教育價值來判定被評定者的行動或學習之價值程度的操作。」(註二十六)蔡培村先生更進一步指出「教育評鑑是利用客觀的科學方法和主觀的價值判斷，對教育的內容與措施進行合理的考核，以期檢討利弊得失，提供改進策略，實現教育目標的一連串歷程。」(註二十七)

就以上所介紹國內外有關教育評鑑的意義界說中，李聰明與陳瑞龍二位先生所提出的評鑑定義，顯然尚未將評鑑視為一種「歷程」，亦未視評鑑為「制訂決策」的重要依據，且仍具有「測驗」的色彩。泰勒的評鑑意義也因過份重視學習結果忽視了歷程概念，可說是早期評鑑的慣用說法。吾人若觀察另外幾位學者所做的評鑑意義，可知他們彼此的看法非常接近，比方說：這些學者都主張評鑑是一歷程而非一產品，須有客觀資料做為主觀價值判斷的依據，目的即在提供改進之道，並進而協助決策的制訂。

上述學者所提出以「決策導向」為目的的評鑑歷程，可說是一種教育計劃的評鑑(Educational Program Evaluation)以別於成品評鑑(Product Evaluation)與學習個體的評鑑(Learner Evaluation)。由於教育計劃評鑑為今日美國評鑑的重點之一，且又與本文討論的評鑑概念關係密切，故有必要將教育計劃的評鑑特色和目的在此做一說明：

## (一) 教育計劃的評鑑與學習者評鑑的區別

前述有關評鑑概念的演變歷史中，曾論及早期的評鑑偏重測量學習個體的態度、性向及學習結果，完全是針對學習者個人的一種評估（assessment）。晚近的評鑑活動着重在評量影響學習的全盤因素，包括了教材、教法及教學情境等，可謂二種評鑑的基本區別。

## (二) 教育計劃評鑑的目的

在司瑰文未提出「評鑑就是決定教育上人、事或物價值的歷程」之前，評鑑人員多半只對受評者提出描述性的資料（Descriptive information），却不作任何的價值判斷。彼時評鑑目的只在描述受評者的現象，無需提供價值判斷；然而，現在一般的評鑑專家都同意司瑰文所說：評鑑的目的之一必須在能決定教育現象的價值。評鑑的另一目的，可說是協助決策人員制訂決策。

## (三) 教育計劃評鑑的特色

計劃性評鑑的特色可從三方面說明：第一個特色是描述性特色，因為評鑑時必須從各方面獲取不同資料以說明計劃的特質；其次為判斷性特色，是基於客觀資料所提出的合理判斷，任何有系統的計劃性評鑑都應該有此種判斷性資料；再者就是「決策導向」的特色，評鑑人員必須明瞭所蒐集的描述性或判斷性資料都將是制訂決策的依據。

至於為什麼要從事計劃性評鑑，一般說來有些評鑑的產生是因為受了決策機構的委託不得不執行，我國目前所實行的評鑑活動多半是屬於委託評鑑；另外有些評鑑的進行目的只在驗證受評者之所以能存在的理由；評鑑最重要的原因還是為了要提出合理可行的改進方案，改善原有缺失，促使進步產生。

從事教育計劃的評鑑可以獲得許多利益。諸如增加教育計劃的正確性、增加對計劃的了解等，同時尚可符合教育績效的要求，增加計劃內各成員間的溝通，進而促進對計劃的支持。然而，此種評鑑也有許多不妥之處，因為要描述計劃內容並做一說明與判斷，往往因此顧小失大，又因為計劃性評鑑與決策制訂有關，故而必然要對計劃中的某些設計提出批評，但衝突也可能就發生了。

不論教育計劃的評鑑利弊為何，在進行評鑑時，評鑑人員必須要了解絕對不應僭越決策人員的權利，尤其應該明瞭評鑑雖然能對教育設施提供深遠有力的影響，但評鑑絕不是解決所有教育問題的萬用靈丹。



總括而言，評鑑的確是一個多重科際整合的複雜歷程。這又可由以下二點加以說明。

#### (1) 從科學觀點探討教育評鑑的意義

上述的評鑑學者都不約而同的認為資料蒐集是評鑑過程中的一個重要階段，亦即收集客觀、實證的資料是評鑑活動中不可或缺的一部分。爲了蒐集資料，評鑑人員必須具備科學的知識與技術；比方說：他必須懂得問卷的設計、晤談的安排、或是邏輯分析法的運用，若缺少了這種客觀的專門技術，絕不可能指導決策人員去接受有關制訂決策的各種資料與改進途徑。進一步來看，蒐集與分析資料又須具備了測量的知識、統計原理的了解、及運用測驗的技術，因爲缺少了這些知識與技術，原始資料的獲得也不過只是一堆無用的資料，對評鑑活動而言真是毫無意義了。事實上，使用這些科學專技，很明顯的是爲了要使評鑑人員在從事評鑑工作時，能將一己的偏見減少至最低程度，同時也儘可能保持應有的客觀態度。豪斯（E. R. House）對於使用科學方法進行評鑑工作有一個很好的解釋。他說：「要之，我們實在得依賴科學的系統方法才能化除我們主觀心理的鬼祟作用。」（註二十八）

#### (2) 就哲學觀點探討教育評鑑的意義

雖然仍有許多學者對評鑑人員應否就其評鑑的對象作價值判斷爭論不已，但事實上我們要避免作價值判斷幾乎是不可能的。舉例來說：在從事「形成性評鑑」的過程中，評鑑人員與教育課程的發展有着不可分割的關係，同時他的主要任務之一，就在協助受評者選擇合宜的規準，以便實際評鑑課程計劃的價值。波爾佛斯也指出：「規準是由知識與價值導出的，價值是一個決定性的因素，因爲有價值才能決定在可能使用的廣泛知識中，那一部份將用做規準的基礎，價值也同樣用來決定何種材料須要被評鑑。」（註二十九）

綜觀以上國內外學者對評鑑所做的意義界定及筆者個人就科學與哲學二個觀點對評鑑意義所作的探究，我們可將評鑑意義作一統合的界定——「教育評鑑是一多重科際整合的複雜歷程，對於受評的教育計劃或成品，必須使用科學方法以蒐集系統客觀的資料，更要利用資料，參照價值規準，作一主觀的價值判斷，並進而提供改進之道，以協助決策的制訂。」

## 肆、教育評鑑計劃(Evaluation Plan)的組成要素

從事教育計劃性的評鑑，若能先有一週密的設計，不但有助於評鑑活動的實施，更有助於指導評鑑人員切實執行評鑑。事先提出評鑑的計劃（Evaluation Plan），能使評鑑人員認清評鑑目的，並發現達成目的的各種方法。此外，評鑑人員若按照計劃的步驟進行，又可不斷考核所蒐集供做決策的資料是否妥當，進而得到執行評鑑的最佳途徑。準此，故有必要將組成計



計劃的要素逐一分析。根據古勒 (D. D. Goeller) 與諾克斯 (A. Knox) 博士的建議，評鑑計劃的組成要素可大致分為七種，分別為：(一) 詳細說明受評計劃 (Program)、(二) 了解評鑑所要服務的對象 (audience) 需要、明瞭受評計劃的議題內容 (issues)、(三) 確定欲蒐集的資料、資料蒐集、解釋評鑑結果及提出評鑑報告。(註三十) 為熟悉各要素的特性，謹作以下的說明。

## (一) 詳細說明受評計劃

欲詳述受評的教育計劃，評鑑人員首先必須了解此計劃的內容與功能，其次可將用做說明的資料分成四類：第一類是描述計劃的整體內容、第二類是描述計劃本身的預懸目標、第三類是關於計劃的現況、最後一類資料在於了解參與計劃的人員有何期望。

描述計劃整體內容的資料，大約不在說明此計劃的特性為何、與其它組織或其它的計劃能否合作、有關教師與人員參與情形、及經費預算的概況；另外也該說明計劃所使用的教學內容與技術是否妥適、有無良好的管理運作及參與計劃的助理人員是否提供了適宜的服務等。第二類資料通常可做為評鑑此一教育計劃如何有效達成預懸目標的證據。評鑑人員一方面必須十分謹慎的參照預懸目標以評量此計劃是否有進步；另一方面要注意這些預懸目標不可限制了評鑑活動的進行。至於第三類描述計劃現況的資料，多半着重在了解不同時間下此計劃所發生的改變，如何改變以及有那些特性變得更為重要，或那些特性已不再重要等。最後一種資料着重在說明參與計劃的成員對此計劃的期望，評鑑人員可觀察各成員、行政人員、或財政機構在此計劃運作時是否表現某種關切之情。理想說來，評鑑人員應將四種資料做一累積記錄以為評鑑的依據。

## (二) 了解評鑑所要服務對象的需要

評鑑所要服務的對象，最主要的應是委託評鑑的機關或人員，其次可能是對此計劃感興趣的人士，財力支持的機關自然也是必須要了解對象，再者即計劃的參與人員了。評鑑人員所以要了解上述人士的需要，最重要的原因就在按其需要提供給他們有關評鑑發現的資料。因此，了解所要服務對象的需要，進一步就可知道他們對計劃的期望及他們希望達到的績效標準，如此方能溝通彼此意見，做好評鑑工作。

## (三) 明瞭受評計劃的議題內容

受評計劃中的主要議題內容往往都是評鑑服務對象認為重要的事件。明瞭這些議題內容的主要目的，就是因為這些議題對

受評計劃的實際運作上有着相當的影響。基本上，評鑑是要將重要議題內容給所要服務的對象作一解釋，供其參考。但是，評鑑人員選擇重要議題的判斷依據為何？到目前為止，美國評鑑界尚無妥切的答案，然古勒博士提供的方法倒也值得參考。古勒認為判斷議題的優先重要次序有三種可行方法：依服務對象所表示對某議題的關切程度列出其優先順序，其次可按受評計劃中的各方面要素定其優先順序，然後可循受評計劃的預懸目標列一優先順序。（註三十一）如何在三種途徑中做一選擇，還得視決策人員的意見而定。毫無疑問的是評鑑人員本身的性格、經驗與專業水準都足以影響對優先順序的判斷，也可以幫助他發現最具優先的主要議題。一般說來，評鑑人員應該仔細研察由服務對象提出的議題，如果認為所提議題內容忽略了最重要的部份，即應坦白相告；如果所提議題無甚重要，也不必固執己見非要服務對象改變原議題不可，因為評鑑人員本身有義務顧及服務對象的需要。透過此種關係，評鑑人員與服務的對象方可更加明瞭評鑑的方向與目的。

#### 四 確定欲蒐集何種資料

評鑑人員需要知道何種資料值得蒐集，且自何處蒐集是一重要的工作。通常資料來源處大致與受評計劃的人事有關，例如：參與受評計劃的人員、現有文獻檔案、或其它有關的研究資料等。所要蒐集的資料應包括受評計劃人員、內容、目標及人際關係等資料；再者，有關此計劃的實施價值及其成果等資料，可自不同的來源處蒐集；至於受評計劃成敗與否的因果分析與推估等資料也應蒐集齊全，資料是否有用，端賴其與服務對象需要及重要議題的關係而定。因此，評鑑人員應該時時將所要服務對象不同的需要牢記在心，才可決定蒐集何種資料方有價值。

#### 五 蒐集資料

資料蒐集與評鑑人員的技術密切相關。評鑑人員通常須先確認資料蒐集的步驟，然後再實地進行蒐集工作。蒐集資料的技術包括可種不同方法，諸如利用問卷調查、意見表、觀察法、晤談法、測驗法、行為分析法、交互作用分析法等等。評鑑人員必須決定選用何種方法蒐集資料與需要蒐集多少資料。值得注意的是所蒐集的資料必須具有效度，也要顧及信度。資料是否需有類化的性質，對評鑑工作並不重要，因為評鑑重點永遠放在特殊的教育現象上，故不須類化至其它情境。

#### 六 解釋評鑑結果

資料蒐集之後，必須加以整理組織，方可得到與議題有關的訊息，然後再將發現的結果提出合理而具意義的解釋。評鑑人員接着可依不同服務對象的需要、評鑑過程中的特殊發展情形及受評計劃所強調的議題內容，來說明資料的意義。此時，評鑑

人員尚無須考慮如何與所服務的對象溝通，應該注意的乃在如何去了解這些發現的結果，使他所做的解釋在日後能為服務的對象接受。更重要的是評鑑人員所說的解釋，不但要使服務的對象了解，更要促進受評計劃的改善與調整。

## (七) 提出評鑑報告

評鑑人員在提出報告之前，必先確定所整理的資料能滿足服務對象的需要，所提的報告資料亦皆詳實無誤；此外，還得知道服務對象的背景，以便提出的報告能為他們所了解。評鑑人員為何需要了解服務對象的背景？主要原因仍在某些人能接受的報告資料不一定能為他人接受。舉例而言；有些人明白成就測驗分數資料的評鑑報告及其意義，有些人則無法接受；有些團體喜歡接受以統計資料為主的評鑑報告，但有些團體根本就無法接受此種統計方式的資料說明。因此，評鑑人員在提出報告以前，必須先確定何種方式最能為評鑑的服務對象接受，否則提出的報告就無法獲得預期的效果了。

評鑑計劃的七種組成要素，是古勒博士的建議，評鑑工作的進行不一定需要考慮到每一個因素，重要的是能彈性的使用這些要素。當然這與評鑑人員本身的素質有關，扼要的說，只要評鑑人員認為他所採用的評鑑方法最能滿足所服務對象的需要，最能達成評鑑的目的，此種評鑑就都有效。

## 伍、教育評鑑模式的比較分析

評鑑計劃 (Evaluation Plan) 是評鑑工作開始前為達到良好效果所做的週密規劃。評鑑模式 (Evaluation model) 却有其不同的功能。所謂模式，是一種結構行為的力量。(註三十二)換言之，模式乃有着一系列次序的行為或步驟，藉此以達其預懸鵲的。蔡培村先生提出「理論模式是根據受評者所建立的一種結構，有其獨特的功能和一系列的實施步驟。」(註三十三)此處所提根據受評者所建立的一種結構，實際上也就是依據受評者的特性，需要及期望所建立的一系列有順序的結構行為，按此結構行為進行，則可達成某種預懸目標。張植珊先生也提到評鑑模式的產生是因為有實際的需要，他認為使用模式的目的「在使評鑑的程序適當地加以概念化，而有助於評鑑的適切進行。」(註三十四)筆者認為評鑑模式代表的意義即在提出事情應該如何進行的參考藍圖，因此，有其優先順序排列的結構行為，以利評鑑的執行。自一九六七到一九七〇年間，美國教育評鑑的學者們先後提出了各種不同的評鑑模式。其中司瑰文與司德克在一九六七年提出的模式，在理念架構上較為相似，他們都認為評鑑必須作價值性的判斷；史達夫賓在一九六八年提出的模式又與艾爾金在一九六九年所提有關協助決策制訂的模式十分接近；而泰勒早期所提出的模式又間接影響了波爾佛斯在一九七〇年提出比較目標與成就間差距結果的模式。有關史達夫

(152)

賓、艾爾金與泰勒的評鑑模式國內已有學者作過介紹，（註三十五）故不再贅述；司瑰文的模式可用司德克的評鑑模式做代表，亦省略不談，下面就將司德克、波爾佛斯及另一學者鄧運克（T. D. Tenbrink）等人的評鑑模式做一比較，並分析其優劣利弊，以爲吾人參考借鏡。

## （一）司德克評鑑模式

司德克（Stake）是美國教育評鑑界中頗具盛名的專家之一，也有人稱他是評鑑界中極富人文思想的學者代表。他所發展出的評鑑模式在美國評鑑中爲一盛舉。司德克十分反對一般人視評鑑爲心智活動的測驗，更反對許多教育界人士以爲進步的科技乃教育的救星他認爲評鑑必須要有價值判斷的資料，也就是要用判斷性資料來說明爲什麼要使用各種心理測量的資料。直言之，司德克認爲不但要有描述受評計劃的資料；更要能對受評的計劃作一價值的判斷。古勒曾說司德克模式比起許多其它的評鑑模式與我們的生活要接近得多。（註三十六）

司德克模式主要是以資料表來說明各種資料的關係與用途。由於司德克認爲評鑑就是要利用描述及判斷的資料對某一教育計劃做詳盡的說明與充份的判斷，因此，其所要蒐集的資料就具有「描述性」（Descriptive）與「判斷性」（judgemental）的兩種特質。這兩種資料又因其本身性質與功能的不同而產生了不同種類的描述性資料及判斷性資料。下表是司德克模式中資料蒐集的說明表。首先必須指出理論架構的重要性，再逐一說明兩種資料表的意義。

〔表一〕 司德克模式資料蒐集的說明表

教育評鑑概念與模式之研究

某教育計劃 的理論架構 Rationale	原先意圖 觀察結果 intents observations		理想標準 價值判斷 standarde judgemente	
	Description matrix 描述性資料		judgement matrix 判斷性資料	

antecedents  
(教育計劃實施前)

Transactione  
(實施過程中)

Outcomes  
(結果)

每一教育計劃設立理論架構的目的在呈現所代表的哲學背景與主要目的。透過理論架構，評鑑人員可以了解此計劃的來龍去脈，認清所要達到的目的，並據此訂出賴以作判斷的理想標準；同時亦可參照理論架構評鑑此計劃的預懸目標是否恰當，進而幫助評鑑人員選出有關的參照團體，俾能共同提出對此計劃作價值判斷的資料。雖然要呈現一個明確的理論架構並非易事，但若缺少了理論架構，就無法使評鑑工作順利完成。因此，如何呈現理論架構是一個十分重要的步驟。

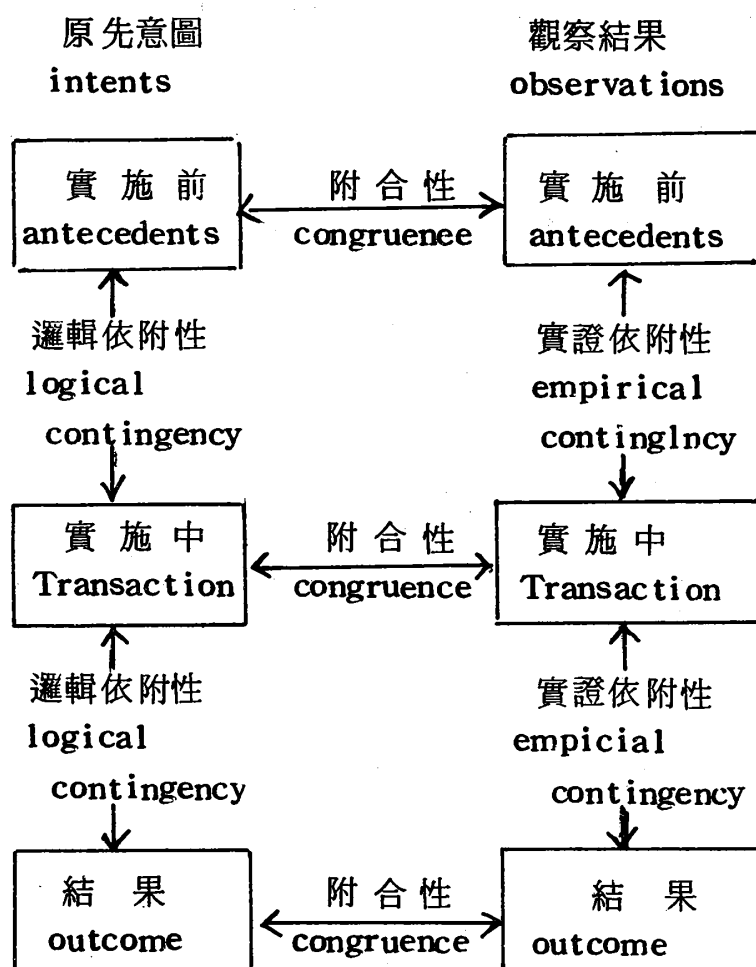
其次討論表一所呈現的評鑑資料表。此表有十二個小空格，每一空格代表一種資料。依照司德克的想法，不論是描述性或判斷性資料，都應有三個層次不同的資料，此即：教育計劃實施前的資料、實施過程中的資料以及實施的結果。實施前的資料，可說是某教學計劃未實施前學生的情形，包括性向、興趣、毅力及先前經驗等，這也是編序教學專家所說的「起點行為」或「起始行為」，這些資料往往與教學結果有關；第二層次實施過程中的資料，是指當某教學計劃實施時，教師與學生間、學生與學生間、家長與輔導人員間的各種交互活動的資料，這些活動的進行，為的是要達到教學計劃的目標；第三層次實施的結果就是教學實施後的長程與短期的結果。例如教學計劃實施後對學生的態度、能力及學業成績產生何種影響；對教師、行政人員及輔導人員又有何種影響；這也是早期傳統的評鑑人員所注意的重點。實施前與實施後的資料較為靜態，而實施過程中的資料當然就屬於動態的資料了。

表一中描述資料又可分為原定意圖（*intent*s）與觀察結果（*observations*）二個階段，亦即指出評鑑人員在蒐集描述性資料時，必須先記載此教育計劃原先意圖設定的目標，再參考所觀察的實情，兩相比照得一客觀資料。所謂原先意圖設定的目標，是教育計劃原先希望對學生產生的一種影響與對教學環境可能有的影響。因此，評鑑人員必須不斷的向參與計劃的教師和課程專家人員蒐集有關其原先意圖設計目標的說明，然後才能參照此資料蒐集觀察性的描述資料。觀察性的描述資料意指評鑑人員運用其個人對教學計劃三個層次所做的直接觀察得來的資料。當然也包括評鑑人員使用種種不同的工具（問卷、測驗等）所獲得的資料。評鑑人員必須依據教學計劃目標來主觀的決定何種資料值得蒐集。

其次再討論如何處理描述性資料的方法。司德克認為評鑑人員在處理三個層次的二種描述資料時，首先必須找出教學計劃實施前、實施中及實施後三層次資料間的邏輯依附與實證依附關係，接着再找出原先意圖達成的目標與實際觀察結果的資料間有無符合的關係。下表為處理描述性資料的方法。

評鑑人員在處理原先意圖與實際觀察這兩種描述性資料時，既然必須找出這兩種資料間有無符合的關係，即應觀察此教育計劃原先希望達成的目標在實際施行時是否確實發生，若答案為肯定，就證明符合性關係存在；否則即不符合。當然，要找出教育計劃實施前、實施中及實施後三個層次的原先希望達成目標資料與實際情況完全相符合是不太可能的。評鑑人員觀察原先意圖與實際結果間的差距，即可依此描述其間的符合程度。值得注意的在於資料相符並不證明此資料可靠，所證明的只是原先意圖的理想確實現而已。至於原先意圖資料的三個層次（實施前、實施中及實施後）必須有邏輯依附性，目的即在指出資料間確有合理的因果關係存在，此種關係也能促進教育措施的改善。評鑑人員的工作就在明瞭資料中某一結果的產生往往是源自產生此結果前的其它情況，他必須依賴自身以往的研究經驗與專業訓練做主觀的判斷，以發現資料間的邏輯依附關係。換言之，若某教育計劃原先意圖的描述資料欄內將學生起點行為列在實施前空格，積極反應列在實施中的空格，良好學習結果在實施後的空格內，評鑑人員就能憑着自身所受的专业訓練判斷這三層次的資料間具有邏輯依附的關係。

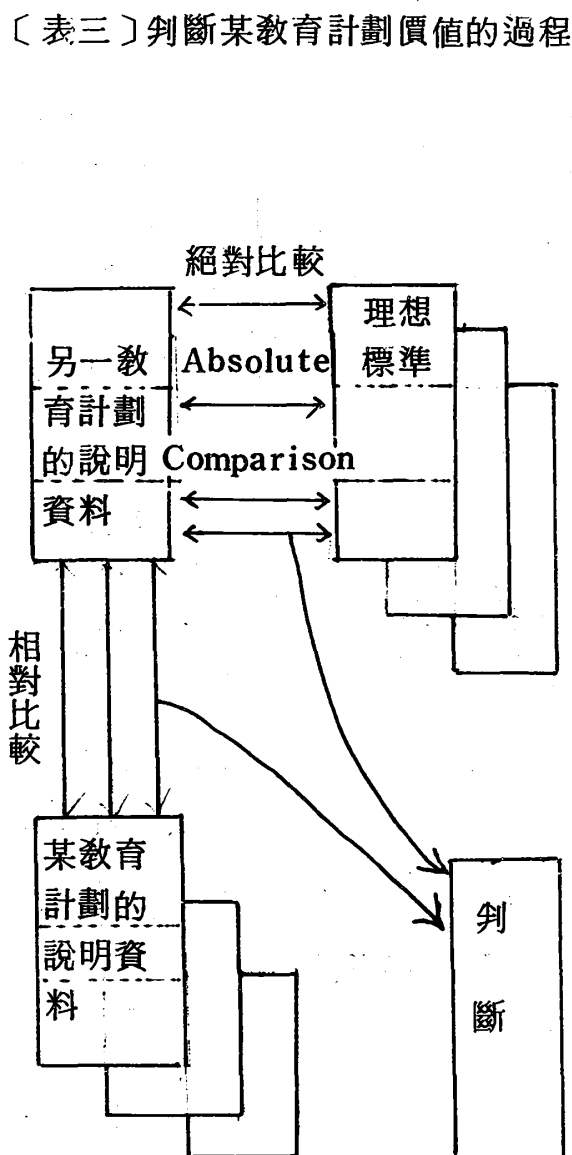
〔表二〕描述性資料處理過程





處理觀察性描述資料時，評鑑人員必須找出其三個層次間的資料是否具有實證性的依附關係。欲求「實證性」(empirical) 依附關係的存在，就必須有實證性資料，例如欲知道學生在某科學方面的快速進步是因為教師採用一種成功的教法，此時評鑑人員就必須用前述各種科學的方法蒐集實際發生的各種資料，以驗證學生成績的快速進步結果的確是教師教法的成功。總之，不論是處理描述性的原先意圖資料或是實際觀察的描述資料，評鑑人員都應該特別注意到此兩種資料間的符合關係及各資料內三個層次間的依附關係。

有關判斷性資料的處理方法也可由二方面來探討。評鑑人員若要判斷某教育計劃的價值為何，可自二個角度去判斷。第一種可先參照一組理想上的標準，將教育計劃與之比較，然後提出絕對的價值判斷；其次亦可參照另一性質相同的教育計劃，兩者做一比較，可得另一種價值判斷。第一種判斷可說是絕對比較性的判斷，是在理想層次上所作的應然式判斷；第二種乃是相對比較性的判斷，是在實際情況下所作的比較判斷，此種判斷也是司瑰文所提出的比較評鑑(Comparative Evaluation)。(註三十七)因此評鑑人員必須依照實施前、實施過程中及實施後三個層次去蒐集一組理想標準的資料，此種資料可向參與教育計劃的一些參照團體蒐集。重要的是，價值判斷資料本身就須評定每一理想標準的重要價值性何在。誠如司德克所說：「教育評鑑上一個合宜的判斷就是一個決定，此判斷將決定何種參照團體提出的理想標準最值得重視，俾做為制訂行政措施的依據。」(三十八)評鑑人員如何判斷某教育計劃的價值。其過程可參考下表而獲致較清晰的概念。



扼要言之，司德克模式主要是為評鑑各種教育上的計劃而設計。評鑑過程中的主要特色在如何蒐集「描述性」與「判斷性」的資料。其提供的理念主要強調評鑑人員必須從各類不同的服務對象處蒐集資料以列成資料表。處理資料時必須顧及教育計劃實施前、實施中及實施後的各種資料間是否有邏輯依附性與實證依附性的關係；接著再找出其原先意圖資料與實際觀察的兩種資料是否具有相當程度的符合性，進而做成絕對比較與相關比較的價值判斷。其最終目的是要為決策人員提供制訂政策的依據。司德克模式最大的優點在於其提供了系統蒐集和處理資料的方法，同時又能考慮到絕對與相關比較的判斷，可說具有相當的效用。但此模式最引人垢病處即在容易引起參與教育計劃的成員間因不一致的價值看法而發生衝突。這也是選擇使用司德克模式進行評鑑時應該注意的一點。

## 〔波爾佛斯(Provus)評鑑模式〕

由波爾佛斯在一九七〇年提出所設計的模式稱做「差距評鑑模式」。(The Descrepency Evaluation Model) (註三十九) 其所以稱做「差距評鑑模式」，主要是因為此模式建議由協同的評鑑人員團體評鑑教育計劃實施過程中其所預懸之標準與實際的成就之間有多少「差距」，獲得此種差距資料的主要目的是要提供給負責人員，由其重新修正原有計劃內容，並再度實施，希望藉此達到一個理想鵠的。此模式有五個順序階段，每一階段中，都有預設標準與實際成就間比較而產生的差距資料。在第一個階段中，預設標準就是最早設計此教育計劃時的規準；其實際成就亦即提出有關此教育計劃設計內容的說明，兩相比較所產生的差距就應向負責人員提出報告；第二階段中的預設標準就是第一階段所提出的成就資料，換言之，就是有關教育計劃內容的說明，實際成就此時就成了由觀察此內容各部份運作情況的資料，負責人員可利用二者間的差距資料重新修訂計劃或改變運作的步驟；第三個階段內的標準是由關於此計劃的實施過程說明所構成，其實際成就即為實施過程在短期的成果資料，比較下的差距資料一方面可用來重新調整計劃，一方面也可因此對施行過程有較妥善的控制；到了第四個階段的預設標準就是此計劃的理想目標，實際成就資料即為此計劃所產生的最後成果；接著第五個階段是由決策人員所做的分析，此時決策人員若須要分析一種以上能產生同樣效果但設計內容却不相同的教育計劃時，就應該使用「成本效益」(cost-benefit)的分析方法以決定何種教育計劃的研究設計能有最大的績效。第一階段內的原設規準之所以確立，目的即在提供其它階段內建立預設標準的依據，同時也為了蒐集教育計劃中三個基本因素的資料，此即輸入(input)、過程(process)、及輸出(output)三方面的資料。下表為波爾佛斯模式中所提出的五個評鑑階段，可對照說明參考：

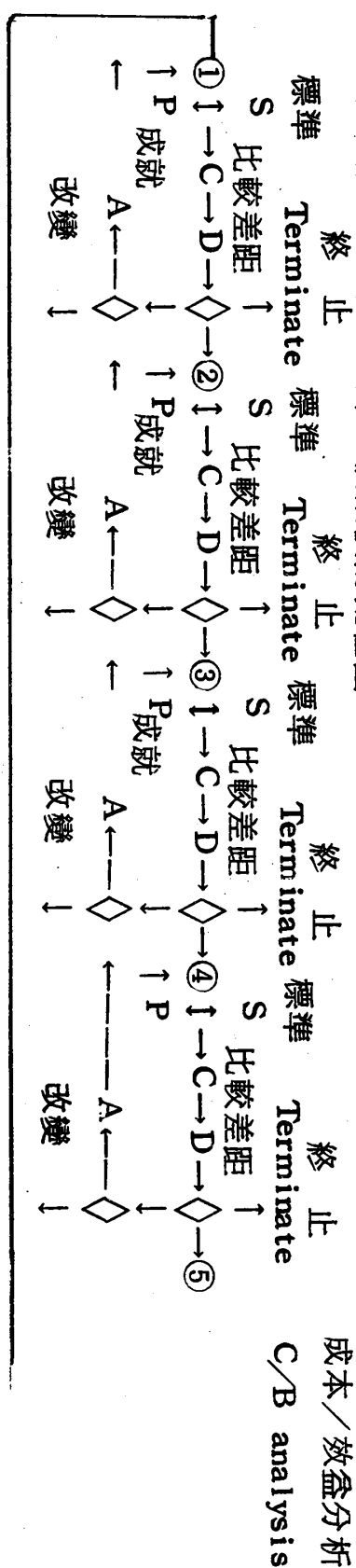
( 158 )

〔表一〕波爾佛斯模式的五個評鑑階段

階段	成就 Performance	標準 Standard
I	原設教育計劃的說明 輸入 輸出 過程	教育計劃的原設規準
II	教育計劃運作情形	原設教育計劃的說明 輸入 過程
III	教育計劃 實施的短期成果	原設教育計劃的說明 過程 輸出
IV	教育計劃實施後的終極成 果	原設教育計劃的說明 輸出
V	實施比教育計劃的成本	使用產生同樣成果的其他教 育計劃的成本

波爾佛斯模式中，前四個階段的差距資料都可供決策人員用作修正原設計計劃的參考，並再將其實施，此種行為構成一個「反饋圈」(Feedback Loop)。「反饋圈」的產生是因為理想標準與實際成就間有了差距的存在。當然，若無差距也就無須評鑑教育計劃。含有反饋圈的評鑑模式包括了四個比較實際成就與預設標準間差距的發展階段，而每一階段又都可能要處理輸入、過程及輸出等方面的資料，因此波爾佛斯又將這種比較的過程用流程圖方式表現出來：

〔表二〕波爾佛斯評鑑模式中五個階段的流程圖



概括而言，第一階段的評鑑目標是要根據原設計計劃的規準評量有關原設計計劃內容的說明是否週全，其與規準所要求的理想是否一致及其與他種教育計劃的競爭能力是否足有餘；第二階段的評鑑目標是要增加此計劃的內容與其運作情況間的符合關係，所謂增加符合關係即意指計劃本身獲得改善；第三階段的評鑑目標最主要在使參與計劃的人員明白此計劃的運作實施對其所產生的成品有何種影響，也就是在此階段評鑑完竣後，參與計劃的人員會知道他們所設計的計劃內容是否達到預期效果，若無，就必須找出原因以便修正此階段的計劃內容以供運作；到了第四階段，評鑑團體人員可採用各種研究方法以明白此計劃的主要目標是否達成；完成上述四個發展階段的評鑑後，就進入使用成本效益分析法的第五階段，目的在比較此計劃與其它計劃的競爭力孰強，如是方能決定有效的資源分配。

波爾佛斯評鑑模式的主要目的在提供決策人員資料，俾憑以決定維持、取消或改進一個教育計劃。波爾佛斯強調協同式的評鑑方法，主張由評鑑人員找出差距的現象與原因並形成反饋的功能，使每一階段都可獲得合宜的修正與改善。我國學者錢卓升先生對評鑑所做的說明頗能將波爾佛斯的模式功能勾勒出來。錢先生說：「評鑑是一段工作——它是工作的尾聲，也開工作的先端；它在探討疏失，也重視欣賞成就；它能綿延慣例，也可以創造新獻；它在評鑑某項工作，而時過境遷，這評鑑的本身，也成為評鑑的對象。」（註四十）將此一段說明參照波爾佛斯的評鑑模式流程圖，即可發現二者的理論架構上有異曲同工之妙，吾人選用波爾佛斯評鑑模式的同時，若能配合錢卓升先生的評鑑界說以進行評鑑工作，必能達到理想的預期效果。

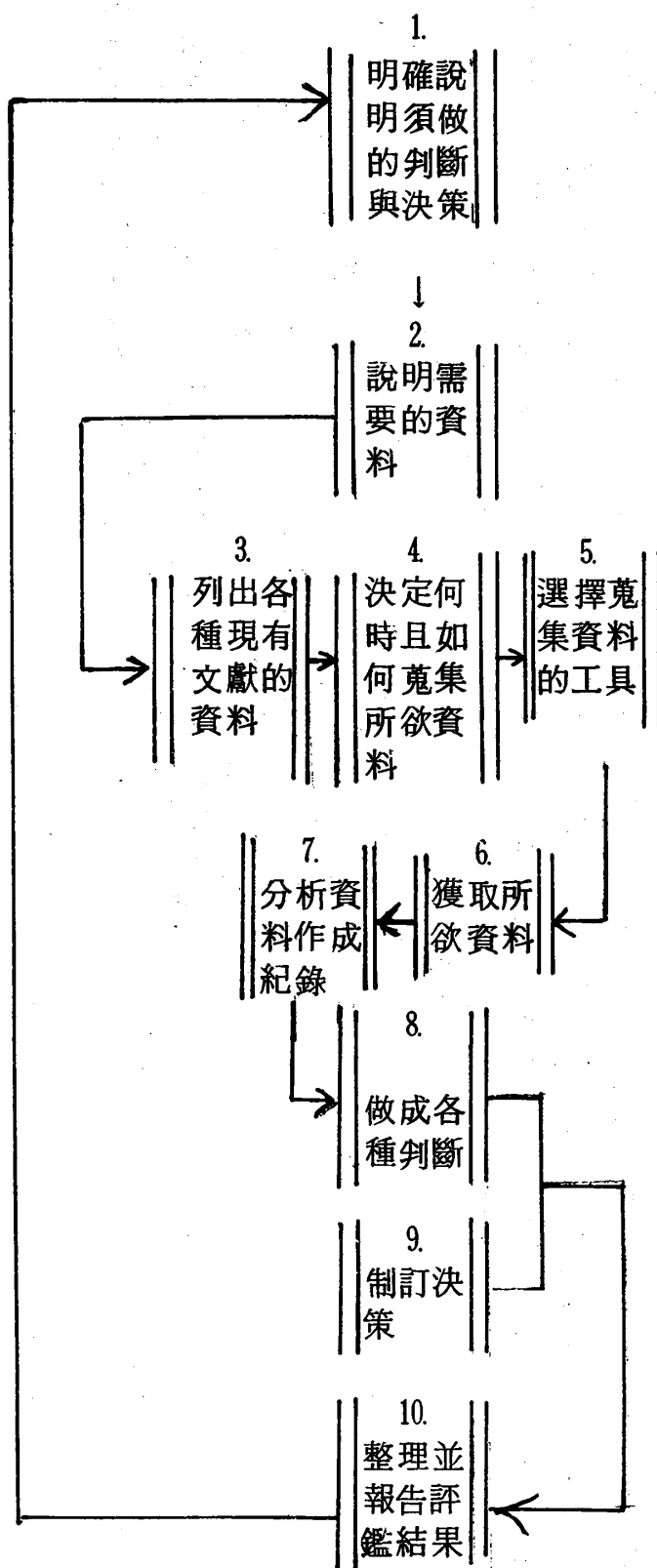
## （一）鄧運克(TENBRINK)評鑑模式

司德克與波爾佛斯的評鑑模式基本上屬於評鑑各種教育計劃的模式。此處要分析的鄧運克模式雖然也是一種教育計劃的評

鑑，然而偏重在教室內教學活動的評鑑，換言之，鄧運克模式主要在為教師提供一個合理客觀的評鑑方式以供改進教學之參考。鄧運克主張「評鑑是一個獲取資料、並利用資料做成判斷以供制訂決策的歷程，」（註四十一）與司德克及史達夫賓對評鑑的解釋頗為接近，皆認為評鑑是一種歷程、是一種利用資料做成判斷的歷程，其終極目的即在協助制訂決策。

鄧運克評鑑模式是由三個實施階段組成，每一個階段又含有一些運作的步驟。三個階段分別為準備期、資料蒐集時期與評鑑期。在準備階段時期，評鑑人員（即教師）首先要確定必須做何種判斷與決策，其次再說明該蒐集何種資料，知道有那些可用的現存文獻檔案之後，再決定在何時、用什麼方法去蒐集資料。資料蒐集階段期，評鑑人員為獲取資料必得使用前面一再強調的各種方法，諸如：量表、測驗等等，資料獲得後就進入第三個評鑑的階段。在這個階段中，將所獲得資料（例如學生的測驗分數）與參照標準做一比較，再根據比較結果加以推測未來或形成判斷，進一步供決策制訂之用。教師所要制訂的決策照鄧運克的建議是應該簡明易行、彈性客觀。（註四十二）簡明易行能省時省力，彈性客觀能週密詳備，換言之，制訂出的決策方可幫助教師獲得理想的教學效果。評鑑階段內最後一個步驟即為報告評鑑的結果。教師最常用的報告方式當然就是成績單，但若能配合類似家長會、訪問或信函等方式。更可將學生的實際學習成果做一更客觀的報告。鄧運克評鑑模式簡單易行，值得教師選為教學評鑑的模式，為能進一步說明，也將其評鑑過程列表於下。

〔表一〕鄧運克評鑑模式的評鑑過程



鄧運克模式最主要的目的在幫助教師處理有關學生及教學上的各種重要問題。此模式特別強調評鑑為一獲取資料的歷程，進而協助制訂決策與做成判斷。鄧運克也指出決策是在各種改變途徑中擇一而行，判斷是要評估教學現況並預測學生未來的成就。此模式的缺點在其可應用範圍嫌小，但此模式的簡單易行及其所強調的評鑑過程可繼續不斷的進行下去，實為其所具特性。教學上若能配合鄧運克評鑑模式，則教師對教學所訂決策與判斷可不斷接受評鑑，俾訂出更佳策略，進而達成教學上的理想目標。

## 陸、教育評鑑中值得爭議的問題——現況與未來

尚未討論教育評鑑中有那些令人爭議的問題以前，筆者先就一個實例來說明一些現象。「美國某中學一位老師，深信只有標準測驗卷測量的分數才是最能代表學習成就的客觀資料。某學期新擔任一班學生的歷史課，學期結束後，甲學生哭喪着臉來見他，告訴他其以往成績都在八十五分以上是優等生，但這學期只有八十三分差二分才能做優等生，希望某老師能看在甲學生平常學習過程的努力表現加給二分。某老師同情的說：『我再查查你的卷子』，過了一會兒，某老師臉帶憐憫、聲調柔和的說：『唉！真沒辦法，你就是八十三，測驗卷的評量是不會錯的！』」在這個實例中，有幾個十分值得討論的現象：第一：某老師堅持教學評鑑就是利用測驗卷測量學習成績；第二：某老師以為評鑑只需測量學習的最後成果；第三：某老師以為測驗所得的分數即評鑑的結果。由這三個現象又可以再推論出幾個相關的問題：第一：評鑑究竟是不是測驗呢？第二：評鑑應是一個繼續不斷的歷程亦或只是產品的評量？第三：量化資料的獲得是否達成評鑑的任務？雖然在探討評鑑意義時，筆者已指出評鑑須借重測驗，但評鑑不是測驗。「某老師認定該學生是八十三分」就是犯了只注意所得的量化資料的錯誤，却因此忽略了必須解釋此資料的實質意義。因為某老師沒有將該生在教學前的資料如過去的學習經驗、對歷史的興趣等與學習過程中的情況如提出有意義問題、對問題的思考與判斷能力及最後由測驗所得的分數資料加以詳盡的說明並做一絕對比較性的判斷；又沒有將該生與班內其它學生的資料做一相對比較之後再做評鑑報告，可說某老師根本不了解評鑑的意義，因此所得有關該生的資料既不完全又不客觀，當然也不可能做一正確的價值判斷了。其次，某老師忽略了評鑑的繼續不斷的歷程特性，所以忽視了評鑑過程中的反饋作用，若某老師未能提供學生反饋的機會，如何能知道所教是否對學生發生影響？又如何修正自身的教學方法，使之獲得改進，並提供有效教學呢？更何況某老師完全依靠測驗工具判斷學生的價值，對某老師來說該學生已不是「張三或李四」，該生只是「八十三分」，換言之，某老師將量化資料視為終極目的，完全忽略價值判斷的實質意義，才會把學生視為「分數」而不再是「人」了。這也正是美國人文主義派教育學者批評最厲的一點。例如美國教育界有名的思想家尼爾波斯曼（Nel-

Postman)，在其一九七九年出版新書內，就指出今日美國評鑑界中，許多評鑑人員完全變成了科學技術的奴役。波斯曼認為「教育評鑑是絕對必須的，」（註四十三）但評鑑人員若只重視用科學的工具去求得客觀的量化資料，以為人的問題可以完全依照物的問題來處理，對教育造成的傷害將是無法彌補的。科學研究的工具，往往不具人味，所得量化數字雖然容易處理，但根本不是評鑑的重心，評鑑人員若不能解釋教育現象中實質問題的前因後果，不能提出教育上理想的價值標準的判斷，則評鑑充其量也不過是在今日科技文明充斥下另一種「蔑視人性」的方法而已。事實上，司德克所設計的模式，就是要為教育計劃的評鑑提供二種性質不同的資料，他雖然贊成評鑑人員必須用各種科學的統計測量方法蒐集資料，更指出非科學性方法獲取資料的同等可貴，目的也就是希望評鑑人員雖能利用科技上的新方法，但絕不可為科技方法所役而成了迷信科技的瞎眼專家。古勒等人也一再強調教育評鑑是一種充滿人味的工作而非冰冷的技術。（註四十四）特別值得一提的是，即使今日美國評鑑界中大力提倡科學方法的學者柯萊與龍氏（Cooley & Lohnes），也提出了類此的建議。（註四十五）他們認為評鑑人員必須儘量避免閉門造車，必須了解教育界中所有的專家都希望對世界提出改進的方案，因此應互相交換有關教育政策的看法與意見，而評鑑人員應負的任務就在扮演組織領導的角色，應將教育上各種不同的貢獻提供給決策制訂的人員以為參考。其次他們建議評鑑人員必須有科際整合的觀念，要具備政治、經濟、管理、哲學、心理等方面的知識，才能與社會大眾建立良好的關係，因為「評鑑是關於人也是為了人才產生的活動。」（註四十六）

由上面的論述可知，今天美國教育界在討論評鑑的現況時，已明顯指出了評鑑未來的發展方向：那就是科學方法與哲學理論齊頭並進，量化資料與價值判斷等量齊觀。唯有如此，評鑑的價值導向任務（註四十七）才可完成，也才能使教育走向一個理想的境界。

## 柒、結論建議

綜合以上對教育評鑑意義、模式及其現況問題與未來發展方向的探究分析，可知教育評鑑乃人類彼此在溝通上無法避免、十分需要且又極為自然的一部份。國內今日推展教育評鑑，從應然層次觀之，絕對有其必要；就實然層次來看，也是勢在必行。然而在推行教育評鑑活動的同時，應如何避免先進國家曾犯的錯誤，來發展適合吾國文化背景與社會需要的評鑑模式，以促進教育績效，改善教育措施，達成教育的理想目標呢？筆者認為首先必須建立正確的評鑑觀念，認清教育評鑑的「動態歷程觀」，亦即明瞭教育評鑑是一個循環不已的繼續歷程。此種歷程為教育提供了不斷改進缺失，不斷追求進步的可能。透過評鑑，人們看見教育措施的妥當與否，發掘錯誤、找尋原因、蒐集資料、解決問題，使人類教育能在此生生不已的循環歷程中朝向理



想價值依歸的發展方向。因此，評鑑人員是身負重任，既不能忽視評鑑對人類教育的消極角色——指出弊端、證明績效；更不應忘却其積極任務——促進改善、導向完美。為達成任務，評鑑人員本身不但必須具備評鑑工作的科學技術，也須充實哲學理論的博雅基礎，更須切記教育評鑑是以「人」為重心的活動歷程，它的終極鵠的在為人類提供完美的教育設施。有了如是了解，才能進一步來談如何做好實際的評鑑工作。要推展教育評鑑工作，筆者認為可以嚐試三種途徑，以下就分別說明。

(一)在師範大學及師範學院內成立評鑑研究中心——評鑑研究中心在國外許多大學都有設立，主要目的即在為當地教育提供評鑑服務。我國在短期內若期望有評鑑專才，最好的方法就是聘請教育學者與心理測驗統計專家組織評鑑研究中心。一方面借重教育學者們既有的專業知能以了解整個教育現象中的複雜問題；另一方面又可倚重測驗統計專家們設計精確可靠的客觀評鑑工具以為蒐集資料之用，兩相配合，必可發揮統整功能，推動教育評鑑工作。

(二)遴選教育學者赴國外進修研究——教育學者本身已具備深厚的學養知能，若能由教育部或國家科學委員會有計劃的遴選有興趣研究評鑑的學者赴國外從事短期進修，進修期間一切經費，由國家負擔，期滿返國即可貢獻所學，如此必能推動教育評鑑的研究工作，並於短期內收到實效。

(三)開設教育評鑑的碩士班課程——(此項辦法的基本假定是師資問題已獲解決。)筆者所以建議開設評鑑的碩士班課程，主要是以師範院校的教育研究所為重心。這是因為教育研究所的研究生本身已具有相當的教育理論基礎與專業技能，對教育問題亦有相當的了解，若提供其修習教育評鑑課程的機會，也較能達到培養評鑑專才的目的。此即美國許多大學只在其研究院設有教育評鑑課程的原因之一。國內如果考慮增設教育評鑑的碩士班研究課程，就長期需要而言，實為獲致所需人才的最佳途徑。

以上幾點建議，雖為筆者個人管見，施行上或有人力、財力與物力方面的困難，但若排除阻礙，付諸實行，則教育評鑑在可見的將來必可耕耘出一片欣欣向榮的園地。

## 附註

( 164 )

註一·Blaine Worthen & James Sanders, *Educational Evaluation : Theory and Practice*, Ohio, Charles Jones Publishing Co., 1973, P 2

註二·黃炳煌·「教育評價概念的分析及其發展」，昨日今日與明日的教育，台灣開明書局，一九七七，頁六八六。

註三·Jack Merwin, " Historical Review of Changing Concepts of Evaluation: " *Educational Evaluation : New Roles, New Means*, National Society for Study of Education, 1969 Ch , P 6 .

註四·Ibid, P 9 .

註五·Ibid, P 10 .

註六·Ibid,

註七·Ibid,

註八·Ibid,

註九·Ibid, P 12 .

註十·Ibid,

註十一·Ibid, P 14 .

註十二·Benjamin Bloom(ed), *Taxonomy of Educational Objectives; Hand Book* [1]: Cognitive Domain, New York, 1956

註十三·Jack Mewin, OP cit., P 19 .

註十四·Michael Scriven, " The Methodology of Evaluation " in Worthen & Sanders, see note 1, P 62 .

註十五·Ibid,

註十六·同註二·頁六八八。

註十七·同前。

註十八·Robert Stake, " Countenance of Educational Evaluation, " in Worthen & Sanders, see note 1, P 125 .

註十九·Scriven, OP . cit. P 61 .

註廿·Malcolm Provus, "The Discrepancy Evaluation Model" in Peter Taylor & Doris Cowley, Reading in Curriculum Evaluation, 1972, P 117.

註廿一·Lee J. Cronbach, "

註廿二·Lee J. Cronbach, "Course Improvement Through Evaluation," in Worthen & Sanders, see note P 44.

註廿三·Daniel Stufflebeam, "New Definition of Evaluation," PDK, Educational Evaluation & Decision Making, Ch2, P 38.

註廿四·Marvin Alkin, "Evaluation Theory Development," in Worthen & Sanders, see note I, P 150.

註廿五·李聰明, 教育評價的理論與方法, 幼獅書局, 台北, 一九七二, 頁三。

註廿六·陳瑞龍, 教育評價法, 高長印書局, 台南, 一九六五, 頁十一。

註廿七·蔡培村, 「CIPP 評價模式之評介」, 教育文粹, 八期, 六十八年六月, 頁一〇八。

註廿八·請參見拙作「教育評鑑應有價值導向的任務」, 今日教育, 三十七期, 六十九年一月, 頁二十五。

註廿九·同前。

註卅·Cooler & Knox, Evaluation in Adult Basic Education: How & Why, 1976, Ch3, P 36.

註卅一·同前。

註卅二·葉維廉等, 中國古典文學比較研究, 黎明文化公司印行, 六十六年十月, 頁二。

註卅三·同註廿七。

註卅四·張植珊:「教育評鑑的實施與展望」, 昨日今日與明日的教育, 台灣開明書局, 一九七七, 頁七〇三。

註卅五·參見註廿七及卅四。

註卅六·此係筆者在美進修時所得自教授之口頭資料。

註卅七·Scriven, OP. cit, P 83.

註卅八·Stake, OP, cit., P 117.

註卅九·Provus, OP, cit., P 117.

註四〇·呂俊甫等著, 教育心理學, 大中國圖書公司, 台北, 六十一年八月, 頁五六六。

註四一·T. D. Tenbrink, Evaluation: A Practical Guide for Teachers, New York, McGraw Hill, 1974, P 8.

(166)

- 註四二·Ibid, P16.  
註四三·Neil Postman, *Teaching as Conserving Activity*, Delacorte Press, New York, 1979, P217.  
註四四·Gooley & Knox, see note30, P58.  
註四五·William Cooley & Paul Lohnes, *Evaluational Research in Education*, Irvington Publishers, Inc., New York, 1976, P349.  
註四六·Ibid.  
註四七·同註廿八。

## 參考書目

- 1 Borich, C. D. *The Appraisal of Teaching Concepts and Process*, Addison-Wesley Publishing Co., 1977.
- 2 Cooley, W. & P. Lohnes, eds. *Evaluational Research in Education*, Irvington, N. Y. 1976.
- 3 Postman, N., *Teaching as Conserving Activity*, Delacorte Press, N. Y. 1979.
- 4 Provus, M., *Discrepancy Evaluation for Educational Program Improvement and Assessment*, Berkeley, McCutchan, 1974.
- 5 Worthen, B. & James Sanders, eds, *Educational Evaluation: Theory and Practice*, Charles Jones Publishing Co., Ohio, 1973.
- 6 李聰明, 教育評價的理論與方法, 台北幼獅, 民國六十一年。
- 7 陳瑞龍, 教育評價法, 台南, 民國五十四年。
- 8 昨日今日與明天的教育, 開明書局, 民國六十四年。
- 9 呂俊甫等著, 教育心理學, 台北市大中國圖書公司, 民國六十一年
- 10 教育文粹, 八期, 六十八年六月。
- 11 今日教育, 卅七期, 六十九年一月。
- 12 教與學, 六十六年三、四月號。