

第二章 文獻探討

本章針對教育評鑑之相關理論與比較教育研究方法，分析與統整國內外之相關文獻，據以作為本研究之理論基礎。本章分為三節，第一節為教育評鑑內涵，第二節為評鑑模式之發展，第三節為貝瑞岱比較教育法之內涵。

第一節 教育評鑑

壹、教育評鑑意義與標準

一、教育評鑑意義

對於教育評鑑較為明確周延的定義，國內外學者有不同的解釋：Stufflebeam(2003)定義教育評鑑為：對於教育現象的選擇、描述以及判斷主張與價值的過程。陳玉琨(2004)提出教育評鑑是對教育活動滿足社會與個體需要的程度做出判斷的活動，是對教育活動現實的(已經取得的)或潛在的(還未取得，但有可能取得的)價值做出判斷，以期達到教育價值增值的過程。蘇錦麗(2004)亦認為教育評鑑係有系統的採用各種有效的方法，蒐集質與量的資料，對照評鑑準則(指標或標準)，以判斷任一教育對象之價值或優缺點的過程，並將結果作為決策之參考。吳清山、王湘粟(2004)歸納教育評鑑的定義為，對於教育現象或活動，就其人員、方案、運作結果，對照評鑑標準，透過質與量的方法，採系統而客觀地蒐集、分析資訊，進行價值判斷，以了解教育成效、改進教育缺失和達成教育目標的歷程。教育評鑑之主要精神有「自願性」、「非官方」、「同僚」、「自我研究」、「目標」與「標準」等六個概念(賴士葆，1994)。蘇錦麗(1995)認為藉由評鑑團體的規劃，透過受評學校內部人員自我評鑑與外來專家的專業判斷，評定該校符合評鑑標準的程度，進而提出其優缺點及應改進之建議，以作為提高行政效率之依據，經由此一評鑑的過程，可以達成提昇教育品質的目的。謝

文全(2004)指出：教育評鑑包括三項要點，一、評鑑是對事務加以審慎的評析；二、評鑑是量度得失及其原因；三、評鑑旨在決定如何改進或重新計畫。綜合各學者之意見，教育評鑑是為一種校內自我管理的過程，透過有系統的蒐集教育活動與現象，以自我改進為本質，進行分析資訊、進行價值評判，以期透過不斷改進，進而能實現教育目標。

二、評鑑標準(standards)

「標準」用來確認評鑑者判斷類別效標優劣的品質。標準可能是統計的結果、百分比、平均數、或是合理的判斷(Hamm, 1988)。曾淑惠(2004)認為評鑑標準是一套由專業人員普遍同意的原則或指引，或是為評核某一特定受評主體是否達成特定目的，而由參與人員相互承認的一套衡量準則。Harvey(2004)也提出「標準」是被用來衡量達成了什麼，通常是指結果的標準。Noddings(1997)指出，標準在教育上應用的意涵有四：1.代表分類與排序；2.該達成的目標；3.各種專業程度的敘述；4.品質控制的準則。

Hamalainen(2003)認為「標準」是迎合機構或是方案的需求或條件的層次，可以被品質保證或認證機構所認證或檢定的。標準也可描述方案的成果，例如被期望的能力、知識、技能或態度；評鑑標準的實施還可以當成學校機構考量課程設計時的資訊參考(information base)。標準可被視為是評鑑過程的要素，因為標準提供了方向、深入的瞭解、以及學校機構決策的實體依據(tangible base) (Tanke, 1992)。Woodhouse(2003)認為運用評判標準(內部或外部)來測量高等教育機構品質和標準，標準可以提供任一個國家或是國際的一機構或更多機構間清楚的比較。

三、評鑑標竿化(benchmarking)

牛津大辭典將「標竿」一詞定義為：標竿是一各固定點(fixed point)或參照點(point of reference)。標竿一詞，到七〇年代應用在企業上時，概念已超越「標示參照點」而已，而擴大到企業進行比較時之測量過程，是企業對某些功能或活動的一些真實的、可信的，且可計數的重要測量 (王保進，

2004)。也就是說標竿或活動之持續測量，可以進一步擴大到一個組織和其他同質機構之比較過程、實務、及表現的一種方法，因此稱為「標竿化」可能更為適切(王保進、黃耀農，2004)。

從組織的觀點來看，標竿化在提升組織的績效，就組織內部的個人而，則需透過標竿化的實踐，提升組織成員的表現績效。換言之，當組織設定標竿後，即嘗試外界新的作法及創意，此時對組織內的成員必定產生衝擊，然而組織績效的有效提升有賴於成員們的改造，因此組織成員需將標竿轉化為標竿化，標竿化的步驟有八：1.決定標竿化的議題；2.組成標竿化團隊；3.檢討現行作業流程；4.選定標竿化夥伴；5.蒐集資料；6.分析目前績效與期望績效間的差距；7.實際採取變革行動；8.評估績效並進行回饋。標竿化的過程之首要步驟，在決定所要進行標竿化議題，然後在遴選於此議題上有卓越表現之機構為「標竿化夥伴」，由於標竿化夥伴通常就是該議題上的最佳實務(王保進、王俊凱，2006)。隨著標竿化在大學品質管理的運用，「標竿」的觀念也逐漸進入品質評鑑機制中，包括美國國家品質獎、澳洲大學標竿化實務、英國學門評鑑評估等，都已經建立了品質的標竿(王保進，2006)。

貳、教育評鑑的類型

教育評鑑類型眾多，不同的分類角度，同一種評鑑可能被歸類到不同的類型，而同一種評鑑也可能同時具備多種不同的類型特色，下列依照評鑑目的、評鑑方法、評鑑對象、評鑑內涵真實性等幾項類型來說明。

一、依評鑑方法分類

Kells(1988)歸納評鑑方法為「自我評鑑」與「外部評鑑」兩種。不論是由學校自己規劃、執行，旨在改善學校自身缺失，以提升學校品質；或由學校以外的團體來執行評鑑，以達到品質保證目標，皆會同時運用「自我評鑑」與「外部評鑑」兩種方式來執行評鑑工作。

(一) 自我評鑑

自我評鑑的過程是為後續評鑑提供充分的資訊過程，是受評單位自我診斷的過程，也是受評單位自我修改的過程(陳玉琨，2004)。Kells(1988)指出，自我評鑑是由機構內全體人員在環境中所進行的評鑑過程，其動機是內發的，性質是自主的、自願的，目的是為促進機構進一步發展與改進，其過程應該以有用的系統觀點來進行，主要內容為二，一為目標達成程度，以供外部評鑑的依據，是外部觸動的自我評鑑(self-evaluation)，另一為系統功能及問題的解決，強調由學校所有的成員參與，在校內高層的領導下，透過內發的動力機制進行評鑑，以達到改善的目的，此種自我評鑑稱為「self-study」。蘇錦麗(1995)認為所謂的自我評鑑係由高等教育本身或指定團體所規劃和執行的評鑑活動，自我評鑑主要基於「自我管制」(self-regulation)的觀念；換言之，各高等教育機構應該自行負責管制，達成內部共識，才能確保品質。

(二) 外部評鑑

外部評鑑的目的著重在品質的保證與績效責任的達成，外部評鑑除了可對內部自我評鑑的效度加以驗證外，也可以讓內部成員感受到來自外部同儕的專業壓力而做必要的調整(Kells, 1988)。外部評鑑有助於提高評鑑結論的客觀性與權威性，專家組成的成員來自社會各界，因而，由他們開展評鑑工作，有助於提高評鑑的客觀性，使評鑑結論易於為大家所接受(陳玉琨，2004)。

綜合上述，當一個機構尚未建立自我管制機制之前，外在的管制是需要的，惟高等教育機構應盡量減少外在的控制力量，以發展學術系統自我評鑑的能力，以達到實施自我管制之目的。唯將外部評鑑與內部評鑑兩者相結合，才能成為一完整的品質系統(蘇錦麗，1997)。

二、以評鑑的對象分類

依照不同的評鑑對象來分，可以分為訓練方案評鑑、學校評鑑、教師評鑑或是評估一般教育方案等。例如學校評鑑時，常見的評鑑模式有學校效能模式、學校認可模式、學校發展性評鑑模式、學校整體評鑑模式等。

參、品質保證系統

近年來高等教育教育評鑑的品質保證系統較常使用的方法有「認可制」(Accreditation)、「全面品質管理」(total quality education)以及「績效責任制」(accountability)，這些制度各有不同的發展理念與限制 Bogue (1998)，分述如下：

一、認可制(Accreditation)

探討各國高等教育品質保證制度，論及美國相關制度時，通常是以「認可」制為探討範圍(楊瑩，2005)。認可評鑑是一種同意機構或是課程符合或是超過教育品質標準的一種狀態，它是一種自願的程序以確保課程品質以及協助課程的改善(Williams, 2003)。認可具備了幾項特徵：1.認可是自願性的民間活動，因此對大學並不具有約束力，也無法控制其行為；2.認可是高等教育機構自我管制的重要工具；3.認可最重要的功能是在判斷高等教育機構之品質；4.認可的本質是一項評鑑的歷程，而高等教育機構的自我評鑑則是此歷程的核心工作；5.認可提供外部諮詢，使高等教育得以擬定本身的研究與計畫；認可分為兩種類型，一為機構認可，另一為專業的學門認可，認可透過受評機構或方案(program)定期的自評(self-study)、接著由外部同儕團體進行訪視(peer review)，以確認受評機構或方案與外部評鑑標準(standards)的符合程度(陳漢強，1997)。

Bogue (1998)認為認可仍有其限制，認可被認為只建立在少數幾項不被大眾認知的標準及過程當中、無法預防問題、面臨不同領域各學門評鑑單位利益整合的挑戰，故提出認可制度未來可能的發展：一為各區域性的

機構評鑑可能轉化為全國性的評鑑；而「認可」評鑑單位可能將依不同型態的大學機構，如：研究型大學、綜合型大學、技術型或社區學院等，進行不同的認可型態；二為「認可」可能會發展成定期的機構外部審議(audit)型態，由外部審查者(external evaluators)定期檢視機構表現與外部標準的符合程度，從而結合績效責任的觀念。

二、全面品質管理 (Total Quality Management ,TQM)

全面品質管理員自於產業界追求產品的品質運動，全面品質管理由美國品管專家戴明博士 (Edwards Deming) 所提出，在看到全面品質管理的優異成效後，日本與美國的教育體系，亦開始引進全面品質管理(莊淇銘、張家宜、張麗卿，2000)。全面品質管理由事先預防，系統、動態、前瞻四大理念為核心，在實際作為上遵循以客維尊、全員參與、品質承諾、持續改進、事實管理的原則，做到持續提升品質並進而創新品牌的目標。然而，對於一些學者而言，要把製造業品管的概念當成是大學品質管理及提昇是有爭議的，有些學者認為，全面品質管理只適合於增進行政方面的品質改進(Bogue, 1998)

三、績效責任制

1980 年代間，英、美等國發現長年的投資並未有效提高學童的學習成就，開始注意教育投資與成效之關係，績效責任便成為教育改的重要課題，以學校為中心的教育改革提升學校績效的重要途徑，增加學校經營與教學權力，責任便相對加重，績效責任概念也就加諸學校之上 (吳清山、林天佑，2001)。教育績效責任就其廣義來看，乃是教育機構及其相關人員和學生自己，負起本身教育和學習成敗的責任；就狹義而言，乃是教育機構及其相關人員負起學生學習成敗的責任(吳清山、黃美芳、徐緯平，2002)。這種評鑑方式的優點在於所公佈的表現指標報告書能讓大學向大眾

展示其績效責任的成果，建立長期活動及成就的趨勢，以及其達成教育目標的進展。有些學者表示，指標建立的目的是在於較容易反映出學校表現的資料內容，而不是在設立找尋證據的規準。因此，無論是大學或政府皆需要大學在活動與成就方面的表現資料，以便讓決策者能瞭解其運作狀況，並建立其品質改進的基礎，達成其設定的目標。然而，如何降低政府的干預與權力的控制、以及如何建立出客觀完善又能適用於不同機構型態的表現指標，則常常是這種績效責任受爭議的地方(Bogue, 1998)。

第二節 評鑑模式之發展

本節包括評鑑模式之意涵，評鑑模式之發展、常見之八種評鑑模式、CIPP 模式，以及澳洲 AUQA 所使用之 ADRI 程序。

壹、評鑑模式之意涵

在英文「牛津字典」中對「模式」的定義為：模式是專業人員概念化

並描述實施過程的總結(summary)、縮影(epitome)或是摘要(abstract)。吳明清(1996)指出，在學術上，模式是指「以抽象的綱要形式來描述、解釋或規範複雜的現象或活動」。蘇錦麗(1995)認為：評鑑模式係指該評鑑資料蒐集的方法、策略與架構為整個評鑑的核心，能具體表現整個評鑑的本質與特色。曾淑惠(2004)表示，評鑑模式是評鑑理論與實務下的產物，它能以流程圖、架構圖、概念圖或是數學模型的形式存在，也能以一套用來描述、執行、報告評鑑的示範性或描述性指引陳述，而評鑑模式需具有可重複驗證的特性，並能有助於評鑑的利害關係人對評鑑工作主要概念與結構的共同認知。

貳、評鑑模式之發展

Stufflebeam 及 Madaus(2000)將評鑑劃分成六個時期(蘇錦麗，2005；吳清山、王湘粟，2004)：

一、改革時期(age of reform, 1980-1900)

以 1972 年為方案評鑑改革時期之開端，為教育評鑑的萌芽時期，因該年為 William Parish 首度以標準化的數據記載測驗的結果。19 世紀是工業革命的世紀，伴隨而來的是經濟和科技的變革，當社會結構改變時，主要的社會改革因此發生教育評鑑為社會改革的一環。在此時期，英國有兩個重要發展值得一提：第一，一些協會或社團致力於社會調查，而且發布若干關於社會問題的報告，並引起社會的討論；第二，為了回應這些私人的報告，政府也建立管理這些方案的機制。

在美國，第一次正式評鑑學校的表現則發生 1845 年的波士頓。這個事件在評鑑歷史上的重要性，乃引其開啟了一項悠久傳統，即以學生的測驗分數作為評鑑學校及教學方案效能的主要資料來源。1887 年至 1898 年間，Joseph Rice 實施了第一次正式的教學方案評鑑，他在數個學區進行學童拼音教學練習的比較研究，利用測驗成績作為測量的標準，並發現每週

花二百分中再拼音練習的學習和只花不到十分鐘的學區，再學習收穫上並無顯著差異，Rice 的研究結果導致教育人員重新檢視並修正拼音教學的方法。該時期還有其他值得一提的是，認可製的基礎或評鑑的專業判斷方法可以追溯到 1800 年代晚期，從那時期開始，認可運動獲得極大的發展，並成為評鑑教育機構是當性的主要工具之一，具有強大的效力及公信力。

二、效率與測驗時期(age of efficiency and testing, 1900-1930)

在二十世紀初期，泰勒發展性的工作開啟了科學管理運動，這是早期人員評鑑的一種方式。在此時期，一些大學系統採用不同的準則(例如：經費、學生輟學率及升學率等)來調查學校和教師的效率，效率為此時追求的目標，隨著測驗理論的建立與開展，以測驗成就成為達成此一目標的唯一手段(蕭霖，2004)。

三、泰勒時期(Tylerian age, 1930-1945)

泰勒在教育領域，特別是在教育評鑑及測驗方面具有重大的影響力，因此被譽為評鑑之父。泰勒創造教育評鑑一辭，將評鑑定義為評估重要目標達成的程度，並將此目標是為教學方案的一部分。泰勒視評鑑概念化為設定的成果與實際成果的比較，強調評鑑學生和方案成效時，應著重在明確目標的評量上，評鑑的重點在學習結果，而非組織與教學的投入量，所以避免了專業判斷的主觀性。

在 1940 年代中期之前，泰勒藉由他的研究和著作，為他在教育上，特別是在社諺語評鑑領域上產生去大的影響力，也連帶的影響了未來的二十五年。

四、純真時期(age of innocence, 1946-1957)

此期學校數量與規模大幅擴張，反倒使教育評鑑的技術停滯不前

(蕭霖，2004)。這時期社會與教育的普遍現象也反映在教育評鑑上。評鑑技術在此時期大幅發展，但是此時期的評鑑工作沒有社會目的，並沒有朝向受益人需求而努力，以及嚴格的檢視社會對這些需求的回應。

五、擴張時期(age of expansion, 1958-1972)

1950年代末與1960年代初之際，為評鑑史上重大的轉變，因評鑑的發展已由一門行業進入專業，並把焦點放在如何符合社會的需求及尋求納稅人稅金的支持。教育評鑑的技術發展迅速，強調教育方案的評鑑與改良是因素評鑑模式的發展是此時期的最大特徵。

六、專業時期(age of professionalization, 1973-1983)

大約從1973年，評鑑的發展開始具體化且呈現專業的取向。在此之前，評鑑人員面臨身分認同之危機，不確定自己應該具備何種特殊的資格。為了打破困進，教育學者致力朝向專業化發展，並在1970年代開始發行了許多相關期刊。在此時期，評鑑者逐漸了解評鑑的技術必須達到效果，在這之前評鑑技術被視為是嚴謹研究的旁之末節，也了解評鑑必須達到客戶的資訊需求、提出中心價值觀的議題、處理情境的事實面、符合廉潔的要求，及滿足誠實的需求。雖然適當的研究方法持續地在發展、方法論學者間有更多的溝通和了解，以及更多新技術的出現，然而在大多數的情境下，評鑑實務並未有多大的改變。

七、擴張與整合時期(the age of expansion and integration, 1983-2001年)

隨著經濟的成長，評鑑領域不斷擴充並更加整合。在評鑑整合方面，最初由社會科學評鑑者組成的評鑑研究學社(ERS)，與主要是由教育評鑑組成的評鑑網合併成立一全新的美國評鑑學會(American Evaluation Association, AEA)，將各學科領域的思想與方法加以整合。在1995年，AEA將大會的重點放在評鑑的國際合作上，邀請了世界各國的評鑑者與會，同

時有超過 20 個國家的評鑑學會成立，各國評鑑期刊也在各國紛紛發行。

參、常用之八種評鑑模式

一、準評鑑(Quasi-Evaluation)

準評鑑過程近似一個真實評鑑的過程，對照主要的科學活動，被認為是一種知識的產出。評鑑以是否及時(on time)、切合目標(on target)、以及合於經費預算(on budget)三個面向的資料結構來判斷，學者稱這種評鑑監督(monitor)性質大於評鑑性質(Sriven, 2003)。準評鑑會有一特定需要評鑑的問題，這種評鑑可能因所定的資訊範圍太窄，無法作為判斷一個方案是否有價值的基礎，雖然在方法上適當，但卻不一定能提出契合問題所需的評鑑資訊 (蕭霖，2004)。

二、目標成就評鑑 (Goal-Achievement Evaluation)

泰勒被認為是這類型評鑑的研究先驅(曾淑惠，2004)。其主要的功用是在確定教育方案達成教育目標的程度。Scriven (2003)認為目標成就評鑑以同樣的程序來看事情的過程，推論(generalizing)人們看事物的角度及當事人的觀點，僅是一種「事實陳述」的研究；劉春榮(2004)目標成就的評鑑者須花費相當多時間在幫助受評單位釐清目標和活動之間的聯繫，運用明確的目的和不標決定成功的標準，評鑑者評估目標被達成和目標被實現的程度。

泰勒式評鑑的第一步驟是清楚界定方案或活動的目標，其提供對教育方案清楚的陳述，描繪出評量目標達成的過程，評鑑人員為方案的參與者設計一個工具，以便評量目標的達成與否。

目標成就特別適用於評估緊密聚焦的計畫，即具有明確的、可支持目標的方案。此種研究能藉由評估預期受益人的需求、發現副作用以及學習過程和結果，以增強其適用性。

此模式之優點是：1.簡單、易懂、容易遵從與執行；2.刺激了測量程序與工具等技術發展；3.清楚的反應出方案的成果以供判斷方案的基礎，可以幫助評鑑人員及其他人看到判斷方案的價值基準(Worthen, Sanders, & Fitzpatrick, 1997，引自於曾淑惠，2004)。缺點是：1.此種模式不足以判斷目標的優缺點與價值；2.當評鑑方案未達成目標時無法知道哪些地方出錯；3.不能證明沒有價值的目標；4.對改進方案或企業體的助益較少(Stufflebeam, 2000)。

三、成果為主評鑑(Outcome-Based Evaluation)

強調以成果為主的評鑑，以作為一個評鑑決策對於支持概念的回應，是屬於「表現管理」的主張。這種評鑑倡導「績效責任」更勝於「改進決策」，輕視評鑑參與過程的考量，以及不同面向上綜合的步驟。此評鑑之優點在於其簡潔，不執行價值批判，是一種管理導向的方法，其有兩大弱點，第一是因為大部分的指標雖容易操作，但是其正確性卻令人質疑，第二是這些指標的正確性容易因環境背景的改變而不適用(Sriven, 2003)。

四、消費者導向評鑑(Consumer-Oriented Evaluation)

此為一種總結性的評鑑，強調消費服務的重要性，消費者一般對於方案或是產品設計的目標不感興趣，而只對方案的改進有興趣。消費者只對其是否符合需求有興趣（與管理者對於成功的定義不同），他們只對方案的所有效益有興趣(Sriven, 2003)。在消費者導向研究中，主張評鑑應從消費者的角度檢視各項教育產品的價值(曾淑惠，2004)。評鑑者被視為開明的/有知識的代理消費者，評鑑者應該要協助消費者(例如：學生、家長、教師及納稅人)確認和評估相似的方案、服務及產出的優點及價值。劉春榮(2004)認為評鑑者蒐集觀察與實驗證據來了解是否需要增加或減

少資訊的使用；結果顯示：直接與關鍵決策者的接觸、及時的資訊、對組織的脈絡具敏感性、以及個人的參與似乎扮演一個關鍵的角色；消費者導向的評鑑者意識到很多因素可能影響到評鑑效用，包括人為的因素：風格、一致性、敏感性。以及脈絡因素：原本存在的條件、組織特性、社區的影響；但最重要的因素可能是評鑑中的潛在使用者。

消費者導向評鑑主要的優點是，它是有力的、獨立的評估，意圖保護消費者不致遭受到劣等的方案、服務及產品，並引導他們支持和採用最佳及最有效滿足其需求的方案。典型代表為 Sriven 所發展的產品檢核表(潘慧玲，2002)。

五、形成性的評鑑(Formative-Only Model)

Cronbach 認為真實世界並沒有總結性評鑑，評鑑所需的是形成性評鑑(Cronbach, 1963, 1982,引自 Scriven, 2003)。Cronbach 認為評鑑應該要審慎計畫、有彈性；並且是要有所獲益的評鑑，並強調要對評鑑活動本身進行反省(Stufflebeam & Shinkfield ,引自黃光雄，2005)。在許多的評鑑的過程中，Cronbach 的論述是真實的，但並不包含諸如消費者報導、陪審團審判及訴求的結果、以及許多方案受到經費凍結的期末報告的這一類評鑑模式(Sriven, 2003)。

六、參與式或角色混合取向(Participatory or Role-Mixing Approaches)

參與式評鑑（將受評者在評鑑過程中輸入）近來廣被使用、增能評鑑(empower)（受評者是共同作者，而評鑑者為教練的角色）或是合作式(collaborative)評鑑。其源自於質性研究的傳統，強調以整體方式了解評鑑，偏好自然探究的方法，並側重參與者的聲音(潘慧玲，2002)。參與式評鑑可以改進消費者樣本的代表性、倫理上的委託、避免事實上的誤差以及其他評鑑品質的問題。類似消費者導向的評鑑(Sriven, 2003)。

七、理論導向評鑑(Theory-Driven Evaluation)

理論導向的評鑑中主張的是，要依據方案理論去定義方案的中心問題，選擇研究方法，並據以蒐集及分析資料(曾淑惠，2004)。主要為解釋一個企業評鑑核心功能成功與否的過程，此為評鑑核心的概念。黑箱作業的評鑑，讓人不知評鑑的成份是什麼，只讓人知道每一個功能。「解釋」對於這類型的評鑑具有加分作用，可促進評鑑改進，但要去發現這些解釋並非是評鑑本身必須要做的(Sriven, 2003)。在一九八〇年代及一九九〇初期就不斷有人將理論導向的評鑑，以邏輯模式(logic model)的形式運用在實務工作中(曾淑惠，2004)。

八、權力模式(Power Model)

權力模式是晚近發展的一種評鑑模式，係受後現代主義影響，批評評鑑者運用社會權力，極大化社會價值所帶來的衝擊。Ernest House 首創此種途徑，此模式採取一種蒐集、處理與分析利害關係人的觀點，增加利害關係人的代表性，以系統化、無誤差、使得每個觀點都成為評鑑結論的一部分，又稱之為審議式的民主評鑑。模式的前導概念包含了三個主要層面：「民主參與」、「進行對話」，以檢視及證實利害關係人的投入，以及「進行審議」，以達成關於方案優點和價值的可辨明的評估，這三個層面在完善的評鑑中被視為必要的(Stufflebeam，引自蘇錦麗 2005)。評鑑者透過與利害關係人的對話和討論，以決定欲陳述的評鑑問題。在此模式中，評鑑者仍處理評鑑的研究並對結果負責，但是亦將利害關係人的觀點、價值與興趣納為研究一個必要的部分。這種評鑑期望集合利害關係人的觀點，以提升結果的效度及評鑑的適切性。在研究方法上仍舊採用傳統的資料蒐集及分析技術，並側重評鑑者的專業，但擴展的評鑑的方法，採取一個額外的過程來蒐集與處理利害關係人的觀點。這些過程

類似焦點團體法，或是一個將利害關係人直接涉入評鑑結論的嶄新作法 (Sriven, 2003)。

九、CIPP 模式

近代的評鑑模式趨向於跨模式的方式，CIPP 在教育評鑑領域中應用的既多且廣。該模式由 Stufflebeam 提出，自提出後一直不斷地修正與補充。CIPP 模式具有廣泛的適應性與適用性，至目前為止，以應用在不同專業領域的教材、人員、學生，以及方案或計畫的評鑑中，其可以用個別或整體的方式，定義多種或是階層式的計畫元素，但是每一個層級所需要的資訊通常會不一樣，因此評鑑人員必須小心地助要道層次的問題，蒐集資訊並聚合成評鑑的發現，符合不同觀眾對資訊的需求(曾淑惠，2004)。

從 CIPP 的英文字母縮寫來看，本模式的核心概念為背景(Context)、輸入(Input)、過程(Process)、結果(Product)評鑑(Stufflebeam, 2003)。

(一)「背景評鑑」(context evaluation)

適用在一個已經定義的環境中，評估需求、問題以及機會，主要目標是：1.描述預期服務者的背景；2.確認預期受益者並評估其需求；3.確認符合需求的問題或障礙；4.確認可用來強調目標需求區域性有利條件和經費的機會；5.評估方案、教學或其他服務等目標之明確度和適切度。

背景評鑑採用的方法及其範圍，包括蒐集母群體成員及週遭環境有關的多樣化資訊，並管理多種的分析型態。評鑑人員需要建構一種調查工具，以調查經過定義的假設，也要檢驗現有的紀錄，以界定母群體的背景資訊。背景評鑑之後，受評單位要持續地運用上述的方法蒐集、組織及報告背景評鑑的資料，因為需求、問題、資產與機會是會改變的。

背景評鑑有許多建設性的用途，其提供學區官員與社會大眾的溝通，獲得對學區的優點、缺點、需要、資產、與優先問題的共有認知與理解，受評單位可以以此提出符應緊急需求的計畫，說服基金的支援單

位，可以說服該選區的選民相信應該投票通過一個增稅的議題，也可以運用評鑑的發現，評估運作計畫的相關性，檢驗計畫的處理如何影響計畫的環境等(蘇錦麗，2005)。

(二)「輸入評鑑」(input evaluation)

輸入評鑑旨在藉由尋求並檢視切合需要的相關途徑，用來確認方案的可利用資源，例如評估競爭策略、工作計畫、實施方案所選擇方式之預算，幫助使用者設計改進事項、發展有利的經費申請計畫書、詳細的行動計畫、紀錄可行的變通計畫、紀錄一方法在其他方案上選擇基礎。

輸入評鑑應該先評估相關的途徑，為受評單位尋找政策的阻礙、財務或法令的限制，以及潛在的資源。輸入評鑑其中一個功能乃協助受評單位避免可能失敗或浪費資源的規劃方案(曾淑惠，2004)。

(三)「過程評鑑」(process)

過程評鑑的本質乃針對計畫執行、過程的文件，以及包括改變計畫某些程序的重大失誤或不良操作等部份，進行持續性檢核。其目標為：
1.提供從業人員實施計畫活動的達成程度；2.幫助從業人員界定實施上的問題；3.周期性地評估參與者接受並實現其角色的程度；4.報告觀察者與參與者如何判斷方案或計畫執行歷程的品質。

過程評鑑的關鍵是評鑑人員，在監督工作計畫的執行時，要檢視相關的策略、工作計畫，及任何之前的背景評鑑，以定義他們必須監督的計畫活動。評鑑人員必須進行觀察活動、提供回饋，並與參與者進行互動，也要發展並執行一套資料蒐集活動的時程表；成果在使用方面，方案或計畫執行人可以用來指引活動、校正計畫、維護績效的紀錄；管理人員可以運用例行安排的歷程評鑑成果，使從業人員隨時準備行動，並切實地負起自己的責任(曾淑惠，2004)。

(四)「結果評鑑」(product evaluation)

發生在方案完成階段，幫助評鑑使用者維持對於迎合學生或其他受

益人需求的關注，目的用來提供方案確認預期與非預期的附加效益、以及是否應該繼續或是需要修改的決策。在成果評鑑中，可以用多種方法運用，結合多種技術，以便廣泛地蒐集成果，交叉檢證多樣的表現 (Stufflebeam, 2003)。

CIPP 模式強調評鑑最重要的目的不在於證明(prove)，而是在於改善(improve)。CIPP 模式將評鑑視為一種改進與專業責任的必要附贈品，這個模式係設計來協助評鑑者及其閱讀評鑑報告者，能夠設計、執行並運用一個完善的評鑑，俾利對其服務與產出，進行持續不斷地改進 (Stufflebeam, 2003)。

十、澳洲之 ADRI 程序

AUQA 發展澳洲高等教育機構審議過程，特別重視內部自我評鑑，必須要相當穩固，並將焦點放在機構實務的檢驗和過程，與機構所制定之願景目標相互校正；計畫與目標該如何轉換成實際的運作；表現指標該如何定義與評鑑；如何透過結構化的檢驗過程確認改進方案。重點目標就在於機構的過程要如何有效的達成所制定的目標(Woodhouse, 2001；引自 Jantti, 2002)；而這些重點目標與澳洲企業卓越架構(Australian Business Excellence Framework, ABEF)的意涵相近，因此將 AUQA 之審議與 ABEF 連結 (Jantti, 2002)。澳洲品質委員會(Australian Quality Council, AQC)於 1980 年代創立，當時的澳洲聯邦政府正是發展與推廣品質管理實務與方針的尖峰時期，其 1987 年發展的 ABEF 提供了結構化的且整合的管理系統，也提供了詳細的持續提升和卓越表現的組織系統、方法、特色的說明(Vogel, 2002)。

ABEF 包含了七個範疇，分別為領導與改革(Leadership and Innovation)、顧客與市場焦點(Customer and Market Focus)、策略與計劃過程(Strategic and Planning Process)、人(People)、過程、產品與服務(Processes,

Products and Services)，以及企業結果(Business Results)這七個範疇組成了一個特殊的結構或是背景，可供組織內的檢驗、問題和分析他們的領導與管理系統。七個範疇向下皆有數個次項目，總共有 22 個次項目，22 個次項目分別為：1.策略方向；2. 組織文化；3. 組織領導；4. 環境與社區參與；5. 了解企業環境；6.計劃過程；7.資源的發展與運用；8. 資訊和數據的蒐集與詮釋；9. 整合決策所需之知識；10. 知識的整理與創造；11.保證；12. 健康、安全和福利；13.市場與顧客的知識；14.顧客關係之管理；15.客戶意識的價值；16. 改革過程；17.管理與改進的過程；18. 產品和服務的品質；19. 成功的指標；20. 指標的穩定性；21. 供應者與夥伴的過程；22.效果與發展。而透過這個 ABEF 這個架構所使用的評鑑方法為四個面向：方法(Approach)、部署執行(Deployment)、結果(Results)、改進(Improvement) (Vogel, 2002)。

(一) 方法(Approach)

此面向指的是目標是什麼，支撐組織方向的信念、計畫和主義。方法包含當計劃未來發展的描述目標、策略和表現指標。

(二) 部署執行(Deployment)

此面向指導是策略、架構和過程之定義，並將方法的概念轉為實務和各部份運作的整合。

(三) 結果(Results)

重點放在如何監測、測量方法與部署執行的結果，以及結果如何傳播。

(四) 改進(Improvement)

重點是檢驗方法與部署執行的達到目標之效果，以及如何檢驗過程提出改進方案。

ADRI 模式提供了系統化的方法來製作一個優良實務機構的文件和評鑑關鍵性的要素。模式反映不僅是機構現在在做什麼，也反映了未來

該如何改進與改革。其是一個讓每個組織成員能夠了解其機構、主義、價值、方針和目標。ADRI 是一個改變管理和品質保證的工具，對於計劃過程有著良好的效果(McGregor, 2003)。

第三節 貝瑞岱比較教育內涵

一般將比較教育定義為「比較各國的教育制度，以達成三個主要目的：促成國際了解、輔弼教育改善或改革、解釋國家變項」(洪雯柔，2000)。比較教育的研究對象包含教育的全部領域，是指進行各個國家或地區中的教育現象與教育問題的比較研究，以發現其相似點和差異點，藉著知識探討的歷程，以加深理解他國及我國的教育與文化(楊思偉、沈姍姍，2000)。

壹、比較教育之發展

比較教育發展脈絡分為四個時期，分別為萌芽期、借用時期、因素分析時期、社會科學方法時期。

一、萌芽期：

對外國教育的研究可追溯至自有旅行者開始前往外國遊訪、民族或國家間的貿易往來、宗教宣傳與戰爭等國際互動。直至 18 世紀，開始有「比較」研究的出現，一種是今昔教育比較，另一種是對各國教育制度的描述分析，企圖做比較(洪雯柔，2000)。

二、借用時期：

本期主要代表人物共有六位，最主要的為朱利安(Julien)，他強調兩種功能，一為教育事實之比較分析，一為教育法則及各項原則的奠定；法國公共教育部長庫辛(Victor Cousin)、美國公共教育之父—曼恩(Horace Mann)、美國的巴納德(Henry Barnard)，以及英國的阿諾德(Matthew

Arnold)，他們皆以外國教育制度的考查制度作為本國教育制度改進的參考。還有沙德勒(Sadler)認為校外事情重於校內的事，並且制約說明校內事物。比較教育部應只借外國教育制度為唯一目的，應以外國先例，觀察研究為基礎，尋求能預測一國教育制度改革結果之理論為首要目標(廖俊傑，2001)。

三、因素分析時期：

本時期的代表人物提倡教育現象的歷史與文化脈絡的動態分析(洪雯柔，2000)。康德爾(Kandel)重視歷史因素與教育傳統、漢斯(Hans)抱持因素決定論，馬霖遜(Verson Mallinson)結合康德爾與漢斯的思想，研究重點為民族性格之重要性(廖俊傑，2001；洪雯柔，2000)。

一、社會科學方法時期：

貝瑞岱即屬於此階段的學者，貝瑞岱處於比較教育研究中意識形態與方法論上新舊交替的時代，貝瑞岱提出四階段的研究模式；其他學者霍姆斯(Brian Holmes)提出批判的二元論以及問題中心法；金恩(Edmund King)提倡觀念架構；諾亞(Noah)及艾克斯坦(Ecksteim)則提出假設驗證法，主張比較教育科學方法論(廖俊傑，2001)。

貳、貝瑞岱比較教育之內涵

貝瑞岱主張，比較教育的對象是學校，認為學校與社會關係的研討是比較教育的重要責任，若無相關的論述，比較教育便失去其預言性的價值，也將因此喪失其目標與基礎。就其觀點而言，比較教育乃是以國家(文化區域)為研究變項進行外國教育制度的分析調查。比較教育的目的最終的目標則在「法則」或「類型」的建立：1.跨國收錄教育方法，自各國教育制度的異同中求意義；2.比較教育有其完足的知識性意義，亦即致力於滿足人們對知識的追求；3.比較教育亦有對實際教育施行與社會科學的貢獻(洪雯柔，2000)。

基本上，貝瑞岱比較教育方法建構可以分為兩個方向分析，一為模式的建構，分為「區域研究」及「比較研究」兩種類型，二是研究步驟的劃分：貝瑞岱將以往由學者所提出的「描述」、「解釋」、「並列」，依照邏輯一貫性與實際研究進行的步驟]，將之融入區域研究與比較研究兩種模式中。區域研究分為兩步驟，描述階段與解釋階段，描述階段是進行教育資料的搜集，而解釋階段是運用其他社會科學的方法對正式教育的資料進行分析；比較研究乃是同時對多個國家或地區的教育現象進行研究，其步驟分為並列階段與比較階段，並列階段式將不同的國家進行初步的對比，據以進行的規準、比較所欲驗證的假設；而比較階段則是對教育所做的跨界同時分析。以下分別將四階段做探究：

一、 描述(description)

描述階段乃是研究一般目的對教育資料的蒐集與編目。蒐集資料之前，則須先進行分類的工作，列出所需搜尋資料的面向，以避免在蒐集過程中有所遺漏，也利於資料的系統收錄與分析(洪雯柔，2000)。資料來源分為第一手資料與第二手資料與輔助資料，第一手資料為政府機關的正式報告書、會議記錄、手冊等。第二手資料為個人或團體的論做、論文、摘要。輔助資料指的是與教育並非直接或顯著相關、卻有某程度相干性的書籍、文章或其他印製刊行的資料。

二、 解釋(interpretation)

詮釋階段，或稱社會分析階段，是根據社會科學的觀點或是人文科學的觀點充分檢討，以詮釋所述教育事實的意義(楊思偉、沈姍姍，2000)。解釋乃是依國別將蒐集而得的學校教育資料就其與社會的相干性。貝瑞岱認為，以「學校」為比較教育的研究核心，人文學科、社會學科則被用以擴展比較教育各類事件的展望。

三、 並列(juxtapotion)

貝瑞岱比較教育研究運用雙欄連貫式列表來並列各國的資料，以尋求

一統的概念與假設(洪雯柔，2000)。並列式針對某一特殊目的所做的各國教育施行狀況細目後，再用以進行比較的定義性陳述來引導，繼此依此主題檢視資料並總結出假設。運用圖表式(又稱垂直式)與文字式(或稱水平式)並列釐清教育以及教育與社會關係之資料，以建構可比較的歸準—比較點，並建立假設。後續的比較階段便致力於以比較點為基準來比較各國的異同，並檢驗假設(沈姍姍，2000)。

四、 比較(comparison)

比較階段是對多個國家同時處理，以證明自「並列」階段的假設。透過排序的歷程更加突顯教育事實，可以分為兩個層級，其一為問題取向，乃是針對某些問題作比較探討，其二為全面取向，主要就兩國或多國的全面相干性作比較，是一種教育全景的呈現；問題取向為全面取向之基礎(沈姍姍，2000)。比較階段所運用的方法可區分為兩種：平衡比較與闡釋比較。平衡比較之過程近似於並列的結構，但其比較的強度較高；平衡比較的核心方法便是系統地進行一次又一次的跨國界研究以彰顯比較因素的特點。闡釋比較著重在對比較要點的陳述，此一方法無法概括出通則，因此不可能推論出法則。當平衡比較無法進行時，闡釋比較能夠對比較料作成較含蓄的分析，也可增進內容分析深度為平衡比較奠下基礎(洪雯柔，2000)。