

教學與評量 --

一種互為鏡像(mirror image)關係

謝祥宏 段曉林

國立彰化師範大學 科學教育研究所

摘要

科學教育的目標除了讓學生獲得知識方面理解外，更重要的是要培養獨立思考與解決問題能力的人。在現今考試領導教學的狀況下，評量的策略、內容及方法應該要多元化才能與教學目標對應。教學與評量，猶如一種互為鏡像的關係；在評量的鏡像中可以反映出教學目標是否達成，在教學實況中則反映出評量的目標；而評量這面鏡子是否可以確實反應教學目標，則涉及評量的有效性問題。因此，本文就評量本質的演變、評量與教學關係及評量策略的探討等議題，探討評量的重要性與影響力。

關鍵詞：教學、評量、公平、科學教育

一、前言

國內長久以來的評量方式，不管是平時或升學考試，大多以概念的理解為主，以紙筆測驗方式進行；而紙筆測驗形式大多又以是非題、選擇題及計算題等方式為主，受限於這些評量形式，無法評量學生解決問題能力及情意方面技能。影響所及，學生的學習逐漸捨去由操作中學習，而改由聽教師講解、習題演練方式去學習，以迎合考試的要求。實際上，科學教育的目標除了讓學生獲得知識方面的理解外，更重要的是要培養一個獨立思考與解決問題能力的人。因此，在現今考試領導教學的狀況下，評量的策略、內容及方法應該要多元化才能與教學目標對應。教學與評量，猶如一種互為鏡像的關係；在評量的鏡像中可以反映出教學目標是否達成，而評量這面鏡子是否可以確實反應教學目標，則涉及評量的有效性問題。因

此，本文擬就評量改變的本質、評量與教學關係及評量策略的探討等議題，探討評量的重要性與影響力(圖 1)。

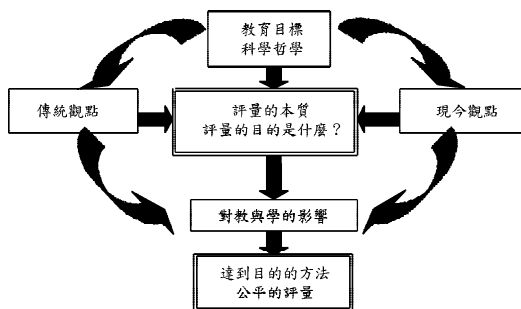


圖 1、評量相關議題。

二、評量本質的演變

根據 Duschl 和 Gitomer(1991)回顧概念改變的模型發展時，指出科學理論的發展是整體性的觀點，亦即一旦科學理論有所改變時，對應的研究方法和研究目的也自然而然地隨著改變。評量的改變，也是同樣的情

形；當評量的理論改變時，它所伴隨而來的目標、方法亦隨著改變。以下從兩方面論述評量改變的本質(圖 1)，一是當代科學哲學的觀點對教學、學習本質的改變，進而影響評量的本質(張文華和郭重吉,1995)。二是基於時代變遷，修改以往不符社會需求的教育目標，而在評量方法、內容、策略也必須改變並與之相配才能達成評量的有效性。有關這兩方面的論述，本文將分述於後：

(一)科學哲學觀點改變

Denzin 和 Lincoln(1994)認為「派典(paradigms)」代表個人對世界觀、世界本質

表 1、傳統派典與建構主義派典比較(Guba & Lincoln ,1989,p.84)

傳統的信念	建構主義信念
本體論(ontology)	本體論(ontology)
存在唯一的實體，且獨立於觀察者。	實體是多元、相對，是社會建構而來
真理被視為與實體同義。	真理被定義為多數人共同建構。
知識論(epistemology)	知識論(epistemology)
二元，客觀的知識論	單一，主觀的知識論
方法論(methodology)	方法論(methodology)
干擾者的方法論(interventionist)	詮釋(hermeneutic)方法論

是單一的，而且知識是客觀獨立存在於觀察者之外。因此，在方法論上主張研究者能藉由客觀觀察，發現事件的真相及事件如何運作。

相對地，以目前頗受教育界肯定的建構主義取向觀點，則認為實體是多元、是由社會建構而成的；知識並非獨立於認知主體之外，知識是研究者與被研究者在互動中創造出來的。觀察者所觀察到的現象，是觀察者帶著某種理論所表達出來的，並非客觀獨立存在於觀察者之外。因此，在方法論上主張

的觀點，可以被當作是一種基本信念系統，並用它來處理哲學上有關本體論(ontology)、知識論(epistemology)和方法論(methodology)上的問題。本體論(ontology)所討論的是實體(reality)的形式與本質是什麼；知識論(epistemology)所討論的是知識與認知主體間關係、知識的本質是什麼；方法論(methodology)，探討如何找出認知主體認為可相信的事物。

然而，不同派典對於本體論、知識論與方法論看法不同(Guba & Lincoln ,1989) (如表 1)，這種不同造成對教學、學習和評量這三者的看法不同。傳統派典，主張實體(reality)

詮釋的/辯證的方法論，由研究者與被研究者重新建構事件真相。

上述這兩種不同派典，對教學、學習和評量這三者的看法有所不同，以下分別以傳統取向、建構主義取向來探討這三者之差異與轉變。

1. 傳統派典取向

Guba 和 Lincoln (1989)認為傳統的評量幾乎是獨遵科學的方法(exclusive scientific)，所持的本體論立場是實證主義的假設 -- 認為具有客觀的實體(reality)存在，而觀察者可以

客觀獨立毫無偏見的對被觀察者進行觀察。由於認為知識是客觀獨立於學習者之外，觀察到的結果與學習者本身經驗無關。因此，「學習」被認為是刺激與反應的連結，「教學」就被認為是不斷透過反覆練習而讓學習者精熟，而「知識」就被當作是不斷的累積事實與基本技能(Doran, Chan & Tamir, 1998; Birenbaum & Cochy, 1996, p.5)。對於評量的結果，被認為就是事實如何運作及事件的真相。

Guba 和 Lincoln (1989)所著第四代評量(Fourth generation evaluation)之前三個世代，都是基於傳統派典，強調以客觀和量化來證實理論及命題時的線性、邏輯化過程，並將洞察力、直覺、創造思考等非理性因素排除於探究過程外。因此，傳統派典下強調評量結果之權威性與客觀性(張文華和郭重吉, 1995)。

在上述傳統派典之學習情境下，是非題、選擇題、配合題等客觀性的評量方法，就成為最常用的評量方式。然而，這樣的評量方式卻衍生出一些問題，如：無法評量學生解決問題能力及情意方面技能、扭曲教學本質及考試領導教學等缺失(Birenbaum & Cochy, 1996)。因此，改變評量方式以導正教學成為很重要課題。

2. 建構主義取向

基於上述取向評量缺失，新派典取代由此產生。從建構主義的觀點來看，Guba 和 Lincoln (1989)主張實體並非客觀的存在於學習者之外(reality are not objectively out there)，實體是人們建構的結果。而新的學習觀著重

提升解決問題及高層次思考的評量。在教學策略上要求能幫助學生發展學習的能力，並在真實的情境中解決問題(Doran, Chan & Tamir, 1998)。

根據 Guba 和 Lincoln (1989)的想法，第四代評量特點如下：就評量的結果而言，並非針對真相如何或事實如何運作所進行之描述，而是評量者與被評量者在情境中意義化後共同建構出來的。所發現的並非事實(fact)而是彼此互動建構的結果，而建構的結果與特定的情境、心理、社會、文化、情境相關；就評量過程而言，評量是被委託者、評量者、被評量者等相關人員共同參與的過程，而不是評量者因其專業性主導，忽略其他參與者，造成被評量者漠視評量結果或不滿意；就被評量者而言，強調對被評量者人格、尊嚴、過程、結果及隱私的尊重。雖然傳統評量也注意到這樣的問題。但第四代評量已達到全面參與(full participate)的層次--被評量者被視為共同參與評量的設計、實施、及結果的解釋等。被評量者被視為人而非物化成研究的對象。

(二)教育目標的改變

不同時代背景下所強調的教育目標不同，因而所要評量的標準、方法亦不同。十九世紀中期，教育的目標是為工業社會培育具有基本技能的人；而在進入資訊化、多元、持續變動的廿一世紀，教育的目標是讓人能獨立獲取知識，並利用它來解決問題(Birenbaum & Cochy, 1996, p.4)。在美國國家科學教育標準中(National Research Council, [NR-C], 1996)揭示了科學素養的重要性，而一位具

有科學素養的廿一世紀公民，必須能夠不斷學習、進行推理、創造思考、做決定及具有解決問題的能力。因此，評量目標將改變以往著重調零碎知識，而較強調知識結構、解決問題能力，並強調教師、學生及同儕參與的評量(如表2)。

表 2：評量標準的改變(NRC,1996,p.100)

較不強調	較強調
評量所有易測量的	評量最有價值的
評量零碎的知識	評量好的結構知識
評量科學知識	評量科學性瞭解與推理
評量學生不知道的知識	評量學生已經知道的知識
評量著重成就	評量成就及學習機會
教師做最後的評量	學生參與自我的評量
評量出測驗專家發展	教師參與發展外在的評量

基於上述教育目標的改變與科學哲學觀點的改變，對於評量本質的看法有所不同。在進入廿一世紀之際，教育的目標，轉變為提升學習者自我思考以解決問題；所要培養的學習者，不是以記憶為主的獲得知識方法，而是強調學生主動建構知識，應用知識；而對評量的觀點，不再是記憶、量化之單一取向，而是以符合情境、學科和學生個別差異的多元化取向評量。

三、評量與教學的關係

教學與評量，猶如一種互為鏡像的關係。在評量的鏡像中可以反映出教學目標是否達成，在教學實況中則反映出評量的目標。評量在教學中扮演怎樣的角色，對教學的影響又是怎樣呢？以下針對這些問題加以

討論。

(一)評量所扮演的角色

根據 Jorgensen 和 Shymansky (1996)主張評量具有執行目標、監控達成目標的過程、明確的改變教師的角色(由知識的傳輸者轉變成思考的促進者)的功能。White 和 Gunstone (1996)認為如果用狹隘的方式來評量學生的理解狀況，則學校所能提昇的理解能力將受到限制，而且是侷限於某個方向。Gipps 和 Murphy (1994) 認為評量具有確認學生學會什麼、未學會什麼及學習困難所在的角色。這種角色的評量，稱為形成性評量(formative assessment)。若把評量當作是提供評鑑的訊息給老師，則它扮演的角色是一種選擇與證明的角色，經常在學習告一段落之後實施，稱為總結性評量(summative assessment)。

換言之，如果希望評量能促進教學，則評量不應放在學期末。評量如果放在教學之後，那麼評量並不能幫助教師作任何教學判斷。評量應是連續的過程，讓教學者經由學習者對評量的回應中獲得一些訊息，以作為教學上修正的依據。因此，評量可改變教與學，評量是教育改革的中心(Jorgensen & Shymansky, 1996)。此時評量將超脫「我們教所要考的內容(we teach what we test)」爭論。

(二) 評量方式對教與學的影響

Carter 和 Berenson (1996)認為評量方式的改革，對學生的學習與教師的教學有決定性的影響。如果教師仍停留在使用選擇題、是非題、填充題的評量方式，則教學將停留於呈現教科書內容，學習也將局限在評量所能達到的範圍。因此，如何改變學生的學習

方式及教師的教學型態，評量扮演很重要的角色。

在 Project 2061 改革藍圖中，回顧課室評量(classroom assessment)、州的評量及國家評量等三個層次的評量實況發現，教室的評量對學生的學習影響最大(American Association for the Advancement of Science, AAAS, 1998)。而一般教室中教師所設計的評量，大多偏向紙筆測驗，著重學生記憶。雖然，要用不同的評量方式才能測出學生在實作表現、成就表現，但證據顯示教師由於未受過有效的評量方法訓練，大部分教師不願改變評量方法。此外，由於時間考量，教師大都侷限於採用簡答、是非題等評量方式。因此，評量策略的研發，也成為科學教育的研究者所重視的課題。

連瑞琦(1998)探討數學科採用實作評量方式對學生數學成就與態度的影響，採用單因子獨立樣本實驗設計進行研究。實驗一組採用實作評量與討論教學；實驗二組採用實作評量與講述教學；實驗三組採用傳統評量與講述教學。並將研究樣本 102 人隨機分派為三組，各組經八次實驗處理。結果顯示實作評量對學生數學的成就提昇，並無顯著幫助。

而宋文菊(1998)則針對國小學童閱讀理解實作評量上的表現進行研究，結果顯示：實作評量對閱讀瞭解能力評量，是一個適切的工具。實作評量可區辨不同國語成就學童之閱讀理解類型。實作評量能測量到學生的閱讀理解思考歷程。曾嘉琪(1995)針對「教學及評量改進班」進行研究，結果顯示：教學及

評量改進班學生比普通班學生更喜愛學習、更主動學習。在教師方面，不同性別、教育背景的教師，其教學的生動活潑及評量多元程度，並沒有顯著差異。但教師的教學歷程因素與教學相關因素，對學生的表現具有影響。

桂怡芬(1996)針對自然科實作評量的效度進行研究，結果顯示：實作評量影響學生學習態度的改變，在實驗時態度趨於積極主動；同時在準備考試的方式與學習焦點上會特別注意有關實驗教學的內容，並有助於學生實驗操作能力的提昇。對教師的影響上，實作評量有助於診斷學生學習困難，提供教師教學上的反省。

林純年(1998)以研究概念圖建構對學習的影響，採用準實驗研究法的不相等組前後測設計，以國小四年級學生四個班為研究樣本，其中兩個班分派為實驗組，另兩個班為控制組。實驗組接受六個單元概念構圖訓練，並在閱讀資料後，實際構圖；控制組則自行閱讀資料，不須構圖，但兩組都接受閱讀測驗。結果顯示：當評量題目類型是較偏重訊息記憶或答案較封閉時，概念構圖學習方法與一般學習方法並無顯著不同。但是，當評量的類型較著重知識整體和關聯時，概念構圖學習方法優於一般學習方法。

Shieh 和 Tuan(2000)研究有關教師教學型態與評量方式關係，採用問卷方式調查學生對於評量、學習進步及評量與教學型態關係，並以卡方考驗分析。結果顯示教師上課方式主要以講解為主時，其評量方式以採用習作、傳統考試為主；以學生操作為主的教

學活動，則其評量方式是以實驗操作為主。由此顯示評量方式與教學型態有著密不可分的關係。

從連瑞琦(1998)研究顯示實作評量對於數學成就提昇、態度的改善並無顯著提昇。但在宋文菊(1998)、曾嘉琪(1995)、桂怡芬(1996)研究卻顯示實作評量對於瞭解學生閱讀能力、閱讀理解類型、提昇學生興趣、教學歷程等項目上有顯著的幫助；Shieh 和 Tuan (2000)研究亦顯示評量方式與教學型態具顯著關係；林純年(1998)的研究亦顯示評量類型若著重知識整體和關聯時，採用概念圖學習較為有效。因此，有關評量對於教學與學習的影響，甚為分歧有待進一步瞭解。

(三)評量取向之教學

Foster 和 Heiting(1994)提出教學與評量是互為鏡像(mirror image)的觀點，認為評量指引教學，教學引導評量；事實上，如果以學生的行為、行動及想法為評量，則教學活動將被評量所引導；並認為即使現今強調的動手做評量(hands-on assessment)，依舊是基於教學與評量是各分開的基本假設。

Gallagher(2000)提出融入式評量(embedded assessment)的想法，企圖將評量融入教學過程中，而不是把評量放在教學後所進行的傳統評量。因此，融入式評量與教學、學習之間的界線是模糊的。根據 Gallagher, Parker 和 Ngwenya (1999)所定義的融入式評量，是指教師利用獲得學生理解的知識，引導教學順序的評量。教師由蒐集學生理解之資料作為教學之用，並藉由這樣的評量方式，教師改善自己的教學。

Gallagher, Parker 和 Ngwenya (1999)提出融入式評量的三大步驟，以說明教師教學時的行動。第一、蒐集有關學生在學習時的想法；第二、深入分析所蒐集到訊息，以瞭解學生困難所在；第三決定下一個的教學步驟。

研究顯示融入式評量，對於提升教師之學習環境有顯著幫助；再者，融入式評量促進學生的理解與應用科學知識上的成功，進而增強了學生學習動機(Gallagher,2000)。就研究結果而言，融入式評量達成當代教育改革的目標 -- 加深學生的理解、應用科學知識能力，是值得一試的評量方法。

四、評量策略的探討

McMillan(2000)認為好的教學可藉由評量激發學生學習動機，並與教學和學習的理論一致。而好的評量是有效的(valid)、公平的及道德的(fair and ethical)。公平包括無偏見(absence of bias)、公平的處理(equitable treatment)、公平的學習結果(equality in outcomes)。針對評量的有效性及公平性問題，分述於後：

(一)評量的有效性

評量的目的在於瞭解實況，但在不同情境下該採用何種評量模式和內容來評量最為有效，是一個值得探討的問題。由於傳統以紙筆測驗方式進行評量，無法反映出教學目標，因而要求改變評量方式呼聲日起。要怎樣才能達到到真確的評量(authentic assessment)，簡單的說，評量應包括學生做功課時最有價值、有意義的事項。Carter 和 Berenson

(1996)認為真確的評量(authentic assessment)包括晤談(interview)、小組解題(group problem solving)、歷程檔案(portfolio)、概念圖、實作評量、開放性問題(open-ended)。

以實作評量(Performance assessment)為例，這種評量是用來測驗學生應用知識和技能於真實情境中的能力，與傳統著重於事實和零碎的技能有所不同。實作評量企圖藉由學生處於真實情境中主動解決問題，以呈現學生的真實能力。以科學教育的目標來看，藉由實作評量能瞭解學生高層次思考能力，這種評量才是不偏廢的全方位評量(陳文典、陳義勳、李虎雄，1995)。根據陳文典等人(1995)進行實作評量的研究結果顯示我國學生傳達與獨立作業能力不足；在察覺問題、規畫工作、及獨立判斷上亦不足；凡涉及敘述方式問題，回答顯得簡略不全；在涉及實驗時，大都忽略其他變因的影響。在採用實作評量可行性上，李虎雄和黃長司(1995)的研究指出實作評量题目的設計不成問題，但評分卻是一大負擔，因此，在採用實作評量時「評分」問題必須加以考慮。

陳華傑(1998)研究發現，採用實作評量，對於解決問題之過程技能和思考智能有很好的評量效果；紙筆評量，對於科學概念的認知及應用有很好的評量效果；媒體評量，利用影帶呈現的現象為題目，由其中獲得資訊，此種方式很容易得知受測者應用所學以解決問題的能力；而心得論述的評量可以得知學生對於綜合性概念及情意方面體會；至於成果發表，則能夠表現學生對於整體事件的評鑑。

黃智信(1998)認為一個真確的評量(authentic assessment)應該是提供一個實際問題，然後觀察學生如何解決它，但實施上考量，受測人數、教師教學負荷下並不可能。該研究構想能否設計一個具實作評量特性，卻又能大量施測的評量方式。並以「密度壓力浮力的測量」、「人體呼出氣體」為主題，分別設計「實驗設計測驗」、「實作測驗」、「填充題」、「選擇題」等四種不同評量工具。結果顯示：「實驗設計測驗」、「實作測驗」在能力的判別上比「填充題」、「選擇題」為高，可用來評量學生在處理問題過程中的缺失。在可行性上發現，實作評量所需時間是「填充題」、「選擇題」的三十倍。

從上述陳文典等人(1995)、李虎雄和黃長司(1995)、陳華傑(1998)及黃智信(1998)等研究顯示，各類評量方式各有其適用性範圍。若評量的目的是在瞭解學生高層次思考能力，則可採用實作評量，但在評分卻是一大負擔；若目的是評量學生在處理問題過程中的缺失，則可採用「實驗設計測驗」、「實作測驗」，但所需時間是一大負擔；若目的是評量學生對於科學概念的認知及應用，可採用紙筆評量；若目的是評量學生應用所學以解決問題的能力，則採用媒體評量；若目的是評量學生綜合性概念及情意方面體會，則採用心得論述；若目的是評量學生對於整體事件的評鑑，則可採用成果發表方式。

(二)公平的評量

科學教育研究者一直致力於研究如何將評量達到公平，但是，近幾年的研究證明要達到這樣的目標，並不是簡單的事情(Parker

& Rennie,1998)。在英國、美國及澳洲的研究顯示，影響公平的評量的重要向度是：學生的背景及世界觀。Popham(1995)認為評量工具的設計，除了以信度、效度作為判準外，另外提出第三個判準 -- 評量的偏見(assessment bias)。評量的偏見涉及有關評量工具是否因受測者的性別、種族、宗教、社經地位的不同，造成對受測者傷害或不公平。

美國國家科學教育標準中(NRC,1996)提出全民的科學素養目標，要達到這樣的目標唯有讓所有的學生都有均等的機會(equal opportunities)學習科學(Gipps & Murphy,1994)。而要達到均等的學習機會，則評量必須是多面向的(multifaceted)，以讓學生能利用個人的學習風格及能力，表現在不同的評量類型上(Doran, Chan & Tamir,1998)。因此，就教師的教學而言，要瞭解不同學生的學習風格與能力以做為改進教學，則評量方法必須因學生而有所變化，如此才能普遍瞭解學生不致造成不公平。

1. 不公平評量的來源

評量工具的公平性，可以從兩方面來討論：評量內容的設計與評分標準。首先在評量內容方面，Popham(1995)認為評量內容如果發生評量的偏見(assessment bias)，將扭曲學生的表現而引發不公平。這種評量的偏見包括：攻擊(offensiveness)和不公平的待遇(unfair penalization)兩種形式。如果評量的內容攻擊某個族群學生，則這項評量的過程是有偏見的，而這種偏見常常出現於將某個族群負面的刻板印象呈現在評量中。例如將男性描繪成高收入與較有聲望的一群(如：太空

人、物理學家)，而將女性描繪成低所得和低聲望的角色(如：家庭主婦、店員)。不公平的待遇，是指評量內容雖不至於有攻擊性，但卻不利於某類學生。例如：讓女性進行不熟悉的競爭性運動。

在評分標準方面的不公平現象，大多發生於開放性問題評分上的不一致。依Popham(1995)所提出的評分者上評量偏見有：慷慨誤差(generosity error)，是指高估學生表現實際上學生並沒有表現很好；嚴厲性誤差(severity error)指低估學生表現；中心傾向誤差(central-tendency error)，認為學生不可能太好或太差；光環效應(halo effect)，認為學生平常表現很好(或很差)，在這項測驗上也是很好(差)。

綜合上述，有關不公平現象的產生，是由於評量設計者，在評量內容的設計與評分的標準所造成。評量除了考量信度、效度之外，評量者的偏見是不可忽視的重要向度。

2 公平的評量策略

為了要使評量變得公平，Parker 和 Rennie(1998)提出要達到公平評量的三個先決條件：

(1)課程的真實度(Curriculum fidelity)。

這項條件是呼應Lawrenz (1991)所提課程、教學過程和評量，三者必須達到一致。Parker 和 Rennie(1998)認為從1960年以來，課程的改革者並未考慮到公平的評量(equitable assessment)；在1990年代，科學教育的教學進入了哲學改變的時期，期望科學能被廣泛的學生瞭解。建構主義的知識論改變了科學教學與學習的策略，更加強調採用小組

合作學習及持續發展語言與溝通技巧。這樣的改變，反映出學生族群的多樣化學習風格 (multiplicity of learning style)；如果這些議題能被注意，透過多樣性的評量風格 (assessment style)，學習風格的多樣性 (diversity of learning style) 將與課程改革的目標達成一致 (Gaffney, 1992)。

(2) 多樣性和機會 (diversity and opportunity)

評量的目的如果在於讓學生真實地展現學習成果，以作為教學改進參考。評量方法必須能讓所有的學生與小組，擁有最大的機會去呈現所有已經知道的、理解的、能做的學習成果 (Parker & Rennie, 1998)。因此，所有阻礙學生展現學習表現的評量應被以避免；所有能讓學生展現學習成果的方法應予鼓勵，才能達到公平的評量。

(3) 價值和倫理 (value and ethics)

要達公正和公平評量的第三項先決條件是：對於評量結果之解釋，在不同群學生不同型式評量和相等的表現 (equivalent performance) 應持等值態度 (Parker & Rennie 1998, p. 899)。

從上述探討有關公平的評量議題中，首先，從教育的目標而言，必須使學生達到學習機會的均等。但這樣的學習機會均等，必須基於學習者學習風格所做的考量，而並不是齊頭式的機會均等。再者，既然強調學生學習時的學習風格以達到學習機會的均等，那麼評量也必須是多面向的，以使學習者在最有利情況下展現出學習成果，如此才是達到所謂公平的評量。最後，就學習的本質而

言，影響學生學習的因素除了學習者學習風格外，學生的背景、種族、宗教、社經地位等都將影響學習成果。因此，在評量時必須考量這些不公平的因素是否產生，才能達到所謂的公平。

五、結語

綜合以上文獻中有關評量議題之探討，對於評量提出以下見解。分別敘述於後：

(一) 評量的本質

「評量」是利用各種不同的方式獲得關於學生學習狀況的訊息，回饋教師，做明智教學決定，以培養學生各種能力健全地發展，達成最大教學效果。因此就提供教師教學改進而言，評量的本質應該是以提升學習者的學習 (Davis, 2000) 為第一要務。

早在多年前在德國 Dr. Bambach 對於「評量」就有所謂「激勵，非評分」的評量觀點。德國歌汀根基礎小學：Hainbundschole 小學，Weiland 女士提倡「沒有分數評量的小學」方案。這些評量改革計畫著眼於認為能力表現評量，造成能力階層、惡性競爭、助長社會不公平現象 (引自陳惠邦，民 89)。然而國內，教師與學生尚且停留於「分數取向」觀點 (Shieh & Tuan, 2000；謝祥宏和段曉林, 2000)，此現象亦亟待進一步探討與改善。

(二) 評量與教學的關係

教學與評量，本是一體的。但是，大多數老師將教學與評量分開。Gallagher (2000) 所提出的融入式評量的觀點，可以說是教學與評量關係的最佳寫照 -- 評量與教學、學習之間的界線是模糊的。在研究中發現教師教

學方式、評量方式、學習者學習方式，三者彼此互相影響(連瑞琦,1998;宋文菊,1998; Carter & Berenson,1996; Parker & Rennie,1998; Shieh & Tuan,2000)。然而，正如 Philip Gammage 所說「沒有人會因為一再被測量而長高」一樣(引自吳麗君,2000)。學習方式的改變，不全然是評量方式所造成。評量方式所引發的各種變化，可能帶動教學品質的改進或產生不同班級氣氛、. . .等，這些才是影響教學、學習成效之因素。因此，在評量的改革運動中找出這些中介因素，才是企圖以評量改進教學的研究者最重要的課題。

(三)公平的評量 -- 多元評量觀點

承上所述，教師的評量方式對於學生學習方式產生影響。如果用狹隘評量觀點，則窄化了學生的學習。再者，學生之學習風格各有不同，執著於某類型評量方式，對於另類學生亦產生不公平(Parker & Rennie,1998)，因此，唯有多元評量，才能與學生多種學習風格匹配(Gaffney,1992)。而多元評量不僅是指評量方式的多樣性，而更重要的是在評量目標、評量對象、評量時機、評量標準的多樣性。

(四)對未來研究之啟示

研究指出教師與學生對「評量之看法」、「對教學與評量關係的看法」不盡相同，學生深受評量方式之影響(謝祥宏和段曉林,1999; Shieh & Tuan,2000)。教師對於評量之看法，本質上是對於知識論、本體論上觀點之不同。因此，要探討教師與學生對評量知覺之交互影響，未來的研究可從教師所持的知識論、本體論之觀點，探討教師評量信念對學

生學習之影響。

參考文獻：

- 1.吳麗君(2000)：交互損益見真章 -- 談英國國課程評量的來時路。課程與教學季刊，3(1)，78-94。
- 2.宋文菊(1998)：國小學童在閱讀理解實作評量上的表現分析。國立臺南師範學院國民教育研究所碩士論文(未出版)。
- 3.李虎雄、黃長司(1995)：美國馬里蘭州實作評量工具在台灣施測的可行性。科學教育月刊，179，41-49。
- 4.林純年(1998)：概念圖對國小學童自我學習科學說明資料之影響。國立台南師範學院國民教育研究所碩士論文(未出版)。
- 5.桂怡芬(1996)：自然科實作評量的效度探討。國立臺北師範學院國民教育研究所碩士論文(未出版)。
- 6.張文華、郭重吉(1995)：科教革新中評量理念的重建。教育研究雙月刊，45，23-30。
- 7.連瑞琦(1998)：實作評量對國小學生數學成就與態度的影響。南華管理學院教育社會學研究所碩士論文(未出版)。
- 8.陳文典、陳義勳、李虎雄和簡茂發(1995)：由馬里蘭州的學習成就評量與其在台灣試測的結果看 -- 實作評量的功能與運用。科學教育月刊，12月，2-11。
- 9.陳惠邦(2000)：在保守的傳統中謹慎前進：德國小學教師評量信念與實踐研究。新竹師院學報，13，245-264。
- 10.陳華傑(1998)：不同評量模式的評量效果

- 分析。國立台灣師範學大學物理研究所碩士論文(未出版)。
11. 曾嘉琪(1996)：臺北市國民小學試辦教學及評量改進班之評鑑研究。國立政治大學教育學類研究所碩士論文(未出版)。
 12. 黃智信(1998)：情境真實程度對評量功能的影響。國立台灣師範學大學物理研究所碩士論文(未出版)。
 13. 謝祥宏、段曉林(1999)：國小自然科教師教學與評量相關變項之研究。第三屆數理教學及師資培育研討會。彰化：國立彰化師範大學主辦。
 14. 謝祥宏、段曉林(2000)：國中學生對於理化科評量看法之研究，中華民國第十六屆科學教育學術討論會論文彙編。台北：國立台灣師範大學主辦。
 15. American Association for the Advancement of Science.(1998). Blueprints for reform-Science, Mathematics, and Technology education. New York: Oxford University Press.
 16. Birenbaum, M. & Cochy, F. J.R.C.(1996). Alternatives in assessment of achievements, learning process, and prior knowledge. Boston: Kluwer academic publishers.
 17. Carter, G. & Berenson, S. B.(1996). Authentic assessment: vehicle for reform. In Rohton, J. & Bowers, P. (Eds), Issue in Science Education (pp.96-106). Virginia: NSTA.
 18. Davis, N.(2000). Using assessment as a tool for educational reform, Paper presented at the annual meeting of the Southeastern Association of Educators of Teachers of Science, Auburn, Alabama, October, 2000.
 19. Denzin, N. K. & Lincoln, Y. S. (1994). Handbook of qualitative research. Thousand Oaks, London: SAGE.
 20. Doran, R., Chan, F. & Tamir, P.(1998) Science Educator's guide to assessment. Virginia: NSTA.
 21. Duschl, R. A. & Gitomer, D. H. (1991). Epistemological perspectives on conceptual change: implications for educational practice. Journal of Research in Science Teaching, 28(9), 839-858.
 22. Foster, W. G. & Heiting, A.W. (1994). Embedded assessment. Science and Children, October, 30-33.
 23. Gaffney, K. E.(1992). Multiple assessment for multiple learning style. Science Scope, March, 54-59.
 24. Gallagher, J.(2000). Implementing needed reforms in science education: research and development to enhance understanding and application of knowledge. Paper present in Second International Conference on Science, Mathematics and Technology Education
 25. Gallagher, J., Parker, J., & Ngwenya, L. (1999). Embedded assessment and reform of science teaching and learning. In preparation.
 26. Gipps, C. & Murphy, P.(1994). A fair test? Assessment, achievement and equity. Open

- University Press.
27. Guba, E. G. & Lincoln, Y. S. (1989). Fourth generation evaluation. Sage publication: The international professional publisher Newbury Park London New Delhi.
 28. Jorgensen, M. A. & Shymansky, J. A. (1996). Assessment in science: A tool to transform teaching and learning. In Rohton, J. & Bowers, P.(1996,eds) Issue in Science Education, NSTA.
 29. Lawrenz, F. (1991). Research matters--to the science teacher, authentic assessment, NARST News, 33,15-17.
 30. McMillan, J. H.(2000).Fundamental assessment principles for teachers and school administrators. Practical assessment, research & Evaluation, 7(8). [Http://ericae.net/pare/getvn.asp?v=7&n=8](http://ericae.net/pare/getvn.asp?v=7&n=8)
 31. National Research Council (1996). National science education standards. National Academic Press, Washington, D. C.
 32. Parker, L. H. & Rennie, L. (1998). Equitable assessment strategies. In B. J. Frase & K.G. Tobin (eds), International Handbook of Science Education, Part Two (pp. 897-910). Virginia: Kluwer academic Publishers.
 33. Popham, W. J.(1995).Classroom assessment What teachers need to know. Boston: Allyn and Nacon.
 34. Shieh, S. H. & Tuan, H. L.(2000). Elementary students' perceptions of assessment of the science class. Paper present in Second International Conference on Science, Mathematics and Technology Education, Taiwan.
 35. Torrance, H. (1995).The role of assessment in educational reform. In Torrance, H.(1995). Evaluating authentic assessment: problems and possibilities in new approaches to assessment. Open University Press.
 36. White, R. & Gunstone. R. (1996). Probing Understanding. New York: The Falmer Press.