

多元評量之解毒與解讀

吳 毓 瑩

本文藉由電腦中文注音輸入法所造成的解讀與解毒二詞之混淆為例，討論多元評量此概念在後現代社會中所需要的解毒策略與解讀歷程。首先作者點出教育界普遍依賴西方理論，以致於對任何盛行的教育策略，都直接聯想到相對應的英文辭彙與立論。是故，當本文企圖解此毒時，鄭重點出多元評量的概念沒有響亮的英文名詞與之相呼應，其豐富的在地意義，正在展開。而在解讀多元評量的過程中，作者指出多元評量除了代表策略多元、價值多元、詮釋多元外，更呈顯出邏輯觀念的多元：若P則Q，與P等於Q兩種邏輯流程的併置。此兩種邏輯觀念的存在，彰顯不同價值觀、不同評量策略、與不同表現媒介。作者認為對於多元評量的討論，除了文獻既存的想法外，亦可含納多元邏輯的論述。是故筆者在文末歸納出一個三元共存圖像，解釋不同專長、訓練、與立場的教育實踐者有不一樣的立足點以實踐各自的教育理想。

關鍵字：多元評量、否證論、後現代、現代、邏輯實證論

本文作者為國立台北師院教育心理與輔導學系教授；學術專長為測驗評量

壹、前言

當筆者在準備本篇論述時，乃是利用漢語拼音法藉由讀音系統讓電腦產出文字。而對於解讀二字的輸入，電腦總是優先輸出解「毒」。此現象給予筆者在構思此文時莫大的靈感，電腦的讀音系統似在提醒，許多概念不只是一要去解「讀」，更要去解「毒」。感謝電腦給的錯字，啓發筆者從解讀想到解毒；另一方面並想到我們常常在復古一些舊日的概念，例如電腦的錯字誤植，如果放在三千年前的周朝時候，可能就是一個神諭。今日之電腦錯字，變成神諭的一個復古現象。現今教育界重視的實作評量所強調的任務取向評量，也未嘗不是老祖宗的科舉之復古。是故本文企圖從復古的意念，思考多元評量的時代意義。

電腦的錯字型態，同時也反映了曾志朗在〈科學人〉雜誌創刊號中的提醒。曾志朗提及（民91），漢字輸入法的不同，會造成不一樣的錯字型態。讀音系統的錯字，與造字系統的錯字，遵循不一樣的形式，因此，若要讓熟悉讀音系統的人去校正自己使用的讀音系統的輸出，與熟悉倉頡輸入的人去校對倉頡系統的輸出，結果效率往往其差無比。但是如果交換測試，則效果就會很好。這個交叉測試的結果告訴我們，人類的認知系統其實是非常封閉的。

我們學習新的事物是爲了適應新的環境，一但學會了，卻又自陷其所造成的藩籬之中，形成當局者迷的現象。（頁I，曾志朗，民91）

因此，電腦的錯字誤植，也讓筆者思索，在評量的實踐歷程中，如何解毒並解讀「多元評量」概念？我們是否會如曾志朗的描述般，陷入安全合理的認知限制中，形成當局者迷的現象？

貳、解毒：多元評量在歷史的演變中

教師教學，需運用測量、考試和考查等方法，考查兒童學習的成果，有時亦可指導兒童作自我評鑑。成績評鑑可以幫助兒童明瞭自己進步的快慢，亦可幫助教師了解教材教法是否適宜，和作爲診斷補救教學的依據。所以，

評鑑的作用，積極重於消極。

教學評鑑的內容，除知識、技能、行為和習慣外，需兼顧態度、理想、情感和興趣等學習成果的考查。

評鑑教學成績的方法，有觀察、紀錄、批判、口試、筆試、量表、表演和作品評量等。教師可按學科性質與評鑑目的，相機配合使用。

以上摘錄的敘述，來自課程標準，出處筆者暫且不表，它所談的是我們最熟悉不過的內容：評量的多元面向（包括認知、技能、態度、還有理想與興趣）、評量的積極性目的、兒童自我評鑑、視時機規劃不同評量方式等概念。如果說這就是目前教育改革在努力的方向，大概也沒有人會反對。去年暑假（民90.8）有同事問筆者，「多元評量你說有理論依據嗎？有人請我講多元評量的理論依據，你說要怎麼講？多元評量有沒有相對應的英文名詞？」面對這個問題，筆者只有很欣喜地告訴他：「多元評量，沒有相對應的英文名詞。就是這樣。」

一、我們以為它有英文的理論

多元評量的想法廣泛為教育實踐者接受，同事確實問了一個很重要的問題——它由來何處？在筆者以及很多人的思考中，不論是說的或是寫的，使用多元評量這個詞，好像很自然，不似真實評量、另類評量、卷宗評量、動態評量、卷宗評量等等，不管如何解釋，都是一個外來名詞，且在使用時，必定附上英文原文，以示其原始出處；甚且關於譯名，還有許多爭議在其中——是實作還是表現？是卷宗還是檔案？是真實還是真切？是另類還是變通？並且動用了許多文章為它解釋（如陳文典，民89；吳鐵雄與洪碧霞，民87；單文經，民87；蘇義翔，民87等）。不過，在眾多名詞當中，「多元評量」這個詞卻獨樹一幟，既不用費口舌解釋它由來何處，亦不用費心為它名詞釋義，它就這麼自然地存在於我們的語彙當中了。

當有人開始問，它是否有理論呢？其實更根本地可能是在問：它是否有英文的理論與解釋？這一問，讓筆者心中一喜，表示這個概念很重要，卻未曾見英文的論述。更表示那常常我們掛在嘴上的名詞，真正是我們的關懷，而不是西方的擔心。當大家都在尋找英文出處時，卻忘記了這可能是誕生自我們社會的需求與語言。與這個現象類似的另外一個例子是人們常說的「考試領導教學」這一概念，廣泛流傳在報章雜誌學生親師之間。然而當學術界

提到此名詞時，都要再加註一個括弧指明是 test-driven instruction，似乎，有相對應的英文，表示其由來有自，儘管我們在小時候自己還是被考的學生時，早已耳熟能詳、參透「考試領導教學」的實際。甚而在二千多年前，從隋朝的科舉制度以降，我們早已知道，考試領導讀書、影響教學、甚而左右仕途。

筆者很高興「多元評量」沒有響亮的英文相對應名詞，讓我們有空間充分思索之。筆者自己使用多元評量這個詞是在民國84年，當開放教育正熱烈著，學習單成爲評量的主流方法，筆者提及：「如果評量的意義沒有革新或轉變，多元也可能變成一種單一的、制式的、規定的多元。如同我們這次在學校拜訪中看到的單一概念——學習單及檢核表一樣，它也許已經形成另一種規定。」（頁95）。而在八十九年出版的九年一貫課程綱要中（教育部，民89），多元評量一詞開始在官方文件中出現，例如語文領域的實施要點：「教學評量建議採用多元化的評量模式。國小階段多採取形成式評量……，亦可採取檔案評量的方式」（頁125）一切是那樣的自然，沒有人想要特別去定義它。

在中華民國期刊論文索引中，自民87起到91年2月，有33篇文章在標題中或關鍵字中出現「多元評量」此詞。如果以"multiple assessment"之詞在ERIC (educational resource information center) 資料庫中尋找，則自1998以降，有14篇文章使用multiple assessment一詞於摘要中（但是沒有任何一篇文章在標題或關鍵字中提出multiple assessment的概念）（見表一）。另外，查詢中華民國博碩士論文摘要系統，顯示自民國87迄今，有126篇博碩士論文在題目、摘要、或是關鍵字中使用多元評量，其中博士論文14篇，碩士論文112篇。相對地，美加博碩士論文資料庫(PQDD, Proquest Digital Dissertation) 則顯示自1998至今，有14篇博士論文使用了此名詞。數據如下表所示：

表一 使用多元評量一詞於題目或摘要或關鍵字之文章（論文）篇數

1/1998-2/2002		1/1998-2/2002	
中華民國期刊論文索引 33	ERIC 9	中華民國博碩士論文摘要 126 (博 = 14; 碩 = 112)	PQDD 14

從表一的結果可以看出多元評量在我國的討論情況遠較美加地區來得熱烈。而當國內人士習慣性地問起「多元評量」出處時，我們甚至於可以直接

指稱回去民國六十四年的課程標準，那就是本節一開始所引之一段話，寫在課程標準第13頁（教育部，民64）。這段話中，如果把第一行中「考查」二字換成了觀察，那麼可能一點也不覺察到這是二十七年前概念。

二、我們可能有的認知侷限

對照今昔，方法的多元，古即有之。僅是當時所謂的量表，今稱之為評量規準；當時的表演、觀察、記錄、評判，今稱之為實作評量；當時的作品評量，今稱之為學習歷程檔案，甚而當時即有「指導兒童做自我評鑑」的想法，可見自評之策略，也不是今日才流行。因此，細查歷史的步履，筆者不禁要問：「究竟什麼在演變？」多元評量的想法，一直是教育實務界、學術界、與行政界努力的目標，三十年來我們重複走著一樣的路程，但重點是對於孩子的學習，我們是否進步、改變了什麼？恐怕才是真正需要去解的毒。

因此筆者認為多元評量在教育界中的使用，有兩個重要的認知限制我們必須去突破。其一是我們已經習慣仰賴外來理論，以致於當多元評量沒有相對應的外來理論時，學者容易輕忽之，以為僅是實務界的一種活動。殊不知沒有外來理論的背景，更顯示此概念與行動乃出自本地社會的需求及語彙，由表一相關文章發表數量上中美兩國的差異，即可強烈感受到。其二是我們以為多元評量乃是教育改革風起雲湧後造成的結果，是相對於傳統評量的後起之秀。然而回顧歷史，我們可能祇是發明了一個好名詞通稱三十年來的評量步履。

正如打字系統的認知限制使得我們不容易糾正使用相同輸入法者的錯字般，我們大家使用著同一套認知結構在解讀多元評量而不自覺有侷限，這是筆者認為在了解多元評量之前必須解的毒。了解這兩項認知侷限後，接著本文討論如何跳脫共用的認知系統、換個邏輯來思索多元評量的意涵。也就是去解開三十年來，教育實踐者所走的同一條路上，行至後現代社會中，有無新的意義？對學習產生了什麼作用？

三、從同儕互評之興起開始解毒起

翻翻配合九年一貫課程而出爐的「國民中小學學生成績評量準則」（教育部，民90），並仔細對照今之文件與昔之歷史（民國六十四年的課程標準），可以發現核心概念都類似，唯一有變動的是參與人員的多元——「……

並得視實際需要，參酌學生自評、同儕互評辦理之。」（頁1）從參與人員多元、同儕互評概念的出現，我們看到了多元評量在「評量方法多元化」這一個面向上歷三十年而不變的政策中，多元主導的興起，亦即頭家換人做做看的企圖。這可以說是今日我們談多元評量時應該照顧的重要觀點，否則，僅是三十年來重述多元評量。

加入同儕互評的策略可代表參與人員多元的開始，除了孩子的參與外，我們還可以想到家長、或是隔壁班的老師一起來參與評量的進行。但是更進一步的意涵，指的是「兒童」不僅是被評的客體，也可以是評量的主體。多元的參與帶來的是多方位的詮釋。例如當廣福國小謝美鳳老師在教「老鼠救獅子」這一課時，與同學一起舉辦了一個戲劇表演。謝老師規劃了評分規準讓小朋友觀賞戲劇一併進行評量，評分面向有：劇情、道具、聲音、肢體動作、與秩序；另一方面謝老師也規劃了一個空間讓孩子對表演的組別寫下評語，於是我們看到了一張這樣的評語：

* 給各組的讚美和建議：

- 第一組 道具似乎準備太少了，不過劇情生動活潑，當台上人在講話時，其他沒有上場的表演者，要離開舞臺，這一點大致上不錯但有一些人需改進。 ✓
- 第二組 肢體動作做得很大，配上很趣的台詞是很不錯，但前面有一些同學跑來跑去，當你還沒有上場時，在一旁也可以隨著劇情笑、傷心，但就是不能沒聲做。 ✓
- 第三組 準備的道具雖然很精緻，但不受允許，當獅子在網裡，不可以在地上說話，這樣觀眾沒看到，可以一邊滾一邊很自然的站起來，或是把陷阱下在保東西，就等法把舞台切。 ✓
- 第四組 這組的道具與讓我嘆為觀止，又豐富又精緻而且還把另一個表演者一邊演一邊把他移到舞臺中央，他們的聲音又有點小聲，最響的是咬蚯蚓清楚。 ✓

（第一組 道具似乎準備太少了，不過劇情生動活潑，當台上人在講話時，其它沒有上場的表演者，要離開舞臺，這一點大致上不錯但有一些人需改進。

第二組 肢體動作做得很大，配上很趣的台詞是很不錯，但前面有一些同學跑來跑去，當你還沒有上場時，在一旁也可以隨著劇情笑、傷心，但就

是不能沒事做。

第三組 準備的道具雖然很精緻，但不過充分，當獅子在網繩裡，不可以在地上說話，這樣觀眾沒看到，可以一邊演一邊很自然的站起來，或是把陷阱下在架東西，就等於把舞台架高。

第四組 這組的道具真讓我嘆為觀止，又豐富又精緻而且還把另一個表演者一邊演一邊把他移動到舞台中央，他們的聲音又有點小聲，最重要的是咬字不夠清楚。)

由於謝老師讓同儕的角色加進評量歷程中，我們發現了孩子的眼睛是這樣看戲劇的：「當你還沒有上場時，在一旁也可以隨著劇情笑、傷心，但就是不能沒事做。」這句話說得讓老師拍案叫絕，特別將它畫線凸顯，表示看戲還要看到後台去。多元的參與會塑造出多元的價值、多元的尊重，這是六〇年代的現代性社會所無法顧及的，只因當時的氛圍滯留在邏輯實證論的信仰中，愈是深奧的、抽象的，才愈能類化至日常生活的不同表現上，如這句話所說：「……知識則是以構成的形式作為基礎，知識的作用就在於把實際的經驗加以統整，使其趨向於統一。要建立一種知識，必須藉助邏輯和數學這種純粹的形式，而與實際無關的知識，才具有絕對的真實性。」（黃光國，民90，頁92）。如果把知識一詞，替換成評量或是測驗，所展現的風貌，正是評量在教育情境中某種面容：測驗總是高高在上，測驗題目的組合總是神秘兮兮且神聖不可揭露，一切測驗內容好像很難用到生活上去，而一份好的測驗通常要動用到好多看不懂的心理計量統計符號來解釋；最後利用計量統計產生出來的抽象數字，具有絕對的真實與權威。那樣的時代，也有著多元的評量策略，如紙筆測驗（選擇、申論、問答）、觀察報告、平時作業等。但當時多元方法的終極目的，不在呈顯多元的價值，而是企圖從多方角度探討某種特質或是表現，而且這種特質（或稱構念 construct）乃是經過理論的分析（形成不同面向或是分量表），透過實徵上相互支援的系統（不同的評量媒介如紙筆測驗或是觀察或是自評），如同房子的四面牆壁一般，所共同架構出來的結構。有效的測驗之最終目的是希望各面向之評量結果能夠輻合統整，以解釋此特質。智力二因論之一般性智力與特殊性智力的結構與魏氏智力測驗就是當時的典型產物，儘管魏氏智力測驗的題目既多元且有趣，有口頭問答、拼圖、說故事等（陳榮華，民86; Wechsler, 1944），但是在那時，多元方法所蒐集的評量證據乃在彼此支持而趨於統整，與多元本身無關，終極目的在輻合與統整。至於透過測驗得到的智力數值，與生活世界無

關，更與街頭小聰明無關。

測驗題目及評量多元方式統整後的最終結果，例如魏氏智力測驗所解釋的智力構念，其實沒有真正地存在於生活世界中，是專家從具體生活萃取出來、成爲理論、再企圖轉而成爲可以普世概化的概念，以及可以具體操作的數值。正因數值的抽象與脫離生活，是故一個學生可以帶著此數值到美洲、澳洲、非洲去，同樣能夠讓當地的教育者捕捉到孩子的狀況。

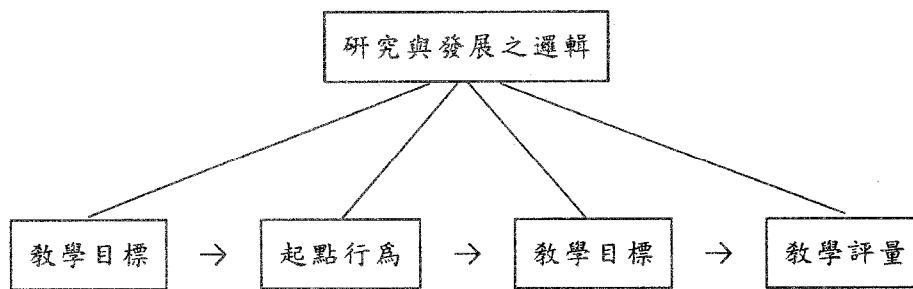
古典測驗理論所努力的多元評量策略，正是處理測驗結果是否能夠統整輻合以測量構念的嚴謹歷程。因此多元評量是教育改革的成果嗎？恐怕不是。三十年前之六〇年代多元評量所盡的責任，毋寧更是九十年代教改所打擊的要害。專家與權威主流的價值取向，藉由多元評量策略而得以實踐，例如多特質多方法的輻合效度證據(Campbell and Fiske, 1959)的大量使用。因此，多元評量本身不代表價值系統的轉變。多元評量的意涵，也可能相當專家主導、抽象取向、與生活世界無關。因此有哈伯馬斯的喟嘆，承認現代社會中生活世界與專家世界的理性溝通，仍是一個未完成的任務(Habermas, 1983, p. 13)。了解多元評量策略亦可能相當傾向現代主義的觀點以構念爲依歸後，筆者企圖從生活世界中的教學過程裡，來解讀多元評量的可能另類面貌。

參、解讀：多元評量在教學歷程中

一、評量的代理角色

上節所引述的謝老師之三年級學生所寫的戲劇表演評語，非常直接與清楚地描述出她所看到的世界。我們無須分數的仲介，便知道孩子看戲的價值取向，甚至分數成了解讀的障礙，例如戲劇表演85分基本上無法清楚溝通它所代表的意義，除非我們先知道戲劇表演在評量什麼，並且知道85分是常模參照概念中的平均分數以上或以下幾個標準差，或是知道85分乃是效標參照測驗中達到什麼樣水準的切分點。一但我們使用分數作爲描述表現的媒介，則對於分數形式的結果之詮釋，皆待許多前提假定的事先釐清，以使結果有意義。而這定出來的分數，與訂定分數的過程（策略），便是邏輯實證論者所指稱的「與實際無關的知識，才具有絕對的真實性。」（黃光國，民90，頁92）此處筆者將之解釋爲具有普世概化的力量。

分數的給予，一方面有其便利：包括清楚的切分點、換算簡便、可排序以比較高低、以及統計上的特質，例如區辨力、難度值、預測力等；另一方面亦有其複雜的一面，包括構念的釐清、評量目標的確立、以及分數的計算由來。欲使表現之後所得到的結果有意義，我們必須同時兩方面皆顧及。這就是測量源起的核心結構，一則要確立構念，使之內涵清晰（例如問題解決能力之內容元素的解析），一則要利用各種媒介使之能夠被操作定義及被測量（例如多組複雜的任務，或是數學考卷中的不同題型）。所以測驗評量具有接近、代理的性質，它「接近」欲測量的構念－問題解決能力，卻永遠不是那個構念，因為真實能力必然在真實生活中展現，考卷或是任務僅是人造的引發來看看表現如何。因此，關於測驗評量的地位，便如1962時，Glaser所提的關係(1962, p. 6)：



圖一 Glaser 1962的教學系統

這張圖呈現的乃是從目標到評量的線性關係，起始於確認目標（有關構念或有關教學），繼而了解起點行為，而後進行教學活動，然後才有可能找到正確的代理人（評量方式）。這就是「若P則Q」的邏輯，亦即，如果問題解決的能力高，則這張數學考卷就會表現好。好測驗的先決條件，須先確認若P則Q的線性是成立的，也就是古典測驗理論效度的邏輯觀。但是這樣的邏輯在實施時，卻碰到誤解與誤用。

若P則Q絕不保證若Q則P，亦即「如果他是我弟弟（P），則他是男的（Q）」，不能保證「如果他是男的（Q），則他就是我弟弟（P）。」因為弟弟與男生不是全等的關係。男生是描述弟弟的一種特質，接近弟弟，但不是弟弟，它可以代理弟弟說明弟弟是性別中的男生，卻不能成為弟弟。這就是評量與其欲測量的構念彼此間的關係。但是評量與評量所代理的構念

間之關係，常常在我們的語言中面貌模糊，我們常常會看到這樣的詮釋：「你這次評量表現好（若 Q），表示你的問題解決能力強（則 P）。」我們稱這樣的想法叫做推論，然則，因為我們先前已確立了評量的若 P 則 Q 之代理關係，所以若 Q 則 P 僅僅是一種「推測」的「臆論」，而不是事實，我們無法保證其存在。透過「若 P 則 Q」我們僅能保證非 Q 則非 P，也就是等價逆推論的存在「如果你不是男生（非 Q），則你必不是我弟弟（非 P）。」所以我們在評量上可以這麼說：「如果你這次評量表現不好（非 Q），那可以說是你的問題解決能力不好（非 P）」。而若 P 則 Q 不能保證若 Q 則 P，亦即不能保證「你如果是男生（Q），你就是我弟弟（P）」，是故「你評量表現好（Q），表示你問題解決能力強（P）」，從邏輯上來說是錯誤的。這也是為什麼實證派典興起之後，Karl Popper(1963)提出否證論來論證實證的意義。

二、若 P 則 Q 邏輯與生活語言之衝突

依照邏輯的原理，在筆者的測驗評量課中，當發下期中考卷時，可能要這麼說：「同學們，這次評量的結果已經通知你們了。考不好的，你自己要知道，表示老師希望你們學的或是說這次評量所評的能力，你還不太好，（非 Q 則非 P）；但是那些表現優異的，說實話，老師不知該怎麼說，因為根據評量本身的代理角色以及邏輯關係（若 P 則 Q），我沒辦法跟你保證你已學到了老師認為重要而且也放在評量過程中去評你們的那個能力（我不能保證若 Q 則 P）。很抱歉，我實在沒有充分理由說你的能力有多好。」這是一個非常符合邏輯的推論，卻是一個不容易使用的描述。

九十一學年度第一次國中基本學力測驗的國文科試題第二十六題的題目，就考驗了學生的邏輯推論是否堅強到可以推翻日常生活的經驗常識。

26. 「凡是兒童搭乘公車可享優待。雯雯可持優待票乘車，所以她是兒童。」
 以上的推論無效，因為優待的對象不只是兒童。下列何者不屬於這種無效的推論？
- (A) 偷竊是犯罪的行為。呆呆犯罪入獄必然是偷了他人的東西。
 (B) 所有的學生都會有學生證。平平是大學生，所以他沒有學生證。
 (C) 有愛心的人都樂於幫助他人。阿益常常幫助別人，他絕對是個有愛心的人。
 (D) 飲食不均衡，健康狀況必不佳。黛玉健康狀況不佳，她的飲食一定不均衡。

此題正確答案是 B，其他都是錯誤的推論。然而若不論邏輯性，每一個選項都貼近我們日常生活的思考，聽起來都很有道理。反思此邏輯，筆者在想，如何可以使它活生生地存在於我們的生活及語言世界中，而讓我們使用起來，覺得概念與經驗間的關係相當順利？也就是若 P 則 Q 無法推衍出若 Q 則 P，可是我們的生活語言卻又習慣於若 Q 則 P（若考試成績好，表示能力佳）的說法，而這樣說又不合若 P 則 Q 的邏輯。因此，如何解套？除非我們能讓 Q 對於 P 的代理關係，作一轉換？是的，P 與 Q 二者間的關係，除了以 Q 來代理 P 之外，是否還有別的可能？例如說，可不可以讓他們相等呢？如果 $P = Q$ ，則 $Q = P$ ，那麼一切問題就較單純了。

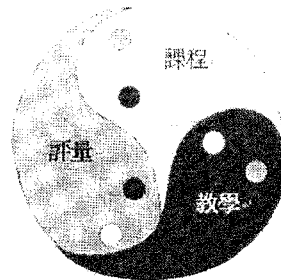
三、從同儕評量中得到啓示

在一個創造力座談會中，參與人員提及一件事：「我女兒在美國第一年讀書，是高一時候。碰到的第一個作業是做一個細胞模型。我們全家出動，用保麗龍，牙籤，小標語、顏料，做了一個惟妙惟肖的模型。交作業那天，她放學回家後，嘟著一個嘴吧。我本以為這個我們全家精心製作的模型一定得到大家的讚賞，可是她說：『媽媽，才不呢，我們同學有人用蛋糕做細胞模型，有人用 pizza 做細胞模型，發表的時候，同學都給他們最大的掌聲，發表後，就直接切了分給大家吃，同學都認為他們做的最好。』」（北區創造力座談會，民 901014）這個事件原意是要提醒在場參與人員創造力的培養，但筆者在其中看到了同儕評量的力量。同學的評鑑，讓這孩子相當在意。同儕評鑑的結果活生生存在於學生的學習生活中，甚而影響著學習興趣。

再回想上一節謝老師進行的活動，原先的設計概念是希望在學完了小老鼠與獅子這一課後，孩子們結合語文能力與對這一課內容的了解，利用戲劇編寫與表演，傳達出孩子們的文字流暢、語言表達、與創意表現。也就是如果孩子們能夠掌握本課內容，並且有良好的語言能力，則他們就能編寫出一個好劇本及清楚的表演，這是一個典型的若 P 則 Q 邏輯。但是從學生的回饋我們可知，學生大概不想那麼多，他的評論直指劇本與演戲本身。從孩子的眼光來看，這個活動本身不具有代理什麼能力的功能，這個活動就是這個活動，評量的過程也無須從這樣的表現去推論什麼樣的能力，他的評量就是發生在當下此刻，亦即就此事論此事的情境中。從他們的眼光來看評量或是測驗，往往呈現出 P 等於 Q 的關係，或是 P 與 Q 是同一件事，中間沒有「若一則」的代理關係，這也是真實評量的觀點：去除評量的代理關係後，讓評量

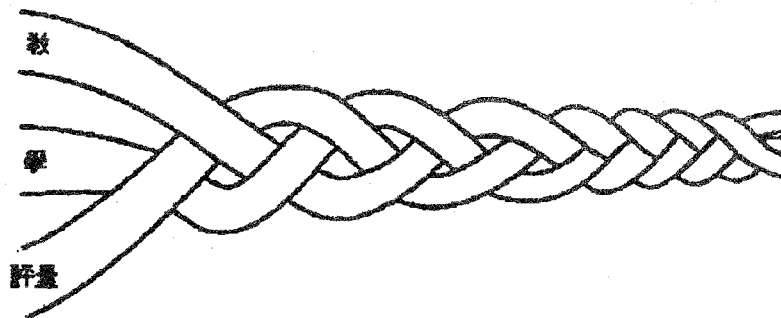
任務本身（例如戲劇表演、做細胞模型、解數學問題、進行科學研究報告）就是教學的重要目標；讓評分規準或說評量項目就是這活動在社會中會遭遇到的評鑑角度；故而我們無須從評量結果去做推論，評量者看到的或是受評者表現出來的，就是教學所要傳遞的重要能力。

在此觀點中，教與學、及評量的關係，有新的形式出現。評量活動不必然是最後監控的地位如1962 Glaser的圖形所示。她可能是類似太極的三元素圖(Gross, 1998, p. 13；王月美，民90，頁35)：



圖二 課程、教學、評量的太極關係

也可能是三股線糾結在一起相互支持的繩索(Puckett and Black,1994,p.34；侯雅齡，民89)



圖三 教、學與評量的繩索關係

四、真實評量的忠實意義

孩子們以P等於Q的簡單邏輯觀念評量他所看到的戲劇表演，是在告訴老師，這是一種對表現結果的看法。這看法，也不盡然惟孩子獨有，即使在若P則Q的古典測驗情境中，例如國中國文科紙筆測驗卷，我們看到老師也

同樣有這樣 P 等於 Q 的思考。如以下題為例：

- () 下列詞語的構成方式，何者和「閒情逸致」相同？
(A) 詩人接物 (B) 海闊天空 (C) 撲朔迷離 (D) 昏天黑地。

這樣的詞性探討加上其他林林總總的題目，就古典測驗理論而言，原意在檢核孩子們是否掌握了語文能力。但是有些認真的老師，忘記了這樣的代理關係，錯以為測驗就是教學目標，以為 P 等於 Q，測驗本身成了教學的源頭：因為基本學力測驗要這樣測，所以老師先考學生這些，幫他們準備類似的考題。這現象林世華亦曾提及（民 89），而我們回顧自己的就學歷程，也早已了然於心。如同目前國中的段考題目，例如國文考卷（以台北市大安區之一國中為例），一律模仿基本學力測驗的題型，50 個選擇題，其他題型一概沒有。其實我們常常自動去掉了測驗的代理角色，將評量與構念劃上等號，讓構念等同於評量策略。

不論是謝老師的學生，或是某一個國中老師，不論是較時髦的戲劇表演實作評量，還是較傳統的紙筆選擇題，我們不可諱言也無須避談在「若 P 則 Q」的古典測驗理論之外，一直有個「P 等於 Q」的邏輯存在於我們的生活當中。而「P 等於 Q」所說的，就是真實評量的意含。真實評量的想法便是評量不必要去代理或是表徵什麼現象，評量本身就是學習的活動。所以，當有老師疑問，我這是真實評量嗎？檢核的條件可以是：

——你認為評量本身就是你想要學生學的東西嗎？

——測驗的內容即是你認為重要的活動嗎？

——你不想去做什麼抽象概化的推論，你想直接就表現結果討論表現品質嗎？

如果以上問題都是肯定的話，那就是真實評量了。雖然戲劇表演或是國文考試所指涉與企圖全等的真實對象（構念）有所不同：國文考卷的真實對象是基本學力測驗——國文考卷等同或盡量模擬基本學力測驗；而小老鼠救獅子戲劇表演之真實對象，是戲劇表演。然而二者所運用的邏輯觀，是同一個——P 等於 Q。

要說國語考卷是一種真實評量，恐怕會讓許多努力教改的人士搖頭否認。確實，在教改浪潮中，通過基本學力測驗並不是我們最終的教育目標（雖然我們在實務上不得不承認這是很多人所認為的國中三年之教育目標），

是故我們從來不認為國文考卷會與新興真實評量概念沾上邊。當真實評量這一辭彙出現時(Wiggins, 1989)，學界與實務界皆熱情擁抱，而根據情境認知學習理論以及杜威的生活教育觀點(Bredo, 1996)，真實評量有了理論依據，指的是評量情境類同或是盡量模擬真實生活情境，而國語考卷的評量情境與真實生活截然不同，所以不是「真實」評量。不過，這樣的認定，忽略了一個基本的疑問，「真實」二字指稱的是「問題解決情境的真實」？還是指稱「評量策略與構念（標的）間關係的真實」？不同的認定，給予國語考卷不一樣的分類：「國語考卷不是真實評量，因為它很不真實於真實的問題解決情境」，或是「國語考卷是真實評量，因為它真實於基本學力測驗」。何者是真實二字的真實意義？

筆者更願意將真實界定為一種關係的忠實程度，而非真實的問題解決情境。唯有將其界定在關係上，我們才有可能以此討論所有具有此性質的評量。「構念」與「測量構念的測驗」間之若P則Q之代理關係，從邏輯實證論興起到現今，已漸漸開始鬆動，而於實務界中產生出構念與測驗的另一種關係——相等，P等於Q；於是評量的代理角色逐漸轉變成具有主體性的角色。真實的意義便是指稱評量的結果與「某個事物」劃上等號，那個事物就是最後要應用評量結果的標的情境，等號就是真實意義之所在。由此觀之，對於基本學力測驗而言，模擬考是真實的，但是對於真實生活中的問題解決能力來講，則模擬考完全不真實。換個情境，進行一個便利商店的買東西活動評量，對於真實生活的問題解決能力而言，是真實的，但是就聯考來講，它就不真實了。因此在討論真實評量之前，須先確認所等同的標的為何，無論是戲劇表演，或是基本學力測驗，凡是企圖與此標的相等的，就是朝著真實的方向行去。

如此定義，我們只需關注隨著時代變遷，標的的變動狀況，而無須擔心是否每變動一種標的情境，就要發明一個新的名詞來描述評量策略的特質。確立真實之意義存在於P與Q的相等關係，則真實評量可以歷年代之演變而不衰。隨著潮流的演變，教育目標也與時求新，二十年前可能是演算能力，十年前是問題解決能力，當今可能是創新能力，五年後可能是資訊整合能力，十年後可能是情緒溝通能力。無論評量（Q）要指稱的構念（P）為何，只要等號是存在的，二者關係是真實的，則隨著標的（構念）的演變，評量方式跟著忠實變動，真實於標的（構念）的展現情境，而非以其他評量方式代理構念，則真實評量永遠有立足之地。

五、多元評量的後現代社會特質

以上討論了評量實務中的兩種邏輯觀念與評量本身的代理與忠實兩種角色後，顯現有時我們以若P則Q的代理關係來處理評量，有時則以P等於Q的全等想法來看待評量，這就是一種多元的看法。雖然多元二字，在字面上的解讀是多種方法，多種策略，然而三十年前，多種方法多種策略已然存在，我們今日感受到的多元，異於往昔者何在？此問，也就是本文在第一節中提出的問題——實施了三十年的評量政策，方式是相同的，但不同的是什麼？今日我們談的多元評量意指什麼？多在哪裡？顯然並不只是方法的多元而已，因為三十年前的多元方法今日看之，一樣領風騷。那麼多在哪裡？

筆者的論點乃在多元指的是不同邏輯概念的同時存在，亦即測驗評量與構念的關係，從若P（構念確立）則Q（設計評量）的代理關係，到P（構念是這樣）等於Q（那就是評量）的全等關係，都為教育現場所實施與接受。不論是代理關係，或是全等關係，皆會涉及多元的評量方式（如紙筆、軼事記錄、觀察）、以及不同面向的能力（學科能力、生活能力、欣賞、團隊合作、或是Gardner的多元智慧面向），這是為什麼我們現在回看三十年前的評量概念覺得好像沒什麼改變的原因。如果以多元方法來論的話，早在公元587年時隋文帝廢除九品中正制，隋陽帝設進士科考試舉才（黃國樞，民82）開始，所採用的論述形式之弱結構題型（呂金燮，民88；Linn and Gronlund, 2000）就是目前流行的實作評量了。實作評量確實是個復一千多年以前之古風的新潮。然而，復古之後，新意何在？在今日後現代社會中，多元評量已指出一條路徑，突破過去評量的單一代理角色，賦予評量更豐富的形象，而有一番新的面貌。此面貌恰好解釋了後現代社會的特質：

1. 主體與他者的對等：在過去現代社會中，主體是欲評量的構念，而他者便是代理或接近構念以便將之表現出來的各種測量方法。若是沒有主體（沒有打算評量的構念），則評量本身即無存在之必要。一切效度之建立過程乃在保證評量作為他者，須盡到烘托支持構念的功能。然而，目前的多元評量已有後現代的味道。多元評量之興起，在說明他者（評量本身）亦有其地位與自身值得存在的意義。例如目前教育界盛行的過關評量，曾遭受批評謂：「教育現象有綜藝節目化的趨勢，過關評量像大地遊戲一樣。如此大玩一場，有達到學習的功能嗎？」雖然言之諄諄，但聽者仍是藐藐，原因很簡單，過關評量有它自己好玩的一面，其地位，不在輔助學習，或是測量某某能力，它就是要讓孩子動一動玩一玩有趣就好。整個實施過程很難釐清這是教學、還是評量、還是課外活動。甚而得到的過關卡，算是數學、國語、還是社會的成績，或不是成績，都很難講得清楚。然而一句話，它就是它自

己，一切關於邏輯推論的質疑都不成問題。過關卡就是一張過關卡，他無須變成代幣去兌換成多少的數學分數，去解釋有多好的能力品質，他可能只是在說：「我經歷了這次的活動，好過癮。」。

2. 不求共識的併置與對話：以往現代社會中，一個評量的價值定位在評量結果得以詮釋構念，以真實並正確表達構念為終極目標。我們或可稱此構念為評量所指涉的實在。而在後現代的多元評量情境中，實在是否存在，已經夠讓人爭議不休了，遑論去指稱實在是否仍有其價值性。例如整個網路所建構的虛擬實境，已沒有什麼可對應的實在，所謂虛擬，便是滑鼠與指頭可以漫遊的實境。所以，多元評量中，不同的評量結果之間，可能毋需利用所指稱的共同實在以相互溝通，因為我們前面已提及，P與Q是對等的關係，或是說Q可以取代P。至於評量結果如何描述一個人呢？靠的是不同評量結果之直接對話。例如謝老師的戲劇評量結果，與作文課的表現可相互參照，再與小朋友連絡簿上的心情敘說互相聯繫，給予一個多角度的描述。但是我們要注意，各個角度的表現結果很可能逗不出一個整體面貌，因為連絡簿上愛談天說地的孩子，可能在戲劇舞台上非常害羞，碰上作文情境時卻又辭彙不夠豐富。所以我們無法三者加權平均來說明語文能力，因為加起來的結果，反而指不出「實在」（構念）是什麼。因此有了新的邏輯關係後，評量有了新的力量，不同評量結果可以相互對話，目的在從不同的觀點了解孩子，而不在統整各方觀點得到共識。

3. 跨越界限的混合體：由於評量與構念間的關係有更大的彈性，所以評量活動也跟著有更大的空間成為她自己而施展她自己。九十年最後一屆大學聯招的數學考卷中的一個選擇題，是個好例讓我們來看看評量內容如何超越她原先被給定的科目限制：

四、數學兼哲學家伽利略於西元1632年出版《對話錄》一書觸怒教廷，後來在他70歲，接受宗教法庭審判且於該年判終身監禁，之後在獄中過世，享年78歲。出版《對話錄》一書到過世是伽利略人生最灰暗的10年。

伽利略年輕時發明十倍率的望眼鏡，並於次年發現木星的歐羅巴衛星，發明望眼鏡到出版《對話錄》算是伽利略人生中的黃金歲月，這段時間之長剛是他發現衛星時年紀的一半。根據上面的敘事，請問下列有關伽利略生平的敘述，哪些是正確的？

- (A) 出生於西元1566年 (B) 在45歲時發明10倍率的望眼鏡
(C) 在西元1610年發現歐羅巴衛星 (D) 在68歲時出版《對話錄》
(E) 於西元1644年過世

這一題所牽涉的數學能力很單純，利用未知數以及線性的加減乘除關係即可，但是學生需要處理的訊息卻龐雜而交錯，牽涉到時間年代的歷史感、關鍵人物的特質、與重要的歷史事件。學生須運用的不只是數學運算能力，更包含線索的解讀能力。如果從評量的代理關係來看這一題，顯然此題無法回應心理計量學者所強調的——「數學题目的閱讀固然重要，但是閱讀能力的強弱不應干擾所欲評量的構念」。很明顯地學生在作這一題數學時，如果不能在閱讀上先解讀線索的話，就一定受限於文字理解而不能進入數學的解題歷程；所以，這題目有嚴重的無關變項之干擾。可是如果我們換個角度／價值來審此題卻會發現他是一個比較接近學科融合的問題，他們不只在測數學，也在測閱讀，也包含了歷史素養與社會敏銳。像這樣的題目沒有回溯指稱那個所欲評量的構念，叫做數學，甚而他們就是自成一格的題目。雖然在大學聯考的結果上，我們還是把這題與其他數學題目相加，共同指涉聯考所打算測量以及大學招生所需要的數學能力。但是類似的題目可以是教學歷程中一個很好的佈題情境。老師出個題目，學生從閱讀、思考、形成想法到討論，至少需要15到20分鐘。而其有趣的部份，在題目自身，如同闖關評量，或是科學研究報告一般，學生之所以投入，乃在任務本身的吸引力，無關其他。

肆、結語：多元評量的特質

今日藉此文與大家分享多元評量的概念。什麼是多元評量？本文乃指在教與學的情境中，評量活動有以下特質：

一、邏輯觀念的多元存在

不同邏輯觀念在教育現象中，皆有存在之空間。從若P則Q的代理關係，到P等於Q的全等關係皆可能出現在教學情境中而形成重要的教學或評量活動。服膺若P則Q者，清楚解析構念的特質以設計可以代理構念的評量方式，著重評量的多元媒介，含有充分的題目、多樣化的題型，共同輻合以建構出欲評量的構念。服膺P等於Q概念者，則還給評量自身的價值，強調評量的真實地位，特別釐清評量所關心的真實是什麼？是科學研究報告？還是推薦甄試入學？還是人文情懷的涵泳？還是基本學力測驗？不同價值觀或

是不同情境下，會有各自的真實評量情境，評量策略與結果解釋永遠忠實（等同）於界定出來的情境，關注評量活動帶給孩子的意義。兩種邏輯觀點皆能在教室中找到演出的舞台。

二、評量與評量之間以及評量與構念之間，存有變化的彈性關係

在若 P 則 Q 的代理關係中，構念與評量乃是主與客的關係；而在 P 等於 Q 的關係中，構念的實在性等同評量活動，故而削弱其權威的力量，取而代之的是不同評量取向間的對話，大家都是具有主體性的主人。當然，我們也要有心理準備，不同評量相互對話的結果，可能出現有對話而無共識的現象，但仍不損多重關係的可能性。

三、學科的界限存有模糊的空間

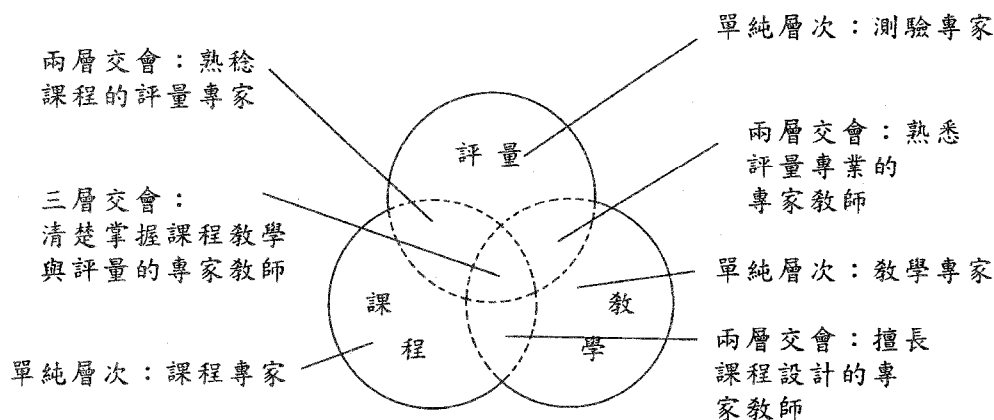
當老師在設計評量活動時，也許會先想到教學目標與學科領域的重要能力指標，對之進行分析，如同我們在分析構念的成份一樣，然後才規劃評量內容與項目，所以特定的評量活動有其清晰的學習領域由來。另外，也有老師在規劃評量時先想到一個有意思的活動，至於這個活動要指向那一個學習領域，就無法清楚劃分了。她所在乎的是孩子在這個評量任務中是否覺得有趣覺得有挑戰力而願意全力以赴。因此我們有時會看到純粹的數學問題，有時也會發現與自然結合的數學步道。在數學步道活動中，融合了語文、數學、與觀察真實情境，對於數學思考的激發歷程，著實不同於書桌前抽象的數學思考。是故評量實務，有時產生自清楚的能力指標與結構分明的學科中，然則有時亦會產生自創意活動的構想中，而此時，便沒有學科的分界了。

四、評量與教學的疆界可相互交融

現代的理性觀念一般會認為評量發生在教學之後，並進而回饋教學。但在後現代的主客角色漸趨混淆之際，有時教學活動讓人摸不清這算教學還是評量，例如學習歷程檔案，或秦代兵馬俑的主題報告。評量任務完成之同時，也是教學告一段落的時候。評量與教學同時發生，評量本身也是孩子學習的過程。

五、課程、教學、與評量交錯的不同立場豐富學習歷程

或有人給予課程、教學、評量各種關係圖如本文前面所引。而筆者的認知乃是如下圖所示：



圖四 課程、教學、與評量交錯的多元關係圖

教育界中，課程、教學、與評量有各自的舞台與支持擁護者。課程專家傾向從課程的觀點來看整個教學現象；測驗評量專家傾向從測評觀點解釋世界；至於教學專家則傾向從教學現場的活動來看待學習。愈是純粹的觀點，在此圖中的位置愈是站在單純層次的範圍。但在複雜的教室現場中，教與學的發生並不是三條路線分開進行的。有的老師站在教學與評量之融合處，教學進行的同時也思索評量活動的設計；也有老師處於課程與教學的兩層交會處，努力構思教學設計同時處理學校本位課程；國家政府大概站在課程與評量交集的地方，規劃許多能力指標來監控教學品質；至於課程、教學、與評量如何連絡以促進學習，恐怕也是很多老師在摸索的。多元的世界便是任一方圓之地，都是你我可以落足且安身立命的角落；況且，還有人尚在這些圈圈之外，提供教育現象豐富多元的思考角度。

伍、分 享

本文之審稿者提出兩個觀點於審稿表格的綜合意見中。茲透過編輯徵得

其同意後，全文錄寫如下：「分享：多元評量承襲 multi-method 的觀念，無須強調其中西理論的源頭。畢竟評量是很難擺脫其樣本的性質，而且構念與評量之間未必是令然的邏輯，有時是概率的關係。」（審稿意見，民91.11.12），筆者相當感謝像這樣分享的啟動，但遺憾不能面對面親身討論之。關於審稿者的第一個觀點，「多元評量承襲 multi-method 的觀念」，筆者另有一番感受。在民國八十四年，開放教育正熱烈時，筆者提及「多元評量」一詞，亦是將之視為「多種評量方法」的概稱。然而，在此名詞逐漸大量使用之情況下，概稱似有成為一種概念的趨勢，而若概念漸次成形，聚集多方說法與解釋之後，它便有成為論述的潛力。證諸筆者多次被詢問「多元評量的英文是什麼？」、「多元評量的理論是不是多元智慧理論？」筆者感受到現場的疑問，不僅僅是將之視為一個「統稱多種方式的名詞」而已，而是在尋找其背後之知識體系。是故有本文之誕生，企圖給予多元評量一個言談的場域。

至於審稿者的另一觀點，「構念與評量之間未必是令然的邏輯，有時是概率的關係。」牽涉到邏輯實證論與測驗評量的理論，若細究其內涵，恐需更長的篇幅與更細膩的文字來處理，筆者在此簡述之。古典測驗理論中評量與構念的關係，就是審稿者所隱含的樣本與母群的關係。評量作為樣本，其所蒐集得來的結果，為樣本統計數（例如閱讀能力85分），乃是在推論母群構念的表現（母數）。是故當利用評量結果來描述構念真正的值時，必須用到測量標準誤與信賴區間，例如「根據此次閱讀能力表現85分的結果，我們有95%的把握這孩子的真正閱讀能力在82分到88分之間。」在使用機率與測量誤差從樣本結果推論回母群時，便是邏輯實證論的證實觀點(verification) (Schlick, 1959; 引自Yu, 2002)。換句話說，證實觀點所採用的統計邏輯是當虛無假設為真時，得到此樣本統計數的可能性，此時，本文提及的若 p 則 q 之邏輯便出現了，亦即若構念的值為真，則就會得到此評量樣本結果（機率就很大）。而現實世界中的構念真值乃是不存在的，因此便反過來利用已存在的評量樣本統計數來推論之。當然，這樣的套套邏輯後續被 Popper 所提出的否證論所駁斥(Popper, 1963)。其間涉及的論證，本文暫不討論。至於證實觀點的若 p 則 q 邏輯，所衍生的，正是審稿者所想到的樣本與母群之間的概率問題。是故迴響審稿所提出的想法謂評量與構念「未必是令然的邏輯」，筆者在此進一步澄清：實則母數與樣本統計數之間的概率關係，類同量結果與構念真實能力間的關係，而此關係與若 P 則 Q 的令然邏輯乃同為邏輯實證論中的證實觀點下之不同說法。以此與審稿者以及曾被本文引發思考的讀者分享。

參考文獻

- 吳毓瑩(民84)。開放教室中開放的評量：從學習單與檢核表的省思談卷宗評量。收錄於台北師範學院主編，*開放教育理念與實務*(頁93-100)。台北：國立台北師範學院。
- 吳鐵雄、洪碧霞(民87)。實作評量問與答。*測驗與輔導*，149，3102-3103。
- 呂金燮(民88)。實作評量－理論。載於王文中吳毓瑩、呂金燮、張郁雯、張淑慧著，*教育測驗與評量--教室學習觀點*(頁169-203)。台北：五南。
- 林世華(民89)。由多元評量的觀念看傳統評量的角色與功能。*科學教育*，231，67-71。
- 侯雅齡(民89)。如何有效運用多元評量的結果。*測驗與輔導*，159，3336-3341。
- 教育部(民64)。國民中小學課程標準。台北：作者。
- 教育部(民89)。國民中小學九年一貫課程暫行綱要。台北：作者。
- 教育部(民90)。國民中小學成績評量準則。台北：作者。
- 陳文典(民89)。實作評量的理念與實施。*科學教育*，231，64-66。
- 陳榮華(民86)。魏氏兒童智力量表第三版指導手冊。台北：中國行為科學社。
- 王月美(民90)。國小小校長課程領導之個案研究---以九年一貫課程試辦國小為例。國立台北師範學院，碩士論文。
- 單文經(民87)。評介二種多元評量：真實評量與實作評量。*北縣教育*，25，46-52。
- 曾志朗(民91)。當局者迷。*科學人*，創刊號，1。
- 黃光國(民90)。社會科學的理路。台北：心理。
- 黃國樞(民82)。談中國科舉制度。*孔孟月刊*，38(8)，34-37。
- 蘇義翔(民87)。實作評量的理論與啓示。*測驗與輔導*，149，3099-3102。
- Bredo, E. (1996). Cognitivism, situated cognition, and Deweyian pragmatism. *Philosophy of Education, 1996*. Retrieved July 11, 2002 from http://www.ed.uiuc.edu/EPS/PES-Yearbook/94_docs/BREDO.HTM.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.
- Glaser, R. (1962). Psychology and instructional technology. In R. Glaser (Ed.), *Training, research and education* (pp.1-30). Pittsburgh: University of Pittsburgh Press.

- Gross, S. J. (1998). *Staying centered: Curriculum leadership in a turbulent era*. Washington, DC: ASCD.
- Habermas, J. (1983). Modernity-an incomplete project. In H. Foster (Ed.), *The anti-aesthetics, essay on postmodern cultures* (pp. 3-15). New York: The New Press.
- Linn, R. L. & Gronlund, N. E. (2000). *Measurement and assessment in teaching (8th ed.)*. Upper Saddle River, NJ: Prentice-Hall.
- Popper, K. R. (1963). Science as falsification. In *Conjectures and refutations* (pp. 33-39). London: Routledge.
- Puckett, M. B. & Black, J. K. (1994). *Authentic assessment of the young child — Celebrating development and learning*. New York: Macmillan College Publishing.
- Schlick, M. (1959). Positivism and realism. In A. J. Ayer (Ed.), *Logical positivism* (pp. 82-107). New York: Free Press.
- Wechsler, D. (1944). *The measurement of adult intelligence (3rd ed.)*. Baltimore: Williams & Wilkins.
- Wiggins, G. (1989). A true test: toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 703-713.
- Yu, C. H. (2002). *Misconceived relationships between logical positivism and quantitative research*. Retrieved November 25, 2002 from <http://seamonkey.ed.asu.edu/~alex/computer/sas/positivism.pdf>.

謝 詞：

1. 本文初稿發表於中華民國教材研究發展學會舉辦之課程與教材改革檢討與展望系列座談會（五）多元評量（民90.11.2）。感謝現場參與討論的教育實踐者，他們是：吳欣黛、呂玉瑛、范信賢、張郁雯、陶知仁、游政男、項必恭、黃敏、莊梅枝、趙曉燕、劉淑雲、歐用生、盧雪梅、蕭福生、以及謝祥宏。
2. 文稿之完成，感謝林怡呈的文獻蒐集、董玉如的文字處理、江長春的圖表掃描，以及系上王俞文、劉淑雲助教在行政上的協助。
3. 感謝國科會計畫之補助89-2413-H-152-001-F14。

（收稿日：2002/9/11；完成修改日：2002/11/26）