

教室作為評量環境的內涵 與其評量新趨勢

吳明隆／高雄市七賢國小教師 高師大教育學博士



教學性評量的目的在於：教學計畫活動要如何配合學生反應而適度改變

壹、教室作為一個評量環境

對教師而言，教室活動是一個令教師非常忙碌的地方。每天於教室中，教師均要對學生行為表現、教學活動的成功、常規管理及班級氣氛等作有效立即的「決定」(make decisions)。

而作決定所需的資訊即是由「教室評量」(classroom assessment)所提供的訊息，這就是評量資訊的搜集。因而對教師而言，教室就是一個評量環境，有效能的教師必須隨時、隨地為作有效決定而對學生一切學習

活動、行為反應、實作表現、情感知覺及社會技能等加以評量。評量雖然被認為是種「技術性活動(technical activity)」，但教師必須重視評量的倫理考量(ethical concerns)，因為教師的評量及所決定之事，對於學生自我知覺、人格發展、學習活動及潛能發揮等均有重要影響。

一、評量的目的

教室評量運用的主要目的有以下六項：(Airasian, 1994)

1. 診斷學生的問題：從評量中得知學生在教室中之學習情形、情緒及

社會性問題所在，以便確認問題，發掘問題根源，進而選擇相關有效的治療或輔導活動策略。

2. 判斷學生的成就表現，以評定學生學業進步情形。

3. 從所搜集的評量資訊中，作為學生的回饋與誘因，以激發學生實作行為改進的參考。

4. 作為安置學生決定的參考，如作為學科分組教學或合作學習分組等的依據，以提昇學生的學習成效。

5. 計畫與實施班級教學決策的依據，以使「教學計畫決定」及「教學過程決定」能適合學生學習狀況，以確保教學活動能密切配合學生的能力。

6. 建立及維持班級的社會性平衡機制，使班級成為正向積極的學習環境，學生能服從命令、遵守常規紀律、培養合作學習及尊敬他人等優良習慣。

Airasian(1989)則綜合相關評量功能，將班級評量的主要類型畫分為三種：

(一)、官方性評量(official assessment)

評量主要關注於教師作為學校科層體制成員之正式職責的行為表現，如評定學生成績、分組、評量學生成

就、解釋標準化測驗成績、親師會議、學生特殊需求安置的確認、學生升級推薦等。此種評量內涵是教師作為一位學校正式職員所應展現的義務行為。

(二)、教學性評量(instructional assessment)

教學性評量主要目的在於計畫及傳遞教學活動。評量所作的決定包含要教什麼內容？如何教？何時教？使用何種教材？課程如何發展？教學計畫活動要如何配合學生反應而適度改變？

(三)、估計性評量(sizing-up assessment)

主要被使用於學年開始時，根據評量資訊決定學生社會、學業及行為特性，以允許教師建立及維持有效的班級社會。使用估計性評量時，教師可快速獲得有關學生人格概略特徵，以有效促進班級教學、溝通及合作學習等。

三種教室評量類型的比較如表一所示。

二、搜集評量資訊的方法

班級每天的教學活動中，教師均要對其學生、教學活動及班級氣氛作決定，這些決定及其決定所根據的內容，是促使班級教學成功有效關鍵。

表一：三種教室評量類型的比較

	估計性評量	教學性評量	官方性評量
目的	迅速提供教師有關學生特性之了解及實用知識	計畫教學活動及監控教學進步情形	完成教科層體制功能，如成績評定、分組及安置
實施時間	開學第一個星期或第二個星期實施	學年期間每天均可實施	學年間固定安排的時間實施
資料搜集方法	大部份屬非正式的觀察	計畫方面以正式觀察及學生報告為主；監控方面以非正式觀察為主	正式測驗、論文、報告、考試及作業
資料搜集型態	認知的、情感的、及心理動作技能	大部份屬認知的及情感的領域	主要是認知的領域
記錄保存	資訊保持在教師的心智中，少數被書寫記錄下來	寫成課堂計畫、監控資訊則未被記錄下來	正式記錄被保持在教師成績登記簿或學校檔案內

資料來源：Airasian，1994，p.10



評量就是搜集、綜合及解釋資訊的一般歷程、以協助教師於班級內作有效決定的參考。搜集評量資訊方法主要有二種：(Airasian, 1994)

(一)、紙筆技術 (paper-and-pencil techniques)

紙筆測驗的評量方式是指由學生寫下他們對問題的反應。當學生參加測驗、作完書寫家庭作業、交出一份書寫報告、畫一張圖或完成一份工作，他們均是提供紙筆式證物給老師，老師據此搜集評量資訊，以了解學生進步、學習情形及洞悉學生不了解之處。至於搜集紙筆資料技術的兩個常用形式：

1. 提供 (supply)：讓學生得以根據問題建構自己反應 (答案)，如簡答或填充題方式。其它如讀書報告、旅行記事、教室主題也是紙筆技巧中提供的另一形式運用。

2. 選擇 (select)：讓學生從給予選項中選擇正確的答案，如選擇題、是非題及配對題等。

(二)、觀察 (observation)

觀察是班級教師使用搜集評量資訊的第二種主要方法。觀察包括對學生活動表現的觀看、聽學生說話、讀書及其討論的事情。觀察活動中教師主要是藉由視覺及聽覺的觀察、以了

解學生行為表現。班級中，教師瞬間的作出明確決定主要是根據教師觀察所獲得的大部份資訊，而非是靠紙筆型式之程序所獲得的資訊。

紙筆和觀察技術各有其優缺點，其間是互補關係。班級決定如果沒有任何有關學生外顯行為、反應、實作表現及互動資訊，則教師決定無法配合實際教育情境；相對的，如果也缺少學生紙筆方面表現的相關資訊，則教師決定欠缺客觀。要建構教室中有意義的評量方式，就同時需要運用這二種評量技術。

此外，教師作決定時也可以從學生以前老師、學校護士及學生父母等人之處獲取有用資訊；不夠，從這些地方所獲得的資訊限制很多，其客觀性、時效性、代表性應一一考量，教師依此資訊作決定時要妥為注意與謹慎。

三、評量信效度的問題

(一)、效度 (validity)

一個良好的評量資訊最重要的特徵是能促進教師作正確決定的能力 (Messick, 1989)。沒有效度，評量資訊就無法導引至正確決定。效度評量時主要考量到三個方面：

1. 效度所關注的最普遍問題是「評量資訊能幫忙教師作出適當的決定

到什麼程度？」

2.效度指的是從評量資訊中所作的決定，而非是評量方法本身。因而以「這個評量資訊是有效的」的表達方式是不適當的。效度有其適用對象，只要能說出此評量資訊在決定什麼及適合什麼群體即是有效的。評量資訊對一個決定或一組學生是有效的，但對其它決定或群體有效與否，並無必然關係。

3.效度只是程度差別，並非全有全無的問題。效度評量常劃分成三個範疇：高效度、中效度及低效度。

(二)、信度(reliability)

良好評量資訊的第二個特徵就是它的一致性。如果不同時間及環境中，對個人或班級的評量結果大同小異，則此評量就具有好的信度。信度評量時主要考量到四個內涵：

1.信度指的是評量資訊的穩定性或一致性，所強調的是「我所搜集的評量資訊與學生行為間的一致性程度如何？可否代表學生行為？」

2.信度並不是關注於搜集之評量資訊的適當性，只要其焦點集中於評量資訊的一致性、穩定性或代表性。評量資訊適當性為何乃評量效度的話題。

3.信度也是程度差別，並非全有

全無的問題。以程度高低表示也劃分成三個範疇：高信度、中信度、低信度。

4.信度是效度的必要條件，但並非是充分條件。一個不一致、無代表性的評量資訊，無法有效提供正確決策所需之有用訊息。

「有效度的評量必定是可靠的，但可靠的評量不一定是有效的」。如果評量資料間歧異性很大，則無法作為教師有效決策的參考。

四、教室評量的倫理話題

(ethical issues of assessment)

評量不只是一項「技術性活動」(technical activity)，而且也是種「人性化的活動」(human activity)。多數人，包含學生、父母、教師、教練、學院入學顧問及各雇主等，除了非常重視評量結果外，更會顯著地受評量結果的影響。

教學是一種兼顧學識及道德的專業，與其它專業人員一樣，其知識、行動及判斷等會於多方面中影響著當事人。由於教師專業特殊性，教師更要受倫理行為標準 (ethical standards of conduct) 的約束，此標準可協助教師保護學生、喚起教師對其決定影響的重視，並負起權威行使的責任。教育領域內倫理規範遵守

顯得格外重要。因為教育不同於其它專業，學生對其是否上學並無選擇權，倫理標準指引著所有教育活動均要把學生視為「一個個體」(an individual)，使學生免於遭受身體或心理上的凌虐、並使學生受到基本尊重，能獲得真實知識，並儘可能為所有學生提供最好的教學服務 (Clark, 1990; Fenstermacher, 1990; Strike & Soltis, 1991)。

教室中的評量應用更應重視倫理標準。班級教師利用其特殊身份地位，獲得有關學生學業、個人、社會及其家庭背景的詳細資料。教師如果未依據評量資訊作客觀決策、或獲取資訊不夠周延、或未考慮學生個人自尊等，則評量所作決策將對學生產生負面的影響。因而教師對其搜集與使用的評量資訊要負完全之責，在作重要決定時，應先確認資訊的信效度及其對學生的可能影響，對於所搜集到的評量資訊，教師有義務確保它的隱私性，確認決定時的限制所在，而不能使用評量資訊以貶低或嘲弄學生個人。(Airasian, 1994)在倫理標準規範訂定上，「全美教育協會」(National Education Association)所定之教育倫理標準，可作為教師應用評量資訊的參考：

1. 在學習歷程中，不能不當地區限制學生之獨立行動。

2. 不能不當地區拒絕學生接觸多元不同的觀點。

3. 對學生的學習進展，不能故意壓制或曲解學科內容適切性。

4. 教師要付出相當職責，以免學生學習、健康及安全等方面受到情境條件的傷害。

5. 不能故意地使學生表現出困難難堪或受輕蔑之感。

6. 不能在種族、膚色、教條、性別、政治或宗教信仰、家庭、社會、文化背景或有關性愛取向基礎上，不正當地

(1). 排除任一學生參與課程學習機會。

(2). 否定任何學生的信念。

(3). 給予任一學生特別優勢或特權。

7. 不能使用與學生特殊的專業關係而獲取私人利益。

8. 除了強制的專業目的或法律所需，在專業服務程序過程中不能任意揭發已獲得之學生資訊。

教室是個複雜的環境，在許多方面均需要老師適時作出決定。在此環境中，不期望教師所作的每項決定均是正確無誤的，這是一種不切實際的

標準，特別是在複雜、多變、不確定教室情境中。但是，教師應被期望且有道德義務提供可辯解的評量資訊來支持教室決定及行動，因而在教室情境中，教師的一舉一動對學生均有舉足輕重的影響。

貳、教室評量新趨勢－實作測驗－直接能力的測量

實作評量(performance assessment)使用的是一種直接的學習測量。健美、跳水、體操領域就是直接實作判斷評量的實例，而以綜合判斷成績來評定比賽者名次。教師能以實作測驗方法評量學生複雜的認知歷程，也可評量科學、社會、數學等學科之態度及社會技巧。為達實作評量目的，教師必須營造適宜情境，以觀察及評定學生學習，如分析、問題解決、實驗、做決定、測驗、與他人合作、口頭講述、作品創作等。這些情境是模擬真實的世界，這些情境可能是一項工作，一種社區活動，或不同形式的進階訓練等。(Borich, 1996)

一、實作評量的功能

實作測驗允許教師觀察學生成就、心智習慣、工作方法、真實世界的價值行為，這是傳統測驗無法測得的。實作測驗也可說是教師觀察及評

量學生完成複雜活動的能力。實作評量不同於傳統教室評量中紙筆測驗評量，其主要特性有：

(一)、實作測驗兼顧過程和作品

實作測驗可同時評量過程和作品二者。如位於Winnipeg之達爾文學校評量學生閱讀過程時，同時兼顧：朗讀時正確讀字的百分比、讀出故事中具代表性而有意義的句子、讀完後能以自己話語說出句子內涵的多寡等項的表現，以綜合判定學生成績。

(二)、實作測驗可融入於課堂活動，教師可將實作評量變為課堂的一部份。事實上，許多實作測驗提倡者常堅持的理念是「實作測驗就是一種有效的教學活動」(Shavelson & Baxter, 1992)。根據此觀點，建構一個適宜的實作測驗，就如同評量一樣，可增進學生的學習經驗。學生在如何表現實作行為中，評量型態可提供立即的回饋，促進教學效果，加強教學與評量活動間的密切配合，使教學活動導向高階(higher-order)及較真實性的行為。

(三)、實作測驗可同時評量情感與社會技巧(affect and social skills)

使用實作測驗不僅可評量高階認知技巧，而且也可評量非認知性結



果，如自我管理、群體合作能力、社會知覺等(Redding, 1992)。情感學習領域所反映出來的是教師要縝密規劃那些需專門實作表現的複雜工作，包括資訊回憶、形成概念、概括推廣及問題解決能力等，其中也包含了心智習慣、態度及社會技能。

如果實作測驗建構得宜，則實作測驗比紙筆測驗更能評量出複雜思考歷程、態度及社會技能；相反的，如果建置得不恰當，則會如傳統測驗一樣，在真實成績評定、信效度方面受人質疑。學者Stiggins(1994)認為實作評量也應如成就評量一樣，重視評量及計分的標準。

二、實作評量發展的步驟

教師發展實作評量以測驗學生，其主要實施步驟如下：

(一)、決定要測驗什麼？ (deciding what to test)

實作評量實施的首要就是決定要測驗什麼？在此就是要訂定行為表現目標目錄，內含知識、技能、態度及其結果指標，以導引教學活動實施。教學決定要教什麼，所要考量的問題有三：

(1).學習者對學科教材所應了解的基本知識及內容是什麼？

(2).當學習者要使用知識或內容

時，其所需要的智識技能又是什麼？

(3).學習者有效地展現知識或內容時，重要的是那種心智習慣或態度？

1.認知領域的實作目標

實作評量的設計通常用以解決下面問題，以導引目標的選擇：

(1).在紙筆測驗中，我們所忽略的是何種基本工作、成就或有價值潛能發展活動？

(2).從何種活動方式的評量中，才能得知學習者有如歷史學家、作家、科學家或數學家等才華，這是傳統測驗所無法測量出來的。

典型的實作技巧由下面二個範疇中引伸：

(1).有關獲取資訊的技巧；

(2).有關組織及使用資訊的技巧。

2.情感及社會領域目標

實作評量不僅需要相關課程以教導思考技巧，而且要發展積極人格與心智習慣(habit of mind)。心智習慣行為包括建設性批評、挫折容忍力、理性尊敬他人、賞識過去重要而有意義之事。實作評量是一種有效的評鑑媒介，可用以評量學習者正向積極的學習態度、心智習慣及社會技巧（如合作、分享、協商）。

下面所列者主要在回答實作目標

範圍內，情感及社會技能領域部份：

(1).課堂中，要培養群體內成功之個人，需要培養學生具有何種人格、態度及重要特質？

(2).一位優秀科學家、作家、採訪記者、歷史學家、數學家、音樂家等所需的心智特徵或性格特質是什麼？

(3).教師應接納何種實作行為表現，方能確知學生有可能展現或發展專業人員的行為特質呢？

(4).作為一位新聞記者、氣象播報員、公園巡邏警察、歷史學家、經濟學家、數學家等需要何種社會技能，才能有效成功的與人和睦相處呢？

(5).要以何種有力論點來說服學生父母，讓學生有機會實作表現，發展以上相關的社會技能？

(二)、設計評量內容(designing the assessment context)

實作工作表現如辯論、模仿嘗試、重新建構歷史、科學實驗、專業工作者角色扮演等，不管情景為何，實作工作所呈現給學習者是一種真實的挑戰。

在實作內容設計範圍方面，Wiggins(1992)提出以下三點作為教師參考：

1. 想要成為似數學家、歷史學家、科學家、藝術家、作家等專業人員於真實世界的生活工作表現情形，要學習什麼內容？

2. 學校教學活動時，需要採用何種方案及工作方式，才可能使學生表現如上述專業人員之行爲呢？

3. 教室中，要學生能創造知識，其所需的專業技能、心智習慣或角色扮演為何？

而在實作工作指標標準則定義原則方面，應注意以下五點 (Resnick & Resnick, 1991; Wiggins, 1992)：

1. 清楚明確的描述工作熟練時所應達到的行爲。

如果工作複雜性很高，學習者不需要懷疑工作是否可以完成，或再提供其所需要的；相對的，學習者所應做的是反複思考、努力完成工作，這就是一種學習過程行為表現。

2. 工作應是有效行為實作表現，使得教師可以歸納學習者知識、思考能力及態度。

教師所挑選的工作要具有適當的複雜性，而在細節說明上要十分詳盡，使教師足以轉化與概括學生表現而獲取結論，如能夠以八至十個重要的實作任務代表完整課程單元，則教師易於評量出希望學生表現的行為結



果。

3.工作要有相當複雜性，得以實施多元的評量 (multimodal assessment)。

傳統評量主要根據書面文字，而實作測驗則藉由不同形式評量以得知學習者學習過程。多元評量比傳統測驗之選擇方式費時，但對於學習者成就而言，則可測出其它評量所無法得知的資訊。

4.工作儘可能有多種解決方法，每種方法均需要實作表現才能獲致效益。

實作測驗並非是種反複練習的方式，它應含蓋較多的簡單工作，這些簡單工作均各有一種解決方法。Resnick(1987)就明確指出：實作測驗是一種沒有固定程序的評量方式，其行動途徑事先並不十分明確；也是一種複雜歷程，不能以一種特殊觀點來發現答案，它包含了不同的判斷及解釋成份。

5.工作需要「自我管理學習」(self-regulated learning) 的功能

實作測驗需要考量個體心智發展，重視評量過程中學習個體的毅力及決心，要求學習者使用認知策略來解決問題。

(三)、確定實作成績計分準則

(specifying the scoring rubrics)

實作評量限制之一就是要花很多時間以評定觀察之行爲。是非、選擇、填充等傳統測驗比實作主題、卷宗及實作表現的成績評定方便很多。此外，評量學生過程之努力、參與情形及態度狀況也甚爲不易。爲克服實作評量計分方式的不便性，可藉由周詳計分系統的規劃設計予以達成，此即爲行爲準則。發展實作測驗的準則所應考量的是：

1.發展完成作品之不同評定準則

學習者於實作測驗中，行爲表現可能有四種型式：

(1).作品：詩、論文、圖、表、展覽、畫展及地圖等。

(2).複雜的認知過程：獲取資訊、組織資訊及使用資訊的技能等。

(3).可觀察到的行爲：身體運動如跳舞、體操或打字、口語表達、使用專門設備如顯微鏡、依一定程序步驟解剖青蛙、等分三角形等。

(4).態度及社會技能：如心智習慣、群體工作及認知技能等。

發展之不同的實作評定準則，經由訓練而有效處理實作成績之計分問題，可提高實作評量計分之信度 (Stiggins, 1994; Sax, 1989)。

2.配合要測量的行爲選擇最合適

的計分系統。

實作測驗的計分方法主要有三種可資運用：檢核表(checklists)、評定量表(rating scales)及整體計分法(holistic scoring)。每種方法均各有其優缺點、限制及適用評定之作品、認知過程、行為表現、態度及社會技能等。

(1). 檢核表

檢核表以條列式目錄方式，包含行為特徵、行為屬性。最適合複雜行為之評定，因以檢核表方式呈現，可將複雜行為分割為一系列、具體明確的實作行為，如解剖青蛙、等分三角形、錄音等活動行為。檢核表計分可以「是否」達到行為指標準則表示，亦可用通過/不通過、+1、-1及0數字符號表示。行為條列中+1表示正向或達到之行為；-1表示未達到或錯誤表現行為；0表示未曾觀察到的行為。

(2). 評定量表

評定量表並非以是／否、通過／不通過來表示實作行為，而是以事先編排之數字來表示。大部份數字型評定量表使用分析式之計分技巧(analytical scoring technique)，稱之為「基本特質計分法」(primary trait scoring)(Sax, 1989)。此種型式就是測驗發展者首先先確認所要觀

察之作品、過程、實作表現之顯著特性、最重要或主要特徵為何？然後，發展者以事先安排之數字(通常為1至5)來表示每個行為特徵之實作表現的程度。使用「基本特質計分法」要考量到二個因素：

甲、可顯示屬性行為最高程度的最重要特徵是什麼？

乙、獲致低分時，確認實作行為主要錯誤在那裏？

評定量表的優點就是其與實作表現間有明顯的相關者，均可具體的計分，不會忽略重要的行為屬性。

(3). 整體計分法

為求實作品質表現，不以個別點數表示實作表現行為程度，而以一整體單一的成績表示，如論文、學期報告、舞蹈、音樂創作之藝術表現行為等，主要根據作品或行為品質給予A到F(或優到丁)不同的等級。

3. 三種計分系統的統合應用

好的實作測驗是學習者可藉由不同主要行為特質來展現學習成就情形，如合作、研究或傳遞等行為表現方式。因而要有效結合檢核表、評定量表及整體計分方法，以達全面而完整的評量。

如上所述，三種計分系統方法均有其優缺點，其間比較可從下列五項

得知：

(1). 建構容易度情形：指建構重要實作行為條目或及格與否之表現屬性所花時間。

(2). 計分效率性：指評定不同種類實作表現及評定整體分數所花之時間。

(3). 信度：指二個獨立評定者對於同一實作行為給予相似成績，或相同評定者於二次分開場合給予之成績差異性，差異性愈小，信度愈高。

(4). 自我防衛性：指成績評定後，向學生及父母解釋成績時，受到他們質疑反問時，自我辯護程度。

(5). 回饋品質：指計分系統對於實作表現優缺點之資訊，呈現給學習者及父母知悉的程度。

(四)、確認測驗限制(specifying testing constraints)

在實際學習活動限制上，實作測驗所要考量的「時間是否足夠」的問題？是否允許學習者更正錯誤？學習者可否參考相關資料或請求他人幫忙。教師在回答這些問題時，要從實際測驗情境著手，Wiggins(1992)認為教師可從下面六點來衡量測驗活動的限制：

1. 時間：學習者準備、重新思考、修正及完成測驗需花費多少時間？

2. 參考材料：當學習者在測驗時，是否要查閱字典、參考書或筆記等？

3. 他人：當學習者參加測驗或完

表二 依據測量指標三種計分系統的比較

	建構容易度	計分效率性	信度	自我防衛性	回饋品質	實作行為最適合於
檢核表	低	中	高	高	高	程序
評定量表	中	中	中	中	中	態度、作品及社會技能
整體計分法	高	高	低	低	低	作品、過程

資料來源: Borich, 1996, p.656

成主題時，可否請求同儕、教師或專家的協助幫忙？

4.設備：學習者可否使用電腦、計算機等工具，以協助其解決問題？

5.工作先備知識：參加測驗時，學習者事先應具備多少相關資訊？

6.計分準則：是否讓學生知道教師對其實作評量所打的成績標準？

Wiggins(1992)進一步建議教師在實施實作評量時，除考量以上因素外，可採用「真實性測驗」(authenticity test)。真實性測驗主要在於回答下面問題：

1.何種限制下，才能於真實世界中使實作者能重新身歷其境與表現機會？

2.何種限制下，才能使初學實作者及生產者有較佳的表現？

3.以上六個因素中，應顧及到幾個因素才是最合適與最真實的情境限制？

三、卷宗評量

(portfolio assesassessment)

根據學者Paulson及Paulson(1991)的定義，卷宗就是一個「述說一個故事」。故事主題依據學生所回答的問題性質而定：教師教學活動期間內，我已經學會什麼？我如何付諸行動？某些卷宗代表著學生自己選擇的

作品，如劇本、樂譜、雕刻、錄影帶、研究報告、故事、模型及照片等，這代表著學習者從學習中建構自己有意義的事物。某些卷宗是教師事先組織的，包含要完成的具體作品或主題等。卷宗正確內涵應由學生自行決定，它是否由教師事先組織編排或由學生自行決定判斷，所要考量的因素有：

1.選擇要放入卷宗的樣品，其所依據準則為何？

2.當個別作品被收集於卷宗時，就立即評量；還是等到學習結束，所有作品均被搜集於卷宗時才開始評量；或以上二者方式均有？

3.在某個學習時段內，學生的成長、進步與改善情形如何？

4.卷宗評量中，同儕、父母、其它教師及學生自己將扮演什麼角色？

Shavelson、Gao及Baxter(1991)建議，卷宗評量時，參考學習者不同主題的作品或工作，才能使卷宗評量更具信度。因而卷宗的建構與評量通常需要間隔一段時間，可見卷宗評量的性質：

1.學生實作行為所代表的是經一段教學過程後，一種對學習累積結果的評定方式。

2.每個細目的評量準則要事先確



認。

3.每個行為準則皆評定完後，就要提供整體評定量表，以評定學習者等級或對其實作表現加以計分。

四、實作測驗與學習成績單

實作測驗成績可與班級其它成績相互結合。實作表現成績轉化方式有以下幾種：

1.以百分位數評定小考、測驗、家庭作業及實作評量的成績，以總分及平均數代表，必要時，得以加權計分方式合計總分。

2.第二個方法是以「總點數的百分比」表示，其步驟如：

(1).確認計分系統的組成成份及每個實作行為成份的加權值。如六個星期的成績評分計劃的實作行為成份及其加權值為：

實作行為成份	加權值(%)
家庭作業	15
目標測驗	20
實作測驗	20
卷 宗	20
課堂作業	15
筆 記	10

	100

(2).在成績記錄簿記載實際獲得之點數。

(3).計算每個實作行為成份點數與其行為表現總點數的百分比，如下所示。

實作行為成份	學習者甲	學習者乙
家庭作業	50/70=71%	55/70=79%
目標測驗	45/60=75%	35/60=58%
實作測驗	33/40=83%	39/40=98%
卷 宗	18/20=90%	15/20=75%
課堂作業	39/50=78%	37/50=74%
筆 記	5/10=50%	8/10=80%

(4).將每個實作行為成份的加權值乘以行為表現點數的百分比。

實作行為成份	學習者甲	學習者乙
家庭作業	71x.15=10.6	79x.15=11.8
目標測驗	75x.20=15	58x.20=11.6
實作測驗	83x.20=16.6	98x.20=19.6
卷 宗	90x.20=18	75x.20=15
課堂作業	78x.15=11.7	74x.15=11
筆 記	50x.10= 5	80x.10= 8

總 分	76.9	77.1

(5). 合計記錄六個星期成績總分，轉化為學校規定之成績表示法，如等第或字母表示法，如得分在90分以上為A或優、得分在80分以上，90分以下為B或甲等。

參、結語

教室作為一個評量環境，就是教師要從所搜集的評量資訊中，作出有效的決定，以達官方式、教學式及估計式的評量目的。評量資訊要重視其信效度，方能使所搜集的資訊內涵能真正反應學生真實行為及表現，因而需應用多元方式的評量技術。為獲致評量資訊的真實性、適當性與一致性，須重視評量技術活動的運用，但教師評量的對象是個活生生的個體——學生，人性的尊重與關懷顯得更為重要，要使教室成為一個成功、有效的評量環境，評量資訊倫理道德的遵守行為，對教師而言應置於教室評量的首要。

在教育改革聲浪中，多元評量、動態評量成為新教室評量趨勢，尤其是實作評量，由於其同時能兼顧學生學習過程及學習結果，不僅可評量學生傳統認知領域，亦可評量學習者的情感及社會技能，其實施可同時融入於課堂活動中，因而在教室評量環

境，已變成另類評量主(alternative assessment)。評量只是協助教師搜集學生資訊的一種手段，不同教室評量方式均有其優缺點及限制所在，為有效、完整、適當的獲取學生學習行為等相關資訊，教師不能只限於用一種評量技術，而應配合評量目的與教學所需，採取多種方式，統合彈性運用，方能達到教室評量的真正目的。

參考書目

Airasian, P. W. (1994). Classroom assessment (2nd ed.). New York: McGraw-Hill, Inc.

Airasian, P. W. (1989). Classroom assessment and educational improvement. In L.W. Anderson (ed.), The Effective Teacher (pp.333-342). New York: Random House.

Borich, G. D. (1996). Effective Teaching Methods (3rd ed.). Englewood Cliffs, NJ: Merrill /Prentice Hall.

Clark, C. M. (1990). The teacher and the taught: Moral transactions in the classroom. In J. Goodloa., R. Soder, &



K. Sirotnik (eds.), The Moral Dimensions of Teaching(pp.251-265) . San Francisco:Jossey-Bass.

Fenstermacher, G. D.(1990). Some moral considerations on teaching as a profession. In J. Goodloa., R. Soder, & K. Sirotnik (eds.), The Moral Dimensions of Teaching (pp.130-151) . San Francisco:Jossey-Bass.

Messick, S. (1989). Validity. In R. L. Linn (ed.), Educational Measurement (3rded.)(pp.13103). New York: American Council on Education/Macmillan.

Redding, N. (1992) . Assessing the big outcomes. Educational Leadership, 49(8), 49-53.

Resnick, L. B.(1987). Education and Learning to Think.Washington, DC:National Academy Press.

Resnick, L. B., & Resnick, D. P. (1991). Assessing the thinking curriculum: New tools

for education reform.In B. R. Gifford,& M.C.O Connor(Eds.). Future Assessments: Changing Views of Aptitude, Achievement and Instruction.Boston: Kluwer.

Sax, G. (1989). Principles of Educational and Psychological Measurement and Evaluation (3rded.). Belmont, CA: Wadsworth.

Shavelson, R. J., & Baxter, G. P. (1992).What we've learned about assessing hands-on science.Educational Leadership, 49(8),20-25.

Shavelson,R.J.,Gao,X.,& Baxter,G. (1991). Design Theory and Psychometrics for Complex Performance Assessment.Los Angeles:UCLA Center for Research on Evaluation, Standards Evaluation, Standards and Student Testing.

Strike, K, & Soltis, J. (1991). The Ethics of Teaching. New York:Teachers College Press.