

科技教育教學評量問題之探討

*李博宏、**王薰巧

*岡山國中、**大東國小

壹、前言

在教育中，評量的用意在於檢視教學目標的達成。教學包含教師的「教」和學生的「學」。過去只要求「教學正常化」，但現在更強調「教學卓越化」。換言之，希望在教師方面，能達到「有效的教學」(effective teaching)；在學生方面，能達到「有意義的學習」(meaningful learning)。所謂「有意義的學習」，係指兒童及青少年學習的內容與方式力求與其日常生活情境貼近且密切結合，同時有所謂的"Authentic Assessment" (真切性評量)。此種評量特別注重在真實的或模擬的情境中進行考查，強調評量的內容與方式應配合兒童當前的生活經驗，使其覺得自然親切，而不致格格不入，且能充分反映其真實的行為。

由於教學決策常因教育環境和狀況以及教育對象之不同而異，因此必須隨時加以修正，而修正的依據，則多半來自評量所提供的具體資料和判斷，故評量實為任何教育計畫實施過程中相當重要且不可或缺的一環，在教學中十分重要；在科學教育評量中，教學評量也佔了極其重要的份量，與學生、教師、家長和社會有密切關係。(鄭湧涇，2002)。

李隆盛(1992)，在美國工藝／科技教育研究文獻回顧與啓示一文中，將各種文獻分為十三個類別，談到其中第九項「評鑑」的看法如下：

很少研究著眼於評鑑學生的學習和教師的知識。

而後李隆盛(1995)針對中美工藝／科技教育研究比較後指出：

國內在七十六年至八十二年間，共有廿六篇研究文獻發表，但文獻內容並未論及工藝／科技教育的評量。

蔣秋萍(民89)所作的學位論文「國民中學生活科技學習評量之研究」，是自工藝／科技教育相關研究所成立以來，第一本針對工藝／科技教學評量所作的碩士論文，最近則有劉桂秀(民92)的碩士論文「國中生

活科技課程實施多元評量之行動研究」，探討生活科技課程實施多元評量的困境，並提出可供參考的解決方案，到目前為止，有關生活科技教學評量之研究實在寥寥無幾。有鑒於科技教學評量相關研究、論述的缺乏，因此，作者依蒐集的文獻還有實務上的觀察所得，探討目前存在於科技教育評量中的問題，並提出建議，茲申述於下。

貳、教學評量

教育的歷程就是評量－教學－再評量－再教學的循環歷程。與評量有關的名詞有以下幾種（胡永崇，1995）：

- (1)測驗：瞭解學生在較為結構性的情境中的行為反應。測驗工具則分為正式和非正式測驗。
- (2)測量：根據某一既定量尺，對於兒童的某一特質加以測定。
- (3)診斷：確認某一病症或障礙狀況的原因，並根據病因做出適當的處理和處方。
- (4)鑑定：根據法定標準，判定兒童接受特殊教育服務的法定資格。
- (5)評鑑：根據某一標準，對兒童身心狀況或教育狀況作價值判斷的歷程。
- (6)評量：以各種相關方法蒐集教育有關的訊息，以作為教育和教學決策之用。

在評量的過程中，至少應該包括量別、獲取和提供可供教育或教學決策的資料、意見和價值觀。所謂教學決策是指任何在教學中，所做的大小決定，例如：課程設計和教學方法的選擇等。前述之「資料」並不只是指評量所得的成績等第而已，它至少還應包括事實描述、記錄、報告等；「意見」則是指看法、思想和觀點，而「價值觀」則包括了對某一事實、做法的批評、價值評估等。

針對學科的學習，現在有所謂的課程本位能力測驗（Curriculum-based Competency testing）。每一個科目或課程，都有其教材大綱及教學重點和教學目標，而教學目標又分為情意、認知、技能三方面。因此，在評量時，課程本位能力測驗應考慮針對該科學習的活動，包括教材內容、行為目標、學習情境、以及學生個人的背景能力等各方面；亦即課程設計、課程

實施、學習情境、個人等因素都應納入考慮。

此外，尚有檔案評量（Portfolio Assessment），係指在學生學習過程中彙集相關的各項資料，針對評量的需要，找出一些切要的項目作成適當的組合，進行成套的評量，而非零碎個別項目的考核，具有持續累積的評量效用。由於評量不只限於靜態的作品，還要評量作品產生的過程以及作品的優劣，是一種整體系統性的評量，因而又連帶出現一種工作取樣系統（Work Sampling System），亦即從工作項目中找出重要的樣本，讓學生操作反應，再進行觀察和記錄。

Guba 和 Lincoln 將評量的演進分為四階段：第一代的評量取向為「測量」（measurement）；第二代的評量取向為「描述」（description）；第三代的評量取向為「判斷」（judgement）；第四代的評量取向則是一種變通的方式（alternative approach）－建構回應評量（responsive constructivist evaluation）。現今教育領域的評量即處於第四代評量，趨向於採取異於傳統的變通方式，學習者才是整個學習歷程的主角，評量者作為一個主觀的夥伴，而評量本身不是目的，而是一個獲得資料的手段，這些資料包括質化與量化的，以了解學習者所學習到的事實和價值，並反映出學習的背景脈絡。（蔣秋萍，1999）。新評量觀與從前的評量有兩個不同想法：（柯啓瑤，2002）

- 1.評量不是指導的終點而是出發點：不是對於教學結果的評量，而是為指導教學的評量、為活化教學指導的評量。因此，評量的意義在於改進教學，促使指導方法有效化。
- 2.注意到情意面的評量：像學習意願、關心、態度等情意面的評量是不適合使用紙筆測驗的。能用紙筆測驗的評量，大致以知識、理解或技能等認知面為中心。但是今後的學校教育，尤其是九年一貫課程，對於情意層面也很重視，因此必須以新的評量觀因應，否則會對九年一貫課程的實施產生極大障礙。

客觀性測驗是能客觀評分的評量方法。學生對於某些問題的解答，可以依據正、誤的基準做二分法的評分，由於解答的正誤預先決定，所以任何人的評分都可以做同樣結果的評分。因此，可以在很短的時間內公平、正確的進行評量，我們將這種客觀性測驗稱為「量的評量方法」。量的評

量最大的優點是其信度高，但效度則多少令人有些質疑。這種量化評量「客觀性」的優點，卻造成了學生學習的問題。其一，由於將學生的知識、理解程度數量化，往往以考試成績表示學生在所屬團體裡的相對位置，也就是由考試成績將學生序列化，所謂第×名之分。因此，考試成績高的學生則給予高的評價。如此一來，學生爲了得到多一分的成績即引發明爭暗鬥、不擇手段，也只爲考試而學習，完全扭曲了學習的真義。

客觀性評量衍生的另一問題即爲理解的淺薄化。到底客觀化測驗能否正確再生知識，或能否正確解決問題，都值得探討。過去，完全忽略理解的實質背景：學生個人的知識和經驗。爲得到考試的高分，不必考慮上課時所出現的科學概念、自然現象和自己的知識或經驗有什麼樣的關聯，只要片斷記憶考試所要問的「正確答案」就可以得到好的成績，因此，學生的理解顯得非常淺薄而且受到限制。而科技教育以往所實施的客觀性評量—紙筆測驗，則落入同樣的窠臼，考試成了填格子的練習，對學生獲得科技的思考並無助益。

根據教育部(2001)公布的國民中小學學生成績評量準則，第四條指出：國民中小學學生成績評量應本適性化、多元化之原則，兼顧形成性評量、總結性評量，必要時應實施診斷性評量及安置性評量。

第六條：國民中小學學生成績評量，應視學生身心發展及個別差異，依各學習領域內容及活動性質，採取筆試、口試、表演、實作、作業、報告、資料蒐集整理、鑑賞、晤談、實踐等適當之多元評量方式，並得視實際需要，參酌學生自評、同儕互評辦理之。

因此，教師在實施教學評量時，應儘量採取多元化的評量方式，避免因僅採用單一評量方式所帶來的偏頗，且對開展學生之多元智慧，亦有較大之成效。

參、科技教育評量問題探討

一、科技教育課程目標

Savage 在「國際科技教育學會」規劃的” Technology System Handbook” 指出科技教育的課程目標爲：(1)瞭解及欣賞科技的發展。(2)建立學生之價值觀，判斷科技對環境的衝擊與影響。(3)發展正確使用科

技資源、程序和系統的知識。(4)培養使用技術、方法，創意解決當前及未來社會問題的能力。(5)開發人類潛能，在科技社會中扮演負責工作、休閒與公民角色(Savage, 1990；引自蔣秋萍，1999)。

依據教育部九年一貫課程綱要(2003)，自然與生活科技學習領域基本理念為：「人類觀察自然，並且研究各種現象變化的道理，於是產生科學；同時對其巧妙的運用，以適應環境、改善生活，於是乃有技術。三者一脈相連，為國民教育必需的基本課程，以學習者為活動主體，重視開放架構和專題本位的方法，以探究及實作方式進行，強調手腦並用、活動導向、設計與製作兼顧、知能與態度並重，以培養國民科學與技術的精神與素養。」

生活科技課程目標在於培養學生利用科技創意解決問題、批判思考科技的影響與衝擊等統整的能力。而生活科技的教學，著重學生透過與真實生活情境切合的活動來學習，需要較高層次的思考能力，工作的複雜程度亦高，若依傳統工藝的評量方式，劃分為認知、情意、技能領域分別加以評量，或以傳統紙筆測驗來評量此一動態、思考的學習歷程，顯然不夠適切，且不足以涵蓋整個科技學習的歷程。科技的學習是心智與雙手交互作用的歷程，在此歷程中內在心智的活動和外在的成品都很重要，科技學習雖植基於智識的理解、思慮之通達，但絕不僅止於此，其尚包括實務的技能。亦即學習科技是將心智形成的概念，透過具體形式展現於外，才能檢視其是否有效。

二、支持多元化評量的原因

依照 Cunningham (1998；引自楊銀興，2002) 的歸納有四個原因：

1. 對傳統客觀化測驗所產生的負面效果的不滿

我國現採用客觀化的測驗來評量學生的學習成就，以作為升級和升學的依據。如此一來卻產生很多負面的效應，因為這關係到教育績效且對學生有重大影響，測驗常常扭曲了課程與教學，造成學生與教師專注於非績效性的行為，以提高測驗的成績，但卻少有教育上的價值。起初學生為了免於重讀一年或失去畢業證書，考試時常倍感壓力。後來漸漸的問題越來越大，當學生表現欠佳時，就好像教師和整個的學校都失敗了一樣。因此，教師為了提高學生的表現水準，只好

扭曲課程的內容，很多學校的教學只是在為要考試作準備而已。教師們花很多的時間教學生要考試的內容，而學生則不斷的練習可能在考試中出現的題目，卻很少從事有意義的教育活動。長期課程扭曲的結果，當然造成很多的問題，最好的解決途徑即是降低傳統考試結果對學生的重要性，不過這卻無法滿足政治上對教育績效的熱切要求，於是就有廢除傳統考試方式，轉而尋求變通式評量的呼聲出現。

2.對現有心理計量理論的不滿

傳統測驗的技術主要是基於心理計量的假設，這些假設包括下列的信念(楊銀興，2002)：

- (1)學生在測驗上所做的反應就如同生理特質(身高、體重)一般，是可以計數的、可以轉換成數字的，並且可以統計的方法加以解釋。
- (2)評量是中性、價值中立的，可以用科學語言來描述學生的表現。
- (3)評量是客觀的，而且是在與真實情境無關之下實施的。

上述信念代表著傳統測量理論以及其符合科學觀的論點。但是最近幾年來，一些相反的看法，如「質的評鑑」、「社會建構主義」(social constructivism)、「後現代思考」等逐漸興起。他們反對傳統評量理論的邏輯實証主義(logical positivism)，認為太過於強調客觀性將會降低對學生學習表現的解釋，因為客觀性意味著情境中立，而評量若是情境中立，可能會忽略對個別學生的特殊考量。

3.學習的社會建構主義模式的影響

過去的學習理論，認為學習是可以一點一滴累積起來的。而現在的學習理論基於認知心理學的觀點，認為學習是學生自己在事實與概念之間去建構他們的知識，發展他們自己的認知概念圖(即建構主義)。從認知的觀點，學生的學習是主動的，且是整體性的學習，評量與學習、教學之間應緊密的結合，評量應該要反映學生整個學習的過程。既然是整體性的學習，評量時就應該提供學生整體的訊息，而不是片段的訊息，這樣才能了解學生的真正能力，也才能夠透過評量提供教學上的回饋。

4.傳統式評量的缺點

傳統式的評量所要求學生：針對題目的敘述選擇對錯，或從選項中選擇一個正確答案，或填寫簡單的字詞、簡短的答案。題目所呈現給學生的訊息相當的有限，沒辦法測出綜合性的能力。由於考試的情境控制得相當的嚴密，又必須在限定的時間內完成，並限制學生不得利用其他的資源。這種的考試方式相當的不真實，跟學生應用所學的知識於日常生活中的經驗差距太遠。雖然傳統式評量有這些的問題，但是往往卻用一次評量的結果來做一些重要的決定，譬如決定在班上的名次、錄取與否、或需不需要參與學習的充實計畫或補救教學計畫等，這樣的決定是不合理的。因此，應該改變評量的方式，以便能真正了解學生會不會。

黃秀文(1996) 認為教學評量改革的原因有四項：傳統評量的負面效應、新近學習理論的提出、政治社會力量的推動、以及種種新教學法的出籠。傳統評量基於計分的簡便，偏向閉鎖性考題。此類考題使整個課程與教學因而窄化，為了掙脫傳統評量的桎梏，有必要改進現行評量系統或建立一新的評量系統。另外，傳統評量的題目常代表著主流文化及社會中高階層的價值觀，這將不利於來自弱勢文化及社會低階層的學生；更有甚者，低階層的學生可能因而接受到低品質的教學；此種文化上不公平的現象，使得傳統評量的改革因而得到另一助力。

多元化評量有下列的優點：(1) 多元化評量要求學生在實際情境中表現他們的能力，兼重評量的過程與結果。(2) 因為評量和教學結合在一起，教師更容易瞭解學生學習的情形。(3) 對學生學習情形非常瞭解，教師能很快的調整他們的教學計畫，教學與測驗之間的距離會縮小 (Frechtling, 1991; 引自楊銀興, 2002)。

綜上所述，可知科技教育若實施多元化教學評量以評鑑學生實際的行為表現，與真實生活相結合，所學的知識與能力可充分應用；所評量的能力是高層次的認知思考、問題解決能力；教師能充分了解學生學習的缺點，兼具診斷的功能；教學與評量結合，有助於教學的活潑生動等。

如此，對學生的實際學習情況方能作一完整的呈現，也才能讓教師根據評量的結果，修正自己的教學策略。

三、科技教育多元化評量的方式

科技教育是公立學校課程中比較新的科目...作為一門可量化的學科，以其難以捉摸的本質，科技的學習常被誤解，或與其他知識混淆。其實，簡要地說，科學是以自然世界為研究對象，而科技是對人造世界的研究（王應文譯，1996）。科技教育由於牽涉到觀念的啟發、創造力及解決問題能力的培養、以及科技程序的教導，也因此，科技教育之教學評量自然不能只以單一評量方式達成。

科技教育可使用的多元化評量方式可如下表：

評量方式	內 容 項 目
實作評量	操作、作品製作
檔案評量	實驗紀錄、報告資料、電腦檔案
紙筆評量	知識、演算、觀念呈現
動態性評量	學生互動、組員分工
另類評量	學生互評

其評量方式說明如下：

1.實作評量：

科技教育教學的實作評量除了評量學生在操作過程中的熟練外，亦需對完成的成品作出評量。但此一過程中目前並無一固定的標準，故評量之結果，常因教師本身的個別差異而不同。

2.檔案評量：

科技教育教學的檔案評量可說是最詳盡的評量方式，但為完成評量所花費的時間相當多，在目前生活科技沒有一個統一的評量標準的情況下，或可對單一學生的學習成就作一完整的評量。由於每位教師所任教的班級太多，對教師而言，檔案評量在某種程度下不易實施。

3.紙筆評量：

科技教育的紙筆評量的主要在檢定學生的認知能力，在科技教育中，紙筆評量除能評量學生之認知能力外，問答及申論題之運用亦可評量學生具備科技素養之程度。紙筆評量在傳統上即屬於前述的「客觀性」的評量，但由於只能測出學生某程度的認知能力，在教育改革的思潮下，將逐漸從最主要的評量方式轉向多元評量的一種，與其它評量方式相輔相成。

4.動態評量：

科技教育的動態評量能評估學生目前的表現水準，了解學生如何達到目前的水準及可能達到的水準。兼顧鑑定與分類、診斷與處方。科技教育中，動態評量可針對學生的操作過程，學習態度等作描述性的評量，由於需對所有的學生進行相當的觀察，科技教育的動態評量亦需教師花費相當時間去達成。

5.另類評量：

所謂的另類評量即是跳脫出固有的評量模式所作的評量，學生互評即是一例，由於並非是以傳統的評量方式實施，科技教師在對學生實施另類評量時，必需多考量其信、效度，若有一個可參考的評量標準，將可免去信、效度不足的缺點。

四、多元化評量的問題

Mehrens 認為多元化評量有下列的問題：缺乏客觀性和科學的嚴謹度。根據科技評鑑辦公室(Office of Technology Assessment, 1992；引自楊銀興, 2002)提出的報告，認為實作評量有技術性品質方面的問題，即題目數少而代表性不足、信度比較差、費用昂貴。據估計在麻塞諸賽州使用實作評量的方式，評量十六歲的高中生，就要花費七百萬美元的經費，實作評量的花費將比傳統式評量的費用高出十倍以上。也因此，教師在科技教育中實施多元評量時，將要花上比傳統評量多上數倍，甚至數十倍的時間，對於教學進度的影響似已過大！

在蔣秋萍(2000)的研究結論中指出：教師們並未著重採用「標準參照評量」，與科技教育學習評量規準不符、教師雖認知變通性評量的重要性高於傳統紙筆測驗，但實際實行時並未著重變通性評量、教師們對

於班級學生人數多，難以深入了解評量學生，以及缺乏適當的評量依據標準和評量表最感困擾。又根據研究的文獻顯示，要改變教師的信念是一件相當困難的事，因此要教師改變傳統的評量方式而改為採用新式評量方式，可能沒有那麼的容易（Borko，1993；引自楊銀興，2002）。

此外，在九年一貫課程中，科技教育與自然統整成一門領域課程，科技教育原可將九年一貫的多元精神展露無遺，但綜觀目前已出版的自然與生活科技教科書，多將科技課程置於每章節末並充作“科技活動”或是置於書末介紹科技教育方面的知識，而且課本的內容繆誤不斷，很難想像這樣能傳達給學生什麼樣的科技概念。對教師而言，又如何從這些“活動”中對學生得到客觀的評量？畢竟教材是課程的主體之一，倘若教材本身就不適當，又如何就評量結果，對教學客觀的回饋以資修正呢？可能對於教材編輯者而言，要統整科技教育與自然課程於一本書中，兼顧科技教育與自然課程的目標，仍在嘗試與摸索階段，在國中小行政上也可能重自然而輕科技，對於擔任自然與生活科技領域課程之授課教師，除非是專任的生活科技教師，否則大概為應付學校定期評量，也多半使用傳統的紙筆測驗，而不會採用新式的評量方式。

在九年一貫課程實施之前，科技教育之教學評量與之前工藝課程一般，採藝能科知識、技能、情意三種不同向度的評量，多偏重技能的學習和成品的評量，學生科技能力的評鑑方式，常被簡化、淺化成紙筆測驗，未盡然合乎課程標準中教學評量實施要點，亦不符合科技教育學習評量的規準與趨勢（蔣秋萍，1999）。九年一貫課程實施之後，情形更為混亂，各學校實施生活科技課程之方式不同，有些是與自然合科教學，有些則是將生活科技課程獨立，以每週一節的方式行之。前者由於將科技與自然視為同一學科，因此科技教師與自然教師均是以班級為單位教授自然與生活科技“一科”，科技教師教導自然課程或有障礙，但自然老師教導科技課程能否將科技教育所要傳達的觀念完整的呈現，則有待研究；而後者則使得科技教師在教學時每每因為時間不足，無論是評量或教學皆無法平順地達成。在此種情形下，科技教育之教學評量是否能確實實施，則又是另一個問題。

肆、結論與建議

九年一貫課程倡促的實施，使得中小學教育產生了不小的衝擊，尤其在課程發展不完備的情形下，科技教育缺乏統一的評量標準，應是目前科技教育在探究教學評量時最重大的問題。由於缺乏統一的標準，使得教師只好使用傳統的評量方式—成品及筆試來得到學生的成績。另一個問題則是自欺欺人的小班教學，除了偏遠地區及少數學校外，一般班級的學生人數在卅五至四十多人之譜，若要針對每一學生進行多元化教學評量，教師可能得將大部份的時間花在評量學生的學習成就上，對教學的影響不可小覷！如此也才造成了教師明知多元評量的重要性，但卻無法落實執行的窘境。

目前國外科技教育學者已針對科技教育發展評量表及檢核表，若國內的科技教育學者能針對國情及課程內容加以增刪，並鼓勵教師參與行動研究，藉以形成一套符合我國科技教育的教學評量規準；如此除了可以減少教師摸索評量的方式及標準所需花費的時間外，也因為有了標準的評量規準，學生的學習成就可以完整的呈現，教師從教學評量所得到的結果，就不再只是一份成績單，並且也是一份足已使教師改進自身教學的一項重要資訊。

此外，建議教育當局能在完成上述評量規準後，針對科技教師舉辦科技課程教學評量之研習，以落實科技教育教學評量規準，期使教師能有效率地完成教學及評量的工作。也希冀教材編輯者，能善加統整科技教育與自然課程，以兼顧科技教育與自然課程，對於教科書內容嚴加把關，減少錯誤；編輯時，每單元的評量方式採多元化設計，方便教師選用。而教師也能將使用教科書的心得與建議反應給書商，朝編出最適合學生的教材的方向努力。

也盼大多數的科技教育學者，除了關注生活科技課程內涵外，對於有關生活科技課程評量的議題，也能加以重視，進行相關研究與探討，以落實生活科技課程評量的功能，彰顯科技教育教學的成果。

參考文獻

- 王應文譯(1996)。科技素養的評鑑。中學工藝教育月刊，29(6)。
- 李隆盛(1992)。美國工藝／科技教育研究文獻回顧與啓示。中學工藝教育月刊，25(8)，2-5。
- 李隆盛(1995)。中美工藝／科技教育研究之比較。中學工藝教育月刊，28(11)，2-10。
- 柯啓瑤(2002)。自然與生活科技的評量。2002年8月23日。引自 http://www.worldone.com.tw/magazine/11/11_06.htm
- 胡永崇(1995)。特殊教育評量的意義與其重要性。國教天地，109，5-10。
- 黃秀文(1996)。從傳統到變通：教學評量的省思。國民教育研究學報，2，1-26。
- 教育部(2001)。國民中小學學生成績評量準則。2002年8月23日。引自 <http://www.edu.tw>。
- 教育部(2001)。九年一貫課程綱要。2002年8月23日。引自 <http://www.eje.ntnu.edu.tw>。
- 國立台灣師範大學工業科技教育系(2003)。系所簡介，2003年3月29日。引自 <http://www.ite.ntnu.edu.tw>。
- 國立台灣師範大學工業教育系(2003)。系所簡介，2003年3月29日。引自 <http://www.ie.ntnu.edu.tw>。
- 國立高雄師範大學工業科技教育系(2003)。系所簡介，2003年3月39日。引自 <http://www.nknu.edu.tw/~ite>。
- 楊銀興(2002)。多元評量的理論與作法。2002年8月23日引自 <http://www2.thu.edu.tw/~teaching/meeting%202/yang.htm>。
- 鄭湧涇(2002)。科學教學評量的理論與實際應用。2002年8月23日。引自 <http://140.122.143.143/doc/evaluate.htm>。
- 蔣秋萍(1999)。國中生活科技學習評量之探討。生活科技教育，32(9)。
- 蔣秋萍(2000)。國民中學生活科技學習評量之研究。國立台灣師範大學碩士論文。
- 劉桂秀(2003)。國中生活科技課程實施多元評量之行動研究。國立高雄師範大學碩士論文。

Gardner, H. (1992). Assessment in context: The alternative to standardized testing. In B.R. Gifford & M.C. O' Connor (Eds.), *Changing Assessments: Alternative View of Aptitude, Achievement and Instruction* (pp. 77-120). London: Kluwer Academic Publishers.