

真實性評量在教育改革中的相關論題

——一個多元文化教育觀點的思考

討論人：莊明貞

國立台北師範學院初等教育學系副教授

近年來，無論中西教育改革皆試圖以評量方式的改革，作為促進教育興革的重要機制。就以我國為例，以大學「多元甄試入學」制度及國中「自學方案」的改革，乃至於在小學推動「教學評量改進班」與「開放教育實驗方案」，其主要目的皆在藉推動多元化評量，來改革課程發展及教學的品質。而以美國為例，其教育測驗與評量革新的動力主要原因為：(1)國家危機意識燃起，希望對教育結構、內容、過程徹底翻修；(2)警覺到國家在世界市場上經濟不利的狀況；(3)受到語言、學習及認知理論興起的影響；(4)試圖想減低或消除測驗對學生所造成負面的影響，特別是對一些文化不利或少數族群者（Torrance, 1995）。本文因此想從多元文化教育觀點，來探討真實性評量的改革（含實作評量、案卷評量等評量方式）在本土教育改革浪潮中所扮演的角色，及其推動評量改革實施後所面臨的相關議題，以求窺知此一評量改革背後所呈現的文化現象及其可能遭遇的難題。

一、評量在教育改革中的角色

主張評量方式必須改革的學者認為傳統的測量理論可以有效地預測學業成就，卻無法有效地測量人類真實能力。因為人類的學習受到許多因素的影響，傳統智力測驗在預測影響人類學習結果的因素，僅能控制生理年齡，但因不同母語系統和社會文化背景殊異會影響學習的結果，以大多數主流文化參照團體常模，來解釋少數族群在智力或學業成就的差異，基本上是忽視影響兩個不同文化母群學習的因素是迥異的事實（莊明貞，民86）。因為傳統標準化測量無法圓滿解釋人類在學習能力的差異是受到學習型態、注意廣度、記憶、發展速度和性向等因素的影響。又因過度倚賴標準化測驗常導致課程窄化，且易流於基本技能與片面瑣碎事實的學習，忽視複雜思考和問題解決能力。以我國各級聯考為例，教師教學常被標準化測驗牽制，考試領導教學（measurement-driven instruction）或者評量領導課程發展（assessment-led curriculum development）的結果，常使教學僅流於機械記憶與練習低認知層次的學習。此現象不僅壓抑了教師的創造力與課程自主性，也貶抑了教師的專業判斷。近來學者因此提出另一測量派典——變通性評量（alternative assessments），以替代標準化紙筆或多重選擇式測驗。

基本上，變通性評量主要受到認知學習理論的影響，主張知識是建構的，學習是個人從新知識和先前知識，所建構的有意義學習；學習有社會文化的內涵，而個人動機、努力和自我尊重會影響學習的表現。換言之，知識的學習一種生活

技能的應用。因此，變通性評量重視知識的真實性應用與自我評鑑標準（Herman, 1992）。此外，變通性評量強調課程本位的能力評量（curricular-based competency testing），與建立不同社會文化常模的建立。

顯然，評量在教育改革中的角色已改變，以往其成品，結果常被視為驅動整個學校系統的動力，現在則需與教學整合，評量的改革意味著教學與課程發展的改進。因此，評量的角色之一即是以其結果以監控學校的學習系統為目的。

二、真實性評量與教學的改進

為有效了解兒童在學校學習的進步情形，現代認知學習論者或認為知識是主動建構出來，學習是個人創造有意義學習的歷程。知識是個人建構的觀點，對傳統的測驗發展造成很大的震撼。首先，是學生如何將所學得的表徵知識去解決新的問題，或表現出真實的能力。在這一方面，「真實評量」相較於傳統的紙筆評量，則較重視解題的內容與歷程是否能解釋評量工作的結果。

真實性評量（authentic assessment）係變通式評量發展形式之一支，又有稱「實作評量」（performance assessment）、或「直接評量」（direct assessment）。其評量方法可以包括以下的技術：書寫測驗、問題解決、實驗操作、展示、表演、作品集、教師觀察、檢核表、問卷，以及團體合作計畫，其實施方式可以是日常教室的形成性評量、總結性評量或特殊教學計畫，其評分是以人的專業判斷為主。

真實性評量要求學生表現從脈絡中學得且與實際生活經驗相符的工作。表現形式多樣，時間常持續一學年且與教學過程密切相關，其並非只有唯一的正確答案，其形式要求教師與學生共同改進學生的工作品質。制式的評量常是以記憶為主，而真實性評量是以實作表現來證明學生了解學習內容（Maurer, 1996）。

真實性評量所要評量的是學生所欲達成的學習目標以及直接教學的結果。評量改革論者雖強調，真實性評量可以改進與延伸教學目標，並可以評估學習者的進步情形，但因為教師在教學過程中對於學生表現過於真實，過於複雜，教師也會遭遇如何去詮釋，引導及評量學生作品或表現的難題（Torrance, 1995）。

以下將配合本次研討主題「實作評量與案卷評量」來介紹真實評量在教室內實施的內涵與方式，及其可能的困難：

（一）真實教室評量：

目標及工具——蒐集學生如何在真實生活中探索、設計、並完成某工作的證據，可從學生的作品、軼事記錄、觀察記錄、師生座談會等資料中獲得。可提供教師多元的角度來了解學生的學習狀況。真實教室評量最常使用的便是案卷評量（portfolios assessment），包括陳列案卷（showcase portfolio）強調學生的選擇及讚賞自己作品能力；文件案卷（documentation portfolio）提供學生依時間順序，表現學習中的證據；歷程案卷（process portfolio）教師安排一較大型、持續進行的設計活動讓學生在每件作品旁註解創作的過程；評鑑案卷（Valencia, 1990）。

- 優點——1.多樣性；2.涵蓋不同文化的資源；3.有彈性；4.適合不同的學生。
- 可能的困難——1.知識的需：教師不僅要對教學知識的全盤瞭解，亦要能

瞭解學生的文化及語言背景。2.傳統的觀念：偏見的產生導因於教師缺乏正確的覺知而誤將主流文化價值判斷加諸學生。因此要達到評量的目的，教師設計活動之前除要反映學生文化背景外，要先瞭解學生對活動熟悉理解的程度。3.社會傳統對工作和遊戲的看法：教師若在校安排競爭情境但對於原住民學生在家合作完成功課的表現，易讓教師以為學生能力較為低落。4.課程未統整：在現行分科課程的架構下，教師的課程負擔過重，較無意願對各科成果進行案卷評量。

(二)實作評量：

- 目標及特徵：1.以真實世界的情境呈現；2.強調與教學活動相結合；3.以各學科知識的本質為根基；4.掌握學生推理運作或問題解決過程。實作評量承襲制式評量與真實教室評量的假設與實務：借用制式評量的目標在於提供決定學習成效及能力的外在評量標準，也與真實教室評量中將真實世界與課程內容相結合。實作評量反對將某領域的知識分解成細支末節加以評量，它強調學習者如何統整資訊及技能，完成真實情境中的工作。另外，實作評量強調以在不同時間和不同情境中，選取學生工作樣本（work samples），作為評量的對象（Garcia & Pearson, 1994）。
- 優點：1.著重整體表現，避免分割技能；2.評量方式動態化；3.師生互動及參與；4.評量情境多樣化；5.評量方式動態化。
- 可能的困難：1.反映多元文化觀點的程度：主流文化工作者，難以選擇反映少數族群文化的教材，在運用實作評量時，究竟能反映多少多元文化觀點？以實作評量作為唯一選擇，反而違反多元標準。2.編輯教材或設計評量項目及標準，若少數／弱勢族群的參與人數不多，可能推動的成效不大；3.數學及科學學科過度相信學習者的語文能力，易影響解釋學生學科表現的評估結果；4.由外來人員擔任評量工作：牽涉到文化覺識、信任、極可能產生評量者詮釋錯誤的問題；5.適應性：有文化及個別的差異。

三、結果效度的論證

所謂結果效度（Consequential validity），是指站在學生的立場，來看測驗對於學生在校內外生活影響的程度，如Linn, Baker, & Dunbar（1991）及Messick（1989）等測量學者咸認為制式測驗結果的誤用，對於那些來自與主流社會具有不同文化、語言和經濟背景的學生而言，可能帶來不利。

傳統上，對測驗效度的觀點，無非是從古典測量理論而來，包括有內容效度與效標關聯效度等。「真實性評量」對效度的解釋，可以延伸為此種測量是否關聯在同一個科目裡，學生的表現能否與真實情境一致，或是否關聯一個學科專家對課程目標如何有效達成的專業判斷。質言之，「真實性評量」對結果效度的解釋，在於能否使學習者對教材內容的反應能與其舊經驗產生聯結，評量係影響教室中的學習歷程。基於此，一個完整的評量，效度的觀點必須回歸到社會文化脈絡中加以探討。

這種結果效度的觀點更能解釋評量即在達成教學目標。心理計量觀點的測驗發展在工具理性的前提下，可以有效地發展涵蓋一組群課程內容且易於計分解釋的標準化測驗，但是否能以論文式、實驗操作、或從事文化或歷史觀點的博多稿

分析以達成教學的目標，則值得質疑？而真實性評量的效度則視學生在學科表現出來的技能是否能具備專業能力或未來生活技能的應用。從這一個延伸效度的觀點加以分析，過去我們認為一種有效的評量工具，可能就不一定是一個理想的評量方式，因為好的評量方法在有效達成教學目標的定義下，就必須能符合真實情境的應用、以及對受試者的公平性要求。換句話說，評量目的與評量結果的解釋，除了在測量學生學業成就的進步外，對於其是否能將所習得概念有效的發展，以應用在往後的生活經驗上才是最重要的。當然真實性評量是否能運用在不同文化、不同語言、不同社經地位的學生身上，還有其他因素尚須考量，光有評量方式的改革，教育資源上卻無任何改變，對一些文化不利的學生仍毫無用處的，Wolf, LeMahieu, & Fresh (1992) 曾提出要達到實作評量的公平性需具備七項條件，茲臚列於後，藉供思考：

- (1)內容—學生是否都能瞭解？表現—學生瞭解的程度為何？表達—要提供什麼來確定學生有能力將所學表現出來。
- (2)真實性評量，即使宣稱是中立的，是否真能避開主流文化的霸權影響？
- (3)是否我們會使用新的參照分數來評量學生的表現，而更增加學生的比較與競爭？
- (4)我們是否能提供專業的發展的機會，以確定教師使用的評量工具公平、適當？
- (5)是否我們能反映社會的真實面貌，以掌握學生群性的特性？
- (6)重視結果的評量方式，能否提供教師教學上詳實的資訊？
- (7)單一的評量方式或制度能否符合學生不同的學習需要？

四、結語

傳統評量文化為了篩選較佳的人進入社會領導階層，測驗被假定為是公平、客觀的，因為考試可以確保社會人力投資在那未來對國家社會有貢獻的人身上。然而標準化的測驗窄化了課程卻是不爭的事實，它也相對地埋沒了一些對國家未來可能有貢獻的人才。因此本文前述評量改革致力擴充課程內容，讓評量的結果能當作更有效達到教育目標的指標。

因為傳統紙筆測驗的評量對課程發展與教學有極有的限制，且引導教師狹窄地傳授學生應付考試的技巧，而忽略教給學生更高層次認知能力，讓教師無法改進教學方法及課程內容。

然而新制評量的改革所衍生的議題，尚包括所費時間與花費較大，導致教師在各科課程分立的多重負擔下，對評量方式改革是否能帶動課程的革新或教學方法的提昇，迄今仍抱質疑態度。再加上評量標準難以界定，又如何能兼顧多元文化背景學生的個別差異與評分公平性，乃至於真實評量結果如何陳述，在在都需要教師專業能力發展的配合，才能使學生有能力去觀察、實驗、詮釋、瞭解所學，並運用知識、技能從事問題解決。以我國現行中小學在既定課程標準的框架下，雖然透過若干以學校為本位的教室評量改革的提倡，在課程安排及教學品質有較大的改進，但教師往往在無法改變國家層次的課程結構，又面對聯考的評量方式未改變前提下，學齡層愈往上，教師擁有的教學及評量自主權也愈見減低。此外，家長在我國傳統科舉文化長期薰陶下，所持有的制式評量信念也是阻礙教師改革的潛在因素。因此未來如何加強教師在真實性評量判斷的公平性，以及建立

評分者信度，使不同評分者在不同時間對學生實作表現的評分有一致性，或建立案卷評量的作品評量基準（benchmarks），以刺激學生提高自我期望的標準，也是在提倡非制式評量改革需反思的重要論題。

參考書目

- 莊明貞（民86），國小課程的改進與發展—真實性的評量。載於「道德教學與評量—多元文化教育觀點」，189-198。台北：師大書苑。
- Garcia, G.E., & Pearson, P.D.(1994). Assessment and diversity. In L. Darling -Hammand (Ed.), *Review of Research in Education*, 20, 337-391.
- Herman, H.L., Aschbacher, P.L., & Winters, L.(1992). *A practical guide to alternative assessment*, VA: Association for Supervision and Curriculum Development.
- Linn, R.L., Bader, E.L., & Dunbar, S.B.(1991). Complex, Performemce-based, assessment:expectations and validation criteria. *Educational Researcher*, 20(8), 15,21.
- Maurer, R.E.(1996). *Designing alternative assessments for interdisciplinary curriculum in middle and secondary schools*. MA: Ally & Bacon.
- Messick, S.(1989). Validity. In R.L. Linn(Ed:), *Educational measurement*. (3rd de, pp. 13~103). New York: American Council on Education.
- Torrance, H.(1995). *Evaluating autheneic assessment: Problems and possibilities in new approaches to assessment*. Buckingham: Open University Press.
- Valencia, S.W.(1990). A portfolio approach tr classroom reading assessment: The Whys , whats, and hows. *Reading Teacher*, 43,338-340 .
- Wolf, D.P. LeMahieu, P.G., & Eresh, J.(1992), Good measure: Assessment as a tool for education reform. *Educational Leadership*, 19(8), 8-13.

建立技職教育體系

文憑證照通通有 一技在身好處多

為因應台灣成為亞太營運中心，培養高素質又有專業的技術人才，我們將強化從國中、高職（綜合高中）、專科，到技術學院、科技大學的第二條教育國道，並落實證照制度，為技職教育打開一片天。

技職教育不但將成為學生及工商企業界的最愛，更將成為中華民國教育體系的驕傲。