

實作評量(Performance Assessment)

理論與實際

彭森明

美國聯邦教育部

一、前言

近些年來，美國積極推展教育改革。其中一項是改進教學成果評量方式。一般人認為目前傳統式測驗方式，如是非及選擇類，偏重在一些片斷，不連接的知識與技能的認知與瞭解，缺乏對整體知識的組合與應用，以及實際應用的考量。長久下來，影響教學不正常的發展，使學習和日常生活脫節，使學生無法有效地將所學所知運用到解決實際問題上，也使學生缺乏高度思考、分析、判斷，與組合能力的訓練與培育。

針對這些問題，測驗學者所提出實作評量方式(Performance assessment或稱authentic assessment)。此種方式，目前甚受重視。有許多州已全面推展此種評量。我國也和馬利蘭州合作，在國內進行一些試驗。另外在許多有關測驗的學術會議中，也是熱門的討論項目。本人相信此項評量方式的理念與推展，對未來教育發展會有很大影響，特此將此項評量方式的理念以及一些問題，扼要敘述如下，供做參考及國內教育研究發展之用。

二、何謂實作評量？

何謂實作評量？它與傳統式測驗有何不同？顧名思意，實作評量與實際製作、運用，與行為有關，著重於將所學所知表現在具體的成果以及應用過程上，如考駕照時的筆試，與傳統性測驗，而路試與實作評量。另外如應用幾何原理製作模型，利用相當器材做科學實驗與報告，對某一社會問題做研究報告，以及進行社區調查報告等，都是實作評量的方式。

除了製作與應用之外，實作評量方式也著重高層次能力如思考、分析、組合、判斷，以及表達能力的啟發，以及思考過程和邏輯推演的程序的考量。因此實作評量的試卷往往與一些狀況，一連串相關問題，或一種事件，可以自由發揮，提供說明、引證、看法與答案。

由上可見，實作評量與傳統式測驗有很明顯的差別：

(一)它要求學生「自做」答案，而不是「選擇」答案，所以對某一問題正確的答案，可能不只一種；

(二)評量不單看成果(答案或成品)，而且也考量過程(如數學演算程序及對問題的

推理引證等），所以能瞭解學生造成錯誤的原因以及學生對某一問題考量的層面、思路過程，以及邏輯方式；

(-)評量的內容，往往不只限於單一學科知識（如代數方程式），而是同時考量多項知識或技能，如容積測量、比率換算、應用某些物理或化學原理，以及報告實作等能力，在同一測驗中同時評量；

(四)評量問題的取材常與實際生活問題有關，以促進學生能運用所學所知去解決與真實生活有關的問題，真正能學以致用；

(五)評量時間有彈性，有些評量，像傳統式測驗一樣，要求在一定時間內完成（如作文及解答程式），但有些評量，時間由學生自訂支配，如在一學期之內完成報告。

另外實作評量與實際教學過程有相當密切的關係，往往可以成為教學過程的一部份，或是教學成果的收集（portfolios），所以有人認為此項評量比較實際，稱之為「實務性評量」（authentic assessment），這也說明了此項評量的一個重要理念：評量是教學的一部份；它的目的在促進教師教學與學生學習。

從以上觀點，可以看出實作評量，其實並非一項新理念。遠自我國古代科舉所採用的方式，如文科的作文及武科的兵器操作，近至目前的駕照路試，醫科臨床實驗，藝能表演以及科技方面的模擬操作等，都是實作評量，它之所以在學科成果評量上被忽略，沒有被研發主要是因為傳統性測驗方式比較容易實施，計分快速、客觀，可以用在大規模的評鑑上。但演變至今，發現傳統性測驗，不管如何改進，都不能盡善盡美，有其侷限。所以實作評量之優點，才重新受到重視，才被做為改進評量之措施。

三、實作評量的好處有那些？

實作評量，如上所述，是針對改進傳統式測驗而提出的辦法，因此它的主要優點，就是沒有傳統式測驗的缺點，這些優點有些在前節已提過，但為加強說明，總結如下：

(-)實作評量不僅考量學生「所知」多少，而且也考量學生是否能「應用」所知，以及運用的程度與技巧。

(-)實作評量能有效地考量學生思考分析、研究、判斷等高層次能力。

(-)實作評量對學生思考過程與學習方式等能做深入考量，尋求學生學習差異的癥結，供改進教學之用。

(四)實作評量由於出題多與實際生活有關，使學生更能體會學習之重要性與實用性。

另外有些學者認為推展實作評量也可以促進課程與教學策略的改進，加強應用與思考能力的訓練，並時時將科學新知，工商需求之技能加入在課程裡。還有教師在此種評量方式中，負有更重要的職責，提高他對學生考核、診斷，與判斷的重要性，也增進他們與學生之間的師生關係（如教學常需指導學生做專題研究），使教師工作更與挑戰性，能提高教師地位與士氣。除此之外，此種評量方式，常由學生自己選擇研究題材，自己支配時間，因此學生對學習工作比較積極與主動。學習的興趣與熱誠會相對提高。同時也會養成學生自治，自律與自學的精神。

總之實作評量的優點與功能受到許多學者、專家與教師們的肯定。尤其是對課堂教學成果評量，最為適用。不過此項評量方式與學習成果的提高，真正成效如何，目前文獻有限，尚待進一步研判，另外如何推展成為全國性大規模評量的方式，如大專入學考試，在技術上也有待進一步研發。一些常見問題，擇述如下。

四、推展實作評量有何困難？

實作評量雖然有許多好處，但實施起來也有許多困難。

(一)試題很難製作，雖然實作評量的方式很多，但若要大程度等用來鑑定學生能力與學習成果，如全國性會考或升學考，試題的製作是一項很困難的工作，因為理想的考題應是難易適中。應俱有挑戰性，但不至於難到無從作答的地步。它不僅要能引發不同的思路，而且也能確切評量所要鑑定的知識或能力。因此試題內容教材與評量目的，必須慎重考慮，必須很明確，以免吃力又不討好，沒能達成應有的效果。

(二)正確、客觀的評分規範也常有困難。前面說過，實作評量是要求學生創作答案，而不是選擇答案。因此往往有不同層次的答案，與不同的邏輯過程。但對這些不同的答案，如何給予評分標準，尤其是有許多不同的人參與評分時，如何達成客觀、公平與一致性，是很費思考的問題。假如沒有很明確規範，評分往往會受到一些主觀思想的影響。

(三)很費功夫與時間。實作評量所需的時間比較長。一來製作時需費神去思考與設計，二來在實施時也需較長時間完成。另外評分時也需要更多的人力。因此實作評量要比傳統測驗方式費功夫與時間。時間就是金錢，所以實作評量要比傳統性評量需要更多的經費。

(四)測驗信度與穩定度的評估尚待進一步研發，有些人認為信度(Validity)可以依測驗與教學與學習目標之相關程度來衡量，只要是測其所應測的東西，則信度高。但穩定度(reliability)的評估卻甚為困難，因為相關因素較多。比如不同的取材內容，即可能會影響學生不同的表現。

另外因為許多教師對此種評量方式不甚瞭解，需要重新接受訓練。教師必先了深刻瞭解之後，方能積極參與，方能推展此種評量，方會有信心與心得，才能有效地向學生家長們說明。所以要推展實作評量方式，許多教師要重新接受在職訓練，當然職前教師也應加強這方面的學習與訓練，這些都是花功夫的工作，而在教師們來做好準備之前，推展實作評量會有困難。

五、結語

總之，實作評量能補傳統式評量方式的不足，加強對高層次能力、思考過程、學習方式，以及運用能力的考量，使教學與學習與實際生活及社會環境有更密切的關係，也使學生對學習更有主動性。但由於此項評量方式，最近幾年才廣受注意，所以在實施上，尚有一些問題，需要進一步研發，例如計分標準，測驗信度與穩定度的評估與提昇方法，以及試題的製作等，都有尚待討論之處。不過許多學者們都相信這些問題遲早會有答案的，會解決的。所以推展實作評量，以提昇教學效能，應是未來教育發展正確的方向。

不過實作評量方式的發展，並不是完全否定傳統式評量的重要性與實用性。其實傳統的評量與實作評量各有千秋。採用何種方式，應依所需評量的知識或技能的性質而定。比如要知學生會不會做乘除的計算，用傳統式評量即可達到此目標，但若要知道學生是否能應用乘除計算，去解決日常生活問題，則需用實作評量方式。另外許多知識與技

能的評量，如駕車常識與技術，宜採混合方式，即筆試與路試兼用。

無可否認，教師對實作評量方式的推展佔極重要的角色，因為實作評量最適合教學成果評量，也與實際教學活動有密切關係。因此除了加強教師對這項評量方式的瞭解之外，也應輔助與鼓勵教師推展此項評量。比如：

(一)輔助教師會作研製實作評量的試題，以減輕教師獨自研製試題的困難與負擔。

(二)社會學科可多要求學生做專題研究報告，教師從旁給予指導。

(三)語文科可加重作文，並多提供討論、演說、辯論，與口頭報告的機會，以培育並考量高層次語文能力。

(四)數學能力或學習成果評量，要兼顧演算過程與最終答案；要多出應用題，而題目內容應儘量和生活狀況或問題有關。

(五)理工科應要求學生做實務研究及報告。

(六)要求學生有系統地收集、整理學習成果、成品，及作業 (performance portfolios)，做為學期結束時整體評量之依據。

(七)在大規模的評量，如升學考試，可增加應用題及問答題，亦可增加專題作文或短文寫作的份量，以考核並藉此培育學生思考、分析、組合與表達能力。目前美國的SAT測驗，即採此種方式。

總之，實作評量是值得推展的評量方式，也是良好的教學策略。期望在學者專家及教師們共同努力下實作評量能在國內推展，提昇教學成果，使我國教育或為國際典範。

參考文獻

- Airasian, P. W. (1988). Measurement driven instruction: A closer look. Educational Measurement: Issue and Practice, 7 (4), 6–11.
- Baker, E. L., O Neil, H. F., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. American Psychologist, 48, 1210–1218.
- Dunbar, S. B., Koretz, D., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education, 4, 289–302.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. Educational Researcher, 18, (9), 27–32.
- Haney, W., & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, whats, and whithers. Phi Delta Kappan, 70 (9), 683–687.
- Herman, J. L., Gearhart, M. G., & Baker, E. L. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. Educational Assessment, 1, 201–224.
- Herman, J. L., Golan, S. (1993). The effect of standardized testing on teaching and schools. Educational Measurement: Issues and Practice, 12 (4), 20–25, 41–42.
- Institute on Education and the Economy (1993). Authentic assessment in secondary education. IEE Brief, 6, 1–4.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based

- assessment: Expectations and validation criteria. Educational Researcher, 20 (8), 15 – 21.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23 (3), 13 – 23.
- National Council on Education Standards and Testing. (1992). Raising standards for American education. Washington, DC: Author.
- Newmann, F. (1991). Linking restructuring to authentic student achievement. Phi Delta Kappan, 72 (2), 458 – 463.
- Resnick, L. (1987). Education and learning to think. Washington, DC: National Academy Press.
- Shepard, L. A. (1991). Will national tests improve student learning? Phi Delta Kappan, 72 (2), 231 – 238.
- Swanson, D. B., Norman, G. R., Linn, R. L. (1995). Performance – based assessment: Lessons from the health professions. Educational Researcher, 24 (5), 5 – 11, 35.
- Wiggins. G. (1989). A true test: Toward more authentic and eduitable assessment. Phi Delta Kappan, 70, 703 – 713.

水竹之居
 吾愛吾廬
 石磷磷亂砌階除
 軒窗隨意
 小巧規模
 卻也清幽
 也瀟灑
 也寬舒

I love my bamboo hut, by water included,
 Where rockery o'er stone steps protruded;
 A quiet, peaceful study, small but fine:
 Which is so cozy,
 So delightful,
 So secluded.