

量度優質學校產出指標： 甚難解決的眾多問題

侯傑泰

《教統會第七號報告書》建議先從 (i) 學生學業成績、(ii) 學生自信、(iii) 家長、教師及學生對學校的總體觀感三項指標的增值開始，然後再研究其他範圍，訂定可以廣泛應用的質素指標，以找出何校較為優質，並鼓勵學校不斷改進。下文將引述外國經驗及研究，討論有效量度產出指標及其相關的困難。

一、局限於表面量度的指標

外國經驗顯示，這些量化的評估，令學校只集中於一些表面及可量度的改變；例如某校大量增加課外活動學會數目，以達致優質學校要有較多不同學會的要求，但學會的實際活動，誰也不管；另一學校每星期均有家長座談會，但實際成效及內容，無人關心。

教育與甚多商品不同，很多教育目標難以量化，準確測量，例如審美眼光、孝敬父母均可能極為重要，但因為這些目標難以測量，故這些不

能準確量度的目標，會漸漸變得不重要，學校只重視簡單及易於量度的指標。

二、非預期的負面後果

外國經驗亦指出各校在追求這些優質指標下，有極大壓力及傾向作一些容易及可觀察的改變，至於這些改變是否與優質教學有關、是否有效、是否最急於改善的項目，教師及教學行政人員均不大關心，他們只在乎這些改變是否易於被人察覺；相對而言，難於觀察的缺點則被人忽略。

現在很多中小學均有甚多極富創意的教學措施，在量化優質教學後，學校會盡量減少創新的冒險，以迎合傳統運作，因為這些新改革，多不能在傳統指標中顯露出優越性。故此容易出錯的新改革會漸漸消失，所有改革均集中在可見到及易測量的項目上。

三、指標無助改善教學

學者批評這些優質指標對改善教學並無幫助，因為這些指標多是一些較宏觀及整體性的評估，例如：我們查問學生是否感到學習氣氛良好，依學生評價來決定何校更為優質，但這些指標除了可以將學校按等級排序外，不能有助我們了解導致良好學習的細微具體因素。又例如我們查問並比較每間學校內好老師的數目，從而決定

何校較優，但好老師的數目，卻無助於明白這些好老師究竟有何出眾的教學法，這些資料才真正有助改善教學。

評估優質學校的工具多涉及宏觀整體性指標，無助於了解導致成功的詳細及具體因素；可是太詳盡的指標，卻又不能收納在一般評估工具內。故優質教學指標一般對改善教學並無實際裨益。

四、不能倚靠少數指標

很多研究者贊成用極大量的指標去檢視學校是否優質，而不同意設計一簡單的教育 GNP 指標，這是基於頗多不同的考慮。

改革是有代價的（trade-off），當我們只有一、二個簡單指標時，不難預見所有改革都朝向這一、二個指標作調整，若果我們只重視好學業成績，則從此所有音樂、體育活動都要讓路消失。當我們過於強調某項指標時，我們必須付出一些未列出或更長期的教學目標作代價，為免偏重單一目標，及令學童發展更全面，頗多學者贊成檢視較多指標。

五、各指標難以合併

當用多個指標比較學校時，我們需訂出一個方法將各指標得分合併，這是一個極為困難，並

涉及統計學及教育哲學的問題。以中學畢業生為例，如何由中學會考成績計出一個總指標並不容易，我們是否只計算中、英、數（其教學理念是中英數最重要）？還是最好成績的五科（可以集中學習較少科目）？抑或九科成績（鼓勵通材教育）？種種措施反映及體現不同的教育理想，何者較優？何者屬劣？難以得到共識。同時要將學生成績及學習態度（家長、教師意見等）得分合併，涉及的理念爭論更多，不是數次會議或求教專家可以解決的問題。

六、增值與優質學校無必然關係

簡單來說，學生能力增值的來源可分為與學校有關（優質學校）及與學生個人有關（優質家庭，例如：高社經地位的家庭，父母關懷照顧）兩大類。某校學生能力有高增值，可以是該校教學表現優良，也可以純因該校收取來自優質家庭的學生所致；即是說學生有增值，可以是學校導致，亦可能是學生自己帶進學校的；另一方面，學生沒有增值，可能是他們帶著破碎家庭入學所致。不考慮學生來源背景（及其他相關因素），可能令學校不再願意接收家庭背景較差的學童。

部分國家（如法國）對有較多低社經地位學生的學校，給予更大經濟支持，那麼在評核學校是否優質時，我們是否需額外加分子那些背景較差的學生，而懲罰（扣分）那些來自較佳家庭的學童？其實在他國同類計劃中，如何依據學生背

景作調整是一大研究課題，並無簡單共識可依循。視乎教統會向那一派學者專家求教，其結論可能絕然不同。

七、增值是一難以處理的概念

增值聽來是一合理且易於具體操作的概念，例如我們將今次得分（或名次）減去上次得分，但事實上這概念涉及甚多統計學上極為複雜、且難以解決的困難。各國經驗只可得出多種相對較佳的學派或處理方法，但各種方法背後所涉及的理念及哲學未盡相同，也難有共識。

簡單來說，例如那些中一成績已是極佳的學生，要求他們與中一成績最差者，在中五時有同等進步，則可能並不公平，就如如要一個 100 公尺賽跑 10 秒的選手，與另一個 17 秒的選手，一致性地要求他們進步 2 秒，對 10 秒選手則極為困難。

學校效能對每一科的作用是否一致？其影響是線性還是曲線關係（例如：對學生關心可能有正面效應，但極度關心可能變成溺愛，對學習有負面影響）？是否所有措施對每一學校影響均一致（例如：課外活動對高收入家庭學生沒有裨益，但對貧苦學童群性發展卻極有好處）？

很多學習態度的指標並不穩定，其信度（reliability）不高。相對而言，學校間差異（學

校效能)則甚小,以低信度工具檢視微小的改變,自是誤差甚大,徒勞無功。這就正如以一把只刻有一、二、三米的尺(沒有厘米刻度),以量度嬰孩的成長,因刻度太疏,難以檢視細微改變。以這些低信度工具作指標,學校的優質次序將主要由運氣決定,每年不同,整個優質評估計劃將成爲抽獎遊戲,這是外國經驗。

增值好像是一個極易明白的概念,其實在研究中,我們並無一套很好的方法去測量及比較。

八、指標並不可靠

指標是否有效可靠呢?外國經驗指出(如:美國評估國家教育水平計劃,NAEP),當學校要營做一個好學業成績的假象時,他們會用種種方法去鼓勵差成績學生不參與公開評核考試。香港的學校可能未致於鼓勵差成績學生不參加中學會考,但他們卻有辦法令大量差成績者不能升上中五,這是否我們希望見到的改變呢?

高自我觀本是與好成績及一些正確學習態度有正相關,但當學生明白其自我觀將用於衡量其學校水平時,則自我觀很快便會失去其應有的特性及效用。這就是正如某著名學校發一問卷,詢問申請入學學生的家長,以了解他們是否關心愛護子女一樣,本來關心子女問卷的得分,確能準確預測學童成績,但當家長知道這是學校收生的

準則，這份關心子女問卷的得分，將失去其原有預測學童成績的能力。同樣道理，頗多在各類學術研究均十分有效的指標，當被公開用以評估學校時，均失去原有效能。

九、需更了解各指標的特性

我們需對各指標的特性，作更多學術研究，以明白其優質學校的關係，舉例來說，以自我觀反映學校是否優質，會遇到甚多動機理論的問題。研究顯示學童填答自我觀問卷時，主要是參考自己與旁人的成績（外在參考系統），及比較自己各科的成績（內在參考系統）。從前者（外在系統）來看，假設校內所有學生，都因學校效能提高而有所改善，對個別學生來說，雖然他的成績有進步，但因所有同班同學成績均有改善，所以他的自我觀極可能無大改變，從後一角度（內在參考）來看，雖然某學生數學成績較前進步，但因他英文、中文、物理、化學也有進步，所以他的數學自我觀也不一定有大躍進。

爲了令學生感到真的有所進步，老師可能不斷強調及引導學生與鄰校（非優質者）比較，以他校作參考點去顯現自己的進步，這方法對高質素學校可能奏效，但根據新近動機理論的主流想法，我們不應鼓勵學生不斷將自己成績與他人比較。簡單來說，要是所有學生均以是否較他人成績更好而獲動機來源，邏輯上則全港必有一半學童成績較他人爲低而失去自信。故此，強調與他

人比較以顯出自己進步，可能提高自我觀，但卻引來其他惡果。

由是觀之，簡單地量化一兩個學習態度概念，冀求以此反映教學成果，不理會各概念間的複雜關係，極可能帶來適得其反的果效。

總結：高瞻遠矚及非量化審核之重要性

由上文討論可知，我們極可能不應該，也不能夠準確訂出一些產出指標，我們不能期望可列出一張客觀產出指標清單，以供教署的質素保證視學小組，作為評審學校的依據，若循這方法而行，不難預見很多學校均依據這些「指示」，增加課外活動學會數目，架床疊屋地設立各種課程教學小組，各種型型式式易為外人觀察的改變遂應運而生，精力花在浮面的櫥窗裝飾上。

建議中的質素保證視學，必須由獨立、高瞻遠矚、明白教學之複雜、有豐富及實際經驗的人士擔任。低水平的評審，自然帶來浮面錯誤的改革，一個只懂整數（1,2,3,4.....）的評審成員，對如何將三個蘋果分給兩個人時，自然建議先將一個蘋果拋進垃圾箱，以便解決問題。

報告書冀求訂出一張可以量化、可以準確測量的產出指標清單，是否永不可能達致？這方向是否違反基本教育原則？在極需依賴質素視學專家的高瞻遠矚、專業能力下，寄望教署的小組能

帶領全港中、小學邁向高質素的使命，是否期望
過高？是否有點兒戲？是否不負責任？

參考文獻

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261–271.
- Berliner, D. C. (1996). Uninvited comments from an uninvited guest. *Educational Researcher*, 25(8), 47–50.
- Bonnet, G. (1996). Effects of evaluation procedures on educational policy decisions in France. *International Journal of Educational Research*, 25, 249–256.
- Creemers, B. P. M., & Reynolds, D. (1996). Issues and implications of international effectiveness research. *International Journal of Educational Research*, 25, 257–266.
- Education Commission. (1996). *Education Commission Report No.7: Quality School Education*. Hong Kong: Hong Kong Government.
- Fitz-Gibbon, C. T. (1996). Official indicators systems in the U.K.: Examinations and inspections. *International Journal of Educational Research*, 25, 239–247.
- Husen, T. (1996). Lessons from the IEA studies. *International Journal of Educational Research*, 25, 207–218.

- Jones, L. V. (1996). A history of the National Assessment of Educational Progress and some questions about its future. *Educational Researcher*, 25(7), 15–22.
- Marsh, H. W. (1990). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. *Educational Psychology Review*, 2, 77–171.
- McEwen, N. (1996). Student outcome indicators in Canada: Uses, issues and solutions. *International Journal of Educational Research*, 25, 219–229.
- Ralph, J. (1996). Issues in the development of national education indicators in the United States. *International Journal of Educational Research*, 25, 231–238.
- Rowan, B., Bossert, S. T., & Dwyer, D. C. (1983). Research on effective schools: A cautionary note. *Educational Researcher*, 12(4), 24–31.