

校長評鑑顯現困境

駐舊金山辦事處教育組

許多州在校長評鑑上已實行多年，依據聯邦教育法輔助建立政策，但許多學校仍在尋找最佳途徑評量校長的工作績效，以及考量該涵蓋學生們的何種成果在考查項目內。

始自 2010 年，至少有 36 州通過法令要求校長接受績效考查，並依據全國州議會聯合會議，逐漸增加考查內容的難度。修正的內容，將原本專注龐大專業能力的評量，改成複雜的學生成就並行審核模組。考查方針一般要求校長評鑑涵蓋固定比例的學生成就（學生表現或學業成長）。涵蓋的比例，由百分之二十(Delaware)至百分之五十(Georgia, Ohio…)皆有。但根據尚未發表的研究顯示，對於校長評鑑與校長們成就、信賴度與績效的相關性研究，並沒有相對應的成長。大多數關注的焦點與研究都只是採用「相似於研究教師之系統」為工具。

Ellen Goldring（那什維爾，梵德必特大學教育與人類學系系主任）審核了 2009 及 2013 年的校長評鑑法，她認為其中關於法令將如何使用的資訊有限，無闡明校長評鑑後懲處及如何回饋；在校長資質管理、資訊分析及建立階段的自治權裡，沒有將校長的演進角色納入調整。

「校長評鑑……只是教師評鑑的一個不受重視的“繼子”。」Goldring 說道，帶出相較於龐大的教師質量研究（例如教學效率評量，MET），校長評鑑研究是如此地稀少與不受重視。「教師評鑑的發展工作已極其龐大…」「…但對校長評鑑投入的心力，卻少得可憐。」

政策擺盪

在 2010 至 2014 之間，全國州議會聯合會議（NCSL）重審法令，並加重校長與教師之責任。2009 年開辦的聯邦「追求卓越」補助金競賽，包含了上述評量系統為申請條件。近期，許多州為了符合「有教無類法案」豁免資格，積極考慮將學生成長列入評鑑校長的重要項目，即使聯邦指導方針將細節交給各州決策。不過，NCSL 也發現缺乏有效度、信度之評量方法，與極少注重評鑑者的訓練。Goldring 將不同州的細節主要分類為三種，下方將一一說明。

“50-50” 百分模型，一半的校長評鑑分數來自學生表現或成長，

並使用這種評量方式為標準測驗。另一半分數係依據州際學校領袖證照核發聯合會規章，並與學校政務員領袖顧問會一同研議。經過這些規章，校長們主要由他們的頂頭上司給予四分制的考績（從無成效至高成效）。

矩陣模型：相較於第一種的百分公式，第二種類似一矩陣。在甘迺迪克州，學生學習佔有 45 百分比，而教師成效佔學生成果的 5 百分比。其他一半的評量分數還包含 40 百分比的領導者作為和 10 百分比的學校相關人員機構回饋（例如：家長、學生、員工、社區機構、鄰里組織等。）

學生「資訊王牌」模型：第三種評量方式，學生資料只佔小部分百分比（在有些州小於 20 百分比），但如果學生成長與成果低，校長無法得到「有成效的」評分。

模型的批評

Goldring 給了幾個這些政策的批評。她認為這規章缺少心理測量其信度與效度，它們主要依據缺乏“360 度”視角的專業評判，而多視角可結合教師對於校長質量與成效的回饋。有些系統過於累贅與複雜，有些在評鑑處分上含糊不清；有的要求停職獲得低分的校長，其他對於該如何處分不在意也無任何表示。對於頗有挑戰性的校長角色，各州也許想考慮不同的評量方式，包含校長是否再訓練他們最有成效的教師，以及專業發展是否符合教學者的需求。

「教育評量及政策分析雜誌」在三月發佈的新研究，似乎也支持 Goldring 對增加評鑑研究與更謹慎對待評鑑的要求。梵德必特大學教授 Jason Grissom 發現，以 34 萬名學生的邁阿密-戴德區學區，即使教師、校長及學生們都相同，以不同統計模型分析出的結果都不同。Grissom 建議一個系統除考量學生表現及成長外，尚須包含一定品質的觀點，這觀點依附：對“好”表現的構成有非常特定的描述；要求證據蒐集、學校視訪、與校長共事者討論，以及家長、教師、學生及社區調查。

「高信度的測試成績，只能表現校長與教師成果的一個短小視角。」國家中等學校校長協會副會長 Dick Flanary 說道。從他的觀點，畢業比例、輟學比例、識字率、教師流動率為更適當的評量方式。

他的團隊以及國家小學校長協會，定義了許多需放入評鑑的其他因素。它們包含專業度的成長與學習；學生成長與成就；學校規劃與

進步；學校文化；專業質量與教學領導能力；學校相關（與周遭）人員機構協助與參與。

許多州正在修改它們對行政人員的評鑑政策。上週，密西根州議會投票通過新的州立教師、校長評鑑綱要。不過施行結果因還在爭論要用哪種評鑑測驗、及測驗的涵蓋範圍而延後。

達拉威有評量教育者（包含校長）的悠久歷史，它們將學生成果放入評鑑，成功申請到 1 億 1,900 萬的聯邦「追求卓越」補助金。在 2012-13 學年推出的修正評鑑系統中，校長在以下五個領域評鑑：遠景與目標；學習文化的創造；管理；專業責任；學生進步。學生資料在最終分析上有“決定效力”，意思就是，當學生成長被認為不符合要求，校長無法得到“有成效”或“高成效”的總成績。一份州內回顧報告發現，在 2012-13 學年，百分之九十七的達拉威行政員得到“符合要求”的綜合評鑑，但這高評價與學生成長沒有太大相關。

更換評鑑方法

那是州內再次修正評鑑系統的其中一個理由，以及加入更多校長實踐指標的細節。達拉威教師與領導者成效單位系主任 Christopher Ruszkowski 表示。修正後的新系統，會在 2013-14 學年於九所校區初步實行。如果在下個月（六月）的州教育委員會批准通過，它們將在下個學年度於全州施行。

達拉威大學教育系助理教授 Jacqueline O. Wilson 說，當州內決定改變學生成長在評鑑裡的比例時，一切有些後退，但達拉威的教育者已經習慣被學生成果所評量。主要考量為，州內是否使用正確的方法評量成效。

「我想，把他們評鑑的優勢放在學生成就上是有點嚇人。」Wilson 說。「但我們州內其他人始終說，如果我真的在 1 到 4 項都分別做得好，……最後應該得出高學生成就的結果。」

目前為止的正面經驗

參加初步評鑑測試的愛佛瑞馬里帝茲中等學校校長 T. J. Vari 表示，目前為止他感受到的是正向的經驗。

Vari 常態性地與他的監督者-區長見面，討論他的進展及督導之前設下的年度目標。如果監督者能幫助他改善表現，他也能對教師做一樣的事，以此推演將使教室內的教學活動更優異，最終改進學生的

表現成果。

「我們為行政人員條列出的四個項目，是該看向正確的方向。」他說，「規範是非常有幫助的。最重要的部分是，校長及其監督者時常會面，正式及非正式地討論學校事務。」

資料來源：2014年5月21日 美國教育週報

Vol. 33, Issue 32, Pages 1, 20-21 Denisa R. Superville 報導

連結網址：“Principal Evaluation Presents Difficulties”

(http://www.edweek.org/ew/articles/2014/05/21/32principals_ep.h33.html)

