

※ 師資培育文獻回顧 ※

如何決定？師資培育方案有效性評鑑的討論

楊雅妃

(國立臺灣師範大學教育學系博士生)

一、引言

在講求績效的潮流下，師資培育方案 (Teacher Preparation Program, 以下簡稱 TPP) 也不得不回應此訴求。政府政策開始將「有效性」(effectiveness) 的概念整合至師資培育方案中，如「邁向巔峰」(Race to the Top, 以下簡稱 RttT) 即把學生成績的進步程度視為教師效能的重要指標。作者希望透過文獻與文件分析，點出這種作法所需面對的關鍵的決定，並依「選擇」(selection)、「估計」(estimation) 與「呈現」(reporting) 三個階段剖析此些決定的利弊得失。最後，作者提出可供各州省視 TPP 評鑑方案的規準，得以使 TPP 評鑑工作更臻縝密與周全。

二、學生成績進步量的出現

過去 TPP 的評鑑主要是針對師培的過程，以及對師培機構的認證；然而，自 RttT 及其他相關政策的出現後，TPP 不再只是「前端作業」，而是必須為學生的學習成果負責。在 RttT 的主導下，各州皆須發展或改善大規模的資料庫，以便將學生成績與教師表現產生連結，藉此才能夠評鑑教師成效與 TPP 的績效。只是為什麼是學生成績的進步量呢？這個指標又是如何出現的？例如，教育生產研究 (education production function studies) 的研究焦點從過去對教師特質的探究轉而找出個別教師對學生成績進步量的影響，這些前人的研究為「把學生成績進步量做為教師評鑑規準」提供合理性基礎。而且因為方法論的進步，使得個別教師的成效能夠被計算出來，例如研究者可以利用加值模型 (value-added model, VAM) 的特性，有效控制其他因素對學生成績的影響力，單獨計算出教師對學生成績進步的影響。只是，採用此制度的各州 (如北卡羅萊納州、紐約市、喬治亞州、俄亥俄州等) 卻都未能提出具體說明，為什麼選擇學生成績進步量做為檢核 TPP 有效性的指標。因此，作者接下來即要進一步了解形成此結果的重要決定。

三、決定的階段性分析

本文作者希望透過前人的研究，了解學生成績如何被納入 TPP 的評鑑當中，並闡述各州做此關鍵決定的影響。作者將此些決定過程分成三個階段：1、「選擇」

即各州選用哪些資料做為 TPP 評鑑的根據；2、「估計」則指州政府選用什麼方法來估算、分析 TPP 成效；3、「呈現」則是州政府如何決定哪些資訊可被釋出與評鑑結果呈現的方式。接下來作者便分別描述研究者如何選擇以及不同選擇所導致的結果。

（一）選擇的決定：什麼可以決定教師成效？

選擇的決定主要處理的是哪些教師、學生、學科和年級的資料應納入 TPP 的評鑑中，此即影響了 TPP 有效性評鑑的完整性。作者在此部分又分為三個類別討論：

1、學科、年級和學年度

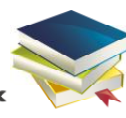
事實上，學生成績與教師成效表現的鏈結取決於各州的學生成績資料庫，縱然《不要讓孩子落後》法案（No Child Left Behind, NCLB）規定各州須於特定年級蒐集特定學科的成績，不過各州實際執行的情形不一，造成後續研究者研究上的限制。例如，佛羅里達州的全面性評量測驗每年檢測學生的數學閱讀、科學和寫作學習成果，不過因為其蒐集的年級幾乎集中於 10 年級以前，使得高中教師的師培方案成效評估沒有相對應的學生成績以做為指標。因此，我們得以知道當州政府擁有越多年級的學科成績，也就保有比較多的彈性進行 TPP 有效性評鑑。不過，各州政府也不會廣納每一個學年度所有的資料，因為他們了解 TPP 是會隨時間變化的。

2、學生

學生成績資料的選用在 TPP 成效評鑑中也是重要的一環。當然，先決條件是學生和教師資料能夠相互媒合，而且因為檢核的是「進步量」，故學生也需有先前的測驗成績。只是各州和各研究者設定的條件不一，所造成的結果也會不同。如田納西州規定學生至少須有 3 次的前測成績，易言之，四年級以下的學生即無法被納入 TPP 成效評鑑之中。又如有的研究者會盡可能選用所有學生的資料；有的研究者則會排除有跳級、休學後復學學生的資料。然而，現有研究少有討論不同選用條件造成評鑑結果的差異。

3、教師

教師的選用牽涉到三個問題：首先，研究者對於教師年資即有不同的看法，如北卡羅萊納州即限定年資 10 年以下才得參與 TPP 成效評鑑，而當長期資料庫資料過多時，又限縮於 5 年以下；紐約市則僅選用年資 1 或 2 年的教師。另一方面路易斯安納和田納西州的研究者則是廣納不同年資教師參與 TPP 有效性評鑑。而且當研究者將研究對象限定於新手教師的同時，可研究的對象也大為減少，故各州



皆會設定研究對象的最低人數。

其次，研究者還需處理同一名學生同時接受不同教師的教學，且可能是同一考科同時包含多個學科、多名教師，如此該如何清楚的區分學生成績受個別教師影響所致？為此，不同研究者採用不同的方法克服此困境，有的研究者以加權的方式解決之；有的則是乾脆避開這些複合性的考科做為評鑑基準。

最後，研究者需要選擇哪些 TPP 進行有效性評鑑呢？理論上，各州只要選擇受州內 TPP 培育而成的教師及其學生成績做為評鑑根據，然各州受他州 TPP 培育而成的教師比例不一，如北卡羅萊納州約莫有 30% 教師是接受非北卡羅萊納州的 TPP；而佛羅里達州則有近五成的比例，若將此些教師排除在成效評鑑之外，那麼評鑑結果是否公允呢？

(二) 估計的決定：什麼樣的分析模式好用？

在這部分研究者即需要選用有助於將 TPP 成效量化的分析模式，以準確估算出個別教師對學生成績的影響。作者研究發現，多數研究者不外乎採以下三種模式：

1. 學生成長模型 (student growth models)

學生成長模型主要是將學生先前與現在成績的差異全部歸因於教師的教學成效。學生成長模型會依據學生成績所反映的學習精熟程度將學生分類，爾後再依此評斷 TPP 的成效。而目前僅有佛羅里達州採此模型。學生成長模型的優點在於它的簡易性與透明性，不過亦有其限制：未能考量學生、教室和學校特性對學生成績的影響力。故麻薩諸塞州僅將此模型當作 TPP 評鑑的一部份，而非全盤採納。綜言之，當資料使用者在運用學生成長模型的分析結果時，因其為單一歸因論，需要更加審慎小心，以避免淪為因果解釋或做為懲罰的依據。

2. 加值模型 (value-added model)

加值模型和學生成長模型的不同在於，加值模型會考量學生本身、教室和學校各種因素對學生成績的影響，並加以區別各種因素的影響程度。而加值模型又可細分為兩種：(1) 直接計算 TPP 的有效性，如路易斯安納州、北卡羅萊納州和紐約市；(2) 先計算個別教師的成效，再取其平均累計為 TPP 的有效性。事實上不管是哪一種加值模型皆可包含共變數或固定效果，但我們不能忽略其背後隱藏的假設，如 TPP 可以解釋些什麼？哪些 TPP 效果是值得被分析比較的。

(三) 呈現的決定：如何詮釋？

由於 RttT 要求 TPP 評鑑結果需要公告周知，此意味著評鑑成果報告書的撰寫需要「親民」，易言之，社會大眾能夠容易理解報告書所傳達的訊息。為此，各州

研究者即需要將複雜的統計數字轉化為易懂的「效果量」(effect size)和分類模式。只是在這簡化的過程中，有些重要資訊便因此而忽略了。

四、結語

將學生成績的進步量做為評鑑 TPP 有效性的基準著實為複雜的過程，其中更牽涉一連串的決定過程，包含選擇、估計和呈現的決定。不過，可以從現有的研究中發現，我們希望藉此評鑑模式得到因果關係的推論，但在資料的選用上並非是隨機分派的結果，故要做因果推論頗為勉強。因此，作者提出一套檢核的規準，希望能夠敦促 TPP 有效性評鑑的改進：

(一) 精確性 (accuracy)

作者認為我們必須承認上述的方式並非估算到 TPP 的真正效果，故為其可信度，研究者必須確保研究的信、效度，如採計學生成績測驗的效度，或搭配運用一些質性的觀察結果。

(二) 公平性 (fairness)

為兼顧公平性，當師培生離開 TPP 而選擇進入富有挑戰性的學校環境時，其所教授的學生成績可能不盡理想，此非 TPP 所能控制。所以作者提醒研究者必須留意此細節，勿將所有的研究結果完全歸因於 TPP 的成效不彰。

(三) 透明性 (transparency)

透明性是最明顯卻最難以達成的規準。因為為了滿足研究結果的精確性和公平性，複雜的統計、分析模型是必須的，相對地，透明性會因而降低。故研究報告的說明與詮釋便需要以深入淺出的方式讓大眾明瞭。

(四) 包容性 (inclusiveness)

現階段的評量模式多為回應 NCLB 的要求，故易忽略 TPP 一些長期且重要的次方案。因此作者建議應將學生投入、畢業率和高品質教學的測量也納入 TPP 成效評鑑中，使得評鑑結果的解釋更為周全。

綜合所述，各類文獻與文件已為 TPP 成效評鑑方案提供合理性的基礎，而作者希望藉由剖析整個方案執行的過程，提醒不同決定可能造成的影響。尤其，我們可以換個角度思考：倘若 TPP 評鑑不以學生成績為指標呢？如果沒有 TPP 方案呢？或若沒有 TPP 評鑑，學生的學習成果又是如何呢？作者希望大眾也能夠關注非 TPP 評鑑的方案。最後作者提出精確性、公平性、透明性和包容性做為評估 TPP 有效性評鑑的規準，希望使 TPP 有效性評鑑施行更為周全。

五、啟示



在現今重視績效責任的時代下，將學生成績的進步量做為師資培育方案與教師效能評鑑的規準表面看似理想，實則仍有許多值得深入探究和思考的空間。故以下分成三點加以討論之。

（一）誰進入了師資培育的範疇？

就本文而言，作者並未討論師資培育方案輸入端的問題，易言之，誰參與了師資培育方案？是個人本身學習與表達能力皆很好的人呢？還是在其他領域屢遭挫敗而不得不選擇教師這條路的人呢？教師如同學生一樣，也有各自的起點行為，若未事先評估或篩選適任者擔任教師，卻課以嚴厲的績效責任，如此一來對師資培育機構或教師個人何嘗公平？

（二）學生的成績可完全拆解嗎？

如作者指出，部分研究因考量到影響學生成績背後的成因甚為複雜，故擇以加值模型控制其他相關因素，進而分析教師教學對學生成績的效果。然而，可否一一窮盡各個影響學生成績的因素並量化與控制呢？例如，一位學生的理化成績不好，可能並非受限於對科學操作的理解，而是受制於國文閱讀的能力或數學計算的能力。又如文中所提，部分測驗科目為複合學科，如社會科是地理、歷史與公民的複合學科，這時又該如何區別各科教師教學的影響力呢？此外，在臺灣特殊的補習文化脈絡下，一個成績代表的不僅是學校教育的結果，也是補習班、安親班或家教共同「陶冶」的成果，該如何將學生的成績歸咎於該科教師的教學效能呢？更遑論是師資培育方案的有效性了。

除此之外，每所學校學生組成的背景大不相同，學生的學習成績可能受到結構性因素影響甚深，如學生的社經地位普遍較低、社區組成中弱勢族群比例較高等，實可預見學生成績進步的侷限性。然新手教師任教學校的選擇權操之在己而非師資培育機構，故已知學生成績進步有限，卻仍以此成績推估、評鑑師資培育方案的有效性，對師資培育機構不甚公允。

（三）以管窺天的限制

有別於美國的背景脈絡，在臺灣的師培制度下，並非每個取得師資生資格者皆能獲得正式的教師工作；且獲得正式教師職缺者僅是接受師資培育課程者的少數。故若以此些少數者的教學成果評鑑整個師資培育方案是否仍具代表性呢？且對師資培育機構公平嗎？例如 100 位的師資生中，僅有 40 名考上正式教師，若僅此 40 名教師的學生之成績做為評鑑師資培育方案的有效性，似乎未能做為師資培育方案成效的證據。

事實上，美國將學生成績的進步量做為師資培育方案有效性評鑑的規準，目



的在於希望能夠建立獲確保師資培育的專業性。然而為達此目的並非只有單一路徑可循，而是需要結合不同研究取徑的特性，綜合診斷師資生在教育現場的處境與問題後，再進一步討論如何使師資培育制度更周全與完善。更何況，我們更應該重視的是師資生在師培機構中究竟學到了什麼？學了多少？真的建立對教育工作的認同感與承諾了嗎？畢竟能夠秉持對教育工作的承諾者，相對地也會比較願意投注心力在教學工作上，而非單以浮泛的數字來證明教師工作與師資培育方案的效能。

導讀文章

Henry, G. T., Kershaw, D. C., Zulli, R. A., & Smith, A. A. (2012). Incorporating teacher effectiveness into Teacher Preparation Program Evaluation. *Journal of Teacher Education*, 63(5), 335-355.

本文引注格式 (APA)

楊雅妃 (2013, 4 月)。臺灣師資培育電子報, 38。取自 <https://tted.cher.ntnu.edu.tw/?p=550>

