

《教育行政與評鑑學刊》

2006年12月，第二期，頁59-82

結合學生學習成果的中小學教師評鑑模式

張德銳

摘 要

結合教師評鑑與學生學習，係當前教師評鑑的發展方向。本文主要論述學生學習結果應用於中小學教師評鑑的可能性。本文先說明學生學習成果運用在教師評鑑的利弊得失，然後介紹國外結合學生學習的二個主要教師評鑑模式。最後，再分析在我國國情下，如何將學生學習成果運用在形成性教師評鑑上。

關鍵詞：教師評鑑模式、形成性評鑑、學生學習

張德銳：臺北市立教育大學教育行政與評鑑所教授

電子郵件：derray@tmue.edu.tw

收件日期 2006.10.24；修改日期 2006.11.21；接受日期 2006.11.23

Linking Student Learning Outcome and Teacher Evaluation for Elementary and Secondary Schools

Derray Chang

Abstract

Linking teacher evaluation and student learning is a new direction for teacher evaluation. This paper discusses the possibility of applying student learning outcome to teacher evaluation for elementary and secondary schools. While the advantages and disadvantages of linking teacher evaluation and student learning are explained, two of the linking models commonly used in US are further reviewed. Finally, this paper analyzes the evaluation context in Taiwan and then proposes an application model for Taiwanese elementary and secondary schools.

Keywords: teacher evaluation model, formative evaluation, student learning

壹、緒論

評鑑是健全的專業服務最基本的構成要素之一。評鑑的目的在搜集資料，從事價值判斷，提供決策之參考。評鑑可以廣泛運用在各種教育領域之中，無論是教育政策、教育行政視導、校務的發展、課程的發展等，均需要評鑑來改進實務或決定其品質或績效。

教師的工作是一個專業的工作，而做為一位專業人員的教師自應接受合理的教師評鑑。教師評鑑便是一種對教師表現作價值判斷和決定的歷程，其步驟為根據教師表現的規準，收集一切有關訊息，以瞭解教師表現的優劣得失及其原因，其目的在改進教師的服務品質和確保教師的工作績效（張德銳，2000）。

教師評鑑的類別，一般而言，可以分為「形成性評鑑」（formative evaluation）和「總結性評鑑」（summative evaluation）。Wheeler、Haertel和Scriven（1993）指出，形成性教師評鑑旨在發現教師教學之優劣得失及其原因，視導教師改進教學，以提高教學效果，達成教學目標。總結性教師評鑑旨在考核教師表現水準的優劣程度，以便作為僱用教師、續聘教師、決定教師薪資水準、表揚優秀教師、以及處理不適任教師的依據，藉以促進學校人事之新陳代謝。

教師評鑑工作在歐美先進國家已經推動數十年，然而在我國中小學推展的相當緩慢，迄今仍在萌芽新興的階段。就教師評鑑的目的而言，我國一向著重總結性目的、輕忽形成性目的，就以民國六十年教育部公布而迄今仍適用的「公立學校教職員考核辦法」為例，便係一種以發放考績獎金為主要目的之總結性評鑑。另外教育部於民國九十二年所發布的「處理高級中等以下學校不適任教師注意事項」以及各縣市政府所發布的各種不適任教師處理辦法亦是一種總結性評鑑。僅有教育部業已於民國九十四年十月二十五日討論通過，九十五學年度開始試辦二至三年的「教育部補助試辦教師專業發展評鑑實施計畫」，係一種以評鑑為手段，以教師專業發展為目的之形成性教師評鑑方案。

就教師評鑑的規準而言，Danielson和McGreal（2000）指出美國的教師評鑑規準向來僅限於教學以及與教學息息相關教師表現，反觀國內教師評鑑，例如「公

立學校教職員考核辦法」，向來過於強調勤惰及行政處理紀錄等，反而忽視了教師工作的主要內涵--教學和班級經營。Danielson和McGreal（2000）亦認為教師評鑑的規準除了教師表現的過程(inputs)外，宜加入教師表現的結果（outputs），亦即學生的學習成效，惟在評鑑學生學習成效時，宜考慮學生學習表現的「基準線」以及教師表現所造成的「附加價值」（value-added）。反觀國內學者（張新仁，2005；潘慧玲等，2004）所研發的教學評鑑規準較重視教師教學表現的過程，相對了忽視學生的學習結果（如在張新仁的標準為精熟學科單元知識、清楚呈現教學內容、靈活運用教學策略、掌握有效班級經營、善用良好溝通技巧、有效引導實驗或實作活動等六項；如在潘慧玲的標準為教師的規劃能力、教學能力、管理能力、評鑑能力、專業發展能力等五大層面共35項指標）。張德銳等（2000、2004a、2004b）的教學評鑑規準雖含有「達成預期學習目標」之規準，但對學生學習結果的重視仍有所不足（如在張德銳2004b的標準為精熟學科、活潑多樣、有效溝通、班級經營、掌握目標、專業責任等六項）。

就教師評鑑的方式，國外學者（Millman & Darling-Hammond, 1990; Peterson, 2000; Stronge, 1997）一致指出必須採取多元的方式，包括行政人員報告、教師自評、教室觀察、教學檔案、學生教學反應、學生學習成果、教師測驗等，俾提供給教師360度的回饋。然而我國「公立學校教職員考核辦法」僅運用出勤文件及行政人員報告。另國內學者所發展的教師評鑑系統，有的重視教室觀察（如馮莉雅、張新仁、吳裕益，2001），有的重視教學檔案的運用（如陳慧萍，1999），有的重視教師自評、教學觀察以及學生教學反應的運用（如張德銳等，2000、2004a），有的雖然在教學檔案系統中強調學生學習目標的達成（如張德銳等，2004b），但是對於學生學習成效資料的檢核及運用，還是相當有限，有待加強。

就學生學習結果運用於教師評鑑的規準及方式上，國外已有許多爭辯。本文作者認為就教學的本質而言，教師的教學工作，除了「教」之外，不能忽視學生的「學」。所以教師評鑑的規準除了繼續強調教師所表現的教學行為外，宜重視學生的學習成果；教師評鑑的方式除了繼續維持教室觀察以及教學檔案等傳統方式外，必須適度的加入學生的學習成果資料作為佐證。此外，隨著教育績效責任時代的來臨，教師很難迴避的必須要對學生的學習表現，負起績效責任。因此，結合學生學習的教師評鑑模式，將勢必成為教師評鑑的主流之一。

因此，展望教師評鑑的發展趨勢，本文主要探討學生學習結果應用於中小學教師評鑑的可能性。本文擬先論述學生學習成果運用在教師評鑑的利弊得失，其次介紹國外結合學生學習的二個主要教師評鑑模式。然後，再論述在我國國情下，如何將學生學習成果分別運用在形成性教師評鑑上。最後再做一個結論。企盼本文能喚起國內教師評鑑界對於學生學習成果的重視，繼而在此一新發展趨勢下，能有更深入更廣泛的研究和論述。

貳、學生學習成果運用在教師評鑑的利弊得失

結合學生學習的教師評鑑，係指將學生學習結果的資料運用在教師評鑑上的一種方式或歷程。換言之，係一種對教師表現作更多元價值判斷和決定的方式或歷程，其不但在教師表現的規準上，強調學生學習成就的達成，而且在教師表現資料的收集上，注重教師的「教」對學生「學」的影響，其目的在引導、增進教師對於學生學習的重視以及確保教師的績效責任。

在教師評鑑上重視學生學習成果的運用，主要來自二股力量。其一，為教師效能的研究。在Marzano、Pickering和Pollock（2001）的後設研究中，他們量化特定教學策略的平均效果。他們發現如果適當的實施，某些教學策略，例如摘要、做筆記、增強學生的努力並給予認可等，可以將學生學習成績提高29至45個百分位數。另外，Sanders和Rivers（1996）的研究亦發現，從三年級開始，分配到高效能教師的三個班級學生，到五年級結束時，他們在田納西州數學測驗的平均成績達到第96百分位數；而另一群成績相當、但分配到低效能教師的三個班級學生，到五年級結束時，他們在同一數學測驗的平均成績僅達到第44百分位數。

這些教師效能研究者極力主張教師需要回饋，不僅限於教學行為的回饋，更應包含教學結果的回饋。適時及大量資訊的回饋，對於任何行業追求改進的努力都是重要的，對於中小學教師亦復如是。例如棒球教練、減重訓練師等專業者，他們固然必須對球員及顧客提供專業的指導，但也必須從球賽的結果以及顧客的減重成效，來改善後續的指導。誠如Schmoker（1999）所言：「當估算付出的努力所換得的結果時，一個人會工作地更有效能、更有效率。更為持久」（p. 2）。

學生學習成果的受視亦來自績效責任運動。吳清山（2001）指出，「績效責

任」(accountability)是指學校相關人員負起學生學習成敗的責任，而績效責任系統則是一套的承諾、政策和實務，以建立並發展良好的實務運作和持續自我評鑑；績效責任本身離不開資料、評鑑、責任、改進等範疇，亦即績效責任係從學校目標、學生表現和學校經營資料以評估學校成效，藉以瞭解責任歸屬，並獎懲相關人員和改進學校教育。

張德銳(1996)指出，九十年代美國聯邦政府所提出的教育改革方案，以老布希總統的「邁向公元2000年美國的教育策略」(*American 2000: An Education Strategy*)及柯林頓總統「邁向公元2000年的目標：教育美國法案」(*Goal 2000: Education America Act*)最具代表性。這兩份教育改革方案，皆以推動全國性的教育改革為優先，並以建立全國性的教育目標為重點，企圖結合聯邦、各州與地方學區的力量，讓所有學生都能接受高水準的教育，俾使美國在未來的國際競爭中能保持優勢。另外，吳清山(2001)指出，柯林頓總統在「學童卓越教育法」(*Educational Excellence for All Children*)的第十一章之「教育績效責任法」(*Education Accountability Act*)，要求學區、學校、教師和學生堅守高的學業標準，以及確保州和學區提供學生高品質的教育。

績效責任運動在當今美國教育的影響力可以顯現在聯邦政府2001年通過的「帶好每個學生法」(*No Child Left Behind Act*)。Tucker和Stronge(2005)指出，此法宣示二個概念，其一是提升教師資格要求為本的「高度合格教師」(*highly qualified teacher*)；其二是學業成就逐年測驗為基礎的「適當年度進步」(*adequate yearly progress*)。前者要求教師至少具備碩士學位、完整取得州政府規定的任教學科執照、以及能證明具備任教學科的專業能力。後者則要求對學生進行「基線評估」(*baseline assessment*)，在高效能教師的教學努力下，學生的學業成就能夠逐年有所進步。當然學生的成效測驗，將作為判斷教師教學對學生學習貢獻的重要依據。

Tucker和Stronge(2005)亦指出，既然大部份教師都會同意自己應對學生的學習負起責任，但整個教育界卻一直迴避採行以評量學生學習為根據的教師評鑑，實有失公允；不能因為傳統行之有年的評鑑策略既溫和又方便，而一直沿用下去，真正的解決方案是發展公平合理的方法，在教師評鑑中把學生學習納入教師的教學成效，這樣才能符應教師的績效責任。

雖然在績效責任運動的壓力下，將學生學習運用在教師評鑑上；支持者固然很多，反對者亦不少。可說是，是否將學生學習結合在教師評鑑上，在美國亦引

起相當的爭議。例如，Kerlinger（1971）力言，學生成就應是教師品質的一個主要決定者。Peterson和Kauchak（1982）亦發現以學生成效的評估作為教師評鑑的依據廣受社會大眾的歡迎，特別是立法人員更重視學生學習在教師品質決定上所顯現的客觀資料。但是，Glass（1974）指出，以商業化、標準化的成就測驗來評鑑教師是既無效、不公平的，且又窄化教學到基本的教育內容和技巧，忽視了教和學的多樣性、創意性、及隨機教育性。Berk（1988）在「學生成就增長並不等於教師效能的五十個理由」一文中指出教師無法掌控學生學習成就分數的五十個因素；因此，以學生成就測驗資料來評估教師表現或效能，基本上是站不住腳的。Fenwick（2001）從學生稟性、社經背景、學校機構等三個層面，來論述學生學習成效的非教師因素。在學生稟性方面，學生個別的學習能力、學習態度、身心發展的成熟度會影響學習成效。在社經背景方面，學生的家庭背景、經濟狀況、先前的知識和經驗、文化價值深深影響學習成就。在學校機構方面，班級規模、學習設備和資源、教職員的工作負荷度、以及學校方案的效能亦會對學生的學習結果產生影響。既然學生學習成效非教師所能掌控的，因此以學生學習成就來評估教師效能並不具正當性。

本文作者綜合McConney、Schalock和Schalock（1997）、Peterson（2000）、Danielson和McGreal（2000）、Tucker和Stronge（2005）、王斌華（2005）的看法，將以學生學習成就評鑑教師效能的利弊論述如下：

一、以學生學習成就評鑑教師效能的優點

在教師評鑑過程中運用學生的學習資料，增強了教師評鑑方式的多元性，可能會產生下列益處：

1. 增加教師評鑑的客觀性

在評鑑時，同時考量教師的教學行為和教學結果，能對教師的教學提供更平衡更豐富的評鑑資訊，有益於對評鑑結果做更正確的判斷。特別是在教室情境裏，應允許教師採用適合情境的教學行為，祇要能增進學習結果的教學行為，應獲得適度的認可，並不一定要強求一致性的教學行為。

2. 提供教師教學更具體的回饋

在教師評鑑的各種方式中，自評可能會有高估的現象；少數的教室觀察與回饋所提供的教師行為樣本亦相當有限。相對於教室觀察的少數觀察樣本，學生學習資料可以提供較長期、較豐富、較具體的回饋資料，足夠教師針對自己的教學

效果做深入的省思，並謀求進一步的教學改善行動。

3. 增強教師改進教學的動機

每一位教師之所以願意從事教學工作，主要係教學成功所獲得的成就感。如果教師能即時獲得學生學習成功的資料，將增強教師教學的自信心與效能感。另外，學生學習結果所呈現的事實，能讓教師更加瞭解自己教學的優缺點，進而產生改進的壓力與動機。

4. 引導教師加強教學評量

教學與評量係教學過程中同樣重要且難以清楚區隔的工作。對於學生學習成果的重視，將有效促使教師重視評量工作，並在評量過程中有效監控學生的學習，達成預期的評量效果。

5. 增進教師評鑑的經濟性

如果能將教師評鑑聚焦在學生學習成果，而學生的學習成果又能長期的建立與取得。評鑑者將可以不必費力費時的經由其他教師評鑑方式，例如教學觀察與回饋、教學檔案的製作與評量，來獲取並不一定比學生學習結果更可靠的評鑑資訊。

二、以學生學習成就評鑑教師效能的限制

以學生學習成果來做為教師評鑑的依據，甚至是惟一的依據，將引起極大的爭議，主要係由於運用學生學習資料時主要具有下列三點疑慮：

1. 教師難以對學生學習負完全責任

教師的品質和努力並不總是直接和學生的學習有關。例如，學生如果自己不努力，再怎樣傑出的老師也教不出好學生。另外，家庭的期望、社經地位、家庭教育的品質、先前的學習成就等等，在在影響學生學習成就。再者，更多的學校因素，諸如學生人數、教室資源、學習環境等亦超出教師所能控制的範圍。因此，教師不必然要為學生的學習成就負起唯一責任。

2. 學生學習成果資料的信效度、可用度有其限制

用學生學習成就資料來評鑑教師的表現目前仍有下列四個難以克服的問題：（1）所要測量的學生成就內涵並不清楚更難有共識。（2）總結性評鑑目的所需良好的（有效的、可信的、可使用的）學生成就測驗目前仍難以廣泛取得，亦即目前教育界所研發的標準化成就測驗大多用來診斷或安置學生用，而不是用來評估教師表現。（3）為遂行總結性評鑑目的所需的施測不但困難而且相當昂貴。

(4) 教師對於不同學生群體、在不同教學時間、教導不同教材內容所顯現的影響力的穩定度甚低，因此很難就某一學生群體、某一教學時間、某一教材內容，就正確判定教師教學績效。就教師評鑑的信度而言，Medley、Coker和Soar（1984）以及Rosenshine（1970）的研究結果顯示，教師對學生影響的效能，在橫跨二個教學學習情境的信度係數約在 .08至 .30之間。以這樣並不算高的信度來看，其效度必然相當有限。

3. 考試領導教學，窄化教師教學的內涵

標準化成就測驗比較容易測量出學科基本學力，較難以測量高層次思考能力以及問題解決能力。況且教師的工作並不全然在學科的教學上，有時亦包括培養健全的公民、展現道德素養、展現與他人合作的群性、培養健康的身心、養成負責的習慣等。因此，如果以標準化成就測驗來斷定教師的成就，將很難不引導教師偏重學科基本學力的教學，而忽視了學生的批判思考能力以及問題解決能力的培養。另外，過於偏重智育的教學，將很難培養學生在德智體群美五育均衡發展。

4. 學生學習結果是教師集體努力結果，難以歸因於單一教師

「十年樹木，百人樹人。」教學工作是一個長期而複雜的歷程，其效果也許在若干年後才能顯現出來。從表面看，學生學習結果是教師一時個人努力的結果，但實際上，學生的學習成果是許多教師多年來集體努力的結果，因此在教師評鑑中，把某班學生學習結果的高低完全歸因於某一位任課教師，是不公平的。從上述論述可知，結合學生學習的教師評鑑有其客觀、多元、經濟、即時回饋、有效增強教師教學表現的優勢，但貿然將學生學習結果應用在教師評鑑上，特別是總結性評鑑，有諸如的弱點和限制，負責規劃和執行中小學教師評鑑的教育人員不得不在運用時要慎思熟慮。

參、結合學生學習的國外教師評鑑模式

歐美先進國家早已開始大量運用學生學習成效的資料來做為教師評鑑的依據之一。例如，王斌華（2005）指出，在1998年12月，英國工黨政府要求實施「績效與薪金掛鉤制度」（Performance Rated Pay [PRP]）。制度的基本內容包括對從事一線的教師進行年度評鑑（或稱績效評鑑），依考試和測驗的結果，考察教師的教學效度，即教師工作對學生的影響程度，重重獎勵表現優秀的教師，從而提

高學生的學業成就。

另外，Tucker 和 Stronge（2005）在《結合教師評鑑和學生學習》一書中，介紹了美國較常用的四個教師評鑑模式，亦即「奧瑞崗州：工作樣本分析法模式」、「科羅拉多州湯普生學區：標準本位模式」、「維吉尼亞州亞歷山卓學區：學生學業目標設定模式」、「田納西州：加值評鑑模式」。Tucker和Stronge的大作，業已由國內學者賴麗珍（2006）完成中譯工作，有興趣的讀者可以參閱。

在結合教師評鑑和學生學習此一領域，仍以美國所發展的諸多模式最為完整週密，較具他山之石之效。而美國常用的四個模式中，又以「奧瑞崗州：工作樣本分析法模式」、「田納西州：加值評鑑模式」分別代表兩個性質相當不同的典型，頗具參考價值。茲對這二個模式進一步說明如后：

一、教師工作樣本分析法

「教師工作樣本分析法」（Teacher Work Sample Methodology [TWSM]）係要求受評教師針對至少二個教學單元的教學目標，提供學生學習進步的證據，亦即單元學習前的測試成績以及單元學習後測試成績的進步指數。然後由評鑑者對受評教師提交的工作樣本進行評估，從而判斷受評教師教學效果的優劣得失。

Tucker和Stronge（2005）指出，TWSM係紮根於「結果本位和脈絡依賴」（outcome-based and context-dependent）的教師效能理論，其設計係要教師回答下列問題：

1. 我要學生完成的學習結果是什麼？
2. 這些學生適合哪些活動和教學方法來達成學習結果？
3. 我需要哪些資源及多少教學時間來實施這些活動或教學方法？
4. 當使用這些教學方法時，適合的評量活動或方法為何？
5. 在幫助學生達成預期學習結果方面，我有多成功？
6. 教學過程中，哪些事情進行順利？哪些出差錯？為什麼？

Tucker和Stronge（2005）認為TWSM要求教師思考、發展、實行、記錄、呈現可作為教師效能證明的工作樣本，其實施步驟如下：

1. 定義所要描述的教學與學習實務之樣本。
2. 在取樣的工作中註明應達成的學習結果。
3. 考慮教學後的結果，在教學前先評量學生的學習狀況。
4. 思考教學及評量宜扣緊應達成的的學習結果。

5. 描述產生教學和學習的情境脈絡。
6. 適應教學情境的需要，擬定教學及評量計畫。
7. 實施經過適當設計且符合情境的教學計畫。
8. 評量教學後的學生成效，並以個別學生為基礎，計算其達到的學習成長。
9. 摘要、解釋及省思學生的學習成長以及其他的學習評量資訊。

Tucker和Stronge（2005）曾以某一小學的生活科學教學工作樣本為例，說明TWSM的運作方式，如下表所示：

表 1 小學生活科學「蜘蛛」單元的教學工作樣本

工作樣本的主要部份	樣本細節
教學背景資訊	<ol style="list-style-type: none"> 1. 學校及教室的描述 2. 學生的族裔分布
單元大綱	<ol style="list-style-type: none"> 1. 單元教學理念 2. 學習目的及目標 3. 教材內容順序圖解 4. 單元教學計畫細節（如教材、教學活動順序、時間分配）
學習評量計畫	<ol style="list-style-type: none"> 1. 前測和後測的程序 2. 教學與評量一致化的矩陣圖解 3. 學生作業的評分說明
工作樣本評量結果	<ol style="list-style-type: none"> 1. 學生作品樣本 2. 每位學生的學習進步淨值（後測分數-前測分數=學習進步） 3. 各學生群組（高中低得分群）的學習進步淨值
教學省思	<ol style="list-style-type: none"> 1. 學習評量之分析－對學生學習結果的摘要敘述及檢討 2. 省思報告－以記述方式討論哪些部份進展良好、哪些出差錯，以及如何在未來單元中改進

資料來源：教師評鑑方法—結合學生學習模式（43頁），賴麗珍（譯），2006，臺北市：心理。

Tucker和Stronge（2005）指出TWSM的運作具有下列優點：（1）它提供自然而然連接教學與學習的合理方法，（2）評鑑過程依賴更多自然發生的學習，（3）TWSM將教學情境視為教學評鑑的主要部份，（4）協助教師將工作聚焦於優質的

教學，（5）促使教學目標與教學實務一致，（6）教與學的連結使學習評量嵌入日常的教學之中，（7）對形成性教師評鑑提供有價值的工具，（8）鼓勵教師省思及參與行動研究。

至於缺點方面，Tucker和Stronge（2005）說明如下：（1）TWSM應用型績效評鑑，很難提供傳統的信效度證明；（2）TWSM鼓勵教師自行發展優質的教學評量方法，是一項實踐上有困難的任務；（3）以複雜的教學樣本欲達到評分者間信度（例如二位評鑑者對教師工作樣本的評分一致性），相當困難；（4）為評量教學不同層面而採用的目標，可能會限制教學內容；（5）TWSM要求的前測策略可能是不必要的限制因素；（6）使用TWSM會使教學的設計、發展、實施及評量，需要投入許多時間；（7）雖然TWSM對教師的工作樣本能提供真實而有深度的評量，但對於做出高利害關係的決定則顯得過窄及欠缺控制；（8）評鑑系統所使用進步分數，仍難以確定教師對於學生的進步有單獨的貢獻。

二、田納西州加值評鑑系統

「田納西州加值評鑑系統」（Tennessee Value-Added Assessment System [TVAAS]）是一種借助統計方法，對教師的教學效果（即教師對學生學習成就測驗的影響程度）做出判斷的教師評鑑方法；加值評鑑法的基本公式是「增值＝輸入－輸出」（value added = output-input）。

王斌華（2005）指出，加值評鑑法堅持認為，無論學生入學成績如何，社會及家長都有權要求學校和教師向學生提供獲得學習進步的機會，亦即每一位學生都能夠並且應該獲得與他們能力相吻合的進步。學校和教師應該採用各種教學方法，努力達到預期的學習結果。

Tucker和Stronge（2005）說明，TVAAS所採用的學習結果係下列三個全州性的常模參照測驗：

1. 田納西州綜合評量方案（Tennessee Comprehensive Assessment Program，簡稱TCAP），適用於三至八年級科學、數學、社會、英語及閱讀等學科。
2. 列入學期成績計算的高中學科課程結業測驗，適用於九至十二年級主要科目範圍。
3. 五、八、十一年級的寫作測驗。

Tucker和Stronge（2005）指出，TVAAS的運作基礎是每位田納西州公立學校學生長期以來的測驗成績，在多數的情況下，資料庫中存有幾年的測驗結果，可

以用來預測未來任何一年學生正常的學習進步分數，於是每位學生之前的學業進步就成為其未來成長的比較標準。教師任教班級學生的平均進步分數，和學校、學區、全州及全國平均進步分數的比較，便可以判斷教師的績效，以及要求教師對於教學效果的省思，並與校長共同擬訂個人專業發展計畫。以下係二位五年級級任教師的統計報告。表2的實例中顯示，甲教師的教學效能好壞參差不齊，其教學專長在數學和閱讀，英語的教學時好時壞，社會和科學則最弱。表3則顯示，乙教師具有高度的教學效能，其成效包括所有學科並且跨越數年之久。

表 2 TVAAS 報告樣本—教學效能參差的甲教師

項目	數學	閱讀	英語	社會	科學
全國常模的進步分數	20.0	13.0	15.0	13.0	16.0
本州的平均進步分數	21.8	14.3	11.9	11.8	18.8
2000年教師的進步分數	18.6	14.4	2.6	2.8	11.2
2000年學區進步分數	17.7	17.0	12.3	9.8	17.0
2001年教師的進步分數	48.7	21.8	29.2	8.2	17.4
2001年學區進步分數	17.1	13.8	11.8	11.7	15.6
2002年教師的進步分數	36.9	19.5	11.9	8.7	8.8
2002年學區進步分數	26.1	13.7	11.5	12.1	14.8
教師近三年的平均進步分數	34.7	18.6	14.6	6.6	12.5
學區近三年平均進步分數	20.3	14.8	11.9	11.2	15.8

資料來源：教師評鑑方法—結合學生學習模式（106頁），賴麗珍（譯），2006，臺北市：心理。

表 3 TVAAS 報告樣本—高教學效能的乙教師

項目	數學	閱讀	英語	社會	科學
全國常模的進步分數	20.0	13.0	15.0	13.0	16.0
本州的平均進步分數	21.8	14.3	11.9	11.8	18.8
2000年教師的進步分數	8.6	18.5	15.7	9.3	6.6
2000年學區進步分數	17.7	17.0	12.3	9.8	17.0
2001年教師的進步分數	29.5	22.8	19.8	14.1	16.4
2001年學區進步分數	17.1	13.8	11.8	11.7	15.6

表 3 (續)

項目	數學	閱讀	英語	社會	科學
2002年教師的進步分數	59.3	23.6	30.5	21.9	40.4
2002年學區進步分數	26.1	13.7	11.5	12.1	14.8
教師近三年的平均進步分數	32.3	21.6	22.0	15.1	21.1
學區近三年平均進步分數	20.3	14.8	11.9	11.2	15.8

資料來源：教師評鑑方法—結合學生學習模式（107頁），賴麗珍（譯），2006，臺北市：心理。

為了加強系統的公平性，TVVAS採取了下列幾個安全措施（Tucker & Stronge, 2005）

1. 學校、學區及教師效能的預測，須依據三年以上且不超過五年的評量資料，以確保統計數值的穩定。
2. 不只有依據TVAAS的資料，單獨評鑑學校、學區及教師。
3. 應用不同媒體促進教育人員了解如何解釋及使用TVVAS。
4. 縮小預測值的全距，用以確保教師、學校或學區對學生學業進步之影響的正確預測。
5. 學生每年應接受新穎、相當、簡明、高信效度的測驗。
6. 被學校特殊教育小組鑑定確認的特殊學生，應被排除在教師效能的分析之外。
7. 出席日數少於150天的學生，應被排除在教師效能的分析之外。
8. 藉著對不當行為的嚴格懲戒，以維護測驗資料的安全性。
9. 對於個人而言，貧窮、英語不流利、父母教育水準及其他變項，每年都相當固定，這些變項的影響程度預測值都不必列入。

Tucker和Stronge（2005）認為TVVAS的運作具有下列優點：（1）TVVAS以複雜的統計模式為基礎，能處理多達數年的學生成績資料。（2）TVVAS被統計專家評為穩固、公平、可靠、有效的系統。（3）能順應學生個別差異，聚焦在進步分數而非根據固定標準的成績表現。（4）教師的資料對於學生的進步及教師影響學生學習結果的能力，提供了相當簡明的評量方法。（5）TCAP具有良好的內容效度，它和田納西州的學校課程密切結合。（6）研究發現TVVAS測定的教師效能和視導人員主觀判定的教師效能，存有正相關。

在缺點方面，Tucker和Stronge（2005）則論述如下：（1）TVVAS涉及複雜的統計分析，使用者需有基本的程式設計及電腦能力。（2）TVVAS模式假定各校及各班學生是隨機分配的，並不利於教導大量低社經地位學生的教師。（3）TVVAS資料有被誤用或錯誤解釋的可能。（4）每年測驗學生一次是重大的時間、經費及人力投資。（5）TCAP對目的複雜的教育只提供了有限的評量方法。（6）TVVAS的評鑑功能仍尚需更多的外部獨立研究所確認。

三、兩種模式的比較

同樣是強調學生學習成果的教師評鑑，但是這兩種模式在評鑑管理的模式上有很大的不同。McConney、Schalock和Schalock（1997）指出，教師工作樣本分析強調學校本位管理以及教師在評鑑歷程中的主導性，而田納西州加值評鑑系統則強調州及學區對教師效能監控的主導性。換言之，在教師工作樣本分析模式，係由教師對學校提出足資證明其教學成果的教師工作樣本，至於是哪一單元的工作樣本，或者何種學生學習成果評量方式，係由教師決定的，而這種評量工具很可能是教師自編的非標準化成就測驗和學生表現的真實評量。反之，在田納西州加值評鑑系統，教師被動的接受由州及學區所舉辦的標準化學生學業成就測驗結果，再經由學生學習結果，研判、推論教師教學效能的績效責任。

此外，Tucker和Stronge（2005）亦從評鑑方法的基礎、學習評量的工具、評鑑方法的運作、評鑑方法的目標等四個向度，來比較奧瑞崗州教師工作樣本分析和田納西州加值評鑑系統的主要特色如下表：

表 4 兩種教師評鑑模式的主要特色

實務問題	奧瑞崗州：教師工作樣本分析法（TWSM）	田納西州：加值評鑑系統（TVVAS）
本方法的基礎是什麼？	以教學前和教學後的測驗評量學生的成長，評量方法依據預期學習結果，選定特定背景和策略。	學生在學習評量上的成長和自己之前的成長率比較，並且依照各班、學校及學區統計成長率。
實務問題	奧瑞崗州：教師工作樣本分析法（TWSM）	田納西州：加值評鑑系統（TVVAS）
學習評量的工具是什麼？	依賴真實的教室評量以記錄學生的學習。	1. 田納西州綜合評量方案 2. 高中學科課程結業測驗 3. 寫作測驗

表 4 (續)

實務問題	奧瑞崗州 教師工作樣本分析法 (TWSM)	田納西州 加值評鑑系統 (TVVAS)
評鑑方法如何運用？	使用進步分數計算學生的學習	使用資料統計分析以產生 1. 學校系統 (學區) 的報告 2. 個別學校層次的報告 3. 個別教師的報告 4. 測驗結果是一項資訊，但不會是廣泛評鑑過程的唯一資訊
評鑑方法的目標是什麼？	加強形成性與總結性的教師省思及自我評鑑	教師專業發展聚焦在加強學生的學習。焦點是學生的成長。

資料來源：教師評鑑方法—結合學生學習模式 (27-28頁)，賴麗珍 (譯)，2006，臺北市：心理。

由以上說明可知，TWSM和TVVAS在評鑑目的上，均含有形成性和總結性評鑑的目的，惟TWSM較強調加強教師自我評鑑及教學省思，而TVVAS較強調提升學生學習成果的績效責任。在評鑑規準上，TWSM強調學生在多層面的學習與成長，而TVVAS僅注重各學科的學業成就。在評鑑方法上，TWSM強調經由教學單元的工作樣本顯示和學生學習結果的連結，而TVVAS則以成就測驗要求教師、學校、學區層層負起績效責任。在學習評量工具上，雖然兩者都著重學生學習的進步分數，但TWSM係運用學生在學習過程中的真實性評量，而TVVAS則實施標準化的成就測驗。在評鑑管理模式上，TWSM強調以教師為主、以學校為本位的評鑑，而TVVAS則透過由上而下的視導和評鑑。

肆、學生學習結果在我國教師評鑑模式上的運用

因為教學是教師工作的首要任務，學生是教師工作的主要對象，所以學生學習結果理所當然應該成為教師評鑑的重要規準之一，教師評鑑過程中亦應運用到學生學習表現資料的檢核。誠如上述，美英兩國在教師評鑑工作，皆非常重視學生的學習表現。但儘管如此，學生學習結果不應該也不能夠作為評鑑教師的惟一依據，而且更應該隨著我國中小學之情境脈絡，做適當的修正與斟酌處理。

就我國國情而言，我國中小學由於尊師重道的傳統，甚少對教師進行實質性的評鑑，而中小學教師對於教師評鑑長期認知不足，甚至有「談評色變」的現象。陳瑞榮（1995，1998）便指出，目前我國中小學教師對教師評鑑缺乏信心、信任，甚至因誤解而產生反感、排斥。由於認知不足，常造成國內教師評鑑工作的推動困難。

其次，國內中小學教師尚難接受教師評鑑和不適任教師處理、教師分級制度的結合。陳聖模（1997）的研究發現，國內教師對於教師評鑑的主要用途，認為應偏重在形成性目的，亦即著重教師發展與成長的功能，而較不認同於總結性目的，亦即教師較不認為以評鑑作為教師升遷、獎勵、處理不適任教師的決定依據。

再者，國內在中等教育階段，升學主義仍盛行，考試領導教學的現象仍所在多有。貿然實施學生學業成就和總結性教師評鑑相結合的政策，將可能使得一方面中等教育階段考試領導教學現象更加嚴重，另一方面小學階段目前相當正常化的教學現況受到波及，其所引起的副作用甚大，不得不加以防範在先。

因此，在衡量我國中小學現況後，本文作者認為現階段學生學習可能並不適用於總結性教師評鑑。但是用於形成性教師評鑑，學生學習結果資料有其價值。一位有心的教師如能在教學前考量學生現有的學習成就，訂定合理的教學目標與教學計畫，然後有計畫的實施教學，再依學生的學習成長狀況，檢視與省思自己的教學成效，然後再另行訂定下一階段的教學目標與教學計畫，當能有效改進教學與提升教學成效。這種「計畫、教學、省思」模式，實是教師教學專業成長的利器之一。

另外，從評鑑學理而言，學生學習結果資料有利於中小學教師評鑑360度回饋的成形。丁一顧、張德銳（2004）指出，360度回饋「是以受評者在中心自我評鑑，而評鑑者則在圓圈的四周，其中，同事看到的面向大約是90度，上級約180度，部屬與外部人員感受的面向約為270度，視導人員所看到的面向必須是360度，然後由人力資源部門或外部專業人員依據四周相關人員的評鑑，並對比受評者的自我評鑑後，向受評者提供回饋，以協助其發展及提昇其能力與績效」（頁3）。因此，受評鑑教師除了進行自我評鑑、同事的教學觀察與回饋、以及接受學校行政人員的教學視導外，如能善用學生學習結果的資料做為回饋線索之一，當有利於發展出360度回饋的評鑑實務，進而經由多評量者的回饋，有助於教師發展及提升教學能力及績效。

除了在評鑑目的上，學生學習成果可運用於形成性教師評鑑之外，在評鑑規

準上，我國中小學教師評鑑規準宜強調學生學習結果的提升。結合學生學習結果，係先進國家在發展教師評鑑規準時，重要的考量。例如美國「全國教學專業標準委員會」（National Board for Professional Teaching Standards, 1989）曾提出教學專業標準的五大核心主張：（1）教師獻身於學生及其學習活動；（2）教師瞭解所教授學科，並且有效地將學科知識傳授給學生；（3）教師對管理和督導學生的學習負有責任；（4）教師有系統地思考教學實務並且從經驗中學習；（5）教師是學習社群的成員。另外美國加州認證委員會和教育廳（California Commission of Teacher Credentialing & California Department of Education, 1997）曾提出六個教學專業標準係：（1）投入和支持所有學生學習，（2）創造和維持有效的學習環境以利學生學習，（3）瞭解和組織學科知識以利學生學習，（4）為所有學生計劃教學和設計學習經驗，（5）有效評估學生學習，（6）發展成爲一位專業的教育者。由上述二個重要機構所設定的教學專業標準便可瞭解美國教育界對於學生學習結果的重視。

惟本文作者認爲學生學習表現的規準並不限於各學科基本學習能力，亦涵蓋問題解決、高層次思考，甚至品格涵養、人群關係、身心健康、藝術鑑賞等表現。誠如邱兆偉（1991）所言，正常化教育的目的係「全人教育」，而五育均衡發展就是全人教育。如何培養德、智、體、群、美五育均衡發展之健全國民，不但是我國國民教育法所揭櫫的教育目的，亦符合多元智慧理論所倡導的原理原則。

另外，學生學習表現的資料亦不限於標準化的成就測驗，亦包含教師自編的非標準化測驗以及學生學習檔案等真實性評量。有鑑於在我國中小學現場，標準化成就測驗資料庫的研發和使用仍相當的不足，如何利用教師自編的紙筆測驗、學生作業、學習單、學生學習自我評估、學生學習表現教室觀察等多元學生學習表現資料來源，來提供教師教學上的回饋，是現階段值得重視的。這種以「教室本位學生成就之教師評鑑」（classroom-based teacher assessment of pupil achievement），係Stiggins（1989）所主張的教師評鑑方式。Stiggins認爲與其投入大量的人力與財力資源來建構大型的標準化成就資料庫，倒不如用來訓練教師如何蒐集運用教室本位的學生成就資料，對於教師教學的改善，有更多實質上的效益。

最後，在評鑑方式上，本文作者認爲奧瑞崗州教師工作樣本分析（TWSM），比起田納西州加值評鑑系統（TVVAS）係更適合我國中小學運用的模式，其理由是：在評鑑規準上，TWSM強調學生在多層面的學習與成長；在學習評量工具上，TWSM係運用學生在學習過程中的真實性評量；在評鑑管理模式上，TWSM強調

以教師為主、以學校為本位的評鑑。另外一個基本的理由是：TVVAS以標準化成就測驗要求教師負起績效責任，在現階段中小學恐將導致考試領導教學的副作用，以及引起對教師評鑑的疑慮和恐懼。

如以教師工作樣本分析做為我國中小學形成性教師評鑑的參考，建議資深教師每三年，初任教師每年做一次評鑑。評鑑時，自選二個以上能代表自己教學內涵的單元，做為工作樣本。每個工作樣本進行下列步驟：（1）描述產生教學和學習的情境脈絡。（2）在取樣的工作中設定應達成的學習結果，而這個學習結果應能反應學生重要的學習內容。（3）考慮教學後的結果，在教學前先評量學生的學習狀況。（4）適應教學情境的需要，擬定適當的教學及評量計畫，惟教學及評量工作宜扣緊應達成的的學習結果。（5）實施經過適當設計且符合情境的教學計畫。（6）評量教學後的學生成效，並以個別學生為基礎，描述其學習進步情形。（7）摘要、解釋及省思學生的學習成長以及其他的學習評量資訊，以設定未來的教學目標以及教師專業成長計畫。當然，受評鑑者在擬訂教學及評量計畫，宜和評鑑者共同討論設定。在實施教學及評量過程中，可以請評鑑者入班觀察與提供觀察後的回饋。在解釋及省思學生的學習結果時，可以請評鑑者提供第三者較客觀的意見。另外，這些所有有關教師工作樣本分析的資料都可以有系統、有組織的放入教師的教學檔案夾，成為教學檔案最重要的部份。

最後，建議教師在工作樣本中所設定的學生學習成果目標，宜和學校校務發展目標做適當的結合，以發揮學校教育目標管理的功能。例如，如果某一所小學刻正推動某一提升學生閱讀能力的三年計畫，則國語文任教學科的教師可以把這三年的教師進修的重點放在閱讀教學，學生學習成果的評量重點亦放在閱讀能力標準化與非標準化測驗的成績。那麼，教師在進行以閱讀教學為工作樣本的分析時，不僅可以呈現出個人對於學校目標的貢獻度，學校亦可以反映出學校整體的績效責任。Iwanicki（2001）認為這種整合學校績效責任以及教師個人績效責任的評鑑模式，是較有效而先進的教師評鑑模式，其不但有利於教師個人專業發展，亦有利於學校的整體校務革新。

伍、結論

世界先進國家教師評鑑的發展，從以考核教師為主的總結性評鑑，走向以形

成性評鑑為主、總結性評鑑為輔的評鑑目的。在評鑑規準上，從評量老師的人格特質、評量教師的教學行為，走向目前所強調的學生的學習。在評鑑方式上，從以往單一的教室觀察，走向結合教師自評、教學檔案、學生教學反應、學生學習結果證據的多元途徑。世界先進國家這一波對於學生學習結果的重視，除了反映教師效能的研究結果以及教師績效責任運動的推波助瀾外，亦源於教師對於教師教學成效回饋與反省的需要。

誠如Danielson和McGreal（2000）所言，一個良好平衡的評鑑系統應包含輸入（教師工作與職責）與輸出（教師欲獲致結果或其欲教學產生學生學習成效）兩種取向，方能檢視教師教學技能及考量教學所產生學習程度的提昇。本文有鑑於國內教師評鑑界對於學生學習結果的相對忽視，旨在提醒對於學生學習結果宜有適當的重視。本文先論述學生學習成果運用在教師評鑑的利弊得失，並說明結合學生學習的教師評鑑模式。其次，則在衡量我國國情後，論述學生學習成果的資料，在我國中小學形成性教師評鑑的目的、規準及方式上，有許許多多可加以運用的空間和途徑。

本文雖強調「教」與「學」在教學歷程的互動性、以及應用學生學習成果資料於改進教師教學的可能性。學生學習結果資料在專業的教師評鑑工作中，有其合理性和必要性，但亦有其限制性，更不是惟一的。誠如Tucker和Stronge（2005）所言，我國中小學教師評鑑界在應用學生學習成果，特別是運用學生成就測驗於教師評鑑歷程時，下列九點注意事項必需謹記在心的：（1）學生的學習評量只是教師評鑑系統的一部份，整個評鑑系統應以多重的資料來源為基礎。（2）評鑑時，必須考慮產生教學與學習的情境脈絡。（3）建議使用學生進步的數據取代固定的成就標準或目標。（4）對同一名學生選擇不同的時間點比較其進步分數，而非比較不同群體的學生。（5）進步分數的陷阱必須避免，例如高學業水準學生進步分數受限所產生的「天花板效應」（ceiling effect）。（6）使用較長的時間架構，以瞭解教師對學生學習的長期影響。（7）使用公平有效的學生學習評量方法。（8）選擇與現行課程最一致的學生評量方法。（9）切勿為了配合測驗而窄化課程又限制教學，除非測驗能真正評量教學內容。

參考文獻

- 丁一顧、張德銳（2004）。360度回饋在中小學教師評鑑上的應用。初等教育學刊，18，1-23。
- 王斌華（2005）。教師評價—績效管理與專業發展。上海市：上海教育出版社。
- 邱兆偉（1991）。正常化教育理念的實踐。載於臺灣省政府教育廳（主編），臺灣省國民小學新進教師教學參考手冊（頁11-18）。臺中縣：臺灣省政府教育廳。
- 吳清山（2001）。教育發展研究。臺北市：元照出版有限公司。
- 陳瑞榮（1995）。工業職業學校教師評鑑模式之研究。國立臺灣師範大學工業教育系碩士論文，未出版，臺北市。
- 陳瑞榮（1998）。職業學校教師評鑑之可行性探討。技職雙月刊，44，38-45。
- 陳聖模（1997）。國小教師對教師評鑑制度之態度研究。初等教育學報，10，417-441。
- 陳惠萍（1999）。教學檔案在教師專業發展上之應用。載於中華民國師範教育學會（主編），師資培育與教學科技（頁183-212）。臺北市：臺灣書店。
- 馮莉亞、張新仁、吳裕益（2001）。國中教師教學效能評鑑工具之分析。測驗年刊，48(2)，17-34。
- 張新仁（2005）。中小學教師教學評鑑工具之發展編製。載於潘慧玲（主編），教育評鑑的回顧與前瞻（頁91-130）。臺北市：心理。
- 張德銳（2000）。師資培育與教師評鑑。臺北市：師大書苑。
- 張德銳、簡紅珠、裘友善、高淑芳、張美玉、成虹飛（1996）。發展性教師評鑑系統。臺北市：五南。
- 張德銳、蔡秀媛、許籐繼、江啓昱、李俊達、蔡美錦、李柏佳、陳順和、馮清皇、賴志峰（2000）。發展性教學輔導系統—理論與實務。臺北市：五南圖書。
- 張德銳、吳武雄、曾燦金、許籐繼、洪寶蓮、王美霞、陳偉泓、曾美蕙、常月如、曾政清、黃春木、白師舜、李俊達（2004a），中學教師教學專業發展系統。臺北市：五南。
- 張德銳、李俊達、蔡美錦、陳輝誠、林秀娟、楊士賢、康心怡、黃柏翔、卓美月、鄧玉芬、管淑華、高紅瑛、邱馨儀、黃旭鈞、高婉妃、江啓昱、張淑娟（2004b）。教學檔案—促進教師專業發展。臺北市：高等教育。

- 潘慧玲、王麗雲、簡茂發、孫志麟、張素貞、張錫勳、陳順和（2004）。發展國民中小學教師教學專業能力指標之研究。教育部委託專案報告。臺北市：國立臺灣師範大學教育研究中心。
- 賴麗珍（譯）（2006）。教師評鑑方法—結合學生學習的模式（原作者：P. D. Tucker & J. H. Stronge）。臺北市：心理。
- Berk, R. A. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. *Journal of Personnel Evaluation, 1*, 345-363.
- California Commission of Teacher Credentialing & California Department of Education (1997). *California standards for the teaching profession*. Sacramento, CA: Author.
- Danielson, D., & T. L. McGreal (2000). *Teacher evaluation to enhance professional practice*. Alexandria, Virginia: ASCD.
- Fenwick, T. J. (2001). Using student outcomes to evaluate teaching: A cautious exploration. *New Direction for Teaching and Learning, 88*, 63-74.
- Glass, G. V. (1974). A review of three methods of determining teacher effectiveness. In H. J. Walberg (Ed.), *Evaluating educational performance* (pp. 11-32). Berkeley, CA: McCutchan.
- Iwanicki, E. F. (2001). Focusing teacher evaluations on student learning. *Educational Leadership, 58*(5), 57-59.
- Kerlinger, F. (1971). Student evaluations of university professors. *School and Society, 99*, 353-356.
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: ASCD.
- McConney, A. A., Schalock, M. D., & Schalock, H. D. (1997). Indicators of student learning in teacher evaluation. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (pp. 162-192). Thousand Oaks, CA: Corwin.
- Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance: An empirical approach*. New York: Longman.
- Millman, J., & Darling-Hammond, L. (Eds.). (1990). *Teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Corwin.

- National Board for Professional Teaching Standards (1989). *Toward high and rigorous standards for the teaching profession*. Detroit, MI: Author.
- Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices* (2nd ed.). Thousand Oaks, CA: Corwin.
- Peterson, K. D., & Kauchak, D. (1982). *Teacher evaluation: Perspectives, practices and promises*. Salt Lake City: University of Utah. (Eric Document Reproduction Service No. ED 228 228)
- Rosenshine, B. (1970). The stability of teacher effects upon student achievement. *Review of Educational Research*, 40, 647-662.
- Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement* (Research Progress Report). Knoxville, TN: University Tennessee Value-Added Research and Assessment Center.
- Schmoker, M. (1999). *Results: The key to continuous school improvement*. Alexandria, VA: ASCD.
- Stronge, J. H. (1997). Improving schools through teacher evaluation. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (pp. 1-23). Thousand Oaks, CA: Corwin.
- Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: ASCD.
- Wheeler, P. H., Haertel, G., & Scriven, M. (1993). *Teacher evaluation glossary*. Kalamazoo: Western Michigan University, Center for Research on Educational Accountability and Teacher Evaluation.

