

美國田納西加值評量系統 及其對我國教師專業發展 評鑑之啓示

林素卿*

摘要

本文探討美國田納西加值評量系統及其對我國教師專業發展評鑑之啓示。透過文獻分析探討此系統之源起、內容、實施、優缺點、應用。研究發現此系統最大的特色是將學生學業成長分為學生及班級（學校）兩部分的歸因，其優點是能順應學生的個別差異，將評量焦點置於學生的進步而非固定標準的成績表現，能作為教學效能的診斷工具，提供教師改進教學及專業發展計畫之依據，使學生的學習成長、教師的教學效能與專業發展間產生更緊密的聯結。據此，本文提出以下建議作為我國推展教師專業發展評鑑之參考：一、以「學生學習成長」作為教師專業發展評鑑的核心概念；二、結合「台灣學生學習成就評量資料庫」的建置，讓學生學習成長、教師教學效能、與教師專業發展間產生聯結；三、將「田納西加值評量系統」評量結果納入教師專業發展評鑑與分享中。

關鍵詞：田納西加值評量系統、教師效能、教師專業發展評鑑

*林素卿，國立彰化師範大學教育研究所副教授

電子郵件：sclin@cc.ncue.edu.tw

來稿日期：2008 年 12 月 29 日；修訂日期：2009 年 2 月 20 日；採用日期：2009 年 2 月 27 日

Tennessee Value-Added Assessment System: Implication for Teacher Professional Development Evaluation in Taiwan

Su-Ching Lin*

Abstract

The main purpose of this study is to investigate the Tennessee Value-Added Assessment System (TVAAS) and to suggest its implication for Teacher Professional Development Evaluation in Taiwan. By reviewing the history, contents, implementations, disadvantages and advantages, and applications TVAAS, we find the distinguishing feature of TVAAS in its ability to separate the annual growth into two parts: that which can be attributed to the student and that which can be attributed to the classroom, school, or district. The advantage of this system is its focus on improvement and not on achievement based on fixed standards and individual differences. Students are expected to grow and improve but not necessarily at the same rate or reaching the same goal. This system also enables administrators and teachers to inform their instructional decision-making using rich data about student learning process. As a result, value-added assessment strengthens teacher's ability to improve student achievement. It helps teachers' professional development. The suggestions for teacher professional development evaluation in Taiwan are also included.

Keywords: Tennessee value-added assessment system, teacher effectiveness, teacher professional development evaluation

*Su-Ching Lin, Associate Professor, Graduate Institute of Education, National Changhua University of Education
E-mail: sclin@cc.ncue.edu.tw

Manuscript received: December 29, 2008; Modified: February 20, 2009; Accepted: February 27, 2009

壹、前言

在「多元化、市場化、資訊化」時代潮流下，教育品質的提升一直是教育改革的焦點。衡諸現今師資培育管道的多元、市場驅力的消費者導向，以及重視教育績效責任等世界潮流，學校如何建立一套系統性的評鑑制度來評鑑教師教學表現，是師資教育改革的重要配套監控機制。教育部為協助教師專業成長，增進教師專業素養，提升教學品質，從九十五學年度開始試辦中小學教師專業發展評鑑計畫（教育部，2006）。藉由實施教師專業評鑑，讓教師在評鑑過程中得到省思與回饋，以協助其專業發展並確立教師的專業地位。自 2006 年四月起，採任各學校、教師自願參與方式試辦，期程為 2 至 3 年，每年檢討其可行性，以決定是否繼續辦理。教師專發展評鑑內容包括：課程設計與教學、班級經營與輔導、研究發展與進修、敬業精神及態度等；評鑑方式主要分為教師自我評鑑與校內評鑑（他評）二類（教育部，2006）。

教育部試辦教師專發展評鑑的基本精神為（教育部，2007）：

- 一、以教師專業發展為主軸；
- 二、鼓勵學校申請試辦，教師自願參加；
- 三、引領學校行政走向教學領導以及學校本位視導與評鑑；
- 四、鼓勵教師以自我省思及同儕專業互動為成長手段；
- 五、以教學和班級經營為主要成長內涵；
- 六、期待學生的學習表現和成效能獲得有效提升。

真正的專業發展是一種深思熟慮的過程，必須有明確的方針和良好的目標來引導，以作為專業發展內容選擇、活動規劃、及成效評鑑的依據。葛斯基（Thomas R. Guskey）（2000）指出任何教師專業發展評鑑應該關注對學生所產生的影響，亦即，判斷教師專業發展經驗或活動成功與否，不能只靠參與者的回應，必須能對學生的學習產生正面的影響，更應將促進學生學習成長列為教師專業發展評鑑的主要目標。教育部所訂定之試辦教師專業發展評鑑計畫之政策，主要目的在幫助教師專業成長，讓教師瞭解自己教學之優劣得失及原因，並透過不斷自我瞭解與省思，掌握成長及提升的契機。然而，教師專業發展的課題尚存在許多問題，例如，教師專業發展活動或模式無法發揮其價值與效果（Corcoran, 1995; Guskey, 1986; Guskey & Huberman, 1995），可能的原因是缺乏清楚明確的目的或目標。試想專業發展活動如不知目標為何，又如何評估其成效。

近年來，美國致力於教師專業發展評鑑結合學生學習模式的探究，如田納西加值評量系統即屬之，期盼學生的學習評量結果能與教師專業發展之間產生直接的關聯。本文首先介紹美國田納西加值評量系統（Tennessee value-added assessment system, TVAAS）；其次，論述其對我國推展教師專業發展評鑑之啓示，最後作一結語。

貳、田納西加值評量系統之介紹

以下分別就加值評量系統的源起、內容、實施、優缺點、應用部分加以說明。

一、源起

加值評量系統（Value-added assessment system, VAAS）是統計專家山德斯（William Sanders）¹所發展出來的一種借助統計方法，直接檢驗學生學業成就，間接對教師教學效能作出判斷的教師評鑑方法（Center for Greater Philadelphia, 2008）。加值評量系統起初並未引起注意，直到 1992 年，田納西州最高法庭要求學校提出一套能將學生學業表現與班級或學校聯結的績效責任制度，山德斯的學生學業成長模式才引起重視，且成為《田納西州教育改進法》（Tennessee's Educational Improvement Act）的一部分（Center for Greater Philadelphia, 2008）。此後，美國許多教育工作者及政策制定者逐漸認同加值評量系統的優點，除田納西州外，目前有超過 21 個州 300 個以上的學區使用山德斯的模式，賓州（Pennsylvania State）和俄亥俄州（Ohio State）甚至全部學區都使用此系統（Hershberg, 2005a, 2005b; Hershberg, Simon, & Lea-Kurger, 2004）。

二、內涵、特色與規準

山德斯所發展的教育加值評量系統（educational value-added assessment system, EVAAS），是美國北卡羅來納州凱利郡（Cary, North Carolina），賽仕電腦軟體（股份）公司（SAS Institute Inc.）的方案之一。加值評量系統是統計方

¹ 山德斯為美國田納西大學生物統計學及量子遺傳學博士，擔任過田納西大學加值研究和評量中心主任，過去 30 年來，他一直是統計顧問和農業、生產、工程和發展工業的顧問。現在是美國北卡羅來納州凱利郡，賽仕電腦軟體（股）公司加值評量和研究之負責人。

法與電腦複雜的組合模式，將預估的學生測驗分數與該學年末所取得的實際分數相比，分析這兩種分數間的差異，藉此瞭解教學對學生學習的影響。再者，此系統也分析各班級、各學校或各學區之測驗分數。各班級測驗分數是以該班所有學生 3 年來的平均分數為主，而學校或學區的分數乃是將各班級的分數加總而成（aggregation）（Tucker & Stronge, 2005）。加值評量系統最大的特色是能將學生每年學術成長（growth）歸因為學生本身或學校兩部分，其分析乃是根據每位學生前 3 年在不同學科的學業成績成長數據，來估算不同班級、學校、學區的學生是否達到，抑或超越、落後該年度的成長值。

一般而言，學生的表現可以用不同的方式加以評量，但重要的是成就與成長。「成就」（achievement）指絕對的等級（absolute levels），從該學年結束時測驗取得；「成長」（growth）則指相對的等級，是該學年結束時學生測驗分數進步的情形。過去常僅根據「成就」將學生和學校加以排名，但此方式最大的問題是忽略學生家庭社經地位與學業成就間的關係。例如，根據教育測驗服務中心（Educational Testing Service）的研究顯示，學術才能測驗（scholastic aptitude test, SAT）的得分與家庭收入之間具有高度相關，家庭收入也會影響學生就業機會、教育年限、教育態度、圖書量、旅行機會等。相對地，家境富裕的學生進入學校時，就已擁有較多影響發展的社會和能力資本（引自 Center for Greater Philadelphia, 2008）。相較於傳統的成就評量方式，加值評量系統對學生較為公平，是一種準確的診斷工具，此系統排除學生一些較為固定的特徵因素（如種族、社經地位）之影響，僅根據學生先前的學術記錄，每位學生只和自己比較，即以自己作為比較的基準點（Hershberg, 2005a）。

山德斯提出以下三點加值評量系統的規準（Sanders, 2003）：

（一）測驗內容：測驗的內容必須與課程目標、教授內容高度相關；測驗範圍必須能測量到極低和極高二極端之學生的成就發展；測驗內容必須有適切的可信度。

（二）統計標準：統計系統必須根據每位學生的測驗結果，提出多元與縱貫性的資料分析；必須能精準預測學校和個別教師對學生學術成長的影響比率；系統必須能分析各種不同來源的評量數據。

（三）提供診斷機會：縱貫性資料的加值分析，除了能提供績效責任的測量外，還必須能診斷出學區或學校低成就與高成就學生成長的比率，使教育工作者能修正課程與教學，讓不同層次學業成就的學生有適當成長的機會。

有些研究（Sanders & Horn, 1998; Carey, 2004）指出，加值評量系統在評量

學生學習成長方面，是一套運用高度技術及構思縝密的統計方法，其產生的縱貫型的資料庫，既能檢視教師及學校對學生學習的影響，又能提供最佳的研究工具和有用的資料。根據此系統的分析，教師教學效能對學生學習所產生的影響，甚於種族、貧窮、父母教育水準、或其他任何因素。換言之，教師是影響學生學業成就的最大因素。因此，優質的教師可以補救或減低社經背景等不利條件對學生學習的影響。

三、實際運作與預防措施

當學校系統決定實施加值評量系統時，必須運用適當的策略，以確保學校和社區已有足夠的準備來面對轉變。教育委員會的成員、家長、行政人員及教師，必須對加值評量系統有基本的認知，例如：如何分辨學生學業成就的進步情形、如何使用此系統、實施時各自的角色為何。再者，持續地與每一群體溝通，是每個學區順利實施加值評量系統的關鍵。各學區如欲實施加值評量系統，可以尋求某些單位的協助，以賓州為例，公共教育運作中心（Operation Public Education, OPE）為賓州大學之大費城中心（Center for Greater Philadelphia）所創立，該中心人員可協助各學區實施加值評量系統，學習如何運用此系統改進教學，並和北卡羅來納州凱利郡，賽仕電腦軟體（股份）公司之教育加值評量系統的專家合作，為教師、學校、學區、社區，提供專業發展和組織的支持（Center for Greater Philadelphia, 2008）。

在實際運作方面，從 1993 年開始，田納西州每年有系統地蒐集各學區所有學生的分數，要求各校以學區報告的格式，摘述三年級至八年級學生在閱讀、英語、數學、科學、社會等五學科的成績，而資料分析乃依據學生學業成績成長或進步的統計，不是固定的標準（fixed standards）。因為資料的蒐集來自每位學生，可獲得全州、各學區、各學校、各班的比較資料，包括各學區進步分數的預測與實際進步的分數、及全州實際進步的分數。再者，加值分析是一種多層次的過程，可提供教育工作者及決策者多種資料與數據。例如：在學生方面，包括：最近三年每學科進步情形；三年來的平均進步成績。在教師方面，包括最近三年來所有任教科目及任教年級之學生平均進步成績。在學校與學區方面，該學年度各學區學生在每學科及每年級平均進步成績；各校、各學區、全國的平均成績比較。田納西加值評量系統的資料庫可依學區的需要提供各種分析的結果，上述幾類結果的分析是最常見的（Tucker & Stronge, 2005: 71-72）。

田納西加值評量系統使用的資料來自全州實施的一項測驗方案，即「田納西州綜合評量方案」（Tennessee comprehensive assessment program, TCAP），適用於三至八年級的科學、數學、社會、英語及閱讀等學科，綜合來自特華諾瓦（TerraNova）的常模參照題目，以及由教師挑選出能充分反應田納西州學校課程之標準參照題目，這兩類題目具高度相關。使用「田納西州綜合評量方案」時，必須採用下列步驟來確認測驗的資訊（Tucker & Stronge, 2005: 83）：

- （一）訂出每名學生在每個科目可進步的分數；
- （二）根據學生過去的學業表現，比較學生的真實進步和預期進步；
- （三）將個別學生的進步分數編入全班、全校、全學區及全州的統計數據中；

（四）對照全校、全學區、全州、全國的平均進步，比較分項統計的資料。

綜而言之，田納西加值評量系統除了比較學生的進步分數與預測分數外，也將學校或學區學生的進步分數和全國的常模作比較。根據每科和每年級進步分數和全國常模進步分數的差距，學校可以得知是否達到全國常模的分數，而各校累計的平均進步分數乃作為評估該校教育成效與否的重要指標。田納西加值評量系統，從絕對的學習成就轉移到學習的進步，有助於學校傑出的教師得到大眾的認可（Tucker & Stronge, 2005: 77）。

為加強田納西加值評量系統的公平性，可採行的預防措施如下：

（一）學區、學校及教師效能評估，是根據最近三年來的評量資料，以確保統計數值的穩定性；

（二）學區、學校及教師效能評估不能僅根據此系統的資料，必須參酌其它資料；

（三）應用不同的媒體，如小冊子、報告、工作坊、錄影介紹等，幫助教育工作者瞭解如何解釋和使用田納西加值評量系統；

（四）「縮小值」（shrinkage）的估計是一種統計的方法，用來正確估計教師、學校或學校系統對學生進步的影響。

易言之，所有教師在所屬學校系統中的表現都是平均值，經過加權後，才是具體的估計值（Tucker & Stronge, 2005）。這類工具能使教師、學校或學區在測驗結果上免於受短期變動（如學生轉入或轉出）或者其它資料分析誤差所影響，而與實際測驗分數有所出入；

（五）學生每年必須接受新穎的、相等的且具高信度和效度的測驗，以確保每年的測驗結果是可比較的，但宜避免為測驗而不斷重覆練習某種特定題型

的試題，或為考試而教學的情形；

（六）被學校特教小組確認為特殊學生，將排除在教師效能分析之外；

（七）每學年出席日數不足 150 天的學生，其測驗分數不列為教師評量的資料；

（八）將維護測驗的安全性列為高度優先，如有不當性使用資料的行為將受到嚴厲懲罰，田納西加值評量系統也包括剔除可疑資料；

（九）因為貧窮、英語不利、父母教育水準及其他變項對每位學生每年的影響都相當固定，所以其影響估計值都不必列入（Tucker & Stronge, 2005）。

四、優缺點

根據塔克（Pamela D. Tucker）和斯壯（James H. Stronge）的分析，田納西加值評量系統的優點，可歸納如下（Tucker & Stronge, 2005）：

（一）以複雜的統計模式為基礎，能處理多達數年的學生成績；（二）被統計專家評為穩定、公平、可靠及有效的系統；（三）能順應學生的個別差異，把焦點放在進步分數而非固定標準的成績表現；（四）對於學生的進步以及教師影響學生學習結果的能力，提供相當簡單的評量方法；（五）田納西州綜合評量方案有良好的內容效度，它和田納西的學校課程有高度的一致性；（六）田納西加值評量系統測定的教師效能，和視導人員以主觀判斷的教師效能之間存在正相關。此外，此系統的公平性亦受到教師肯定，可作為個別教師層級績效判斷的依據，改變教學的實務和提升學生成就（Hershberg, Simon, & Lea-Kruger, 2004），亦可提供教育政策制定者作為提升教師教學效能、學校與學區辦學績效改進之參考（Drury & Doran, 2003）。

相反的，田納西加值評量系統具有以下的缺點：（一）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；（二）其模式假定各校及各班學生是隨機分配組成；（三）資料有被誤用或錯誤解釋的可能；（四）每年必須為學生的測驗，投入許多的時間、金錢和人力；（五）田納西州綜合評量方案，對目前複雜的教育只能提供有限的評量方法；（六）田納西加值評量系統所分析之結果（例如教師效能是影響學生最大的因素），目前尚無法從實證研究中得到證實（Drury & Doran, 2003; Tucker & Stronge, 2005）。此外，麥克卡菲立（Daniel F. McCaffrey）、克拉伍德（J. R. Lockwood）、可瑞茲（Daniel M. Koretz）、漢謨頓（Laura S. Hamilton）指出，此加值評量系統在田納西州雖然是非常有用的診斷工具，但若作為決定人事的依據，則可能有太多的不確定性和

不公平性，若僅據此評鑑教師效能，是非常不恰當的（McCaffrey, Lockwood, Koretz, & Hamilton, 2000）。貝瑞西（Gerald W. Bracey）（2004）亦指出此系統將學校教師排名，甚至比較整個學區的教師分數是不適切的。

雖然，部份學者對田納西加值評量系統提出許多的質疑和批評，但仍有學者呼籲：「縱使學校無法評量教師在課堂中，所做的全部事情，但不表示我們不應評量可以評量的事」（Tucker & Stronge, 2005: 82）。亦即，田納西加值評量系統雖有其限制性，但有其存在的價值，可作為教育工作者更一步瞭解教學實務，尋求改進的工具。

五、提升學校辦學的績效

美國《帶好每一位孩子》（No Child Left Behind, NCLB）法案通過，意謂著高利害關係（high-stakes）測驗紀元的來臨。學校行政人員背負著學校的績效責任，正面臨前所未有的挑戰。教室是學生學習的場所，但是督學和校長對它所知有限，迄今公立教育的系統尚無法提供適當工具有效地評鑑學校的績效。行政人員需要一套有效的工具，才能測量和評鑑課程實施、學生學業成就和教師專業發展的實務。誠如赫胥堡（Ted Hershberg）、賽門（Virginia Adam Simon）、李·可華葛（Barbara Lea-Kruger）所言，除非可以明顯地改進課堂教學和學習方案，否則學生將不可能達到高標準的要求（Hershberg, Simon, & Lea-Kruger, 2004）。田納西加值評量系統使用加值方法獨立出（isolates）教學對學生學習的影響，提供詳細的班級層級資訊，作為瞭解課程與教學問題有效的工具及預估學生適當年度進步（adequate yearly progress, AYP）的根據。

田納西加值評量系統主要分析學生測驗的預估分數，和該學年度結束時所真正測得的分數，如果真實分數高於預估分數，則屬高效能的教學，反之亦然。同時，此系統也考慮其他相關因素，如學生先前考試得分的模組（pattern）與同班學生的分數狀況。如果同班同學的學業都比以前進步，而某位學生的學業成就卻表現低於預估的分數，這可能是該名學生個人因素所造成，非受該班教師教學效能所能影響。加值評量系統強調教學該如何有效地促進學生每年度的成長，根據此系統分析的數據，教師可以定期開會討論如何改變教學方法，形成學習組織的社群。再者，適當年度進步測量如一種模糊不清的快照（snapshot），雖無法提供較具體優勢和弱點的圖像，但此系統聚焦於個別學生和縱貫性資料的分析，能幫助教育工作者降低學生成就的落差，評估教學革新的情形，進而改善教師教學和學生學習。

赫胥堡、賽門和李·可華葛以田納西州瑪利由市（Maryville）的瑪利由中學為例，作為學校運用加值評量系統改變教學實務和提升學生成就的成功案例（Hershberg, Simon, & Lea-Kruger, 2004）。根據加值資料分析和診斷，指出瑪利由中學七年級的學生，數學成就測驗得分最低的 20 位學生，相較於全州基準點的 15 級分（15 scale score points），平均只獲得 12.5 級分。校長葛宜芬（Joel Giffin）和該校教師藉由學生背景的瞭解，發現這些學生具有以下共同的特徵（引自 Hershberg, Simon, & Lea-Kruger, 2004）：

（一）來自低社經地位的家庭；

（二）單親家庭；

（三）放學後沒有大人在家；

（四）沒有錢買學校用品；

（五）缺少協助者或完成家庭作業的積極誘因。前三項是學校無法改變的事實，該校教師只好從後二項加以改善。首先，募款讓這些學生可以買學校用品。

其次，增加在校期間的數學課程，以幫助這些學生完成作業、提供輔導、給予鼓勵和獎勵、培養自信心及確定他們已作好上課的準備。經過學校的輔導，這些學生有顯著的進步且測驗結果表現傑出，加值分析後，其得分是全國常模的 360%，相較於全州基準點是全國常模的 100%而言，這是非常大的成就。

加值評量系統本身雖不能改進學生學業成就，但由上述的例子說明，當教育工作者瞭解它的功能，運用所獲得的資料來改善教學，和促進教師專業發展，就能夠提升學生學習成長。加值評量系統在美國各學區推行多年，也造成各公立學校以下的改變（Hershberg, Simon, & Lea-Kruger, 2004; Hershberg, 2004）：

（一）學科和學年具體的專業發展模式，幫助教育工作者改變其教學實務和改進學生學習，及提供師資培育單位作為培育教師的參考。

（二）研究顯示，未來 4—5 年內，所有學校將無法達到《帶好每一位孩子》法案的目標，加值評量系統在此歷程中可能扮演主要的角色，亦即，提供另類的測量，確認學校目前預估的適當年度進步曲線，以建立成長的軌道，使學生在往後幾年的學習能達到精熟的程度，此也被稱為另類的「成長標準」。

（三）多數教師和行政人員肯定加值評量系統的優點，願意接受其作為評鑑個別教師層級績效之依據。

六、結合教師評鑑與專業發展

田納西加值評量系統的基本目的在滿足《田納西州教育改進法》對績效責任的要求，其作法是確認教師、學校、學區對促進學生學習成長方面，是否達到前3年所預估的成長值。雖然，田納西加值評量系統的資訊不作為評估任何層次效能的唯一指標，但是因為學校及學區的資訊會被告，因而造成大眾要求持續改進的社會壓力。在教師層次方面的統計資料不會公告，只是提供教師及其主管單位作為擬定個人專業發展計畫的參考（Tucker & Stronge, 2005: 73-74）。此系統如何和教師評鑑結合呢？現行的《田納西州地方評鑑模式》（Tennessee State model for local evaluation）始於1988年，是根據一套教師效能研究而訂定的最低標準能力和指標。傳統的評鑑程序包括觀察前會議（pre-observation）、完成被認可的評鑑文件、觀察後的會議。田納西州教育委員會於1995年，要求重新評估此模式，並引進《全國初任教師標準》（national standards for beginning teachers），修訂其評鑑過程，強調運用多元方法、根據全國標準指標、和學生學業表現資訊來評鑑教師。

田納西州亦參考各州甚至加拿大的評鑑模式之辦法，設計該州的《評鑑和專業成長綱領》（Framework for evaluation and professional growth），此綱領的評鑑重點在評量教師發展和改進學生學業表現的能力，並落實以下三項目的：

- （一）績效責任：確定班級教學與學校效能；
- （二）專業成長：為某個領域的專業成長提供焦點，聚焦於提升教師的能力，以改進學生學業表現；
- （三）凝聚學校結構：增加並聚焦校內如何改進學生服務的對話（Tucker & Stronge, 2005: 133-134）。

此綱領主要的評鑑包括「綜合評鑑和專業成長」（comprehensive assessment and professional growth）及「焦點評鑑與專業成長」（focus assessment and professional growth）。「綜合評鑑和專業成長」適用於評估新任（實習）教師，但也適用於需要視導人員或行政人員提供結構性訊息的有經驗教師，此類評鑑目的在提供教師專業表現的完整圖像和未來成長的焦點。「焦點評鑑與專業成長」，則只雇用有專業執照的專業人員，根據先前評鑑、教師自評、學生學業表現的資料，確認目前的表現水準，再加上行政人員所提供的建議，擬定教師專業成長計畫。教師的專業成長計畫，必須包括以下內容：

- （一）根據不同評量工具所蒐集到的學生成績資料及績效標準擬出該加強

的領域；

- (二) 陳述專業成長的目的和目標；
- (三) 摘要行動計畫，包括完成的時間表；
- (四) 確認計畫實施後，用於衡量進步和成長的評鑑方法和規準；
- (五) 以教師專業成長對學生學業表現的影響為重點，說明該計畫的預期效益 (Tucker & Stronge, 2005: 135-136)。

七、作為教師評鑑的隱憂

雖然，田納西州採用田納西加值評量系統作為教師評鑑的指標之一，然而貝魯 (Dale Ballou) (2002) 指出以此系統作為教師評鑑的隱憂。因為他認為加值評量系統的價值是用於達成診斷目的，而不宜作為考核績效的依據。教育單局如以此系統來評鑑學校辦學績效和教師教學效能，並伴隨著人事的決定，甚至用於支持地方決策，將是非常不恰當的作法。貝魯甚至認為此評量工具是不明確的，不可靠的，根本無法達成評量目的，更遑論以此作為教師評鑑的依據。貝魯所持的理由如下：

第一，測驗本身不是學生能力或成就測量的理想工具：學生在特定測驗上的表現，部分由於真實能力，部分乃受隨機影響 (random influence)，例如：測驗期間的分心、當天的情緒狀況、測驗題目選擇等。當我們企圖去評量學生的成就等級，測驗所產生的誤差足以造成問題。

第二，學生的測驗分數可能受其他因素所影響，而非僅來自於學校教學的品質：加值評量系統最大的優點，乃是基於學生個人學習成長，而非成就等級。事實上，加值評量系統雖然一開始就將家庭因素、種族、學生初始成就等級列為控制變項，但還是不夠。因為相同的因素（如家庭因素）可能影響學生初始的成就等級，也可能影響學生進步的比例。除非加值評量系統能明顯地控制所有影響學生學習的因素，否則無法說明學生學習成長乃是學校辦學績效或教師教學效能所造成的結果。

第三，類似的測驗分數不見得能相互比較：加值評量系統必須能以有意義的方式比較不同學生的成就，也需要確認測量學生的成就所用的量尺 (scale) 是一種相等的單位。換言之，當學生的成就測驗分數，不論從 15 分到 20 分，或是 25 分到 30 分，每增加的 5 分必須是代表相同的意義，否則，無法以學生進步的分數作為教師效能或學校績效的依據。基此，以此系統分析的結果作為教師和學校相關效能 (relative effectiveness) 的推論即屬謬論。

基於上述三點理由，貝魯指出以田納西加值評量系統作為教師評鑑的指標是不恰當的。雖然，加值評量系統背後的理念非常單純，認為只要根據學生進步來評鑑學校和教師效能即可，事實上，此系統還有很多無法控制的變數，若以此作為教師評鑑的依據是令人隱憂的。

麥克卡菲立、克拉伍德、可瑞茲與漢謨頓（McCaffrey, Lockwood, Koretz & Hamilton, (2000) 亦指出，田納西加值評量系統恐無法解決績效的問題，因為有太多不確定性和不公平性，同理，若據此評鑑教師效能是不恰當的。貝瑞西（Gerald W. Bracey）（2004）也認為田納西加值評量系統，對教師效能採常模參照測量，將各校的全體教師排名，但教師效能的解釋卻又採標準參照，相互矛盾。事實上，常模參照測量將整個學區的教師分數進行評比並不適切，因為能力較差的教師在較差的學區中，比同能力的教師在較強的學區中，可能得到較有利的排名。

參、對我國教師專業發展評鑑之啓示

一、以「學生學習成長」作為教師專業發展評鑑的核心概念

雖然，田納西加值評量系統作為教師評鑑有其限制性，但也提供我國推展教師專業發展評鑑另一種思維，教育工作者常困惑於教育現場有太多無法掌控的因素，有太多無法改變的事實，特別是對於社經背景不利，家庭問題眾多的學生，其衍生學習和行為上的問題，常讓許多教育工作者感到無能為力。但從前述學校運用田納西加值評量系統改變教學的實務，和提升學生成就的成功案例中，我們可以清楚地看到，學校雖然不能解決所有的社會問題，但應該給每位孩子每年合理的成長期望，透過學校與教師的努力，孩子的學習與成長會有顯著的進步。

我國教師專業發展評鑑從呼籲、理解、尋求共識、小規模嘗試、整合意見到實際試辦，前後歷經 10 年，除了顯示教師專業評鑑的重要性，也凸顯教育部對實施教師專業評鑑審慎與重視的態度。經過 10 年的嘗試與溝通，教育部確立以「教師專業成長為主軸」的教師評鑑規劃，提出「提升教師專業素質、建立教育評鑑制度」的呼籲，其精神與田納西州一致，田納西州評鑑的目的在於提升學校辦學績效、增進教師專業成長、增加和聚焦校內專業對話。然而，該州採用明確的參考標準，要求以田納西加值評量系統作為教師評鑑的指標之一，

將學生的學業成長作為教師評鑑與專業成長的核心，及擬訂教師專業發展計畫的依據。

前已提及，若干研究指出，透過加值評量系統分析出教師效能是影響學業成就最大的因素。因此，高品質教師可以彌補社經背景不利條件對學生的影響，對於孩子的學習成長，學校有太多可以著力的空間與希望。學校因「學生」而存在，而提升「學生的成長」是推動一切教改的理由，也是教師專業發展評鑑的核心價值。因此，本文建議將學生學習成長列為判斷教師專業發展評鑑成效的重要指標。

二、建置「台灣學生學習成就評量資料庫」，使學生學業成長、教師教學效能、與專業發展產生聯結

田納西加值評量系統是以學生學業成長為評量的重點，強調每位學生只和自己競爭，而非彼此的競爭，這種理念對社經背景不利的學生有積極的鼓勵作用，使教師和學生都有共同努力的目標。然而，如欲推展田納西加值評量系統需要有以下三種條件：第一，測驗內容必須與課程目標、教授內容高度相關；第二，必須能根據每位學生的測驗結果，提出多元與縱貫的分析，精準地預測學校和教師對學生學術成長比率的影響；第三，必須診斷出學區或學校低與高成就學生成長的比率，使教育工作者能修正課程與教學，使不同層次學業成就的學生有適當的成長機會。

台灣是否有適合的資料庫可推展田納西加值評鑑系統呢？目前國家教育研究院籌備處正積極針對國民中小學學生學習成就建立常態性的資料庫，稱為「台灣學生學習成就評量資料庫」（Taiwan assessment of student achievement, TASA），為能與相似資料庫進行比較，並參考國外資料庫建置之核心科目，其施測科目為國文、英語、數學、自然等四科，以測量學生的語文與數理能力為主要考量，故將施測科目分為語文能力與數理能力兩大類，其中語文能力包括國文與英語兩科目，數理能力包括數學及自然兩科目。為能完整測量學生的各項學科能力，每位學生將接受上述四種學科能力測驗，並填寫學生背景問卷及各科問卷，以利不同學習內容指標的建立及測驗結果的交叉分析。

該資料庫的目的在於：（一）建立國民中小學及高中職學生學習成就長期資料庫，以追蹤、分析學生在學習上變遷的趨勢，進而檢視目前國家教育體制與政策實施的成效；（二）瞭解國內學校教育的現況，作為教育部研訂課程與

教學政策，縣市政府教育局（處）及學校推動補救教學的重要參據；（三）分析學生在學習成就上的表現差異，藉此評估學生未來在學術方面的發展與社會期許；（四）提供完整、標準化的量化資料，以供國內外相關研究人員深入探討學生學習成就的相關政策議題（國家教育研究院籌備處，2008）。

我國在建置「台灣學生學習成就評量資料庫」的同時，若除了根據九年一貫課程綱要課程能力指標發展標準化測驗工具，測驗學生個人的學業成就外，也能蒐集學校與班級層級的資料，又能在系統的設計上掌握以下三大原則：第一，為每名學生在每個科目的測驗訂出進步的分數；第二，根據學生過去的學業表現，比較學生的真實進步和預期進步；第三，能將個別學生的進步分數編入全班、全校、全縣及全國的統計數據中，對照全校、全縣、全國的平均進步，比較分項統計的資料。如此一來，透過加值評量系統的分析，學生評量結果就能提供各校各班明確的資訊，與教師效能間產生直接的關聯，進而作為擬定教師專業成長計畫的依據。

然而，此系統可能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、其模式假定各校及各班學生是隨機分配組成、資料有被誤用或錯誤解釋的可能等。我國如欲實施此系統必須在系統的設計上避免上述缺失，此有賴測驗專家與統計專家的共同努力。在實施過程必須謹慎規劃，確保學校和家長有足夠的準備，包括如何分辨學生學業成就的進步、如何使用此系統、實施時各自的角色等。再者，必須有專責單位為教師與學校提供專業發展訓練和系統的支持。

三、將田納西加值評量系統之評量結果納入教師專業發展評鑑

專業發展活動是有意義、有目標的過程，不是為改變而改變，其模式是多元的，然而專業發展活動必須展現努力的成果，蒐尋相關資料，並加以分析，提供更明確改進的訊息，使專業發展方案更有提升教學效能的價值，以引導教師達到專業發展的目的。

田納西加值評量系統，雖不能改進學生學業成就，但教師如果能夠瞭解它的功能，運用此系統分析的資訊，改善自身的教學，增進自我的專業發展，必能改進學生的學業成就。換言之，教師宜將學生的學業成長作為專業發展的目標，根據此系統所分析的結果，找出造成學生無法達成預期成長的因素，進一步瞭解哪些因素是學校全體師生必須共同努力去解決的？哪些是教師個人所應改善的？如果屬教師個人所應強化的領域，學校和教師必須共同擬定專業發展

方案，其內容包括完成的時間表、評量成效的方法和規準，及預期達成的效益等。再者，此系統能根據學生進步的情形，指出各校教學效能良好的教師，這些高效能的教師可以分享其教學經驗與策略，學校亦可根據此系統分析的結果，形成各類教師學習組織社群，定期開會共同討論如何改善教學。

然而，前已提及，有學者認為此系統的價值乃用於達成診斷目的，但不宜作為考核績效的依據。教育單局如以此系統來評鑑學校辦學績效和教師教學效能，並伴隨著人事的決定，甚至用於支持地方決策，是非常不恰當的作法。因此，在許多疑慮尚未解除之前，我國如欲採用此系統，必須以診斷教學，提升教育品質為目的，不宜作為評鑑學校績效與教師效能之依據。

肆、結語

教師是教室的靈魂人物，其教學的品質牽動著學生學習的興趣與成效，而教師專業發展是教師持續學習與研究的具體展現。近年來，我國為協助教師專業成長，增進教師專業素養，提升教學品質，從呼籲、理解、尋求共識、小規模嘗試、整合意見到實際試辦，可見教育部對實施教師專業評鑑之審慎。然而，教師專業發展評鑑必須提供更明確的改進訊息，方能提升教師的教學效能，達成專業發展的目的。美國田納西加值評鑑系統是一種教學效能的診斷工具，可作為教師改進教學及專業發展之依據，其作法可提供我國推展教師專業發展評鑑的參考。此評量系統可以清楚看到學校應順應學生個別差異，將評量置於學生的進步而非固定標準的成績表現，應給予每位孩子每年合理的成長期望，並確認每位孩子都能達成其預估的學業成長，並將提升學生的成長視為教師專業發展的核心價值與評鑑成效的重要指標，使學生的學業成長、教師的教學效能與專業發展間產生更緊密的聯結。

參考文獻

- 國家教育研究院籌備處（2008）。台灣學生學習成就評量資料庫。2008年10月2日，取自 <http://tasa.naer.edu.tw/brief.htm>
- 教育部（2006）。教師專業評鑑宣導手冊。台北市：作者。
- 教育部（2007）。試辦中小學教師專業發展評鑑計畫說明。2007年9月2日，

- 取自 http://www.chehjh.kh.edu.tw/others/96teacher/BE_96.ppt
- Ballou, D. (2002). *Sizing up test scores. Summer 2002/ education next*. Retrieved November 1, 2008, from <http://www.educationnext.org>
- Bracey, G. W. (2004). Serious questions about the Tennessee value-added assessment system. *Phi Delta kappan*, May, 716-717.
- Carey, K. (2004). The real value of teachers: Using new information about teacher effectiveness to close the achievement gap. *Thinking K-16*, 8(1), 3-32.
- Center for Greater Philadelphia (2008). *Value-added assessment services*. Retrieved November 1, 2008, from http://www.cgp.upenn.edu/ope_brochure.html
- Corcoran, T. B. (1995). *Transforming professional development for: A guide for state policymakers*. Washington, D. C.: National Governors' Association.
- Drury, D., & Doran, H. (2003). The value of value-added analysis. *Policy Research Brief*, 3(1), 1-4.
- Guskey, T. R. (1986). Staff development and teacher change. *Educational Researcher*, 15 (5), 5-12.
- Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks, California: Corwin Press, Inc.
- Guskey, T. R., & Huberman, M. (1995). *Professional development in education: New paradigms and practices*. New York: Teachers College Press.
- Hershberg, T. (2005a). *Value-added assessment: Powerful diagnostics to improve instruction and promote student achievement*. Retrieved October 1, 2008, from http://www.cgp.upenn.edu/ctr_pubs.html
- Hershberg, T. (2005b). Value-added assessment and systemic reform: A response to the challenge of human capital development. *Phi Delta Kappan*, 87(4), 276-283.
- Hershberg, T., Simon, V. A., & Lea-Kurger, B. (2004). The revelations of value-added: An assessment model that measures student growth in ways that NCLB fails to do. *School Administrator*, 61(11), 11-14.
- McCaffrey, D., Lockwood, J. R., Koretz, D., & Hamilton, L. (2004). *Evaluating value-added models for teacher accountability*. An education report by the Rand Corporation prepared for the Carnegie Corporation. Retrieved October 2, 2008, from <http://www.rand.org>
- Sanders, W. (2003). *Beyond no child left behind*. Retrieved October 2, 2008, from <http://>

[/www.sas.com/govedu/edu/no-child.pdf](http://www.sas.com/govedu/edu/no-child.pdf)

Sanders, W., & Horn, S. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256.

Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning*. Alexandria, Virginia: Association for Supervision and Curriculum Development.