

大學通識教學評鑑的後設評鑑研究

林劭仁

基於後設評鑑系統化特質及對評鑑專業發展的重要性，本研究旨在進行大學通識教學評鑑之後設評鑑，探討通識教學評鑑符合後設評鑑標準的程度。以一所國立大學師生為研究對象，研究方法先以問卷調查 425 學生，並配合教師焦點團體座談蒐集資料。主要結果發現在學生方面：學生自信能勝任教學評鑑工作、大致能掌握教學評鑑的內容、冀望評鑑結果能適度公布；在教師方面則認為教學評鑑有其限制，提升教學品質還是要聚焦於教師專業與教學知能提升、應發展適用不同領域學科的版本，以及教師對教學評鑑態度仍待改變等；整體上師生均肯定後設評鑑的重要性。研究最後亦根據後設評鑑結果，提出後續推動通識教學評鑑的具體建議。

關鍵字：教學評鑑、後設評鑑、學生評鑑教師、通識教學

作者現職：國立臺北藝術大學藝術與人文教育研究所副教授

通訊作者：林劭仁，e-mail: srlin@academic.tnua.edu.tw

壹、前言

大學通識教育兼重「廣博」與「統整」的價值，強調學生在博雅教育薰陶下養成自我探索與多元批判能力，並提升文化與人文素養，達成全人教育的理想。而新世紀通識教育更由普及邁向深化卓越，加上歐美國主要大學對通識教育課程持續進行變革，也引起國內大學對通識課程與教學的探討。民國 101 年第二輪的系所評鑑中，通識教育也成爲此波評鑑重點，未來在大學追求學術競爭力與教學卓越的氣氛下，通識教育從課程規劃到教學實踐都必將越來越受到教育工作者的重視。

通識教育的目標需透過教學活動來實踐，故教師的教學品質實扮演重要角色。中華民國通識教育學會(1999)曾主張通識教學評量可從對學生的評量，以及對教師的評量兩方面分別進行(引自蕭揚基，2005)。前者指的是教師在教學過程中評量活動的實施；而後者指的就是教師的教學評鑑。近年來教學評鑑已廣爲應用在國內、外大學校院中，Wilson(1998)就推測大學終將會全面運用學生評鑑教師來提升教學品質。而國內至 2001 年約有近 83%的學校已實施學生評鑑教師制度(張德勝，2005)。近幾年除了於教育部委託專業評鑑機構實施校務、系所，及各種專案評鑑外，大學教學卓越計畫的推動，也促使大學開始更有系統的推動教學評鑑(鄭博真、黃義良，2009)，並陸續規範於校內教師評鑑相關法令中。目前國內在教學評鑑上相關研究頗多，特別是評鑑量表的發展，因是學生評估與判斷教師表現的重要工具，諸多學者投入發展出兼顧信、效度考驗的量表，皆具有實證基礎(林珊如，1999；孫志麟，2004；張郁雯，2003；葉連祺、董娟娟、楊世英、陳仁海、蕭芳華，2005；湯誌龍，2006；黃義良、鄭博真，2011；潘靖瑛，2004；鄭博真、黃義良，2009)。

除了教學評鑑工具的研發外，評鑑制度本身也應完整的被評鑑，即爲後設評鑑(meta evaluation)實施的重要。站在「評鑑的評鑑」的角度，對原級評鑑進行價值的判斷，以了解其優缺點，提出改善的建議。也因爲後設評鑑具有評估評鑑績效、改善評鑑技術、監控評鑑過程、呈現評鑑現況等功用，爲有效達成評鑑目的，後設評鑑會是評鑑發揮功效必須具備的條件，也是評鑑專業化的重要關鍵(Worthen, Sanders, & Fitzpatrick, 1997)。

因此，若以後設評鑑的觀點評估通識教學評鑑，則核心問題至少包括：整體上評鑑制度是否合理適用？評鑑是否真能反映通識課程的教學情形？程序上評鑑過程是否顧及所有利害關係人的看法？評鑑實施程序或方法上是否適合，以及教師態度是否會影響教學評鑑？學生是否能夠認真而理性作答等。本研究企圖以一所國立大學爲個案，運用後設評鑑來探討通識教學評鑑的相關議題，希望能深入了解問題焦點，再與相關研究結果印證，而較不考量多校間比較。據此，研究目的除分析大學通識教育教學評鑑與後設評鑑之實施內涵與方式

外，最重要是對單一原級評鑑進行後設評鑑，檢視其通識教學評鑑實施情形，並提出結論與建議。

貳、文獻探討

一、通識教學評鑑的內涵

教學評鑑是以教師為評鑑對象，而以教學為評鑑範圍，透過回饋式的程序，對教師教學與學生學習給予價值的判斷（黃政傑，1999）。故教學評鑑應包括那些內涵？曾淑惠(2002)曾將評鑑分為規劃、實施、結果運用三階段，完整的說明了評鑑工作內涵。規劃階段工作包括成立評鑑規劃團隊、定義並確認評鑑關係人、利益衝突及問題、背景分析、獲得評鑑關係人正式同意、選擇評鑑人員、取得經費預算，及規劃評鑑程序；實施階段包括目的與程序管控、建立評鑑標準及資訊蒐集策略、發展蒐集工具、資料管理與分析、溝通評鑑發現、依標準形成判斷、導出結論與建議、撰寫報告；結果運用階段包括公布結果、報告推廣回饋、效益分析、評估影響與衝擊、進行後設評鑑。故運用至教學上，完整教學評鑑的內涵至少應包括教學準備、教學進行、教學評量三大內容，並具體評估教師教學準備是否充分、教學進行是否適當，以及教學結果是否有效（林天祐，2006）。

既然以教師為評鑑對象，為顧及客觀及倫理性，教學評鑑的資料就不能只依靠一、二次的橫斷資料，包括專家指導、同儕惕勵、學生及家長回饋等，都必須盡量蒐集以為判斷依據，故像是校友評鑑、同儕評鑑、檔案紀錄、教師自我評鑑，以及學生評鑑教師等皆為可行方式（蕭揚基，2005）。不過因教學評鑑品質不易量化，難脫離主觀判斷方式，若將教學評鑑作為教學改進之用比較不會有爭議，但如果做為教師教學績效評量之用，受限於評鑑的目的與形式，則應以做為「門檻」為宜（林天祐，2006）。

由於學生為教學主體，教師教學成效對學生的影響也最深，故不少研究提出學生評鑑教師是教學評鑑最主要、最直接、也是最有效方式（湯誌龍，2006；潘靖瑛，2004；Marann, Barbara, 2003; Tollefson, Chen, & Kleinsasser, 1989）。至於應有的內涵或項目，陳琦媛（2007）歸納國內 41 所大學學生評鑑教學工具，發現教材內容、教學方法、教學態度、教學能力、成績評量，與教師勤惰是主要內涵項目。在通識教育上，孫志麟(2004)認為通識教學品質必須考量教學準備、教學內容、教學方法、教學評量、教學態度、教學溝通、教學責任，與教學效果等八個項目。與蕭揚基（2005）的教學態度、教學方法、教材內容、教學設備、學習態度等五向度；黃義良、鄭博真（2011）的教材教法、師生互動、學生評量和學生學習四構面；及潘靖瑛（2004）的教學設計與準備、教學方法、

主題文章

教學態度、學習評量，及課程總評，有頗多相似之處。不過潘靖英還因應一般課程、藝術課程，與體育課程的不同而有所區隔。整體而言，包括教學準備、教學態度、教學方法、教學內容、教學設備、教學溝通、教學評量等，皆是通識教學評鑑的重要內容。

二、後設評鑑的理念與模組

最早提出後設評鑑此名詞的 Scriven 認為後設評鑑就是所謂「第二層級評鑑」(second-order evaluation)，即「評鑑的評鑑」。將原來評鑑者變成受評者，對該評鑑活動及評鑑者表現進行價值判斷(蘇錦麗、楊瑩、王偉中、呂鴻德、詹惠雪，2001；Scriven,1981; Scriven,2001)。林劭仁(2007)整理學者定義，認為後設評鑑就是對一項進行中或已完成評鑑進行價值判斷的歷程，有系統的描述問題、蒐集資訊和分析資訊，並重視原級評鑑的效用性、可行性、適切性、精確性。因此，後設評鑑對決策者而言，可協助其確認方案評鑑的結論，評估方案評鑑的績效，以獲得新的評鑑結果和知識，作為選擇變通方案的參考。對評鑑研究而言，後設評鑑可協助獲得科技整合的學術訓練，增進進行評鑑時的專業能力，擴展評鑑研究的領域(游家政、曾祥榕，2004)。

Stufflebeam 提醒後設評鑑重要概念，除必須蒐集及評估原級評鑑的重要訊息，包括原級評鑑計畫、評鑑工具、相關資料、評鑑報告，及評鑑者的公正性等。也必須認真考量運用訪談、調查等方式蒐集相關人員的意見，最後提出對整個評鑑敘述性及判斷性的訊息與建議(Stufflebeam,2001a)。而評鑑人員應該要能從後設評鑑的執行過程學得分析評鑑的經驗與能力，且能在未來工作中持續改進評鑑實務(Sanders, 1999)。

實際推動後設評鑑時，可採以下程序或步驟：1.安排與利害關係人互動、2.組織專業的后設評鑑團隊、3.定義後設評鑑的問題、4.確認將使用的後設評鑑標準、5.建立規範或約定書以利評鑑管理、6.蒐集適當可用的訊息、7.必要時可利用實地訪視、訪談、觀察與調查等方法蒐集資訊、8.分析量化與質化的資訊、9.逐一分析評鑑與標準的符合情形、10.利用書面或口頭報告傳達後設評鑑發現、11.必要時協助利害關係人解釋或應用評鑑發現(Stufflebeam,2001a)。這些步驟提供後設評鑑推動時有所依循，頗具實務參考的價值。

潘慧玲、王麗雲、張素貞、鄭淑惠、林純雯(2007)曾整理後設評鑑關心的問題與面向，區分為四個模組；分別是診斷性、改進性、結論性，解釋性模組。診斷性模組主要關切原級評鑑的設計與實施，及判斷符合計畫執行的程度；改進性模組用以判斷符合評鑑標準的程度；結論性模組在分析評鑑辦理的結果；而解釋性模組則著重於界定影響評鑑結果的因素。此分類涵蓋了後設評鑑從規劃設計、執行，到結果的實施歷程，也說明推動後設評鑑的方法與型態，及進行價值判斷之標的，不致模糊真正關心的焦點。

三、後設評鑑標準的應用

既然後設評鑑的重點在對原級評鑑進行價值判斷，那麼判斷的基礎就非常重要，此乃評鑑標準存在目的（林劭仁，2008；曾淑惠，2002；蘇錦麗，1997）。Stufflebeam (2001a)認為要有效將後設評鑑問題聚焦，關鍵方法在於將現實狀況與專業標準相互比較。故進行後設評鑑時，評鑑標準為不可缺少的重要工具。

美國「教育評鑑標準聯合委員會」(The Joint Committee on Standards for Educational Evaluation) 擁有二百多萬名會員並是 12 至 15 個專業團體認同的組織，且是「美國國家標準局」(American National Standards Institute, ANSI)認定在教育評鑑領域最具公信力的標準制訂團體。於 1994 年利用專家群體決策建立共識，發展「方案評鑑標準」(The Program Evaluation Standards)，分為效用性(Utility)、可行性(Feasibility)、適切性(Propriety)，和精確性(Accuracy)，共四類 30 項標準，為後設評鑑標準研究上最具權威與參考性的指引(Joint Committee on Standards for Educational Evaluation,1994)。而聯合委員會也為此標準發展出指導原則(Guiding Principles)，已廣為應用至諸多方案評鑑的實務中(Grasso,1999; Sanders,1995; Stake & Davis, 1999; Stufflebeam,2001b)。第三版的方案評鑑標準亦於 2011 年出版，修改及更新了部分標準及說明，並新增一類績效性(Evaluation Accountability)標準，下含三項標準，特別強調總結性及形成性後設評鑑的重要(Yarbrough, Shulha, Hopson, & Caruther,2011)。聯合委員會也建議方案評鑑標準的應用必須考量不同地區及國家的文化背景，不需要無批判或無條件的採用，才能符合評鑑時的實際需求。

本研究係探討通識教學評鑑的執行狀況，故原級評鑑為通識教學評鑑，再引用聯合委員會的方案評鑑標準，考量新版剛出版，仍待學界充分討論，故以舊版標準作為判斷原級評鑑符合程度的依據，屬於前述後設評鑑的改進性模組。但因調查對象為學生與教師，會分析評鑑參與人員對評鑑實施結果的意見與反應，故也具備結論性模組的特質。

參、研究方法

本研究以後設評鑑探討一所大學通識教學評鑑實施現況，除欲判斷原級評鑑符合評鑑標準的程度，也企圖分析評鑑辦理的結果，依潘慧玲、陳文彥(2011)的分類，屬於後設評鑑的改進性與結論性模組。研究設計上，先利用問卷調查蒐集學生自評表現，並對照評鑑標準，進行改進性與結論性模組；接著召開教師焦點團體座談，就調查結果提出意見以交互檢定，深入蒐集教師對評鑑結果的意見，屬於結論性模組。

一、研究對象

主題文章

以某國立大學學生為調查對象，母群體約 2490 人，採立意取樣於 100 年 11 月至 12 月二週期間，透過教師協助由研究助理進入教室發放問卷。最後發出問卷 482 份，抽樣比例約 19.36%，回收整理得有效問卷 425 份，有效樣本回收率約為 88.2%。

此外，研究者邀請學校教務長、承辦教學評鑑業務之課務組組長、通識教育委員會主委、教學與學習資源中心主任，於 100 年 12 月初舉行焦點團體座談，歷時約 70 分鐘，交叉驗證所蒐集資訊，也彌補問卷資料之不足。

二、研究工具

研究工具以聯合委員 1994 年版之方案評鑑標準為基礎，為使標準的敘述易被學生接受，研究者逐項調整修正，兼顧原有標準語意並符合學校通識教學評鑑實際狀況。最後刪除不適合學校現況的 4 題，保留 26 題，再以專家及學生提升內容效度，共邀請 10 位評鑑領域專家學者及該校師生（表 1），提供量表修正意見，研究工具各題項的調整情形請參考表 2。

表 1 量表修正成員一覽表

編號	服務單位/職稱	專長與背景
1	政治大學教育系/教授	教育評鑑、課程評鑑
2	台北市立教育大學教育行政與評鑑研究所/副教授	教育行政、教師評鑑
3	中正大學課程研究所/副教授	比較教育、教育社會
4	中原大學教育研究所/助理教授	教育行政、師資培育
5	該校副校長	通識教育、舞蹈教育
6	該校教學與學習資源中心主任	教育社會學
7	該校課務組組長	音樂教育、課程設計
8	該校藝術與人文教育研究所	研二學生
9	該校美術系	四年級生
10	該校音樂系	三年級生

表 2 研究工具各題項調整對照表

原方案評鑑標準	調整後之後設評鑑標準	主要調整方向
U1、確認評鑑利害關係人	有確認並涵括所有應參與對象並重視需求	
U2、評鑑者的可靠性	學生有能力勝任並填答評鑑問題值得信賴	界定評鑑者為學生
U3、資訊的範圍與選擇	問卷所蒐集到的資訊是很適切的	將資訊範圍界定在評鑑問卷
U4、價值的確認	我能清楚說明下判斷時的原因和價值基礎	
U5、報告的清晰性	刪除	評鑑報告沒有對學生公布

U6、報告的及時性與傳播	評鑑結果有及時提供給師生了解	當事人界定為師生
U7、評鑑的影響	評鑑結果是有影響力的且可持續運用	
F1、務實的程序	評鑑過程務實可行，並盡量減低干擾	
F2、政治的可行性	有徵詢與考量師生意見盡量避免爭議誤解	相關團體界定為師生
F3、成本效益	評鑑方式在資源運用上有效率	
P1、服務導向	評鑑方式能幫助師生獲得充分服務	相關參與者界定為師生
P2、正式的協議	刪除	教學評鑑無簽訂書面協定
P3、受評者的權利	評鑑設計有顧及教師的權利與福祉	受評者界定為教師
P4、人際間的互動	我在評鑑過程中會尊重受評教師尊嚴避免威脅與傷害	界定評鑑者為學生，受評者為教師
P5、完整和公正的評估	評鑑方式能夠完整且公正檢驗紀錄教學優缺點	受評方案界定為教學
P6、結果的公佈	有適切的公布評鑑結果讓師生獲得相關資訊	界定受評評鑑影響人為師生
P7、利益的衝突	遇到利益衝突時能公開且誠懇處理	
P8、會計的責任	刪除	教學評鑑沒有經費分配與開支
A1、方案文件	學校有正確而清楚的描述評鑑方式	
A2、情境分析	學校有分析並說明評鑑背景因素及可能造成的影響	
A3、目的與程序的描述	學校有清楚描述評鑑的目的與實施程序	
A4、具辯護力的資訊來源	刪除	資訊來源為學生
A5、有效的資訊	評鑑資訊是有效的可適當的解釋結果	
A6、可信的資訊	評鑑資訊是可信的	
A7、系統化的資訊	學校能有系統檢視蒐集的資訊並修正錯誤	
A8、量化的分析	學校能適切分析量化(問卷勾選意見)資訊	
A9、質化的分析	學校能適切分析質化(文字書寫意見)資訊	
A10、正當化的結論	學校有提出正當且可受公評的評鑑結論	
A11、公平的報告	評鑑結果公正不受個人情緒和偏見的扭曲	
A12、後設評鑑	整個教學評鑑方式也應該受到評鑑	

問卷分成二個部分，第一部分蒐集學生對共同科教學評鑑「適合度」意見；第二則想了解學生對共同科教學評鑑實施成效「同意程度」。信度分析採 Cronbach α 之內部一致性法。經計算後第一部分的 Cronbach α 係數為.943，第二個部分為.973，顯示兩問卷工具的內部一致性信度頗佳。

主題文章

焦點團體座談時為使會議進行時更為聚焦，研究者逐項說明標準的調查結果，請教師提供意見，以利交叉驗證。接著探討更深入的議題，聚焦於對現有通識教學評鑑量表的看法、學生評鑑教師的看法、是否滿意通識教學評鑑現況、有效的通識教學評鑑應如何實施等。

問卷之量化資料以 SPSS 18.0 進行分析，以遺漏值檢驗各題項學生遺漏情形，了解抗拒或難以回答情況；以因素負荷量(factor loading)檢測試題同質性；以百分比與平均數(Mean)呈現學生意見上的集中狀況；並以平均數偏離檢驗方式(邱皓政, 2006)，將各題項與整體平均數進行差異之顯著性考驗，增加判定認同程度的客觀依據。焦點座談資料則先由研究助理全程錄音並完成逐字稿，再由研究者整理與分析資料，以大寫英文字母依序將參與者編號，字母後則呈現發言時間，如(A1702)代表 A 教師於下午 5 點 2 分之發言。

肆、結果與討論

一、問卷調查結果與討論

(一) 教學評鑑各題項適合度分析

第一部分蒐集學生對現有教學評鑑每項題目適合度意見，並考驗平均數差異的顯著水準，結果如表 3。各題項遺漏值檢定結果最高僅.7%，顯示學生應能了解題意。學生自我檢視學習狀況共 5 題，發現第 2 題「我課餘每週花費於這門課時間」平均數(M=3.64)最低，差異達顯著($t=-13.09, p<.05$)。可能學生認為花費時間多，並不一定代表學習狀況就會好，故不想讓老師知道花費時間，以免作為判斷成績或學習效果的依據。

在對教學共 14 題的適合度意見上，平均數較高集中在前 4 題，像是「教師教學認真，不無故遲到早退或缺席」、「我覺得老師對本課程有充分的準備」、「教師授課或示範講解的表達與說明清楚有條理」等，差異皆達顯著，前半段題目的分數普遍高於後段。仔細分析可發現這些排序較前題目和教師教學都有直接的關係，包括教師的教學準備、教學態度、授課內容、教材使用、發問技巧等，都是針對教學實質內涵的題目。而後段題目重點則在對教師及課程的評價，或推薦程度，和教師教學表現較無直接關係。可看出學生認為重要教學評鑑題目集中於教學直接的表現，符合前面文獻探討提及教學評鑑應有的重要向度，顯示學生頗能掌握教學評鑑內容的重點，這對學生評鑑教師的可信度及學生成熟度皆有激勵作用。

表 3 學生對教學評鑑量表適合度之統計結果

題項	非常適合	適合	有點適合	有點不適合	不適合	非常不適合	平均數	t 值
	%							
一、學生學習狀況								
1.我修讀這門課的性質	18.0	47.6	24.1	6.1	2.2	1.9	4.67	7.571*
2.我課餘每週花費於這門課的時間	4.2	24.7	31.5	18.0	14.0	7.6	3.64	-13.09*
3.我上這門課的情形	15.0	41.8	29.1	8.5	4.0	1.7	4.50	4.755*
4.我上這門課的態度	9.8	41.9	33.4	9.3	3.9	1.7	4.39	2.137
5.我對自我學習上的要求	7.6	43.7	34.4	9.3	3.4	1.7	4.38	1.622
二、教學意見								
1.教師教學認真不無故遲到早退或缺席	43.6	38.8	11.9	2.5	2.1	1.1	5.16	10.32*
2.我覺得老師對本課程有充分準備	43.5	36.8	15.6	2.1	0.9	1.1	5.16	12.32*
3.教師授課示範講解表達與說明清楚有條理	34.2	41.7	17.0	4.6	1.4	1.1	4.99	6.481*
4.教師鼓勵發問並樂於傾聽或與學生討論	32.4	40.0	20.3	3.9	1.8	1.6	4.92	3.522*
5.我覺得老師能重視學生反應並調整他教學方式或進度	28.4	39.8	21.5	6.9	2.3	1.1	4.82	-.117
6.老師對課程內容完整規劃並於開學時即將課程大綱、教材內容發放或上網	28.8	44.3	17.6	5.9	1.6	1.8	4.87	1.299
7.課程內容難易適中各主題間有良好聯繫	22.0	43.2	24.9	7.3	1.1	1.4	4.73	-2.877*
8.本科採用的教材或講義恰當有助於促成專業知識的完整學習應用	26.3	42.8	21.3	6.4	1.6	1.6	4.81	-.586
9.我覺得老師能適時指引促進學生思考	23.7	42.8	24.1	6.7	1.1	1.6	4.76	-2.217
10.我覺得老師能帶給我們啟蒙及成長的經驗	25.1	36.8	25.3	9.0	2.1	1.8	4.68	-4.538*

主題文章

續上頁

11.老師考評方式公平合理能兼顧學生上課及課後的努力程度	23.3	39.9	25.8	7.8	1.6	1.6	4.71	-3.422*
12.我覺得作業份量課後練習難易適中並配合教學內容有助學習	20.3	39.7	27.3	8.8	2.1	1.8	4.62	-5.941*
13.我覺得本科目老師是我上過的課程中教得較好的幾位之一	25.7	33.1	23.9	9.4	3.4	4.4	4.55	-6.672*
14.這是一門值得學習的課程，有機會我會推薦給學弟妹們	30.4	32.0	23.0	6.7	4.1	3.7	4.67	-4.193*

* $p < .05$

(二) 教學評鑑後設評鑑分析

問卷第二部分利用後設評鑑標準檢視通識教學評鑑實施情形，各題項遺漏值檢定結果最高僅 2.1%，代表學生應能了解題意。因素負荷量最後一題較低，但整體而言數值顯示內部同質性頗高，應可測出一致的概念。平均數則以效用性標準為最高($M=4.23$)，以精確性標準($M=4.15$)為最低，結果請參考表 4。

效用性標準共 6 題，在確保評鑑能讓使用者獲得有用的資訊。平均數較高的題項為「我能清楚說明下判斷時的原因和價值基礎」($t=8.491, p<.05$)、「學生有能力勝任並填答評鑑問題，值得信賴」，顯示出學生對自己能否勝任教學評鑑充滿自信，認為能清楚明確下判斷。檢視百分比數值較集中於「有點同意」以上，合計超過八成，符合平均數排序結果，學生自覺學生評鑑教師是可信的。較低題目為「目前的問卷所蒐集到的資訊是很適切的」($t=-3.287, p<.05$)、「評鑑結果是有影響力的且可持續運用的」，顯示學生對目前教學評鑑的影響力、及所蒐集到訊息的適切性同意度較低。

可行性標準共 3 題，用來確保評鑑確實可行、謹慎，且具成本效益。結果發現「評鑑過程務實可行並盡量減低干擾」與「有徵詢與考量教師學生意見盡量避免爭議與誤解」排序較高，在務實可行與避免爭議上的同意度較高，但在「評鑑方式在資源運用上是有效率的」則相對低很多($t=-4.517, p<.05$)。顯示學生頗同意教學評鑑是務實可行的，也能徵詢教師與學生意見避免誤解，不過資源運用的效率上仍有改進的空間。

適切性標準共 6 題，是確保評鑑合法、合乎倫理，並顧及參與者與受評者福祉。最高是「我在評鑑過程會尊重受評教師尊嚴，避免威脅傷害」($t=8.508, p<.05$)，同時也是所有題項中次高者。這結果和前面效用性標準的結果

頗為一致，可共同解釋，亦即學生自認會以成熟與理性的態度面對教學評鑑，評鑑過程也會顧及教師尊嚴避免傷害。最低是「目前評鑑方式能幫助教師與學生獲得充分服務」，顯示出學生對教學評鑑能提供的實質服務仍有更高期待。

精確性標準共 11 題，確保評鑑能提供充分有效訊息，來判斷受評方案優缺點與價值。平均數最高為「整個課程評鑑方式也應該受到評鑑」($t=8.887, p<.05$)，同時也是所有題項中最高者($M=4.68$)。顯示學生認同教學評鑑後設評鑑的必要性。其他像「學校有提出正當且可受公評的評鑑結論」、「學校有清楚描述課程評鑑的目的與實施程序」等，排序上都不是很前面，這些題項主要在檢視評鑑過程中的背景說明、資料分析、評鑑結果解釋，這應該和學校沒有對學生公布評鑑結果，以及學生不甚了解教學評鑑行政執行面有關。學校目前教學評鑑的結果僅提供給教師本身及系所主管參考，對於教學評鑑成績不理想的教師，則由系所主管了解並提供協助。張德勝(2005)的研究亦指出，國內超過九成的大學校院並沒有將評鑑結果公布，或讓學生知道，主要原因來自於教師反對。而第 17 題「學校有分析並說明課程評鑑背景因素及可能造成的影響」($M=4.00$)，在所有題項中敬陪末座，都顯示學校在結果公布與說明宣導仍有努力改進的空間。

表 4、學生對教學評鑑後設評鑑之統計結果

題項	遺漏檢定 %	平均數	標準差	因素負荷量	t 值
U、效用性(utility)		4.23			
1.有確認涵括所有應參與對象並重視其需求	1.1	4.27	1.07	.75	2.407*
2.學生有能力勝任並填答評鑑問題值得信賴	.9	4.33	1.03	.62	3.852*
3.問卷所蒐集到的資訊是很適切的	.9	4.07	1.07	.78	-3.287*
4.我能清楚說明下判斷時的原因和價值基礎	.9	4.54	.99	.60	8.491*
5.評鑑結果有及時提供給師生了解	.9	4.14	1.21	.75	-.982
6.評鑑結果是有影響力的且可持續運用	1.1	4.05	1.19	.79	-3.619*
F、可行性(feasibility)		4.18			
7.評鑑過程務實可行並盡量減低干擾	1.1	4.25	1.04	.79	1.846
8.有徵詢考量師生意見盡量避免爭議與誤解	1.6	4.25	1.11	.78	1.911
9.評鑑方式在資源運用上是有效率的	1.4	4.03	1.12	.77	-4.517*
P、適切性(propriety)		4.21			

主題文章

					續上頁
10.評鑑方式能幫助師生獲得充分服務	1.1	4.05	1.10	.82	-4.268*
11.評鑑設計有顧及教師權利與福祉	1.8	4.19	1.09	.77	-.347
12.我在評鑑過程中會尊重受評教師尊嚴，避免威脅與傷害	1.4	4.54	1.05	.62	8.508*
13.評鑑方式能夠完整且公正的檢驗、紀錄教學的優缺點	1.4	4.13	1.15	.85	-1.323
14.有適切的公布評鑑結果，讓師生獲得相關資訊	1.4	4.13	1.17	.79	-1.165
15.遇到利益衝突時能公開且誠懇處理	1.6	4.20	1.14	.78	.908
A、精確性(accuracy)		4.15			
16.學校有正確而清楚的描述評鑑方式	1.4	4.12	1.16	.80	-1.899
17.學校有分析並說明評鑑背景因素及可能造成的影響	1.1	4.00	1.20	.82	-5.238*
18.學校有清楚描述評鑑目的與實施程序	1.1	4.08	1.16	.78	-2.936*
19.評鑑的資訊是有效的可適當的解釋結果	1.4	4.06	1.12	.87	-4.716*
20.評鑑的資訊是可信任的	1.8	4.06	1.16	.84	-4.182*
21.學校能有系統檢視所蒐集的資訊修正錯誤	1.8	4.13	1.11	.86	-2.149*
22.學校能適切分析量化(問卷勾選意見)資訊	1.8	4.18	1.08	.87	-.528
23.學校能適切分析質化(文字書寫意見)資訊	2.1	4.16	1.07	.87	-1.427*
24.學校有提出正當且可受公評的評鑑結論	1.8	4.03	1.18	.85	-4.331*
25.評鑑結果公正不受個人情緒和偏見扭曲	1.8	4.14	1.22	.75	-.794
26.整個教學評鑑方式也應該受到評鑑	1.1	4.68	1.15	.43	8.887*

* $p < .05$

二、焦點團體結果與討論

焦點團體座談目的在校正與彌補問卷調查，故研究者先將學生問卷結果說明，再與教師探討原因。討論結果經分析後彙整出「內在」層面及「外部」層面：

(一) 教學評鑑內在層面的問題探討

教學評鑑究竟能對教學產生多大的助益？教學評鑑制度本身實質的功能與限制歸納為「內在層面」問題，包括：

1. 教學評鑑能維持教學品質的最低門檻

教師認為教學評鑑訂定出共通性的標準，並依標準逐項檢視教學品質，這能維持教學品質的最低門檻，但屬被動的改進方式，不宜過度擴大教學評鑑的使用及解釋範圍。真正要提升教師教學品質，必須回歸到教師專業知識與教學知能中。教學評鑑是維持品質的最後一道防線，能防止教師不逾越評鑑標準的規範，但卻不能確保教師真正提升教學品質。

教學評鑑在提升教學品質上有一定限度的，要改善教學還是要落實在教師教學專業知能上。…目前我們會辦理領域會議，邀請各領域教師相互教學觀摩，以專業對話提供交流，以團體討論代替個別約談，並且鼓勵優質課程替代懲罰。共同科教師的教師專業發展活動非常密集，近一年也沒有教師低於平均 3.5 分的門檻。(C1725;1742)

2. 教學評鑑內容應能有適用不同領域課程的版本

問卷調查的數據看出學生對現有量表的適合度尚稱滿意，在教師座談中也有類似的看法。不過大學中存在領域的多樣性，學門專業知識的差異促成教學上多元評量的運用，當然教學評鑑量表也必須符合課程多元的性質。大學通識教學評鑑應顧及多元課程發展適用的評鑑量表。

現在這個版本還蠻適合共同科來使用，內容涵蓋的範圍也算完整，只是用字遣詞上可以不用那麼口語化。(D1750)

學校正發展新的修正版，我建議可以進一步將類別標明清楚，哪些是教學方法，哪些是評量方式的，讓填寫者確切知道評鑑時的面向。(B1731)

學校的通識核心課程有些是理論加上實務，因此會有學科與術科的學習。一般適用學科的版本可能在某些地方不適用於術科。(D1736)

(二) 教學評鑑外部層面的問題探討

外部層面意指那些外來的影響因素，主要來自於教師與學生的態度，因對教學評鑑的詮釋與解讀角度不同，可能會影響教學評鑑的成效。

1. 教師應正向的看待教學評鑑

並非所有教師都認同以學生評鑑教師做為教學評鑑的方式，座談中發現主要原因在於大學教師堅持專業自主及維護教師尊嚴的心態，以及不信任學生評鑑的成熟度，此結果和前面學生調查結果不一致。主要是教師認為自己教學努力，堅守師生分際，但不見得會受到學生的喜愛。

主題文章

教學評鑑有個前提，是老師怎麼看這件事情，…有些老師會自認是專業教學者，學生能客觀的提出意見嗎？這中間少了一些信任。另外，大學老師教學有義務但也有權利，教學評鑑是對老師的不信任，不尊重老師。如果這種心態沒變，這條線越不過去，教學評鑑怎麼改都是沒用。(A1710)

目前共同科教師大致上可接受教學評鑑，且樂見教學評鑑提供的回饋意見。事實上教學評鑑無論在實務、理論上，都站得住腳，任何教學都應該接受檢視，無論學科或術科。但是當老師根本就在排斥這件事情時，就不可能認同能對他產生改變，就不可能因此改變教學。(D1718)

其實教師如果不認同教學評鑑，認為內容不合乎實際，那就要提出自己可以認同的方式，這才是你的專業，才能去說服別人，而不要因為不願意改變，就認為這不適合。(A1722)

學生評鑑教師具爭議性，張德勝（2000）認為主因在於部分教師認為會侵犯教師教學自主權、違反尊師重道精神、引起教師間猜忌，降低對學生的要求。再加上評鑑結果公布與否也一直備受討論，學生調查結果與教師座談發現，許多題項認同度低和學生不了解評鑑過程和結果有關。不過教師的態度若涉及教師專業，那麼以正向態度去面對教師評鑑是否應為教師專業的表現？Danielson(2007)認為教師教學表現除了在規畫準備、教室環境、教學過程三個基本面向外，還有專業責任(professional responsibility)。此面向包括進行教學回饋、撰寫教學紀錄、參與專業社群，以及展現專業成熟等。故以正面態度面對學生評鑑教師，也應是教師開放心胸、專業責任的展現。Chang(2004)的研究就證實，那些對學生評鑑教師態度越正向的教師，也會獲得學生較高的評量分數。

2. 學生面對教學評鑑成熟度的提升

學生能否理性客觀的面對教學評鑑是很重要的因素。前面問卷調查結果發現，學生對自己是否有能力勝任評鑑工作甚有自信，並自認能顧及教師的尊嚴。座談中教師雖然大致認同評鑑用意，但學生不理性的表現會使得教學評鑑效益大打折扣。

事實上就曾發生學生在教師點名紀錄中明明多次缺課，但在學期末填寫教學評鑑量表時卻勾選很少缺課。部分學生「嚴以律人，寬以待己」的態度降低了客觀性，不認真填寫評量表，確實影響教師對學生評鑑教師方式的信任。(E1710)

教學或點名嚴格的老師很容易有教學評鑑兩極化的現象。有些嚴格老師如果教學認真，人緣也好，那他教學評鑑的分數不見得會低。

因為這還牽涉到老師個人特質的問題，會反應在老師行為及言語的使用上。(B1717; A1720)

學生若不滿意某些課程內容，但因為教師給分或作業負擔很鬆，他們知道必須讓某些老師「存活」，課程才能繼續開下去，就是所謂的「營養學分」，學生間可能會默默形成共識。(A1728)

目前全校的課程評量都有提供文字書寫的空間，學生可以自由發言。不過教學評鑑辦了這麼多年，學生辱罵或難聽的詞眼始終存在。(B1714)

3. 修課人數會影響教學評鑑的公信力與正確性

通識課程相較其他學院開設的課程，每班學生人普遍較多。但學校中課程多樣，有些個別指導課甚至只有一個人，樣本小除了易受極端值影響外，教師也易猜測，影響量表的客觀性。但通識課程班級人數至少都會有三、四十人以上，甚至有接近百人的課程，較不受樣本小的限制，結果的參考價值也會較高。

目前通識課程平均人數都會在三、四十人以上，教學評鑑結果扣除極端值後樣本數也不會少太多，可信度會比較高。(D1748)

學生曾提出教學評鑑沒有效果的說法。因為如果上課人數少，那老師一查就會知道，評鑑的可信度便受到影響。甚至我們課務組也接過老師電話，詢問是哪一位學生寫出的意見，老師如果這樣做，學生其實會有壓力。(B1736)

焦點團體結果將教學評鑑區分出內在與外部兩個層面的問題。類似於黃毅志、巫有益(2003)將影響教學評鑑的因素分成，對教學品質真正有影響的「實質因素」，與某些對教學品質並非有影響的，像是教師給分的寬嚴、學生預期分數、先前興趣、學習動機等「偏誤因素」。教育部100年度大學教學卓越計畫滿意度問卷調查結果也發現，全部31校同意度最低就是「教學評鑑制度僅以學生教學評量為依據」(教育部, 2011)，顯示教學評鑑推動時必須先真正釐清它的定義與定位，不宜過度擴張與放大效用，以免因誤用而降低評鑑的實質功能。綜合此些研究結果，焦點團體發現內在與外部層面的後設意見，顯露出教學評鑑必須重視偏向「門檻」的本質，以及教師與學生認同度的影響。

伍、結論與建議

一、結論

(一) 評鑑量表內與教學實質相關之題項適合度較高，對學生評鑑教師的可信度有激勵作用

研究發現學生認為評鑑教師教學題項的適合度，優於學生自我學習狀況評鑑的題項。而平均數較高的「教師教學認真，不無故遲到早退或缺席」、「我覺得老師對本課程有充分的準備」、「教師授課或示範講解的表達與說明清楚有條理」等前幾項，都和教學實質內容有關。而較低的則幾乎為學生學習狀況自評，及推薦或評價老師的題目，平均數偏離檢驗均達顯著。顯示和教學實質或直接相關的題目適合度較佳，學生了解教學評鑑重點何在，對學生評鑑教師的可信度及成熟度具有激勵作用。

(二) 學生肯定後設評鑑重要性，對勝任教學評鑑自信高，也對教學評鑑沒有公布結果的認同度低

經後設評鑑調查結果，效用性標準方面學生對自己能否勝任教學評鑑充滿自信。可行性標準顯示評鑑過程尚稱務實可行，也能盡量避免爭議，但對資源運用效率同意度較低。適切性標準顯示學生自認會理性面對教師評鑑，顧及教師尊嚴。精確性標準中「整個課程評鑑方式也應受到評鑑」，在所有 26 個題項中最高，顯示學生同意推動教學評鑑後設評鑑的必需性。另外對評鑑結論、結果解釋、評鑑實施程序等認同度較低，此因經焦點座談探討後，認為和學校沒有公布評鑑結果應有直接關係。

(三) 釐清教學評鑑的適用範圍與限制，重視師生對教學評鑑的正向態度

焦點團體發現教師對教學評鑑執行的意見大致可彙整成「內在」及「外部」層面。內在層面主要是論及教學評鑑本身實質功能與限制，教學評鑑僅是維持教學品質的最低門檻，真正要提升教學品質，還是必須回歸到教師專業知識與教學知能中。另外，教師也認為教學評鑑內容應能有適用不同領域課程的版本，應顧及領域的課程性質發展出合適的評鑑工具。

外部層面主要在教師與學生的態度上。教師會受維護專業自主與尊嚴，以及對學生評鑑成熟度的影響。教師和學生間的信任基礎是教學評鑑的關鍵，學生面對教學評鑑的成熟度必須持續提升，包括學校對評鑑結果的公布。此外，課程選修人數會影響教學評鑑的正確性，但通識教學評鑑因學生樣本數普遍較多，較不受極端值影響，參考價值也會較高。

二、建議

(一) 研議教師與學生能接受的教學評鑑結果公布方式

目前許多學校並沒有公布評鑑結果，主要是擔心評鑑結果公布後對教師造成壓力，甚而遭致其他負面影響。不過公布評鑑結果讓大眾檢驗，學生的學習態度及評鑑成熟度也同時會受到考驗，可能因此提升學生責任感。學校應考量這之間的利弊得失，或配合獎勵措施與審查機制，做出教師與學生均能接受的公布方式。

(二) 持續協助教師與學生更深入了解教師評鑑

部份教師對學生評鑑教師方式仍有疑慮，因此建議各大學應持續協助教師與學生更深入了解教學評鑑的內涵與功能，利用會議或研討會多與教師學生溝通，除化解教師疑慮外，也增強學生評鑑知能，逐漸提升對教學評鑑正向態度。另外，對教學評鑑執行層面的諸多措施，包括評鑑方法、評鑑時機、資料分析方法、評鑑工具類型、結果後續處理等，也應盡量讓師生了解。

(三) 發展適用不同通識領域課程的評鑑版本

大學通識課程涵括領域甚廣，教學評鑑工具應秉持「同中求異」的理念，在一套共通性版本的基礎下，依領域或學科特性調整出最適合的評鑑內容。目前雖有不少研究投入工具編製，也具信、效度之實證研究基礎，但較集中於共通性版本的編製，建議除發展各適用版本外，也應進一步針對不同版本進行信、效度的考驗，增進教學評鑑工具的實用價值。

(四) 釐清教學評鑑限制，鼓勵教師教學成長提升品質

教學評鑑是被動的教學改進方式，能督促教師通過門檻，教師專業與教學知能成長才是目的。學校應有專責推動教學成長單位，針對教師成長提供協助與諮詢服務。推動教師專業成長社群，學校中開設相同性質課程的教師可組成分享團隊，可共同建立課綱與教材，交流授課心得與教學技巧，分享資訊與資源的經驗，共同成長，應是具體實質助益方式之一。

(五) 持續推動教學評鑑之後設評鑑

教學評鑑也應受到評鑑的觀念已受到教師與學生重視，大學不僅只在通識教學上，更要能對所有課程的教學評鑑都定期的實施後設評鑑。學校可指定一專責單位負責後設評鑑之研究、規畫及執行工作。但推動後設評鑑時應牢記 Sanders (1999) 的提醒，亦即要批判一項評鑑工作是很容易的事情，然而如果沒有針對批判意見進行最終討論，那後設評鑑則不可能完整。

參考文獻

- 林天祐 (2006)。評鑑小辭典—教學評鑑。評鑑雙月刊, 3, 22-23。
- 林劭仁 (2008)。教育評鑑：標準的發展與探索。臺北：心理。
- 林劭仁 (2007)。師資培育後設評鑑標準發展之研究。國立編譯館館刊, 35(2), 41-53。
- 林珊如 (1999)。大學生評鑑教學量表：編製與效度考驗。教育與心理研究, 22(2), 295-322。
- 邱皓政 (2006)。量化研究與統計分析。臺北：五南。
- 孫志麟 (2004)。探討師範學院通識教育的教學品質：評鑑的觀點。國立台北師範學院學報, 17(1), 329-354。
- 教育部 (2011)。「100 年度獎勵大學教學卓越計畫」教學滿意度問卷調查結果。財團法人台灣評鑑協會執行, 未出版。
- 陳琦媛 (2007)。我國公立大學教師教學評鑑之研究。國立政治大學教育系博士論文, 未出版, 臺北。
- 游家政、曾祥榕 (2004)。教育評鑑的後設評鑑。教育資料集刊, 29, 53-94。
- 黃政傑 (1999)。落實教學評鑑的實施。教師天地, 99, 39-45。
- 黃義良、鄭博真, (2011)。大學通識課程學生評鑑教師教學量表驗證與影響因素考驗。臺中教育大學學報, 25(1), 139-161。
- 黃毅志、巫有益 (2003)。影響教學評鑑得分因素之探討-以臺東師院為例。臺東師院學報, 14, 347-370。
- 曾淑惠 (2002)。教育方案評鑑。臺北：師大書苑。
- 張郁雯 (2003)。大學生教學評鑑量表之發展研究。教育與心理研究, 26(2), 227-239。
- 張德勝 (2000)。「學生評鑑教師教學」制度之比較：以師範學院為例。載於國立新竹師範學院 (主編), 八十九學年度師範學院教育學術論文發表會論文集, (頁 182-207)。
- 張德勝 (2005)。臺灣地區大學校院「學生評鑑教師教學」制度之研究。師大學報教育類, 50, 203-225。
- 湯誌龍 (2006)。屏東科技大學學生評鑑教師教學之工具修訂。臺東大學教育學報, 17(1), 1-33。
- 葉連祺、董娟娟、楊世英、陳仁海、蕭芳華, (2005)。大學學生評鑑教師教學量表之編製。測驗學刊, 52(1), 59-82。
- 潘靖瑛, (2004)。通識教育課程學生評鑑教師教學問卷之發展：以慈濟大學為

- 例。測驗學刊，51(1)，79-102。
- 潘慧玲、王麗雲、張素貞、鄭淑惠、林純雯（2007）。高級中等以下學校試辦教師專業發展評鑑之後設評鑑規劃研究。教育部託專案報告。臺北市：國立臺灣師範大學教育政策與行政研究所。
- 潘慧玲、陳文彥（2011）。中小學試辦教師專業發展評鑑全國診斷性後設評鑑之研究。中等教育，62(3)，70-88。
- 鄭博真、黃義良（2009）。大學教師教學評鑑量表發展之研究。教育與心理研究，32(2)，57-80。
- 蕭揚基（2005）。大葉大學通識教育教學之自我評鑑。研究與動態，12，191-220。
- 蘇錦麗（1997）。高等教育評鑑：理論與實際。臺北：五南。
- 蘇錦麗、楊瑩、王偉中、呂鴻德、詹惠雪（2001）。八十六學年度大學綜合評鑑試辦計畫成效評估之研究。新竹師院學報，14，127-162。
- Chang, T. S. (2004). Results of student ratings: Does faculty attitude matter? 教育心理學報，35(2)，183-200。
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. (2nd Ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Grasso, P. G. (1999). Meta-evaluation of an evaluation of reader focused writing for the Veterans Benefits Administration. *American Journal of Evaluation*, 20, 355-371.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.
- Marann, B. & Barbara, F. (2003). Assessing the teaching quality of accounting programmes: An evaluation of the course experience questionnaire. *Assessment and Evaluation in Higher Education*, 28(2), 135-145.
- Sanders, J. R. (1995). Standards and principles. In W. R. Shadish, D. L. Newman, M. A. Ssheirer, & C. Wye (Eds). *New direction for program evaluation*, (pp. 47-52). San Francisco: Jossey-Bass.
- Sanders, J. R. (1999). Metaevaluation of “The effectiveness of comprehensive, case management interventions: Evidence from the National Evaluation of the Comprehensive Child Development Program ”. *American Journal of Evaluation*, 20(3), 577-582.
- Scriven, M. (1981). *Evaluation thesaurus*. CA: Edge Press.
- Scriven, M. (2001). Evaluation: Future tense. *American Journal of Evaluation*, 22(3), 301-307.

主題文章

- Stake, R., & Davis, R. (1999). Summary evaluation of reader focused writing for the Veterans Benefits Administration. *American Journal of Evaluation*, 20, 323-344.
- Stufflebeam, D. L. (2001a). The Metaevaluation Imperative. *American Journal of Evaluation*, 22(2), 183-209.
- Stufflebeam, D. L. (2001b). Guilding and judging evaluations. Evaluation checklist: Practical tools for . *American Journal of Evaluation*, 22, 71-79.
- Tollefson, N., Chen, S., & Kleinsasser, J. (1989). The relationship of students' attitude about effective teaching to students' ratings of effective teaching. *Educational and Psychological Measurement*, 529-536.
- Wilson. R. (1998). New research casts doubt of student evaluation of professors. *The Chronicle of Higher Education*, 44(19), 12-14.
- Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practical guidelines*. (2nd ed.). White plains, NY: Longman.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruther, F. A. (2011). *The program evaluation standards: A guild for evaluators and evaluation users*(3rd ed.). Thousand Oaks, CA: Sage.

A Study of Meta- evaluation on Teaching Evaluation of General Education

Shaw-Ren Lin

Because of the systematic nature of the importance of professional development in meta-evaluation, I conduct a meta-evaluation on teaching evaluation of general education in a public HEI in Taiwan. The questionnaire survey and focus group are used for data gathering. The result shows that the students are self-confident when evaluating the instruction, and that they know the main contents of instruction and are interested in the result. For the quality of instruction, the best practices are not to use teaching evaluation but to enhance professional knowledge and teachers' instruction skills. The different versions for diverse domains should be developed. Teachers' attitude should be more positive. Finally, we propose some concrete suggestions for further application on teaching evaluation of general education.

Keywords: teaching evaluation, meta evaluation, student ratings of instruction, instruction of general education

Shaw-Ren Lin, Associate professor, Graduate Institute of Arts and Humanities Education, Taipei National University of the Arts

主題文章