

網路彈性化學生評鑑教師教學系統 對評鑑行為與結果的影響

曾正宜* 陳舜芬**

學生評鑑教師教學是高等教育中教師教學表現的重要指標。若要讓評鑑結果能反映個別課程的特殊性，且能在所有異質的教學實務之間產生可比較的基準，評鑑工具便需要有讓教師編輯問卷的彈性。本文以國立清華大學所使用網路彈性化學生評鑑教師教學系統為例，藉著比較同一位教師開設同一門課的評鑑資料，來了解紙本量表與網路量表之間模式的轉換是否改變了學生填答的行為與結果。結果發現評鑑模式不同雖對學生填答行為產生一些影響，但不影響評鑑結果，而關鍵在於行政單位如何善用評鑑結果讓它成為改進教學的動力。

關鍵字：教學評鑑、學生評鑑教師教學、網路教學評鑑

* 作者現職：國立清華大學師資培育中心副教授

**作者現職：國立清華大學師資培育中心教授

壹、研究背景

目前國內大學中普遍以學生評鑑的方式來評量教師的教學表現，而評鑑的工具絕大多數都採全校統一的問卷（張德勝，2005）。由於這種一體適用的問卷無法反映多元化知識領域的性質與教師個人化的教學需求，不但在校園中常可聽見對量表題目脫離教學現實的抱怨，所得到的評鑑結果就教師教學表現而言也常缺乏代表性與正當性。然而分殊化的教學評鑑工具雖顧及課程的獨特性，但所產生的結果卻往往難以作為行政決策的依據。於是如何能兼顧「教師教學診斷」以及「行政決策」的需求是教學評鑑業務上一大挑戰，而目前國內針對這個問題的相關研究仍非常有限。

國立清華大學自 94 學年度起啟用「網路彈性化學生評鑑教師教學系統」。該系統最大特色在於利用網路資料庫讓教師能自己選擇或編寫評鑑題目，不僅讓評鑑工具能更適切地反映各課程的教學性質，也能提供共同比較基準作為行政決策的參考。本文除了探討彈性化教學評鑑系統的相關理論之外，並就該系統所產生網路問卷與紙本問卷在施測後所得評鑑結果作深入分析，以了解這種評鑑模式對於學生填答評鑑問卷的行為與結果有何影響。

貳、學生評鑑教師教學的理論

一般而言，國內外大學皆以學生評鑑作為教師教學表現的重要評量方式，除了因為它是信效度兼具的評鑑工具之外（Marsh, 1987; Wachtel, 1998），易於短時間內大量收集資料也是重要的考量。然而不少學者們卻也對於使用這種制度將反而阻礙教師提升教學品質的可能性提出了警告。例如 Hinton（1993）認為因為「有效教學」的定義眾說紛云，學生評鑑量表的設計便時常缺乏理想的基礎。而 Ornstein（1990）則認為教學評鑑中往往無法妥善地反映出教學中藝術與情感的層面。此外，更有人認為學生評鑑量表中所反映的「有效教學」，其實往往是一種保守而傳統的教學思維，常導致被動的學生、主動的教師以及知識的物化（reification），若教師為評鑑結果而特別調整教學方式的話反而會造成更差的教學（d'Apollonia & Abrami, 1997; Wilson, 1988）。這些對於學生評鑑制度的批評主要針對的是以一套有限且被過度延伸詮釋（overly generalized）的教學原理毫無區辨地施用在各種屬性迥異的課程上，因而塑造出制式而粗糙的教學價值觀。要避免這些缺點則必須讓評鑑工具的設計能有因課制宜的彈性，讓有效教學是依特定教學情境而定義，而非局限於一組在所有屬性迥異的課程之間硬找出的最大公約的教學概念。

一、教學評鑑量表的目的

教學評鑑最主要的目的在於蒐集學生對於教師教學的意見以提供以下四種活動的參考：教師自我教學診斷、學校執行人事決策、學生選課、以及學者研究教學效能與課程發展相關議題（Abrami & d'Apollonia, 1990; Abrami, d'Apollonia, & Cohen, 1993; Marsh, 1984），其中，又以前兩種對於教師的專業發展最具直接的影響。然而，因為針對「教學診斷」與「學校人事決策」為目的所發展的學生評鑑工具在本質上有著「因課制宜」與「共同比較基準」等需求上的矛盾，為達成這兩種目的所開發出來的評鑑量表不管在題目設計或資料分析上都有相當大的差異，如何兼顧這兩種目的便成為量表設計上的首要考量。

二、學生評鑑工具的類型

根據 Arreola (2000) 的整理，學生評鑑工具從最差異化到最制式化可分為以下五種類型：

- (一)教師自編型 (instructor-constructed form)：教師為特定課程編寫評鑑問卷。
- (二)教師自選型 (instructor-selected form)：除了幾題總評題目之外，量表其他題目由教師從校方提供的題庫中自由選擇。這種類型也常被稱為「自助餐式」系統 (cafeteria system)。
- (三)混合型 (standard form-optional items)：量表中包括一組所有課程都通用的題目，及一組教師可自題庫中自由選取的題目，有的量表也可以加上教師自編的題目。
- (四)多款選一型 (multiple standard form)：校方準備數種量表，由教師擇一使用。
- (五)單一標準型 (one standard form)：全校使用一份共同的分量表。

目前國內各大學普遍使用的都是單一標準型的評量表。這種評量表在評分與施行上都較為容易，且能提供人事決策時（如升等或傑出教學獎評選）一致的評比標準，但其缺點是無法反映不同課程的特質。例如以論理講述為主的課程所使用的教學方法與技巧可能與討論辯證的課程很不相同。此外，一套評量表代表一組優良教學的標準，而全校採用單一量表意味所有課程的教學均應遵循單一的標準，這與教學需因情境、對象或內容差異而調整的基本原理相左。因此，就教學診斷的目的而言，一個合理的評鑑系統要能有效反映不同課程的特性，才能協助教師從評鑑結果中得到真正有助於改進教學的回饋。

然而，完全依個別課程特性而設計的評鑑量表不僅將加重教師的工作負

擔，所收集到的資料也無法互相比較。Arreola (2000) 甚至批評這種評量表不過是在教學評鑑壓力下的產物，不具任何實質評量的意義。有些學者建議採用教師自選型（或稱自助餐式）評量表，因其最能針對課程的特性選擇適合的題目。然而，完全開放教師自選的評量表未必能達到改善教學的目的，因為有些教師不知如何選擇適當的評量題目（Ory & Weities, 1991），或只選擇能獲得高分的題目，而不在乎是否能從中得到改進教學的回饋（Theall & Franklin, 2000）。於是，在這五種類別中，Arreola (2000) 認為，「混合型」量表既保有一組共用的核心題目來確保量表能反映有效教學的共同特質，也允許教師自選或自編適合其課程的題目來了解學生對特定教學行為的意見，不僅在理論上最周全，也是最能兼顧課程差異化與共同評比基準需求的量表。就評鑑實務而言，要讓學校中數百名教師都能為特定課程編輯評鑑量表則必須要利用網路資料庫，將評鑑由紙本問卷改為網路問卷才能做到。

三、紙本與網路問卷及教學評鑑

一般而言，以網路問卷蒐集資料具備了便宜、快速、高互動性（可依回答內容跳題）、能接觸廣大甚至隱藏的族群、簡化鍵入資料的程序及避免鍵入錯誤資料等優點（Cole, 2005），而隨著環保意識高漲以及網際網路工具蓬勃發展，許多學者們認為網路問卷是最具前景的資料蒐集方法（Schillewaert, Langerak, & Duhamel, 1998）。

許多研究發現，網路問卷的回收速度較快，但紙本問卷的回收率較高。而在實際回收的問卷中，網路問卷的漏題率較低，並且在開放式問題中能得到較長且較詳細的答案（Cole, 2005, Kiesler & Sproull, 1986, Kwak & Rafler, 2002, Schaefer & Dillman, 1998）。除了這些表面的填答行為之外，就問卷所得到的資料品質而言，大多數的研究都指出光是紙本與網路模式的差異並不影響填答的結果（Carini, Hayek, Kuh, Kennedy, & Ouimet, 2003; Layne, DeCristoforo, & McGinty, 1999; McCabe, Diez, Boyd, Nelson, & Weitzman, 2006）。例如，McCabe 等人（2002）比較在兩種模式中填答的中輟率、漏答率與亂答率（如所有項目都填同一答案）時發現網路模式與紙本模式的答題品質並無顯著差異。而 McCabe 等人（2006）的研究也發現在填答網路問卷的過程中填答者自由中斷填答（如將問卷視窗關掉去做其他事）數分鐘之後再繼續的行為並不會影響填答結果的信度與效度。然而，一些研究指出不同的調查模式的確可能產生不同的調查結果，例如網路問卷因有較高的隱匿性，所以常能得到較誠實且不受社會性期待（social desirability）影響的答案（Kiesler & Sproull, 1986; McCabe, Boyd, Couper, Crawford, & D'arcy, 2002），特別是在回答具有高度敏感性的問題時，填答者往往在電腦面前比在紙本上更願意提供誠實且私密的回答（Turner et al., 1998; Wright, Aquilino, & Supple, 1998）。

因網路問卷具有上述特性，在國內外許多學校中也逐漸出現由網路填卷取代課堂中以紙本問卷施測的趨勢（Chang, 2004）。許多學者也開始探討使用紙本問卷或網路問卷作為教學評鑑工具之間的差異，而回收率往往是這些研究首要檢驗的變項。Centra（1979）認為教學評鑑問卷的回收率必需達到三分之二才能有效反映學生意見。一般而言，紙本評鑑都能達到這樣的回收水準，但網路評鑑的回收率在不同研究中卻有很大的出入。大多數直接比較這兩種評鑑模式的回收率的研究都顯示紙本評鑑的回收率高於網路評鑑（Ardalan, Ardalan, Coppage, & Crouch, 2007; Avery, Bryant, Mathios, Kang, & Bell, 2006; Sax, Gilmartin, & Bryant, 2003）。Hmieleski 與 Champagne（2000）調查了美國 200 個有良好網路設備的學校，在 105 個回應的學校中，有 67% 的學校使用紙本評鑑並表示他們有至少 70% 的回收率，而使用網路評鑑的學校則有 20-90% 不等的回收率。其他關於網路評鑑回收率的研究也顯示出類似分歧的結果（如 Moss & Hendry, 2002）。

其次，評鑑模式的不同是否造成評鑑結果的差異也是很重要的議題，而文獻中則呈現出不一致的結果。就量化的資料而言，許多研究顯示學生填答評鑑的結果不會因為評鑑模式的改變而有所不同（如 Ardalan et al., 2007; Avery et al., 2006; Leung & Kember, 2005; Liegle & McDonald, 2005; Vispoel, 2000）；但 Chang（2004）的研究則顯示紙本評鑑的結果會高於網路評鑑，而 Carini 等人（2003）的研究卻顯示利用網路問卷所得到的分數會稍高於紙本問卷。另一方面，就質性資料而言，文獻中對於模式差異所產生的影響則有相當一致的發現，即學生在網路問卷中較願意提供較多質性的回饋（Ardalan et al., 2007; Hmieleski & Champagne, 2000; Layne et al., 1999）。Hmieleski 與 Champagne（2000）指出，在評鑑某一管理課程時，使用網路問卷的學生所寫回饋意見的數量是使用紙本問卷者的四倍。而 Ardalan 等人（2007）的分析則指出學生在網路問卷中較願意針對他們的意見提出較豐富的討論與解釋。

最後，關於學生對於兩種評鑑模式的偏好，研究發現多數學生較喜歡網路評鑑的方式（Layne et al., 1999）。但教師們大多仍喜歡在課堂中施行紙本評鑑，因為教師們擔心網路評鑑會得到較低或較不可信的評鑑分數，且認為讓平常很少出席的學生也能填答是不公平的（Dommeyer, Baum, Chapman, & Hanna, 2002）。

綜觀上述文獻分析可知，模式的差異的確可能影響學生填答評鑑問卷的行為或結果，但這些研究大都使用相同制式化的問卷，除了未探討讓教師參與問卷設計所帶來的影響之外，對於答題品質以及教師們對網路評鑑的疑慮（即讓平時不上課的同學也能參與評鑑造成評鑑不公的感受）也少見有系統的分析。目前文獻中對於網路彈性化評鑑模式的研究極為有限，而清華大學所開發並於

94 學年度首度使用的網路彈性化評鑑系統提供了很好研究的機會，以下是對這個系統簡單的描述。

四、網路彈性化評鑑系統

網路彈性化評鑑系統將所有評鑑題目儲存在資料庫中，藉由一選題的介面讓教師們在電腦上編輯教學評鑑問卷。問卷中除了通用的基本題與總評題之外，教師們可以將題庫中合適的題目或自己編寫的題目加入問卷中。教師編輯完評鑑表後系統會自動產生專屬該教師任教課程的評鑑問卷供學生在網路上填答。

(一)量表結構

系統將所有題目以四個向度（內容、方法、評量及態度）及總評題的結構來呈現。相較於 Marsh（1991）從題目中所擷取出的四個第二階因素向度（課程資料、教學者、親和及課程負擔），系統將教學者與親和併成教學態度，並用「教學評量」來取代「課業負擔/難度」。除了因為「課業負擔/難度」不宜當作教學評鑑項目（Abrami, 1985, 1989）之外，更重要的是課堂中作業/測驗的設計，評分方式的公平性，與教師所給的回饋等都與教學效果息息相關，也都比課程難易度對教學效果有更直接的影響。此外，在孫志麟（2004）的八大向度中也有教學內容、教學方法及教學評量等三個向度，該八大向度中的「教學效果」與「總評」相當，其餘的「教學準備」、「教學態度」、「教學責任」、「教學溝通」在此四加一的向度中，都分別整合於「教學態度」或「教學方法」之下。

(二)問卷題目性質分類

系統所產生的問卷依各評鑑題目所指涉的教學層面將題目分為「基本題」、「自選題」、「總評題」以及「技術建議」等四種類型。另外，為了幫助教師蒐集到學生對課程更完整的意見，問卷中也加上了「教師自訂題」與「開放式問題」兩種類型。其中，基本題所反應的是優良教學品質必備之基本原則（例如 Feldman（1976）所整理出包括激發學生的學習興趣、講課清楚易於了解、精通學科知識、充分備課等）。自選題則都是較為具體且可能會有領域差異的題目，提供教師選擇適合自己課程性質的題目。總評題是指在一個學期之後學生對於教師與課程的整體評價。技術建議則是在教學過程中一些針對教師教學方式或行為等細節的提醒（如音量太小、講話速度太快等）。最後，自訂題是讓教師們能夠藉評鑑的機會來徵求學生們對其教學中特定設計、內容或方式的意見；而開放式問答則是讓學生能自由表達他們對於課程的想法與感受。表 1 是這六種題型之性質與題數的概述。

表 1 彈性化教學評鑑網路問卷題型結構

題型	性質	題數
基本題	優良教學之基本原理，全校通用不可替換	6 ¹
自選題	與優良教學相關較為具體的描述，從題庫中自由選擇	0-5
總評題	對課程與教師表現之總體感受，全校通用不可替換之評鑑指標	2
自編題	教師自己編製題目來蒐集學生對特定教學事項的回饋	0-2
開放題	學生自由表達對課程或教師的意見	1
技術建議	關於教學中技術性事項的建議，學生自由勾選	17 項

為了解清華大學所發展網路彈性化評鑑系統對於學生評鑑教師教學的行為與結果有何影響，除了檢驗評鑑問卷的信效度之外，本研究藉著比較在 93 年（紙本問卷）與 94 年（網路彈性化問卷）由同一位教師所教授同一門課所得到評鑑結果的資料，來探究以下問題：

- （一）學生上課的出席率是否與其填答評鑑題目的分數有關？
- （二）學生填答評鑑題目的品質是否因評鑑模式的不同而異？
- （三）學生填答評鑑題目的分數是否因評鑑模式的不同而異？
- （四）教師是否會特別挑選或編寫能得高分的評鑑題目？
- （五）技術建議題的勾選狀況是否與評鑑分數的高低有關？

叁、研究方法

一、施測程序

93 年紙本問卷的施測是在學期結束前 3 週（為期 2 週）開始由系所助理在上課結束前 10 分鐘到教室中施測。94 年網路問卷的施測則是在學期結束前 7

¹ 95 學年度之後的量表中基本題改為 5 題。

週便通知教師到「校務資訊系統」中選擇或自訂題目，2 週後教師修改問卷的功能停止，學生開始上網填答，填答的時間約為 4 週。在評鑑期間教務處行政人員會持續發出電子郵件提醒學生上網填答。凡填完全部修習課程之評鑑問卷的學生將可獲得一枚精美的清大書籤為獎勵。

二、樣本分析

本研究以教務處所提供 93 學年度第二學期（紙本評鑑）及 94 學年度第二學期（網路評鑑）一般課程的評鑑資料為分析樣本。為了讓分析有相同的比較基準，這些資料在分析之前都先經過以下兩個條件的篩選：

(一)在 93 年與 94 年皆由同一位教師所連續開設的同一門課程。

(二)每門課的總填答人數至少為 5 人以上²。

在這兩個條件之下，本研究所分析的樣本數為 492 門課（表 2）。誠然，教師每年在教同一門課時所用的方法與內容可能未必完全相同，且修課的學生也不一樣，但 Hativa（1996）的研究指出教師教學通常有很高的穩定性，整體而言，同一門課隔年評鑑分數的變動性不大，故可以作為評鑑模式差異的比較基準。

表 2 分析樣本概述

學年度	總課程數	總資料筆數	連續課程數	連續課程資料筆數	連續課程數 (n>5)	重複課程 (n>5) 資料筆數
93	1,204	29,858	540	15,693	492	15,474
94	1,288	29,742		15,280		15,113

肆、資料分析

資料分析分為網路問卷之項目分析，以及網路彈性化評鑑與紙本評鑑之間的模式差異分析兩大面向來進行。

一、網路問卷之項目分析

項目分析主要在於了解整體問卷與個別題目的品質，包括信效度檢測及鑑別度分析。由於網路問卷的主體結構是由基本題所構成，而自選題與自訂題則

² 清華大學傑出教學獎的評選一般只參考至少 5 人以上填答的課程。

因教師而異，無法作鑑別度的檢視，故以下的分析皆以基本題為主，其結果列於表 3 之中。

表 3 網路問卷選擇題之項目分析總表

題號	題目	M	SD	與 T1 相關	與 T2 相關	因素 負荷	極端組 T 檢定	效度 係數 ³
T1 ⁴	我覺得這位教師教學表現優異	3.19	.79	1.00	.84	.88	155.61***	
T2	我覺得這是一門品質優良的課程	3.14	.81	.84	1.00	.90	166.68***	.56***
C1	內容豐富且組織完善，符合教學目標	3.15	.78	.76	.81	.89	162.66***	.53***
M1	講解清楚而有條理	3.08	.84	.76	.78	.88	185.40***	.60***
M2	在課堂中與同學有良好的互動	3.03	.87	.63	.65	.79	254.31***	
A1	充分準備課程內容及教材	3.25	.74	.71	.71	.85	149.58***	.48***
A2	能尊重並友善對待同學	3.29	.73	.64	.65	.81	142.02***	
E1	評量方式能合理反映教學重點及同學的學習成果	2.99	.82	.62	.66	.79	193.47***	.50***

***p<0.001

(一)信度檢測

積差相關的分析結果顯示，基本題之間以及基本題與總評題之間都存在著 0.5 以上的中高度顯著相關，顯示量表內題目之間存在高度的內部一致性，其 Cronbach's alpha 值為 0.94，而其折半信度為 0.92，顯示問卷題目具有高度的可信度。

(二)效度檢測

效度檢測主要以建構效度與效標關聯效度兩種方法來檢測。

³ 以 492 門重複課程中各題目的平均分數求得 93 及 94 年相同題目間的相關係數。

⁴ 94 年問卷題目題號分別為總評題 (T1-T2)，內容題 (C1-C2)，方法題 (M1-M4)，態度題 (A1-A3)，評量題 (E1-E2)，自訂題 (S1-S2) 等，其中 C1、M1、M2、A1、A2、E1 為基本題。

1. 建構效度

問卷題目雖然依「內容」、「態度」、「方法」、「評量」、「總評」等類別來設計，然而主成份因素分析的結果卻發現在所有基本題目之間只存在一個因素(其 Eigenvalue 是 5.78，解釋總變異量為 72.07%)，且所有題目的因素負荷量都在 0.79 以上。另外，所有題目與兩題總評題之間都有 0.6 以上的高度相關 (Cohen, 1988)，而兩題總評題 (對課程與教師的總評) 之間的相關係數更高達 0.84，說明了雖然問卷試圖以不同的題目來描繪教學中的不同面向，但學生的填答卻只反應出同一個潛在因子，即對課程的總體印象。

2. 效標關聯效度

94 年問卷中的 8 題基本題裡保留了 5 題與 93 年問卷相同的題目，於是若以 93 年各課程的平均評鑑分數作為效標，並與 94 年同一位教師同一門課在這些相同題目上所得的平均數作積差相關，則可看出利用紙本與網路問卷評量是否得到一致的結果。積差相關的計算結果顯示同一個題目在 93 年與 94 年問卷中所得分數間的相關係數 (效度係數) 都接近或超過 0.5 的高度相關 (Cohen, 1988) 且達到 $p < 0.001$ 的顯著水準，顯示紙本與網路問卷之間有顯著的關聯效度。

(三) 鑑別度

各題目的鑑別度主要由各題的標準差及極端組 t 檢定來分析。所有基本題之中只有 2 題的標準差略低於 0.75 (分別為 0.74 與 0.73)，顯示一般而言，題目能反映不同的意見。極端組 t 檢定將每一個基本題中的前 27% (高分組) 與後 27% (低分組) 以獨立樣本 t 檢定來檢視兩極端組之間是否有顯著差異。結果發現所有基本題兩極端組之間的差異都達 $p < 0.001$ 的顯著水準，表示題目具有不錯的鑑別度。

二、模式差異分析

紙本問卷與網路問卷之間的模式差異分析主要從出席率、答題品質、分數差異以及自選/自訂題與基本題之差異等方面的比較來了解不同的填答模式是否對學生的答題行為或答題結果產生影響。

(一) 出席率

從 93 年與 94 年學生在「我的缺席次數」一題的回答中發現，整體而言 94 年的缺席率高於 93 年，卡方檢定的結果也証實其中的差別達顯著水準 ($\chi^2 = 315.68, P < 0.001$)。這樣的結果有三種可能。第一，94 年學生的缺席率確

實較高，然而因無特別理由或資料顯示 94 年學生缺席率會顯著高於 93 年，故這個可能性先予以排除。第二，實際缺席狀況沒差別，但網路問卷較隱匿的感覺容易讓學生誠實的報告其缺課的日數，故產生統計上的差異。從 93 年資料中缺席率一項的未答率高於 94 年的 10 倍以上的結果看來，93 年的紙本問卷似乎比網路問卷更容易讓較高缺席率的學生選擇不填答這一題以逃避愧疚感，而網路問卷中填答者較願意誠實地回答私密或難堪的問題也符合文獻中的發現。第三，網路問卷讓較多出席率低的學生也能填到教學意見調查問卷。從表 4 中可見 93 年裡缺課 4 次以上的填卷者佔了總填卷數的 8.5%，而 94 年則佔了 14.4%，顯示 94 年中高缺席率者的填答比率確實高於 93 年。

表 4 九十三與九十四年學生填答缺席狀況總表

	全勤 (%)	1-3 次 (%)	4-5 次 (%)	6-10 次 (%)	11 次以上 (%)	未答 (%)	總計
93 年	7,349 (47.6)	6,665 (43.2)	978 (6.3)	252 (1.6)	98 (0.6)	102 (0.7)	15,444
94 年	6,752 (44.7)	6,175 (40.9)	1,496 (9.9)	465 (3.1)	216 (1.4)	9 (0.06)	15,113
總計	14,099	12,838	2,472	716	313	119	30,557

就教學評量的結果而言，這個差別所引發的實質問題是，高缺席率填答者所填的評量分數與低缺席率者是否有不同。若以 93 及 94 年相同的總評題「整體而言，本課程之教學品質優良」來作為比較基準，由表 5 中可看出，出席率越低者在總評題中所填答的分數也越低，且標準差越大。將 93 年與 94 年的資料以及這 5 個缺席水準用 2x3 ANOVA 分析，其結果顯示年度、五組缺席率及交互作用皆達顯著水準，若以 Cohen(1988)的標準來看，缺席率的效果量(effect size)也達低度關聯強度(η^2 在 0.01 到 0.059 之間)。可見缺席率高者的確會比缺席率低者給予較低的評鑑分數(見表 6)。

表 5 九十三與九十四年各不同缺席狀況的學生填答總評題的平均分數及標準差

	全勤 (SD)	1-3 次 (SD)	4-5 次 (SD)	6-10 次 (SD)	11 次以上 (SD)	平均 (SD)
93 Q15 ⁵	3.26 (.78)	3.17 (.79)	3.03 (.82)	2.93 (.93)	2.80 (1.13)	3.20 (.80)
94 T2	3.30 (.76)	3.10 (.78)	2.82 (.87)	2.68 (.86)	2.53 (1.05)	3.14 (.81)

⁵ 93 年問卷題號為 Q1-Q15

表 6 九十四年缺席次數與總評題的變異數分析表

Source	SS	df	MS	F	η^2
年度	24.817	1	24.817	39.980***	.001
缺席	482.889	4	120.722	194.484***	.025
年度*缺席	49.422	4	12.355	19.905***	.003
誤差	18,809.356	30,302	.621		
總計	19366.484	30311			

***p<0.001

(二)答題品質

就答題的品質而言，文獻中常以「總答題率」、「漏答率」、「一致性填答」等指標來描述答題的品質（Kwak & Rafler, 2002; McCabe et al., 2002）。

1.總答題率

總答題率為所有填答學生佔所有修課學生的比率。據教務處的統計，93 年第二學期的填答率為 73.45%，而 94 年第二學期則為 70.11%，二者以卡方檢定的結果達顯著差異水準（ $\chi^2=132.59$ ），說明這兩年的填卷率的確有所不同，然而如前述文獻探討所示，70%的回收率已與一般紙本的回收率相當，3%的差異並不具太大實質意義（前後兩年填卷狀況見表 7）。

表 7 九十三與九十四年填卷人數統計（一般、實驗、體育課程三種問卷總計）

	實填卷人數	未填卷人數	應填卷人數	填答率
93 年	35,975	13,992	48,976	73%
94 年	32,813	13,001	46,805	70%
總計	68,788	26,993	95,781	

2.漏答率

漏答率為問卷中學生漏答題目的比率。為能在共同的比較基準上作對照，避免兩種版本題數與內容的不同造成比較上的偏差，故漏答率的比較以兩個版本中題目內容相同的 5 題⁶為準。在 93 年這 5 題所有答題數中，漏答的總題數為 751 題，而 94 年則為 538 題，兩年的漏答率皆不到 1%，可見不論是紙本與網路問卷，學生漏答的情況都不算嚴重，且其間的差異也不大。

3.一致性答題

一致性答題指學生在問卷中從頭到尾都填同一個數字，即答題的標準差為 0 的狀況。雖不是所有一致性答題都是倉促答題的結果，但若在這兩個模式之間發現顯著性差異，則代表模式的差異可能引發答題行為的改變。因為 93 與 94 年的題數不同，所以用三種方法來比較一致性答題的比率。(1)93 與 94 年皆以所有題目來計算。(2)在 93 年題目中選擇與 94 年基本題性質最接近的 8 題來計算。(3)以 93 年的前 8 題與 94 年的基本題來計算。不管是用哪一種比較方法，93 與 94 年的資料中，一致性答題的比率都達顯著差異，表示學生填答網路問卷的確比填答紙本問卷時更容易出現一致性填答的現象（見表 8）。

表 8 九十三與九十四年一致性填答樣本數之比較與卡方檢定

93 年 (N=15,444)			94 年 (N=15,113)			χ^2 (df=1)
比較題目	一致性 填答數	比率%	比較題目	一致性 填答數	比率%	
15 題	4,027	26.07	8-15 題	5,534	36.62	394.89***
相似的 8 題	4,857	31.45	基本題 8 題	5,714	37.81	136.53***
前 8 題	4,596	29.76	基本題 8 題	5,714	37.81	221.37***

***p<0.001

(三)分數差異分析

針對模式差異是否影響填答分數的問題，若以 93 年與 94 年問卷中 5 題相同題目所得分數為比較基準，則單因子變異數分析的結果顯示有兩組題目的分

⁶分別是「內容豐富且組織完善，符合教學目標」、「講解清楚而有條理」、「教師對每一堂課都認真準備」、「評量方式合理地反映教學重點及同學的學習成果」、「我覺得這是一門品質優良的課程」

數達到顯著差異(Q8-M1, Q15-T2)。然而因樣本數很大,且其效果量都與Cohen (1988)所提出低度關聯強度的最低水準($\eta^2=0.01$)還有一段距離,故可說這兩年利用不同問卷所測得的評鑑分數之間並沒有具實質意義的差異(見表9)。

表9 九十三與九十四年相同題目之間分數的差異分析

93-94 題目	93 Mean (SD)	94 Mean (SD)	SS 組間	SS 組內	df	F	η^2
Q1-C1	3.13 (.72)	3.15 (.78)	3.07	16957.20	1,30335	5.49*	.000
Q6-E1	2.99 (.84)	2.99 (.82)	.01	20494.58	1,30000	.01	.000
Q8-M1	3.17 (.83)	3.08 (.84)	54.96	21210.22	1,30378	78.72***	.003
Q11-A1	3.26 (.76)	3.25 (.74)	.38	17125.10	1,30352	.68	.000
Q15-T2	3.20 (.80)	3.14 (.81)	23.58	19562.15	1,30421	36.68***	.001

* $p<0.05$, *** $p<0.001$

(四)自選題/自訂題效果

教師可從題庫中挑選0-5題的自選題,或自己撰寫0-2題的自訂題來編輯評鑑問卷。在94年的15,113筆資料中,約有14%包括了自選題,6%包括了自訂題。為了瞭解教師是否會刻意選擇或編寫能得到高分但與改進教學無關的題目,本研究以相依樣本事前比較的方式,針對同時填答基本題與自選/自訂題的樣本,逐一檢驗各自選/自訂題與8題基本題之平均數之間是否達到顯著的差異水準。從表10中可見,在5題自選題與基本題的比較中有4題達顯著水準,其中有3題達Cohen的低度關聯強度,但所有達顯著差異的比較中只有一題自選題的分數高於基本題的平均分數。在兩題自訂題中,有1題自訂題分數顯著低於基本題平均,另一題則沒有顯著差異。從這個分析中可知,教師在挑選自選題或編寫題目時,並未特別選用能得高分的題目,而是真的想知道學生對某一教學面向的反應,符合以評量結果作為教學診斷參考的精神。

表 10 自選題/自訂題與基本題之間的事前比較

題目	自選/ 自訂題	基本題	SS 組間	SS 組內	df	F	η^2	關聯 強度
C2 ⁷	3.21	3.20	.37	442.76	1,2076	1.74	.001	-
M3	3.17	3.20	1.25	614.34	1,2193	4.45*	.002	-
M4	3.13	3.19	8.49	636.12	1,1898	25.33***	.013	低
A3	3.30	3.20	22.58	473.97	1,2090	99.58***	.045	低
E2	3.10	3.20	23.41	627.08	1,2149	80.23***	.036	低
S1	3.25	3.26	.13	385.74	1,1063	.35	.000	-
S2	3.20	3.25	1.33	177.73	1,604	4.53*	.007	-

* $p < 0.05$, *** $p < 0.001$

(五)技術建議題

教學中技術性細節雖然與課程整體規劃、內容與教課品質無直接關係，但學生填選這些項目的行為卻清楚地表達了他們對於課程的感受。表 11 是所有填選各技術建議的人次統計以及填選人數佔總人數（29,742）的百分比，表面上看來在 17 個項目中填選率超過 10% 的只有 2 項，有 5 項填選率不到 2%，整體填選率似乎不算高，然而若將每一堂課中每人平均填選率（班級所有項目的填選總數/班級人數）與總評題的分數來作積差相關，則發現每班每人平均填選數與兩個總評題分數之間都有著頗高的負相關（ $r = -0.58$ ）。換言之，如果選擇題是讓學生表達他們對於教師是否展現優良教學的認同程度，則這些技術性建議則可視為他們對於教師教學中有所不滿的反應。

⁷ C2、M3、M4、A3 與 E2 為自選題，S1 與 S2 為自訂題。

表 11 教學技術性建議項目、填答人次與填答比率

項目	人次 (%)	項目	人次 (%)	項目	人次 (%)
1. 加大音量	2,499 (8.4)	7. 咬字更清楚些	1,750 (5.9)	13. 放慢講課速度	3,200 (10.8)
2. 放慢投影片放映速度	1,236 (4.2)	8. 字跡力求清晰	1,287 (4.3)	14. 按照課程大綱進行教學	620 (2.1)
3. 多提供講義及參考資料	3,723 (12.5)	9. 將教材放上網	2,361 (7.9)	15. 多舉實例	2,460 (8.3)
4. 再充實課程內容	2,594 (8.7)	10. 準時上/下課	1,768 (5.9)	16. 減少調課/缺課	393 (1.3)
5. 加強情緒管理	407 (1.4)	11. 立論力求持平	320 (1.1)	17. 保持政治/宗教中立	287 (1.0)
6. 加強要求學生的上課態度	1,109 (3.7)	12. 認真監考	175 (0.6)		

伍、研究發現

從基本題與總評題的分析結果得知本研究所使用的評鑑系統具有良好的信效度以及鑑別度，而文獻中也都認為網路評鑑可以是紙本評鑑很好的替代模式，然而許多教師們對於從紙本到網路問卷的轉換仍有疑慮。其中最常見的便是網路問卷讓許多平時不來上課的學生也能填答，對教師的教學恐怕難以作出公正的評鑑。本研究發現網路問卷中學生的整體缺席率確實比紙本問卷來得高，可能是因為網路的隱匿性讓學生較容易在網路問卷上誠實報告其缺課次數，或者網路問卷的確收集了更多來自高缺席率學生的填答。進一步分析結果發現高缺席率的學生所填答的分數比低缺席率的學生更低。這個結果正面的解釋是因為低缺席率學生在課堂上收穫較多，所以對老師與課程有較正面的評鑑，而其負面的解釋則可能是學生因缺課而覺得收穫少，或者高缺席率是對教師教學表現或課程設計不滿的反映。若為後者，則這些學生的填答仍該被視為對於教師教學表現的有效反應，似乎不宜因為缺課而剝奪其表達對教師教學之意見的機會。另一種情況是學生因個人問題而常態性缺課，其行為無關教師的教學表現。不強制填答的政策將可讓這些無法對於教師教學作出公正評量的學生放棄評鑑，而缺課學生若對教師並無不滿，應也不致於會作出報復性填答的行為。由於清華大學並無點名制度，缺席狀況是以學生在問卷中填答的結果為準，自然無法排除學生未誠實填答的可能，但這是所有問卷調查都存在的隱性

誤差。若要藉統計方法或以一些原則來排除某些學生的填答結果（如排除高缺席率、成績不佳學生或離群值等資料）則需考慮學生表達意見的權利或教師操弄評鑑結果的可能性。

其次，兩種填答模式有非常不同的填答情境，而其對於學生答題品質的影響也略有不同。首先，網路問卷 70% 的回收率已與文獻中一般紙本問卷的回收率相近，雖略低於本研究中紙本問卷的回收率，但差異並不大。就漏答率而言，網路問卷與紙本問卷的漏答率都不到 1%，基本上沒有什麼差別。紙本問卷與網路問卷中各有約 30% 及 38% 的學生在各題問填答結果的標準差是 0，且這兩個比例的差異達顯著水準，顯示在網路問卷模式中，學生常將不同課程的評鑑問卷一次連續填完的行為可能增強一致性填答的傾向。因這種答題傾向讓教學各面向之間的差異無法確實地呈現，以致於影響了問卷的效度。然而，從紙本問卷中仍有 30% 一致性填答的結果看來，一致性填答可能反映出學生對教師的整體教學表現在填答前已先有定見，並將此定見投射到所有題目中，而對於個別題目是否適切描述該課程並未仔細思考的現象（Cashin, Downey, & Sixbury, 1994）。如若如此，則總評題似乎是教師教學表現的合理指標。而根據 Cashin 等人（1994）所言，總評題一般包括「課程」、「教師」以及「學習」三種，若要從中擇一作為教師教學評鑑的指標，則以「教師」的總評題分數為最直接且具表面效度。

此外，從相同題目在紙本與網路模式的分數之間未發現有意義之顯著差異的結果看來，模式的差異並未影響學生答題的結果。這個結果與文獻中大多數的研究結果相符，說明網路評鑑可以是紙本評鑑可信賴的替代模式。然而前述一致性填答比例偏高的現象卻容易讓網路問卷產生的結果較為粗糙，於是需輔以質性的回饋以及技術建議的選項，讓學生能提供更明確的意見來幫助教師改進其教學。本研究發現，一個班級中每人平均點選技術建議項目的次數與總評題分數之間有接近高度的負相關，說明這些教學上技術細節雖不足以代表教師教學表現的優劣，但仍可能影響學生對於該教師或課程的感受。

最後，從自選/自訂題的評鑑分數未必高於基本題的結果看來，只要教學評鑑不是以全部題目的平均分數作為教師教學表現的指標，願意花時間挑選或編寫評鑑题目的教師並不會特別選擇或編寫能得高分的题目來影響學生填答的結果，而是真的想藉這些題目來蒐集學生對某特定教學設計或方式的看法。當然，這個結果也可能是因為上述學生對於教師教學與課程的定見而影響了自選/自訂題的答題，但不論原因為何，自選/自訂題的設計並不會影響評鑑結果的可信度。

值得一提的是，雖然紙本與網路問卷都對所有修課學生施測，但網路自由

填答的方式可能會造成某些學生因具有某種特性，而有系統性地被這種評鑑模式遺漏，進而影響研究的結果。這種網路評鑑中系統性遺漏樣本的可能性有三。首先，網路設備不足造成弱勢學生填答的阻礙。其次，研究顯示，大學生某些特質可能與他們填答網路教學評鑑問卷的意願有關，例如資訊或科學相關科系、成績較好或大二學生較願意填網路教學評鑑問卷（Layne et al., 1999）。第三：不願參與教學評鑑或對獎勵辦法不感興趣者（例如許多準畢業生）。理論上這三種因素對於自由填答的網路問卷所可能產生的影響會比紙本問卷更大。然而，就第一點而言，清華大學具有豐富的電腦網路資源與設備，師生平時使用電腦網路從事教學與學習的活動也非常頻繁，網路設備不至於成為學生填卷的阻礙。至於第二、三點，雖然有報導指出得高分的學生較容易給教師較高的評鑑分數（Johnson, 2002），而他們又較願意上網填問卷，所以可能對評鑑結果造成偏頗。然而得高分的學生應該也較會上課填寫紙本問卷，所以就填寫問卷而言，得高分者在紙本與網路評鑑上應該沒有差別。至於科系與年級等因素應該跟填答的評鑑分數之間沒有直接關連。此外，本研究中 70% 的回收率已經具有很高的代表性，即使這些因素真的會對研究結果造成偏差，其影響也將非常有限。所以雖然本研究在實質上可能不能稱為完全隨機抽樣，但這個問題對於本研究的結果並沒什麼影響。

陸、討論與建議

在教學究竟該視為一門藝術或科學的討論中，Centra 和 Bonesteel（1990）認為凡有效的教學皆具有某些特質。不同研究者也許會從不同角度或方法來描繪它們，但這些特質提供了有效教學的科學基礎，而教學中的藝術成份便是在此基礎之上，由教師依內容與對象不同所展現出他們對教學的洞見與創意。由此觀之，教學評鑑除了要能反映有效教學的共通特質，也需賦予教師們施展其教學藝術的空間。而今教學已日益強調適性與個別差異，教學評鑑自然也需要更多的彈性才能真正反映教師的教學表現。於是，要提升大學教學品質及激勵教師認真執教，一套能同時反映優良教學的基本原則以及學科領域或課程的特殊屬性將是不可或缺的工具。國立清華大學所發展的網路彈性化評鑑系統便是藉著基本題、自選題、自訂題、總評題、技術建議與心得感想等題型，讓評鑑結果能同時提供行政決策與因課制宜的教學回饋。

這套系統的施行對於教師與學生而言有不同的意義。就教師而言，教師們實際參與問卷編輯的比率仍有待提升，其原因可能包括學期中無暇或無心挑選或編寫題目、只在意教學總分但不在意個別題目、找不到更適合的題目、或根本不在意評鑑結果等等。在執行上可藉由加強對教師的宣傳、向教師徵求題目

以擴充題庫、或分析學生所填寫的質性回饋內容來調整題目的設計等等方式來提升教師的使用率。就學生而言，當填卷行為不再受到課堂上施測的時空約束時，行政單位便需要提供學生一個好理由讓他們願意花自己的時間與精神來填卷。常見提升學生填答率的措施包括不斷以 e-mail 提醒學生填答或由對學生具有影響力的人（如學務長、學生會長或學生會代表等）寄 e-mail 提醒學生填答（Ardalan et al., 2007）、填卷者可以抽獎（Bosnjak & Tuten, 2003; Cobanoglu & Cobanoglu, 2003）、享有選修下學期課程或看成績的優先權（Layne et al., 1999）、或給填答者加分（Avery et al., 2006）等。這些鼓勵或強迫性的措施固可提升填卷率，但也可能有道德上可議之處。若要讓教學評鑑成為改進教師教學的力量，最根本的辦法是在學校、教師與學生之間建立起互信與互重的良性循環。換言之，學生需要感受到填卷不是為學校做行政業務，而他們所填答的意見對教師的教學是真的有影響且是被重視的。教師們也需要感受到教學評鑑不只是為了給教師的教學表現貼標籤，而是真的可以從中看到善意的建議或回饋。有些學校以公佈課程評鑑分數提供學生選課參考的方式讓學生感受到填卷的價值，但除此之外應該還有更能利用評鑑結果促進師生溝通，並為雙方創造出切身價值感的作法，至於如何進行則有賴行政人員與研究者集思廣益了。

參考文獻

- 孫志麟（2004）。**教育政策與評鑑研究-追求卓越**。臺北：學富。
- 張德勝（2005）。台灣地區大學校院「學生評鑑教師教學」制度之研究。**師大學報：教育類**，50(2)，203-225。
- Abrami, P. C. (1985). Dimensions of effective college instruction. *The Review of Higher Education*, 8, 211-228.
- Abrami, P. C. (1989). How should we use student ratings to evaluate teaching? *Research in Higher Education*, 30, 221-227.
- Abrami, P. C., & d'Apollonia, S. (1990). The dimensionality of ratings and their use in personnel decisions. *New Directions for Institutional Research*, 43, 97-111.
- Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1993). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology*, 82, 219-231.

- Ardalan, A., Ardalan, R., Coppage, S., & Crouch, W. (2007). A comparison of student feedback obtained through paper-based and web-based surveys of faculty teaching. *British Journal of Educational Technology*, 38, 1085-1101.
- Arreola, R. A. (2000). *Developing a comprehensive faculty evaluation system*. MA, Bolton: Anker Publishing Company.
- Avery, R. J., Bryant, W. K., Mathios, A., Kang, H., & Bell, D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations? *Journal of Economic Education*, 37, 21-38.
- Bosnjak, M., & Tuten, T. L. (2003). Prepaid and promised incentives in web surveys. *Social Science Computer Review*, 21, 208-217.
- Carini, R. M., Hayek, J. C., Kuh, G.D., Kennedy, J. M., & Ouimet, J. A. (2003). College student response to web and paper surveys: Does mode matter? *Research in Higher Education*, 44, 1-19.
- Cashin, W. E., Downey, R. G., & Sixbury, G. R. (1994). Global and specific ratings of teaching effectiveness and their relation to course objectives: Reply to Marsh (1994). *Journal of Educational Psychology*, 86, 649-657.
- Centra, J. A. (1979). *Determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Centra, J. A., & Bonesteel, P. (1990). College teaching: An art or a science? *New Directions for Teaching and Learning*, 43, 7-15.
- Chang, T. (2004). The results of student ratings: Paper vs. online. *Journal of Taiwan Normal University Education*, 49, 171-186.
- Cobanoglu, C., & Coganoglu, N. (2003). The effect of incentives in web surveys: Application and ethical considerations. *International Journal of Market Research*, 45, 475-488.
- Cohen, J. (1988). *Statistical power analysis for the behavioral science*. Hillsdale, NJ: Lawrence Erlbaum.
- Cole, S. T. (2005). Comparing mail and web-based survey distribution methods: Results of surveys to leisure travel retailers. *Journal of Travel Research*, 43, 422-430.

- d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*, 1198-1208.
- Dommeier, C. J., Baum, P., Chapman, K. S., & Hanna, R. W. (2002). Attitudes of business faculty toward two methods of collecting teaching evaluations: Paper vs. online. *Assessment & Evaluation in Higher Education, 27*, 445-462.
- Feldman, K. A. (1976). The superior college teacher from the students' view. *Research in Higher Education, 5*, 243-288.
- Hativa, N. (1996). University instructors' ratings profiles: Stability over time, and disciplinary differences. *Research in Higher Education, 37*, 241-265.
- Hinton, H. (1993). Reliability and validity of student evaluations: testing models versus survey research models. *Political Science & Politics, 26*, 562-569.
- Hmieleski, K., Champagne, M.V. (2000). Plugging in to Course Evaluations. *The Technology Source*. Retrieved March 22, 2007 from <http://ts.mivu.org/default.asp?show=article&id=1034>
- Johnson, V. E. (2002). An A is an A is an A and that's the problem. *New York Times, Education Today*, April 14, 14.
- Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly, 50*, 402-413.
- Kwak, N., & Rafler, B. (2002). A comparison between mail and web surveys: Response pattern, respondent profile, and data quality. *Journal of Official Statistics, 18*, 257-273.
- Layne, B. H., DeCristoforo, J. R., & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education, 40*, 221-232.
- Leung, D. Y. P., & Kember, D. (2005). Comparability of data gathered from evaluation questionnaires on paper and through the internet. *Research in Higher Education, 46*, 571-591.
- Liegle, J., & McDonald, D. (2005). Lessons learned from online vs. paper-based computer information students' evaluation system. *Information Systems Education Journal, 3*(37), 1-14.

- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology*, 76, 707-754.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.
- Marsh, H. W. (1991). Multidimensional students' evaluations of teaching effectiveness: A test of alternative higher-order structures. *Journal of Educational Psychology*, 83, 285-296.
- McCabe, S. E., Boyd, C. J., Couper, M. P., Crawford, S., & D'arcy, H. (2002). Mode effects for collecting alcohol and other drug use data: Web and U.S. mail. *Journal of Student on Alcohol*, 63, 755-761.
- McCabe, S. E., Diez, A., Boyd, C. J., Nelson, T. F., Weitzman, E. R. (2006). Comparing web and mail responses in a mixed mode survey in college alcohol use research. *Addictive Behaviors*, 31, 1619-1627.
- Moss, J., & Hendry, G. (2002). Use of electronic surveys in course evaluation. *British Journal of Educational Technology*, 33, 583-592.
- Ornstein, A. C. (1990). A look at teacher effectiveness research--Theory and practice. *National Association of Secondary School Principals Bulletin*, 74, 78-88.
- Ory, J. C., & Weities, R. (1991). A longitudinal study of faculty selection of ICES student evaluation items. *Paper presented at the 72nd Annual Meeting of the American Educational Research Association*, Chicago, IL.
- Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*, 44, 409-431.
- Schaefer, D. R., & Dillman, D. A. (1998). Development of a standard E-mail methodology. *Public Opinion Quarterly*, 62, 378-397.
- Schillewaert, N., Langerak, F., & Duhamel, T. (1998). Non-probability sampling for WWW surveys: A comparison of methods. *Journal of the Market Research Society*, 40, 307-322.

- Theall, M., & Franklin, J. (2000). Creating responsive student ratings systems to improve evaluation practice. *New Directions for Institutional Research*, 83, 95-107.
- Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, 280, 867-873.
- Vispoel, W. P. (2000). Computerized versus paper-and-pencil assessment of self-concept: Score comparability and respondent preferences. *Measurement and Evaluation in Counseling and Development*, 33, 130-143.
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23(2), 191-211.
- Wilson, T. (1988). Student evaluation of teaching forms: A critical perspective. *Study of Higher Education*, 12, 79-95.
- Wright, D.L., Aquilion, W. S., & Supple, A. J. (1998). A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly* 62, 331-353.

The Impact of a Web-Based Customizable System for Student Evaluations of University Teaching on the Behaviors and Outcomes of Students' Evaluations

Jeng-Yi Tzeng * Shun-fen Chen **

Student evaluations of teaching effectiveness have been a generally accepted practice among universities in Taiwan. The current practice of using a set of evaluation items for all courses may generate comparable outcomes for administrative purposes, but often falls short in honoring the diverse nature of teaching. This paper presents a web-based customizable evaluation system in National Tsing Hua University, which features a combination of core items, optional items, self-designed items, global items, technical suggestions, and free comments. Comparisons of the evaluation data generated from the paper-based questionnaires and the web-based evaluation system conclude that the web-based system, with more powerful customization functions, is a suitable replacement for the paper-based questionnaires, but the key to the success of the system lies in the interpretation and the utilization of the results.

Keywords: student evaluation of university teaching, teaching evaluation, web-based teaching evaluations

* Associate Professor, Center for Teacher Education, National Tsing-Hua University

** Professor, Center for Teacher Education, National Tsing-Hua University