

大學質性評鑑的挑戰：系所與外部專家 評鑑觀點的一致性分析

陳慧蓉

摘 要

研究目的

質性評鑑受到質性研究的理念影響，質性研究的信實度已從過去客觀中立事實的認定，轉而追求詮釋與推論的確實性與可靠性，亦即質性研究者能從被研究者的立場詮釋，使雙方意見具有一致性。我國系所評鑑採行認可制，由熟悉大學治理的專家進行質性品質的判斷，以確保高等教育品質。雖然系所評鑑運用了質性研究的理念，以資料的三角校正等方法設計評鑑流程，然而，質性評鑑的品質仍然受到各界關注。由於評鑑專家是在社會情境脈落之下進行評鑑，專家本身的社會文化及個人信念，有可能帶入評鑑過程中。為尊重受評系所的意見，我國評鑑設計了申復機制，系所可藉由申復讓專家重新考量。本研究的目的是探討系所自評與外部專家評鑑觀點之異同，從訪評意見、學校申復、及申復回覆的意見往返，探討學校與外部專家意見的一致性，了解大學質性評鑑的挑戰。

研究設計／方法／取徑

本研究運用內容分析法，分析 2006 至 2010 年第一週期系所評鑑 1,156 個班制的申復報告及其相關回覆。

研究發現或結論

研究發現，系所申復報告主要說明大學提供資訊或質疑專家引用資訊的確實性（佔 68.4%），次要是質疑評鑑結論與決定的公正性（佔 13.4%），第三是說明系所情境脈絡或質疑專家未考慮學校情境（佔 12.0%）。而專家對於申復的回覆，主要說明評鑑報告的結論與決定的（佔 48.4%），次要是說明資訊確實性的考量情形（佔 31.0%），第三是詳細說明評鑑推論（佔 12.7%）。交叉檢視申復及回覆的往返內容，結果顯示系所評鑑運用了增進質性研究信實性的方法設計評鑑流程，然而，專家在執行上仍然面臨了四項威脅：一、未能考慮情境脈絡；二、檢視的是結果而非動態歷程；三、專家價值觀的涉入；四、評鑑歷程缺乏長期投入與互動。

研究原創性／價值性

良好的大學質性評鑑，需要大學與專家具備相當的評鑑專業知能，才能從根本審視學校問題，促進學校改善。本研究以分析當前大學質性評鑑的執行情形及面臨的挑戰為例，提出增進質性評鑑品質的建議，以期作為改進評鑑機制及擬訂相關政策之參考。

關鍵詞：質性研究、外部專家、高等教育、系所評鑑、信實度、確實性、一致性

CHALLENGES OF QUALITATIVE EVALUATION: DIFFERENT VIEWS OF REVIEWERS AND UNIVERSITY STAFF IN THE PROGRAM ACCREDITATION OF HIGHER EDUCATION

Karen Hui-Jung Chen

ABSTRACT

Purpose

Qualitative research is a method of inquiry employed in the qualitative evaluation. The quality of qualitative research was shifted from emphasizing on reliability and validity based on neutral facts to trustworthiness of inferences based on the consensus of qualitative researchers and participants. Taiwanese higher education adopts accreditation approach for educational quality assurance. Having experts with deeply understanding of the field being evaluated, peer review is able to make fair judgment and comments in external evaluation to assure the higher education quality. Although the accreditation approach applies the conceptions of qualitative research methods, such as triangulation for data collection during an evaluation procedure, the quality of qualitative evaluation has been challenged. It has been noticed that reviewers often conducted evaluations in a situated context and inevitably brought their own socio-cultural history and beliefs into the evaluation process. In order to remedy this problem, an objection and appeal system has been developed in Taiwanese higher education evaluation. The institution being evaluated can file an objection if they considered that the evaluation result is inconsistent with the facts. This study aims to investigate the major issues emerging from the objection reports and the corresponding replies, as well as to examine the gap between the educational quality of peer reviewers and institutional staff based on the objection reports. It provides an exploration of challenges of qualitative evaluation of higher education.

Karen Hui-Jung Chen (corresponding author), Assistant Professor, Department of Education, National Taipei University of Education.

E-mail: karen3117tw@gmail.com

Manuscript received: October 22, 2015; Modified: January 31, 2016; Accepted: April 18, 2016

Design/methodology/approach

Applying content analysis methods, this study analyzed the objection reports of 1,156 programs during the 1st cycle of the program accreditation from 2006 to 2010.

Findings

Results showed that the institutions being evaluated tried to alter the conclusion of the on-site-visit reports by questioning the credibility of on-site visit report and providing more evidence (68.4%), weakening the justification of the review (13.4%), and asking reviewers to consider university context (12.0%). However, reviewers often repeated the same conclusion of the evaluation in the replies (48.4%), illustrated their considerations for data incredibility and inconsistency (31.0%), and provided detailed explanation for the evaluation reasoning (12.7%). Cross analysis of the opinions between university staff and reviewers revealed that the reviewers applied similar strategies to reply the objections of institutional staff. Four threats to credibility and consistency exist: 1. The university context was not fully considered; 2. The on-site visit examined only the isolated phenomena but not dynamic processes; 3. The evaluation judgments could be influenced by personal values; 4. The conclusion was not drawn based on long-term observations and interactions.

Originality/value

To conduct best practice of qualitative evaluation of higher education, institutional staff and external reviewers are required to have appropriate evaluation knowledge and skills. Making reflections from the suggestions of evaluation reports, institution can find out its weakness for institutional development, and make improvements for better educational quality. Through the inquiry of this study, the challenges of qualitative evaluation can be better understood, enabling future improvement of the evaluation process and making better educational policy of higher education.

Keywords: *qualitative evaluation, peer reviewers, higher education, program accreditation, trustworthiness, credibility, consistency*

壹、緒論

質性評鑑是重要的評鑑方法，可以針對個別特質蒐集詳盡資訊，深入了解方案的差異性（吳芝儀、李奉儒，1995；Patton, 1999）。質性評鑑與量化研究不同，量化研究從建立假設、運用演繹方法進行考驗，而質性研究以多元的研究方法進行探究，透過文字與圖像的分析、或是運用觀察及晤談，再歸納各項資料得到詮釋性的結論（郭秋永，2011；黃政傑，2004；Silverman, 2013）。現行大學校院的系所評鑑及校務評鑑，除了量化資料之外，更設置了質性的評鑑指標，評鑑委員以質性方法蒐集學校資料，如：文件檢視、簡報、實地觀察、晤談與座談等方式，深入了解學校的概況，進而提出評鑑結果與改善建議。

質性評鑑的品質，受到了評鑑專家與專業組織的重視。Scriven (1969) 首先提出「後設評鑑」(meta-evaluation) 的概念，對評鑑進行評鑑，以確保評鑑的品質。而後，美國教育評鑑標準聯合委員會 (Joint Committee on Standards for Educational Evaluation, JCSEE) 於 1982 年提出「方案評鑑標準」(The Program Evaluation Standards)，並於 2011 年修訂為第三版 (Yarbrough, Shulha, Hopson, & Caruthers, 2011)，主張以五個層面檢視評鑑的品質，包括：效用性 (Utility)、可行性 (Feasibility)、適切性 (Properity)、精確性 (Accuracy)、及績效責任 (Evaluation Accountability)。

我國大學系所評鑑採行質性評鑑，運用了質性研究的資料三角校正等方式設計評鑑流程，然而，質性評鑑的品質，仍然受到各界的關注。我國系所評鑑的執行，是根據「大學評鑑辦法」，由教育部委託財團法人高等教育評鑑中心基金會，以六年為一週期，進行一般大學校院的系所評鑑及校務評鑑；同時，為了維持評鑑的品質，教育部根據該辦法的第七條，委託第三方單位（臺灣高等教育學會）對系所及校務評鑑進行後設評鑑，作為改進評鑑流程及相關制度的參考（教育部，2013a；楊瑩、楊國賜、劉秀曦、黃家凱，2014）。雖然剛開始實施第一週期系所評鑑時，後設評鑑顯示評鑑委員的專業性仍有改善空間，但是目前大部分受評學校對於評鑑的滿意度已逐漸改善，並且認同評鑑的重要性（楊瑩等人，2014）。另一方面，國內研究者也從質性研究的角度，提出檢討改進與加強評鑑知能的建議，如王保進（2011）指出，系所評鑑的過程並非缺乏學科專業知識，

而是需要具備評鑑專業知識與評鑑倫理，以確保實地訪評的品質。林天佑（2004）則認為專業的評鑑行為，應包括具備學術專業、及評鑑的過程與方法受到評鑑社群的認同。

由於質性評鑑受到質性研究的理念影響，因此本研究從質性研究典範的轉移，探討質性評鑑的品質。質性研究的信實度，已從過去客觀中立事實的認定，轉而追求質性評鑑的詮釋與推論的確實與可靠性，亦即質性研究者能從被研究者的立場詮釋，雙方對於評鑑詮釋與推論的具有一致性（Cronbach, 1988; Kane, 2006）。有鑒於此，本研究透過系所申復報告的內容分析，從訪評意見、學校申復、及申復回覆的意見往返，探討學校與外部專家意見的一致性，以檢視系所質性評鑑的挑戰。研究問題有三：

問題一、對於辦學表現，系所與專家的訪評意見有哪些不一致的看法？亦即學校申復報告有哪些議題？

問題二、對於系所提出的申復，外部專家的看法為何？亦即外部專家回覆申復有哪些議題？

問題三、交叉比對系所申復與外部專家回覆的內容，雙方的看法一致性為何？對於質性評鑑威脅的意涵為何？

貳、文獻探討

以下分別從質性研究、後設評鑑的角度分析質性評鑑品質的相關論述，並探討系所評鑑的實施與挑戰之相關文獻。

一、從質性研究的角度探討質性評鑑的品質

（一）質性研究的典範與信實度

質性評鑑是當前普遍使用的方法，將社會學和人類學的質性研究方法運用於評鑑領域。由於質性評鑑是從質性研究的概念發展而來，以下探討不同研究方法典範的研究信、效度，幫助了解質性評鑑的品質。

早期質性研究是在後實證主義典範下發展，受到了量化研究的影響。後實證主義的典範，本體上認為有真實的存在，但不容易掌握，只能夠接近實在；在研究方法上，主張避免個人價值的涉入，以便呈現客觀的真實

（潘慧玲，2003；Denzin & Lincoln, 2011）。在此典範下的質性研究者，引用了量化研究的想法，認為只要遵循一定的方法與操作程序，將可提高研究的信、效度。Maxwell（1992）將效度分為三種類型：1. 描述的效度（Descriptive Validity），是研究者對事實報導的正確性；2. 解釋的效度（Interpretive Validity），指研究者能夠從受試者觀點出發，描述受試者對研究現象所賦予的意義；3. 理論的效度（Theoretical Validity），是指理論解釋能夠解釋現象。在此典範之下，採用長時間觀察及三角檢證法，運用多元方法蒐集資料，並透過外部專家檢視資料，提高研究結果的信、效度性（Stufflebeam, 2001）。

隨著社會研究方法典範的轉移，後實證主義逐漸轉變為建構主義、批判主義、以及實用主義，各研究方法的典範對於研究品質的看法也有所不同。建構主義主張的本體論，認為固定客觀的實在並不存在，每個人因為不同的社會情境產生不同的看法，因而建構不同的實在。同時，研究者與被研究的價值觀會涉入資料的蒐集與詮釋，兩者的互動共同建構了研究結果。在此典範之下，質性研究的品質，是透過研究者與被研究者在過程中不斷地對話，才能建立共同的詮釋結果（潘慧玲，2003）。而後，批判主義興起，主張歷史實在論，認為人類社會所謂的「實在」，其實是透過長期的社會、文化等因素形成，研究者對事物的理解，已經不是單純的主體對於客體的認識，而是主體與主體之間在社會文化脈絡之下的重新建構，是一個複雜的互動過程，因此需要持續的對話與辯證，以便接近實在（陳向明，2009；黃光國，2001）。質性評鑑的確實有效性，從過去追求客觀中立的事實，轉變為論證的確實有效性，研究報告與實際研究的相符程度（Cronbach, 1988; Kane, 2006）。近年來混合研究方法（mixed method research）逐漸興起，重視實用主義，價值是重要的角色，在本體上主張以經驗、效用決定實在，因此在研究方法上主張質、量並重的多元方法，質性研究的品質被認為是推論的品質，也就是研究方法上採取開放性的觀察，捕捉到多樣化的內部觀點、產生詳細厚實的描述（宋曜廷、潘佩妤，2010；Creswell & Clark, 2007; Johnson & Onwueabuzie, 2004）。此典範之下的研究者更進一步以實用主義的概念說明推論的品質，認為推論的品質包括概念的一致性、詮釋的同意度，以及以情境化意義去理解的因果關係（Teddlie & Tashakkori, 2009）。

不同研究典範各有不同的信念及增進研究品質的方法（表 1），「研究的信、效度」稱呼，已經難以涵蓋各家看法（陳向明，2009；Creswell & Miller, 2000）。詮釋主義學者 Guba 與 Lincoln 於 1994 年提出「信實度」（trustworthiness），以確實性（credibility）、轉移性（transferability）、可靠性（dependability）與確認性（confirmability）來取代內在效度、外在效度、信度與客觀性（Guba & Lincoln, 1994）由於本研究以質性資料為主，因此本研究採用「質性研究的信實度」來指稱質性研究的品質，以「確實性」取代效度，「可靠性」或「一致性」取代信度。

（二）質性研究的信實度對質性評鑑的意涵

受到不同典範的質性研究信實度的影響，質性評鑑品質的定義與改進方法，也隨著時代轉變。大學認可制運用了質性研究的特性，由大學進行自我評鑑，再由外部專家實地訪評，從內、外部觀點檢驗質性評鑑的信實度。另一方面，外部專家運用多元質性方式蒐集資料，如：晤談、現場訪視等。然而，專家實地訪視的過程，主要是基於一種假設：專家是客觀中

表 1 不同研究典範的信念及增進研究信實度的方法

項目		後實證主義	建構主義	批判主義	實用主義
不同研究 典範的 信念	本體論	有客觀實體存在 但無法充分瞭解	多元建構 的實體	長期社會 文化形成 的實體	經驗效用 形成的實在
	知識論	發現可能是 真實的	研究發現 是創造與 建構的	反映價值 立場的	視研究階段而 定，價值是 重要的角色
	方法論	假設與實驗	詮釋辯證 及脈絡考 量	對話與 辯證	多重方法
增進研究 信實度的 方法	研究者的角度	三角檢證	以證據反 駁	反思	開放的觀察
	參與者的角度	參與者檢證	長時間投 入	與參與者 合作	參與者對詮釋 的同意程度
	研究以外人們 的角度	外部人員的檢視	從厚實豐 富的描述 建構理解	外部同儕 的審視	了解推論、 從詳細的描述 中理解

註：整理自宋曜廷與潘佩好（2010）、黃光國（2001）、潘慧玲（2003）、謝志偉（2007）及 Creswell 與 Miller（2000）。

立的，可根據其學術專業適當地蒐集、詮釋資料，判斷方案的實際狀況（House, 2011）。在此假設之下，提高信實度的方法為充分掌握方法和程序，運用多種方法蒐集資料、多個資料來源、多位研究者共同詮釋，不需考慮專家的個別差異或情緒因素（Denzin, 1978; Greene, 2011）。

然而，此項假設隨著研究典範的轉移逐漸受到質疑，由於大學評鑑也是一種質性評鑑，實施品質端賴專家的判斷，歸納文獻得到大學評鑑品質的四個威脅：大學的情境脈絡、檢視結果而非動態過程、外部專家的價值涉入、缺乏長期投入與互動。第一個威脅來自大學的情境脈絡，由於每個方案都有其獨特的社會文化與經濟條件，這些因素相互交織，影響了方案的施行。因此，評鑑必須以各校的情境脈絡，了解該所大學的教育目的、課程、教學及學習成效（House, 2011; Kane, 2004）。

第二個威脅為評鑑檢視的是結果而非動態過程。由於實地訪視的焦點是教學與學習的結果，評鑑將這些視為獨立的現象，可以透過探訪而蒐集資料。但是，其實教學與學習是動態的過程，與大學情境息息相關。然而目前的評鑑設計，專家只能根據成果進行判斷，容易忽略動態的因素與情境的變化（Kane, 2006; Lather, 1993）。

第三個威脅是價值觀的涉入。質性評鑑是一種知識建構的過程，在資料詮釋與結果判定中，不可避免地加入個人價值觀（Abma, 2006）。外部專家在審查過程中，有可能因為自身不同的信念和經驗而產生不同的判斷，因此 Greene（2011）認為與其期待產生客觀的詮釋與推論，不如尋求評鑑詮釋的信實度；亦即蒐集各種證據，發覺其中的因果關係，並且以有意義的說明與推論作為評鑑結果。此外，評鑑本身就是一種價值判斷，每個判斷都反應了不同的價值觀、以及利害關係人的不同角度，所以質性評鑑很難被視為客觀的判斷，而是透過各種證據的蒐集與詮釋，呈現出一種有說服力的論述（Lather, 2001; Thomas, 2004）。

第四個威脅是缺乏長期投入與互動。質性研究需要研究者直接、長期地投入現場，與被研究者充分互動、對話，以了解真實情形。大學實地訪視是透過專家近距離觀察、與不同利害關係人對話，在持續與動態的關係中詮釋資料並進行推論（Greene, 2005; Schwandt, 2002）。專家審查論述的確實性，並非找出客觀的真理，而是在情境脈絡下了解大學的行動，從

校準、整合和詮釋證據中建立令人信服的論述（Smith, 1992; Teddlie & Tashakkori, 2009）。但是，外部專家與學校人員的直接互動主要來自實地訪視，但是時間相當短暫（僅為 2 天），不易充分互動，可能影響評鑑品質。

二、從後設評鑑的角度探討評鑑的品質

後設評鑑是一種對評鑑所進行的系統性評鑑，用以決定該評鑑的品質，並且提供改進方向。後設評鑑是由 Scriven（1969）首先提出，Stufflebeam（2001）更進一步提出操作型定義，是對於評鑑的一種描述、獲得、與判斷的過程，用以引導評鑑的實施及報導評鑑的優缺點。後設評鑑依照執行的時間與功用，可分為形成性與總結性的後設評鑑，形成性後設評鑑是在評鑑進行之前及執行過程中，對評鑑的設計與實施提出回饋改善之建議，偏重於改善導向；而總結性的後設評鑑則是對評鑑的優、缺點提出判斷，雖具有改善的功用，但較偏向績效責任（游家政、曾祥榕，2004；Scriven, 1969）。

後設評鑑長久以來受到各界的重視，常見的方案評鑑後設評鑑標準有二，一是美國教育評鑑標準聯合委員會（JCSEE）的方案評鑑標準，於 1982 年首先提出，1994 年及 2011 年後續修訂，目前為第三版（Yarbrough et al., 2011）；另一是由美國評鑑學會（American Evaluation Association, AEA）所訂定的美國評鑑學會評鑑人員指導準則，於 1995 年提出、2004 年修訂，共有五大原則與 25 項規範，其中五大原則為系統性探究、評鑑者能力、公正誠實的評鑑過程、評鑑過程尊重所有參與評鑑的團體、對社會大眾福祉的責任（AEA, 2004）。此二標準中，以 JCSEE 版本最為廣泛運用，被納為美國方案評鑑的國家標準。

JCSEE 方案評鑑標準的 2011 年版本，以五種屬性標準檢視評鑑品質。不同於前兩版，新增了第五個屬性「評鑑的績效責任」，主張以內、外部的後設評鑑檢視評鑑過程與成果（Yarbrough et al., 2011）。五大屬性包括：效用性、可行性、適切性、精確性、及績效責任。其中效用性的目的在於增加評鑑過程與結果所公布訊息的價值性，確保提供有用的資訊以滿足評鑑關係人的需求；可行性的目的在於提高評鑑的效能，而且能夠考量不同立場的關係人之需求，增加合作機會；適切性的目的在於確保評鑑是

適切、合法、正確及公正的，所有關係人的權益應獲得合理保障；精確性的目的在於提高評鑑陳述與發現的可靠性及真實性，以增進評鑑的品質；評鑑的績效責任，其目的在運用自我改善與績效責任的後設評鑑觀點來執行評鑑。

三、我國大學系所評鑑的實施與挑戰

我國大學系所評鑑由高教評鑑中心執行，以下說明實施情形及面臨的挑戰。

（一）我國大學系所評鑑的實施

我國一般大學校院的評鑑，可分為校務評鑑與系所評鑑，採取的是「認可制」（*accreditation*），由熟悉大學校務治理的專業同儕進行質性的品質判斷，以確保高等教育品質（王如哲、楊瑩、劉秀曦，2012；蘇錦麗，1997）。認可結果分為通過、有條件通過、及未通過等三種，認可期限為六年。第一週期系所評鑑的執行時間是從 2006 至 2010 年，共有 81 所大學校院及軍事校院的系所接受評鑑（高教評鑑中心，2012），而第二週期是從 2012 至 2016 年，改採雙軌制，有 34 所大學校院為大學自我認可（*self-accreditation*），或自辦外部評鑑，其餘持續接受高教評鑑中心的外部評鑑（教育部，2013b）。

系所評鑑程序分為四階段，第一階段是系所自我評鑑（*self-evaluation*），系所根據評鑑項目，自我檢視教育目標的達成情形，並將結果整理為自我評鑑報告書。第二階段為實地訪評，由外部專家組成實地訪評小組，蒐集系所表現的資料，包括座談、教學現場訪視、文件檢視、問卷調查及晤談，再將資料與系所自評報告比對，撰寫訪評報告初稿。第三階段是結果認定，訪評報告初稿經過兩階段的認定，先由相同學門的專家組成「學門認可初審小組」審核，再由跨學門專家組成「認可審議委員會」裁定認可結果。第四階段為後續追蹤，針對需要改進的項目，給予系所一年的改善期限，追蹤改進成效（高教評鑑中心，2012）。

（二）我國大學系所評鑑的申復與申訴機制

我國系所評鑑採取同儕檢視的外部評鑑，然而，為了尊重受評系所的權益，設計了申復與申訴的機制（高教評鑑中心，2015）。申復是在實地

訪評完成後、結果尚未公布前，受評系所可針對訪評報告初稿提出修正意見。若受評系所認為報告內容或評鑑程序不符，可提出申復，由高教評鑑中心召集實地訪評小組查證相關資料，並回覆結果。

申訴是在評鑑結果公布之後，受評系所若不認同評鑑結果，可提請重新裁量。為表示公正獨立，申訴是由「申訴評議委員會」（申評會）主持，委員會成員與實地訪評小組不同，由 9 位具有法律或教育評鑑的專家組成。申訴程序分為兩個階段，第一階段是申評會召開審查會議，根據申訴的項目與理由進行查證，結果分為申訴有理與駁回。若申訴有理，則進入第二階段，將重新召開「學門認可初審小組會議」，再次審查書面資料，做出以下建議：維持認可決議、重新實地訪評、或修訂認可結果，再將建議提報「認可審議委員會」議決之後公布。

綜合以上，申復的審理主要是實地訪評小組，針對訪評意見重新審視，是訪評委員與大學系所的再次對話與溝通；而申訴是經由第三方多位成員（非實地訪評小組、或受評系所人員）審查，包括法律與評鑑專家（申評會）、相同學門專家（學門認可初審小組）、跨學門專家（認可審議委員會）等，重新審視評鑑資料，以保障受評系所的權益（表 2）。由於本研究的目的是比較受評系所與實地訪評小組對於辦學品質的看法，因此針對申復報告加以分析。

（三）我國大學評鑑的品質挑戰

我國系所評鑑採行認可制，雖然評鑑過程中蒐集了大量、多元的資料，所有判斷均以資料作為依據，但是資料的信實性，對於系所與外部專家都是一大挑戰（王保進，2010）。整理我國高教評鑑之文獻，將評鑑資料信實性的挑戰分成五類，說明如下。首先是文件的確實性。外部專家可

表 2 我國系所評鑑申復與申訴機制的審理者與提出者

機制	審理者	提出者
申復	實地訪評委員	大學系所人員
申訴	第三方多位成員，含法律與評鑑專家（申評會）、相同學門專家（學門認可初審小組）、跨學門專家（認可審議委員會）	大學系所人員

藉由實地訪視檢視大量的系所文件，以直接了解辦學情形。但是系所通常提供大量資料，或選擇表現優良的部分呈現，也有可能巧妙地包裝評鑑資料，將缺點以正面的描述敘寫，或是盡量隱藏問題，因此文件的全貌性與真實性受到挑戰（曾淑惠，2009；黃政傑、張嘉育，2010）。而實地參觀由於時間有限，很難瀏覽所有設施，通常會引導外部專家參觀較好的部分。外部專家如何在有限時間內有效地蒐集分析資料以了解全貌，十分不容易。評鑑外部專家的評鑑品質，也是高教評鑑所面臨的挑戰（監察院，2010）。

第二是互動的確實性。理論上外部專家可在主管及教師座談，與學校人員直接互動以蒐集資料，但其確實性會受到一些因素影響，如：主管座談時間較短，難以一窺全貌，而且由於與會人員的權位不對等，發言會傾向保留而影響資料蒐集（黃政傑，2004）。

第三是晤談的信實度。對於大學端而言，雖然晤談是個別受訪且在密閉空間進行，理想上具有隱私性，但仍有其限制。如：受訪師生可能擔心被影響而不願表達真正想法；而受訪學生的代表性也是挑戰，雖然全體學生都可被抽選，但實際卻常因學生無法出席而影響抽樣的公正性（林天祐，2004；曾淑惠，2009）。對於外部專家而言，質性晤談技巧也是一大挑戰，如何在短時間內掌握關鍵問題，並從不同關係人的晤談中蒐集重要資料，需要有相當的質性晤談技巧。一方面與受訪者建立信任的關係，一方面能夠旁敲側擊地釐清某些議題，進行中立客觀的判斷，這些都需要評鑑與質性晤談的專業（楊國賜，2010；顏善邦，2010）。

第四是評鑑報告撰寫的信實度。外部專家透過評鑑報告，以文字的描述呈現大學辦學的觀察與綜合判斷結果，並對大學提出改善建議，是一種質性資料的呈現。但是在評鑑報告的撰寫上，容易出現一些問題，如：文字表達的流暢性與邏輯性不足，報告所指的問題未能對應改善建議；又可能因為內容過於簡單，缺乏完整說明，致使系所對於改善建議無法充分了解，因此評鑑報告應朝具體明確，且具有精確性（王保進，2010；劉維琪，2011）。

第五是系所學門的多樣性。系所評鑑前，需進行系所的學門歸屬，以邀請適當學門專業的外部專家進行評鑑。雖然高教評鑑中心在第一週期系

所評鑑歸納出 49 個學門（劉維琪，2010），但是由於有愈來愈多的跨領域學系，或是新興學系，難以歸類，或是由於專門領域愈分愈細，難以邀請到適合專長的委員，增加了系所評鑑的困難度（吳清山、王令宜，2007）。

參、研究設計與實施

一、研究對象

本研究的研究對象為我國第一週期系所評鑑提出申復的系所。該週期受評班制共計 3,165 個，提出申復有 1,156 班制（37%）（高教評鑑中心，2015）。若以班制為統計單位來看申復的結果，接受申復的有 116 班制（10%）、部分接受 115 班制（10%）、維持原建議 925 班制（80%）；若以申復項目為統計單位，共計 8,366 個申復項目，接受申復 842 項、部分接受 886 項、維持原建議 6,638 項。為能詳細比較每一項目中，系所與專家的意見往返情形，本研究以申復項目作為分析單位。

二、研究方法

本研究運用「內容分析法」（content analysis）進行資料分析。內容分析法是將蒐集到的文獻及資料進行分類，再以量化分析的方式加以統計。由於該方法是依照現有的資料進行分析，有系統地將資料進行分類編目，而非單純地蒐集資料內容，因此具有客觀性及系統性（王文科、王智弘，2013；Krippendorff, 2013）。本研究分析的文件是第一週期系所評鑑的相關資料，包括 2006 年下半年至 2010 年的訪評報告及申復報告書，所有資料均自高教評鑑中心網站下載（高教評鑑中心，2015）。

本研究以兩階段方式進行，第一階段是建立編碼表，主要參考 JCSEE 於 2011 修訂之「方案評鑑標準」的「精確性標準」擬定主要類目。JCSEE 的方案評鑑標準共含五大屬性：效用性、可行性、適切性、精確性、及績效責任（Yarbrough, et al, 2011）。然而，本研究擬探討質性評鑑的信實度，尤其是確實性與可靠性，與方案評鑑標準的精確性標準較為相關，與其他四個標準較無直接相關。而精確性標準共含 8 個面向：公正的結論與決定、確實有效的資訊、可靠的資訊、明確的方案與脈絡描述、資訊管理、

健全的設計與分析、明確推論、傳播與報導。其中兩個面向（資訊管理、健全的設計與分析）在本研究不適用，因為我國大學系所評鑑是由高教評鑑中心統一管理評鑑資訊、制定評鑑架構與分析方法，因此捨棄此二面向，並加上新類目「其他」，共計七個類目。

其次，研究者反覆閱讀申復報告及其回覆，據以建立編碼表的次要類目，並將編碼表分為學校申復報告以及審查回覆兩個層面。所建立號碼共分五段，第一段號碼為內容出處，包括：訪評報告、大學申復、及申復回覆，其代碼分別為 V, U 及 R。學校的申復原因共分為七大屬性（U1.....U7），再細分為 19 個細項（U11.....U73），包括：U1 學校對於評鑑結論與決定的公正性之疑問、U2 學校對於資訊的確實有效性的說明或疑問、U3 學校對於資訊可靠性的說明或疑問、U4 學校對於方案情境脈絡的說明或疑問、U5 學校根據評鑑進行詳細的評鑑推論說明或疑問、U6 學校對於訪評報告的傳播與報導的相關問題、U7 學校其他議題；而專家申復的回覆則分為七大屬性（R1.....R7），再分為 19 個細項（R11.....R72），包含：R1 專家對於評鑑報告的結論與決定的說明或疑問、R2 專家對於資訊確實有效性的考量說明或疑問、R3 專家對於資訊可靠的考量說明或疑問、R4 專家對於方案情境脈絡的考量或疑問、R5 專家以詳細的評鑑推論說明或疑問、R6 專家對於傳播與報導問題的回應或說明、R7 專家的其他議題。本研究編碼表的範例請見表 3。

表 3 本研究的編碼表（範例）

屬性	學校提申復的原因	外部專家對於申復的回覆
可信的資訊	U3. 學校對於資訊可靠性的說明或疑問	R3. 外部專家對於資訊可靠的考量說明或疑問
	U31. 學校要求專家澄清資料來源	R31. 專家要求學校澄清資料來源
	U32. 學校要求專家有一致性資訊	R32. 專家表示學校提供資料矛盾
		R33. 從不同資料均顯示結論一致

編碼的第二段號碼為年度，如 98 年度下半年為 802；第三及第四段號為學校及系所代碼，第五段為該內容於該系報告中的項目順序號碼。如：U3-502-10-05-03 為大學在 95 年下半年所提申復書，關於資訊可靠性的說明或疑問 U3，由學校代碼 10、系所代碼 05 所提的第 3 項意見。

第二階段是確定資料分析的一致性。本研究編碼表的審查，是由兩位教育評鑑教授進行，再參酌其意見予以修訂；至於資料的編碼，則是由研究者與兩位評鑑專家（評鑑實務經驗平均 5 年）進行一致性試驗，並將結果與前述審查編碼的其中一位教育評鑑教授討論。第一次分別選取 60 個申復項目及回覆（共計 120 項）進行編碼，平均一致性為 .83 及 .80。重新討論後進行第二次試驗，分別抽取 30 個申復項目及回覆，一致性提升至 .87 及 .85。三人針對不一致再次討論，第三次試驗分別抽取 30 個申復項目及回覆，一致性達到 .92 及 .90。

肆、研究結果

為瞭解大學與外部專家對評鑑看法的差異，以下分析大學申復以及外部專家回覆申復的內容、並且交叉檢視兩方觀點的差異。

一、大學申復的主要議題

分析大學提出申復的內容，七大屬性中，以「U2 學校對於資訊的確實性的說明或疑問」出現的頻率最多，共有 5,722 項（佔 68.4%），如圖 1 所示。檢視該屬性的細項分布，頻率最高的是學校提供更多詳盡的資料說明辦學成效，希望獲得專家採納而改變決定（U21）、或表示先前已經提供足夠資料，請專家予以考慮（U23）、及學校正在改善，希望專家了解（U24）；另一方面，學校也質疑專家引用資料的確實性（U22），詳見圖 1。

頻率出現第二多的屬性是學校質疑評鑑結論與決定的公正性（U1），共有 1,119 項（佔 13.4%）。檢視原因，主要是學校認為訪評結論不合理（U11），包括學校認為專家建議難以達成、不符合學校情境、或是質疑專家如何達成訪評結論。其次是學校透過申復報告表示日後會加強改進

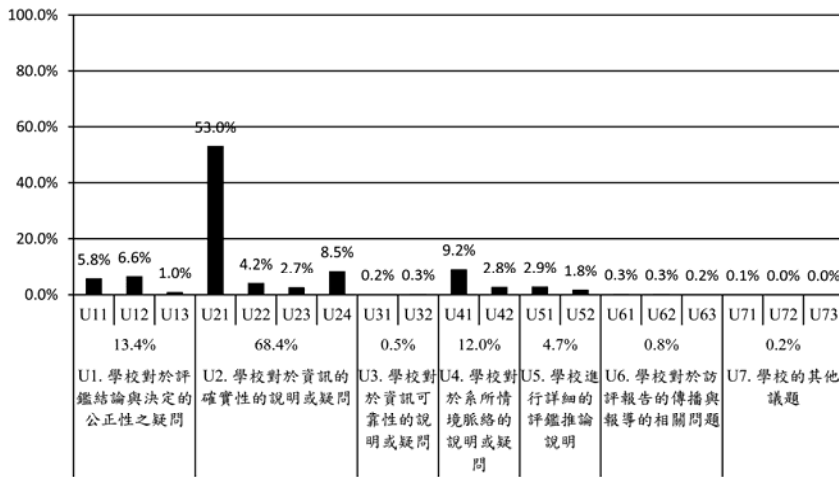


圖 1 大學提出系所評鑑申復的主要內容

（U12）、以及學校僅表示反對訪評的結論，但無具體證據（U13）。此處值得注意的是，學校雖然同意訪評結論，但卻又提出申復，推測主要原因是提出申復的時間點在評鑑認可結果公布之前，因此學校希望藉由認同專家意見，或是表示積極改進的決心，爭取評鑑過關的機會。

第三多的屬性是學校對於系所情境脈絡的說明或疑問（U4），共有 1,005 項（佔 12.0%）。學校透過申復報告，說明系所或學校的特殊情境（U41），例如系所是新設立的，或是系所學科的特殊性。也有學校表示該系所符合一般規定、或是該系所的表示是配合教育部、或按該校的校級規定辦理（U42），希望外部專家予以考慮。其他屬性依照出現頻率之排序為：學校進行詳細的評鑑推論說明（U5, 4.7%）、學校對於訪評報告的傳播與報導的相關問題（U6, 0.8%）、及學校對於資訊可靠性的說明或疑問（U3, 0.5%）。

二、外部專家回覆申復的主要議題

分析外部專家對於大學系所申復的回覆內容，發現七大屬性中，以「R1 專家對於評鑑報告的結論與決定的說明」出現的頻率最多，共有 3,271 項（佔所有回覆項目 48.4%），如圖 2 所示。檢視細項的分布，主要是專

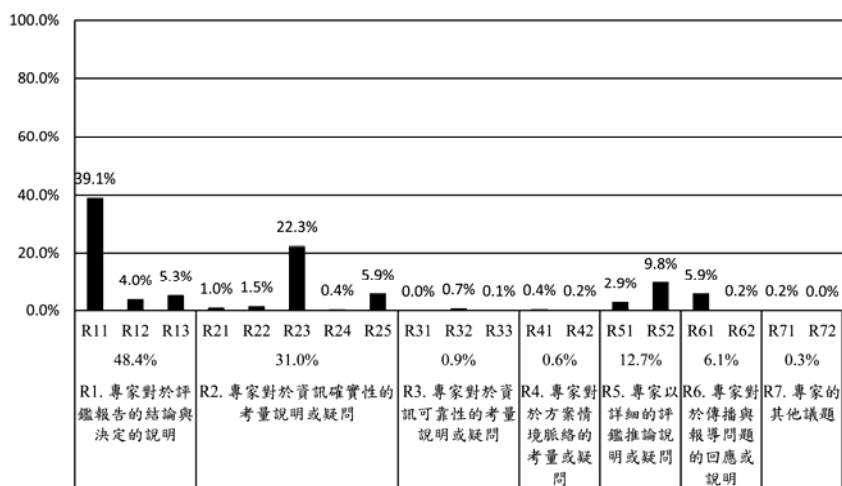


圖 2 外部專家回覆系所評鑑申復的主要內容分析結果

家簡單重申訪評報告的結論（R11）、專家認為學校雖然提出申復，但實際已同意訪評結論（R12）、專家表示學校針對的是訪評報告中的狀況描述，並非建議，不須申復（R13）。

頻率出現第二多的屬性是 R2 專家對於資訊確實性的考量說明，共有 2,593 項（31.0%）。檢視細項的分布，主要是專家表示學校提供的資料時間不對，為訪評後的改善資料（R23）、表示學校提供的資料不足，無法確實反駁訪評建議（R25）、表示學校未針對議題提供資料（R22）、訪評當日學校未提供相關資料（R21）、表示學校引用資料錯誤（R24）。

第三多的屬性是 R5 專家以詳細的評鑑推論說明，共有 1,066 項（12.7%）。包括專家詳細說明資料的計算方式及推論（R51）、詳細說明資料的蒐集與推理（R52）。值得一提的是，第四多的屬性為「R6 專家對於傳播與報導問題的回應或說明」，由於專家與學校溝通上的問題，造成了學校對於訪評建議的誤解，所以建議未來應將溝通與報導納入培訓課程之中。檢視細項的分布，包括：專家表示學校摘述並非改善建議之原意（R61）、專家表示乃用字遣詞，並不影響結論（R62）。

三、交叉檢視大學申復與外部專家回覆的觀點

以下根據前述學校最常提出申復的八個細項，分析專家回覆申復的內容，了解雙方看法的差異。

（一）學校提供更多確實的事證

學校最常提出申復的內容，認為外部專家所持的資料不足，未能充分了解學校作法，因此傾向於提出更多事證，展現辦學成效（U21, 53.0%），包括提出數據或具體作法，但未詳加說明推理或計算方式。然而，分析外部專家的回應主要是重申訪評結論（R11, 19.1%）、表示學校提供的資料時間並非評鑑時間範圍內（R23, 10.2%）、詳細說明訪評報告資料的蒐集與推理（R52, 5.2%）。

（二）學校說明系所的特殊情境

學校提出第二多的申復內容，為說明系所的特殊情境或獨特的學科本質（U41, 9.2%），包括：新設立的系所、學科的特別性質、因應系所專業領域而產生的特別規劃、因應該校規定或是教育部政策而產生的規劃、及系所為跨領域學科。如：系所在申復報告中表示，由於成立時間較短，尚無畢業生，無法充分呈現辦學成效。

訪評當時成立剛滿一年，部分課程尚未開設，亦尚無學生完成本所學業……

許多籌備過程中所規劃開設之課程及應辦理之學生學習活動，尚未完全實現……（U41-502-07-18-03）

又如：建議評鑑應考慮該系的特殊學科性質，給予適當的彈性空間。

因本學院性質較其他大專院校不同，因此在評鑑的標準與審查資料上，勢必會有所區分……然而，後續實際的審查過程卻是依照各校相同之項目進行評鑑……並沒有當初所說明的彈性空間……（U41-502-08-13-01）

外部專家回覆此類申復的方式，主要是重申訪評結論（R11, 2.8%）、以詳細的評鑑推論說明或疑問（R52, 2.1%）、說明學校提供的資料時間

不對（R23, 1.1%）。另一方面，相較於系所主張自身的特殊情境，專家主動提及考量系所情境的比例相對較少（0.1%），包括：建議系所以學生學習能力的訂定教育目標、以及規劃課程與教學。

（三）學校不認同訪評報告的結論

學校提出第三多的申復內容，是質疑評鑑結論與決定的公正性，認為訪評結論不合理（U11, 5.8%），包括：訪評結論與系所實際運作不符、訪評建議不符合學校情境而難以達成、或是訪評建議有理，但礙於超越系所管轄的校級規定而難以改進。如：

本系通識課程大學部學生至少須修習 30 學分……因通識課程、講座的師資安排皆由本校通識中心主導負責，故系上只能被動採建議方式……（U11-502-11-02-02）

外部專家對於此類申復的回覆，大都是簡短重申訪評結論（R11, 2.1%）、表示學校資料是訪評之後的改善情形，無法採納（R23, 1.0%）、專家詳細說明資料的蒐集與推理（R52, 0.6%）、以及認為系所在申復所摘述的訪評意見與改善建議並非原意（R61, 0.1%）。如：

……該系並未針對外部專家之問題進行回應。本訪評意見的內容為「教師研究能力」之檢討，而非申覆意見所述之「教學技巧與教學成效」之檢討……（R61-602-05-03-01）

（四）學校質疑專家引用的資料有誤

學校質疑資訊的確實性，認為專家引用的資料有誤（U22, 4.2%），包括認為訪評報告引用的資料錯誤、訪評晤談的抽樣有所偏差或是樣本過少，晤談內容屬於個人言論，無法代表全體利害關係人的看法。

此項建議或為訪談不同學年學生導致的誤解所致……此乃本系新增之課程，如委員當日訪談對象為……學生，其所屬學年度確實沒有這個科目……（U23-502-07-02-02）

專家對於此類申復的回應，主要是簡單重申訪評結論（R11, 1.4%）、詳細說明資料的蒐集資料與推理（R52, 1.0%）、說明學校提供的資料為訪評之後的改善作法（R23, 0.4%）。

（五）雙方溝通產生問題

學校認為與專家的溝通產生問題（U61, U62, U63, 共 0.8%），包括不了解審查意見、以及訪評意見不明確。雖然此項比例不高，但需要重視與改善，以避免雙方誤解。包括詞語不明確、或是指稱的標的物不清楚，如：「少部分」、「經常」。

.....學生認為少部分課程內容太過淺顯.....（V61-502-17-01-03）

學校提出疑問：

本系課程內容係考量一般學生理解力及認知程度而設計.....不知所指課程內容太淺為何種課程？希有進一步資訊以利本系參考改進.....（U61-502-17-01-01）

外部專家對於此類型的回應，主要是重申訪評結論（R11, 0.3%）、詳細說明資料蒐集與推理過程（R52, 0.2%）、表示學校提供的資料不足，包括未提供具體的辦法、描述、或具體數字。

（六）評鑑資料的一致性

系所對於專家意見的資訊來源提出疑問、以及質疑訪評意見不一致（U31 & U32, 0.5%）。如：

懇請說明「畢業生表現未提供相關資料」，因為下一段又說「畢業生的升學與就業率情況良好」，兩者似有矛盾之處.....（U32-502-07-02-04）

外部專家的回應方式，主要以詳細的評鑑推理回應（R52, 0.2%）、重申訪評結論（R11, 0.1%）、認為與學校的溝通產生問題（R61, 0.06%）。值得一提的是，專家回覆申復內容中，也質疑學校資料的一致性，認為學校提供的資料前後矛盾。如：

該系申復書提及師資與圖書期刊「未完全平均」，卻又說「未有失衡」，似有所矛盾。（R32-601-06-03-02）

雖然學校因應評鑑需要準備的資料量十分龐大，而且由於時間急迫而有所疏漏，但是，長期耕耘的教育成效若因一時的資料誤植而引起誤會，將十分令人惋惜。因此需要重視各種資料之間的一致性，如：數據在自我評鑑報告、現場訪視資料、及申復引用的必須相同，避免錯誤產生。

伍、討論

大學系所質性評鑑為質性研究的一種，在 Creswell 與 Miller（2000）的架構下，外部專家可視為主體，大學系所是客體。在這過程當中，透過主、客體之間的互動，外部專家蒐集分析了多種資料，建立了系所教育品質的質性論述。雖然大部分系所認同系所評鑑的品質，平均申復比例 37%，且有逐年下降的趨勢（高教評鑑中心，2012），但為精進評鑑品質，本研究參考質性研究的信實度觀點，探討影響系所評鑑品質的可能原因。

一、影響系所評鑑品質之可能原因

（一）部分評鑑資訊的確實性及可靠性不足

質性研究注重資料的三角檢證，提高研究結果的信實度，而我國系所評鑑運用了此概念設計評鑑機制，從多種資料來源、多樣資料類別、多位研究者等蒐集學校表現資料（Denzin, 1978; Greene, 2011）。實際的執行上，第一是以蒐集多種方法蒐集資料，包括：大學自評報告、文件檢視、座談、晤談、現場教室觀察等方法；其次是多種資料來源，盡可能地囊括所有關係人的意見，包括行政主管、教師、職員、學生、畢業校友、及畢業生的雇主，由不同角度檢視學生學習成效。第三是多位評鑑者共同檢視與詮釋質性資料，避免個人意見影響評鑑結果，實地訪評小組委員為 4 至 6 人。

雖然以三角檢證的方式設計系所評鑑的資料蒐集，但本研究發現，資料的確實性與可靠性仍然是申復的重要議題。大學申復最常見的內容，是認為外部專家需要更多確實有效的資料，因此傾向於提供更詳盡的資料說

明辦學成效，或是以詳細的評鑑推理說明該系的課程設計、教學及學習成效。然而，提供更多的資料卻不一定獲得外部專家的認同，因為所提供的資料未必全是確實、可靠的資料，例如：提供的數據或作法並不在評鑑所規範的時間範圍（確實有效的資料需要在實地訪視的前兩年半至前半年）；相對地，系所也對訪評報告資料的確實性與可靠性提出疑問，認為外部專家訪談的抽樣有偏差、或是評鑑結果的判斷過於重視晤談而忽略其他形式的資料。

上述分析印證了 Abma（2006）質性評鑑的價值觀威脅。由於質性評鑑是一種知識建構的歷程，在資料詮釋與評鑑結果判定的歷程中，會因為個人價值觀而有不同的解讀。因此，質性評鑑不應追求客觀中立的事實，而是需要強調評鑑的詮釋與推論的確實性（Greene, 2011; Kane, 2006）。然而，本研究的結果顯示，系所認為的確實有效的資料，在專家的眼中不見得成立；而專家認為透過晤談及書面資料等多種資料，應足以讓學校接受評鑑結果，但是結果卻不然，因此仍然需要加強資料的信實性。

（二）部分訪評報告未能確實掌握情境脈絡的意義

我國系所評鑑為了重視大學的情境脈絡，在評鑑中特別設計了三種機制。第一，實地訪評專家的專長必須與受評系所相同或密切相關，以確保背景訓練類似，能從相同的著眼點考量。其次，系所可根據該校特殊情境自訂教育目標，據以規劃課程與教學，以達到預定的學習成效；而外部專家依循各校目標檢核成效，不做學校之間的比較。第三，各系所於實地訪評前進行自我評鑑，並將結果送交實地訪評小組。透過系所內部觀點，充分表達執行情形及所處脈絡，提供專家參考。

雖有上述設計，但是申復結果顯示，系所希望專家考量學校情境脈絡的比例仍然不低。前述文獻指出，質性評鑑的威脅是未能充分考量情境脈絡（House, 2011），以及缺乏長期投入與互動（Greene, 2005）。質性研究重視情境脈絡的意義掌握，研究者能夠從被研究者所處的情境，詮釋其行為及表現。然而，根據社會建構理論，人的行為是社會文化下所產生，對於外界現象的詮釋與判斷，不免會涉入自身的價值觀，因此需要藉由不斷的對話與辯證，產生接近被研究者認同的詮釋。但是反觀系所評鑑的實地訪視，為期 2 天，時間短暫，而且資料多為片段結果，無法完整呈現所

有歷程，導致影響外部專家的判斷。此外，系所與專家的會談，是在非自然情境、面臨被檢視壓力下的互動，其真實性及情境脈絡考量的困難度，比一般自然探究的質性研究要高。

（三）雙方意見的往返未能充分對焦

質性研究中，重視的是研究者長期沉浸於研究現場，與被研究者有充分的觀察與溝通。我國系所評鑑的機制設計上，有三項設計，增進系所與外部專家之間的溝通。首先，在實地訪評前，外部專家可以針對系所自我評鑑報告書，提出待釐清問題，而系所可在實地訪評前，針對問題特別備齊相關資料，以供外部專家檢視。第二，實地訪評期間，外部專家可透過多種形式與評鑑關係人溝通，包括：座談、晤談、或是教學觀察等。第三，是外部專家若對某部分資料有問題，可以直接請學校相關人員進到訪評現場，進行面對面的溝通。雖然有多種溝通管道的設計，但是本研究結果顯示，大學系所與外部專家的意見往返仍然存在一些問題，歸納為兩種：首先是雙方關注的議題未能對焦，從委員認為主要原因是：學校對訪評意見的曲解、或是答非所問。其次是雙方關注的議題能夠對焦，但是看法不同，學校的主要考量是訪評建議難以執行，而委員則認為學校提供的資料不充分。

前述文獻指出，質性評鑑的信實度應以推論的品質進行判斷，研究方法上應以開放性觀察、詳細厚實的描述（Creswell & Clark, 2007），顯示出概念的一致性、詮釋的同意度、以及以情境意義理解因果關係（Teddlie & Tashakkori, 2009）。本研究結果發現，專家與系所人員雙方互動時間較少，而且溝通形式以書面資料居多。雖然實地訪評可以直接進行外部專家與大學人員的直接溝通與釐清，但由於行程安排十分緊湊，再加上資料繁多，需在短時間之內翻閱，因此不容易有充分的溝通，大部分有賴於書面資料的說明。然而，書面表達表達有其限制，從文字推測對方的想要表達的意義，若無豐厚的描述，容易有所誤解。未來是否能在時間與經費的考量之下，如何增加外部專家與學校的互動，讓雙方的意見往返能夠對焦，這將是一個重要議題。

（四）部分訪評報告結論的合理性可再增加

質性評鑑不同於一般質性研究，在於必須做出判斷並且提供改善建議，因此評鑑報告的結論十分重要。本研究發現，雖然大部分學校認同訪評報告的結論，但仍有部分結論未獲得認同，如：學校認為結論難以達成、或不符合學科特性。然而，訪評結論為系所辦學改善所依據的重要外部意見，而且會被持續追蹤改善情形，甚或影響系所的未來發展方向。

前述文獻指出，質性評鑑有情境脈絡威脅，以及缺乏長期投入與互動的威脅。若希望評鑑是有建設性的，唯有長期投入了解受評系所的立場與情境脈絡，考量系所可運用的資源，促使評鑑結果的運用，並且不斷地回饋改進。「評鑑產生使用」近年來受到評鑑社群的重視（Mark & Henry, 2004; Patton, 2008），而效用性標準也是教育評鑑標準聯合委員會的方案評鑑專業標準之一（Yarbrough et al., 2011），可見評鑑需要在實務提供資訊給意圖使用的關係人。評鑑不僅可以評估績效，也可促進學校整體發展，成為組織內不斷提供改進回饋的重要機制。雖然評鑑有其改善的功用，但是，大學評鑑產生評鑑結果運用的最重要關鍵，在於大學認同評鑑的結論，願意運用評鑑意見進行回饋改善。因此，評鑑的結論若能以該校情境為出發點，提出更具合理性的結論，將會大幅增加評鑑結果的運用性。

二、綜合討論

綜合上述，我國系所評鑑運用增進質性研究信實性的方法，有對應的增進品質設計。然而，由於質性評鑑的品質，端賴專家對於質性資料的蒐集、詮釋、及結果判斷，所以從專家執行評鑑的角度，系所評鑑面臨了四項威脅，包括：未能考慮情境脈絡、檢視的是結果而非動態歷程、專家的價值觀涉入、評鑑歷程缺乏長期投入與互動。本研究從前述的分析結果，發現這些威脅的確出現在申復報告當中，造成質性評鑑的挑戰（表 4）。

首先，系所評鑑的設計運用了資料的三角校正的概念，以多種資料類別、多方資料來源、多位評鑑者詮釋資料等方式進行。但是從前述的分析

表 4 比較一般質性研究增進信實度的方法、我國系所評鑑增進品質的設計及面臨的挑戰

一般質性研究增進信實度的方法	我國系所評鑑增進品質的設計	由申復報告的分析，歸納系所評鑑的挑戰
資料的三角校正	多種資料類別、多方資料來源、多位評鑑者詮釋資料	質性評鑑的價值觀威脅。雙方認定的資料確實性不一定對對方成立。
情境脈絡的意義掌握	專家與系所具有相同專長、系所可根據情境自訂教育目標、並可運用內部觀點進行自我評鑑，提供專家參考	缺乏長期投入與互動的威脅。實地訪評僅為 2 天，且資料多為片段結果。
以對方立場思考而進行詮釋及提出結論	評鑑培訓設有評鑑報告課程	質性評鑑的情境脈絡威脅。可從關係人立場出發，多考量結論的合理性，增進評鑑結果運用。
長期沉浸研究現場進行觀察與溝通	以不同形式增進溝通與了解 1. 實地訪評前可針對自評報告提出待釐清問題。 2. 專家可透過多種形式與評鑑關係人溝通。	檢視結果而非動態歷程、及缺乏長期投入與互動。系所評鑑互動時間短、為書面溝通、多種溝通管道的限制性。

結果，發現系所評鑑仍然受到了價值觀的威脅，由於專家與系所人員的雙方立場不同，單方認定確實有效的資料，對方不一定認同，因此評鑑資訊的確實性為挑戰之一。

第二，系所評鑑的設計運用了情境脈絡意義的掌握，包括：外部專家與受評系所具有相同專長、受評系所可根據所處的情境自訂教育目標、可運用內部觀點進行自我評鑑，並提供外部專家參考。但是，這樣的設計，仍然受到了未能掌握情境脈絡及缺乏長期互動的威脅，部分學校認為專家未以受評系所的立場進行評鑑。推測可能的原因，是因為實地訪評時間僅為 2 天，且資料多為片段結果，無法見到所有歷程，不易掌握情境脈絡下的意義。因此如何在有限時間掌握完整歷程，除質性研究的知能之外，評鑑專業知能的增進亦十分重要。

第三，系所評鑑以協助學校回饋改善為主要目的，以系所立場思考提出結論，因此高教評鑑中心要求所有專家在實地訪評之前，須接受培訓課程，了解評鑑報告的撰寫準則。但是，分析結果顯示學校對於評鑑結論的公正性與合理性的質疑比例不少，因此未來可以設計相關的評鑑培訓課

程，增加專家的評鑑專業知能，能夠從利害關係人立場出發，多考量結論的合理性，以增進評鑑結果的運用。

第四，系所評鑑運用多元方式進行專家與系所之間的溝通，但是，這樣的設計仍然受到質性研究的威脅。由於系所評鑑溝通互動時間短、多為書面溝通、或是多種溝通管道的設計仍有其限制，如：系所人員由於權位不等，發言有所保留。因此，專家與系所雙方的溝通仍然有其限制。

陸、結論與建議

本研究探討系所與專家對於系所評鑑觀點的差異性，並從質性評鑑的精確性瞭解面臨的挑戰，得到的重要結論與建議如下：

一、結論

（一）系所申復報告以說明或質疑資訊的確實性最多

系所申復報告的內容，頻率出現最高的是「學校對於資訊的確實性的說明或疑問」，佔 68.4%，包括學校提供更多辦學資料，促使專家改變決定、及學校說明已接受建議，並在改善當中。頻率出現第二多的屬性是「學校質疑評鑑結論與決定的公正性」，佔 13.4%，包括學校認為專家建議難以達成、不符合學校情境、或是質疑專家如何達成訪評結論。頻率出現第三多的屬性是「學校對於系所情境脈絡的說明或疑問」，佔 12.0%，包括系所說明所處的特殊情境、學科的特殊性、或表示是配合教育部或校級規定辦理，符合一般原則，並無特殊之處。

（二）專家回覆系所的申復以說明評鑑報告的結論與決定最多

專家對於申復的回覆，頻率最高的是「專家對於評鑑報告的結論與決定的說明」，佔 48.4%，包括專家簡單重申訪評報告的結論、認為系所雖然提出申復，但實際已同意訪評結論。頻率出現第二多的屬性是「專家對於資訊確實性的考量說明」，佔 31.0%，包括專家表示學校提供的是訪評後的改善資料、學校提供的資料不足，無法確實反駁訪評建議。第三多的屬性是「專家以詳細的評鑑推論說明」，佔 12.7%，包括專家詳細說明資料的計算及推論、或是資料的蒐集與推理。

（三）系所申復最出現的三類屬性，專家給予回覆的方式類似

交叉比對申復及其回覆，探討學校與外部專家對於大學評鑑意見的一致性，分析結果顯示系所申復內容最出現的三類屬性，外部專家給予回覆的方式類似，最常出現的方式是重申訪評意見的結論，第二與第三分別是專家表示學校提供的資料時間不對、外部專家以清楚的評鑑推理回應，詳細地說明訪評報告的資料是如何蒐集與推理。

（四）大學質性評鑑的四項可能威脅

我國系所評鑑運用增進質性研究信實性的方法，在評鑑的歷程中有維持評鑑品質的設計，如蒐集多元資料及參考系所自我評鑑的內部觀點等。但是，由上述分析，系所評鑑仍然面臨了四項威脅，包括：未能考慮情境脈絡、檢視的是結果而非動態歷程、專家的價值觀涉入、評鑑歷程缺乏長期投入與互動。

二、建議

（一）全面提升評鑑專業知能，並增進質性資料的提供與分析詮釋的品質

系所評鑑為質性評鑑，質性資料為評鑑結果判斷的重要依據，然而上述研究結果顯示，不論是系所提供的資料、或是訪評報告引用與詮釋資料的信實度，仍有改善空間。因此，學校人員及外部專家，需要全面提升評鑑的專業知能。近年來「評鑑專業知能的建立」為重要議題，著重評鑑專業態度、知識技能與行為，包括評鑑資料的蒐集、分析詮釋、成果運用、及評鑑倫理（鄭淑惠，2010；Cousins, Goh, Elliott, & Bourgeois, 2014）。其中質性資料的蒐集詮釋及情境的分析能力，更是受到重視，被囊括於Stevahn 與 King 的「方案評鑑人員的必要技術與能力」（Stevahn, King, Ghore, & Minnema, 2005）……以及加拿大評鑑學會（Canadian Evaluation Society, CES）「六大評鑑專業能力」之中（李懿芳，2012；CES, 2010）。

（二）將各校的特殊性納入系所評鑑的設計與實施

我國第一週期系所評鑑，是以五大評鑑項目與 38 個評鑑效標檢視系所辦學成效，雖然載明學校可依各校特殊情境脈絡，調整評鑑項目與效標，但目前為止，仍未有系所提出變動。2014 年高教評鑑中心新修訂的第二週期系所評鑑項目，已明訂各校可依照學校特色，增訂特色項目或特

色效標；然而，大部分學校仍然抱持著觀望的態度。因此建議將各校的特殊性納入系所評鑑的設計與實施，讓各校可以發揮特色，不須擔心增加評鑑項目被扣分，也能被外界適當的監督。

（三） 建立討論案例，作為評鑑專業課程的教材

本研究分析第一週期的申復資料，了解學校與外部專家對於評鑑看法的差異情形，並運用質性研究方法以及後設評鑑的精確性概念，檢視評鑑執行的信實性。由於所有資料均來自申復報告，是大學與外部專家的真實反應，可做為評鑑課程之討論案例。如：報告宜使用較精確的詞彙、質性資料蒐集宜注重多元資料的整合，不宜偏重晤談。此外，學校評鑑準備人員的評鑑概念也需要加強，可利用本研究的分析結果，設計申復課程，如：提出申復需針對訪評意見提供相關資料，且須呈現具體事實或證據。

（四） 增進評鑑結果的運用，以評鑑促進系所改進

評鑑關係人能夠運用評鑑結果，是執行評鑑的驅動力，因此評鑑須分析評鑑關係人的立場，並將焦點關注於他們所重視的議題之上（Patton, 2008）。大學系所評鑑的最主要目的是促進大學改善，因此評鑑若能讓學校人員透過評鑑檢視辦學表現，並且增進其對於評鑑過程與結果的運用，將會增加大學回饋改進的成效。因此，實地訪視在蒐集與解釋資料之餘，若能顧及訪評報告內容能夠被學校了解、建議能考量學校的情境脈絡，將有助於評鑑結果被學校使用，提升學校的改善意願。

（五） 以大學自我改進為重點，完善大學內部品質保證機制並建立評鑑文化

要解決質性評鑑的根本問題，應該回歸到評鑑的原始目的，協助學校精進辦學品質。亦即大學若能建立完整的內部品保機制，則可減少外部評鑑對於學校內部運作的干預，大學與外部專家的誤解亦不復存在。我國大學系所評鑑在 2012 年施行大學校院自我認可，各校可根據特色自訂評鑑指標，自聘外部專家檢核，建立學校品保系統。但是，大學辦學品質的維持，需將品保機制深植於學校內部系統中，因此最終目標是建立大學的評鑑文化，促使大學自發性地永續維持教育品質。Preskill 與 Boyle（2008）主張透過組織成員的互動，建立共同的評鑑價值觀，形成運用評鑑回饋改進的氛圍，從個人評鑑專業知能的建立，逐步擴展到組織評鑑文化的形

成。近來各校開始推動「校務研究」，透過大學內部的蒐集與分析資料，幫助學校的運作、發展與決策，這是形成評鑑文化的第一步（王麗雲，2014；彭森明，2013）。未來若能推動評鑑文化，將使大學內部品保機制更加完善。

大學校院的系所質性評鑑，需要大學與外部專家具備評鑑專業知能，並從根本審視學校評鑑的真義，才能深入質性評鑑的問題。本研究透過分析當前大學質性評鑑的執行情形及面臨的挑戰，提出增進質性評鑑品質的建議。由於研究者僅採第一週期系所評鑑的申復報告為研究對象，並未納入第二週期系所評鑑或是校務評鑑的資料，而且我國大學評鑑為質性評鑑的一種，受限於我國高等教育的情境脈絡，並未進行其他國家高等教育質性評鑑的比較，因此推論有限。日後研究相關議題，建議可以將多種質性評鑑的資料納入研究範疇，以期作為未來改進評鑑機制、設計評鑑課程、以及擬訂未來評鑑政策之參考。

參考文獻

- 王文科、王智弘（2013）。**教育研究法**。臺北市：五南。
- [Wang, W. K., & Wang, C. H. (2013). *Research in education*. Taipei, Taiwan: Wunan.]
- 王如哲、楊瑩、劉秀曦（2012）。臺灣高等教育評鑑的回顧與展望。**臺灣教育**，674，20-24。
- [Wang, R. J., Chan, Y., & Liu, H. I. (2012). Retrospect and prospect of higher education evaluation in Taiwan. *Taiwan Education Review*, 674, 20-24.]
- 王保進（2010）。專業化評鑑之效度總體展現－評鑑報告撰寫。**教育評鑑雙月刊**，25，16-21。
- [Wang, B. J. (2010). Validity of professional evaluation: Evaluation report writing. *Evaluation Bimonthly*, 25, 16-21.]
- 王保進（2011）。第一週期系所評鑑結果之後設評鑑：評鑑報告內容分析。**評鑑雙月刊**，29，9-14。
- [Wang, B. J. (2011). Meta-evaluation of the 1st cycle program accreditation. *Evaluation Bimonthly*, 29, 9-14.]
- 王麗雲（2014）。透過校務研究進行自我評鑑與自我改進。**評鑑雙月刊**，47，19-23。
- [Wang, L. Y. (2014). Conducting self-evaluation and self-improvement via institutional research. *Evaluation Bimonthly*, 47, 19-23.]

- 吳芝儀、李奉儒（譯）（1995）。**質的評鑑與研究**（原作者：M. Q. Patton）。臺北市：桂冠。（原著出版年：1990）。
- [Patton, M. Q. (1990). *Qualitative evaluation and research methods* (C. Y. Wu & F. J. Lee, Trans.). Beverly Hills, CA: Sage.]
- 吳清山、王令宜（2007）。我國大學評鑑：挑戰、因應策略與發展方向。**課程與教學季刊**，**10**（4），15-30。
- [Wu, C. S., & Wang, L. Y. (2007). University evaluation in Taiwan: Challenges, strategies, and developments. *Curriculum & Instruction Quarterly*, 10(4), 15-30.]
- 宋曜庭、潘佩妤（2010）。混合研究在教育研究的應用。**教育科學研究期刊**，**55**（4），97-130。
- [Sung, Y. T., & Pan, P. Y. (2010). Applications of mixed methods research in educational studies. *Journal of Research in Education Sciences*, 55(4), 97-130.]
- 李懿芳（2012）。加拿大評鑑人員資格認證機制及其對我國教育評鑑專業化之啟示。**教育政策論壇**，**15**（4），1-24。
- [Lee, Y. F. (2012). An exploration of the Canadian professional designations project and its implications for educational evaluation in Taiwan. *Educational Policy Forum*, 15(4), 1-24.]
- 林天祐（2004）。教育評鑑實施過程與方法的專業化。**教育資料集刊**，**29**，27-52。
- [Lin, T. Y. (2004). The professional inquiry of processes and methods of educational evaluation. *Bulletin of Educational Resources and Research*, 29, 27-52.]
- 財團法人高等教育評鑑中心基金會（2012）。財團法人高等教育評鑑中心基金會2012年報。取自<http://www.heeact.edu.tw/public/Data/3491672571.pdf>
- [Higher Education Evaluation & Accreditation Council of Taiwan (2012). *2012 annual report of HEEACT*. Retrieved from <http://www.heeact.edu.tw/public/Data/3491672571.pdf>]
- 財團法人高等教育評鑑中心基金會（2015）。第一週期系所評鑑結果。取自http://www.heeact.edu.tw/sp.asp?xdurl=appraise/appraise_list.asp&ctNode=491&mp=2
- [Higher Education Evaluation & Accreditation Council of Taiwan (2015). *Evaluation result of the 1st cycle of program accreditation by HEEACT*. Retrieved from http://www.heeact.edu.tw/sp.asp?xdurl=appraise/appraise_list.asp&ctNode=491&mp=2]
- 教育部（2013a）。大學評鑑辦法。取自<http://edu.law.moe.gov.tw/LawContentDetails.aspx?id=FL041732&KeyWordHL=&StyleType=1>
- [Ministry of Education (2013a). *University evaluation act*. Retrieved from: <http://edu.law.moe.gov.tw/LawContentDetails.aspx?id=FL041732&KeyWordHL=&StyleType=1>]

教育部（2013b）。教育部試辦認定大學校院自我評鑑機制及結果審查作業原則。
取自<http://edu.law.moe.gov.tw/LawContent.aspx?id=GL000677>

[Ministry of Education (2013b). *Guiding principles for self-accreditation university review*. Retrieved from <http://edu.law.moe.gov.tw/LawContent.aspx?id=GL000677>]

郭秋永（2011）。混和研究與質量爭議。**東吳政治學報**，29（1），1-64。

[Kuo, C. Y. (2011). Mixed research and the qualitative-quantitative debate. *Soochow Journal of Political Science*, 29(1), 1-64.]

陳向明（2009）。**社會科學質的研究**。臺北市：五南。

[Chen, X. (2009). *Qualitative research of social science*. Taipei, Taiwan: Wunan.]

彭森明（2013）。**高等教育校務研究的理念與應用**。臺北市：高等教育。

[Peng, S. (2013). *Institutional research in higher education: Concepts and applications*. Taipei, Taiwan: Higher Education Press.]

曾淑惠（2009）。技職校院中評鑑倫理兩難之探析。**教育研究與發展期刊**，5（3），209-239。

[Tseng, S. H. (2009). An exploration on ethical dilemma in technological and vocational school evaluation practice. *Journal of Educational Research and Development*, 5(3), 209-239.]

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。

[Yu, C. C., & Tseng, H. J. (2004). Meta-evaluation of educational evaluation. *Bulletin of Educational Resources and Research*, 29, 53-94.]

黃光國（2001）。**社會科學的理路**。臺北市：心理。

[Hwang, K. K. (2001). *Philosophy of the social sciences*. Taipei, Taiwan: Psychological.]

黃政傑（2004）。質性教育評鑑之探討。**教育資料集刊**，29，95-118。

[Huang, J. J. (2004). A study on qualitative evaluation of education. *Bulletin of Educational Resources and Research*, 29, 95-118.]

黃政傑、張嘉育（2010）。我國大學系所評鑑之問題分析與改進方向。**教育政策論壇**，13（2），43-76。

[Huang, J. J., & Chang, C. Y. (2010). The problems and prospects of university-department evaluation in Taiwan. *Educational Policy Forum*, 13(2), 43-76.]

楊國賜（2010）。談評鑑委員的角色與晤談技巧。**教育評鑑雙月刊**，25，13-15。

[Yang, K. S. (2010). Role and interview tips for reviewer of higher education accreditation. *Evaluation Bimonthly*, 25, 13-15.]

楊瑩、楊國賜、劉秀曦、黃家凱（2014）。100 年度大專校院校務評鑑後設評鑑研究之分析。**高教評鑑與發展**，8（1），1-40。

[Chan, Y., Yang, K. S., Liu, H. H., & Huang, C. K. (2014). An analysis of the meta-evaluation studies of university institutional evaluation in 2011 in Taiwan. *Higher Education Evaluation and Development*, 8(1), 1-40.]

監察院（2010）。糾正案文。取自 http://www.cy.gov.tw/sp.asp?xdURL=.%2Fdi%2FMessage%2Fmessage_1.asp&ctNode=903&msg_id=3113。

[The Control Yuan of the Republic of China (2010). *Correct text on evaluation of higher education*. Retrieved from http://www.cy.gov.tw/sp.asp?xdURL=.%2Fdi%2FMessage%2Fmessage_1.asp&ctNode=903&msg_id=3113]

劉維琪（2010）。大學分類評鑑是限制不是鬆綁。**教育評鑑雙月刊**，27，4-5。

[Liu, W. C. (2010). Conducting evaluating by university classification is a restriction to universities. *Evaluation Bimonthly*, 27, 4-5.]

劉維琪（2011）。系所對評鑑委員的期待。**教育評鑑雙月刊**，29，6-7。

[Liu, W. C. (2011). Faculty's expectations for program reviewers. *Evaluation Bimonthly*, 29, 6-7.]

潘慧玲（2003）。社會科學研究典範的流變。**教育研究資訊**，11（1），115-143。

[Pan, H. L. (2003). The evolution of social research paradigms. *Educational Research & Information*, 11(1), 115-143.]

鄭淑惠（2010）。建構優質化學校：評鑑能力之探析。**教育研究月刊**，192，55-66。

[Cheng, S. H. (2010). Constructing high quality schools: Studies on evaluation capabilities. *Journal of Education Research*, 192, 55-66.]

謝志偉（2007）。教育研究典範的未來趨勢：混合方法論（mixed methodology）介紹。**屏東教育大學學報**，26，175-194。

[Hsieh, C. W. (2007). Educational research paradigm of future: The introduction of mixed methodology. *Journal of National Pingtung University of Education*, 26, 175-194.]

顏善邦（2010）。談外部專家的角色與晤談技巧。**教育評鑑雙月刊**，25，9-12。

[Yien, S. P. (2010). The role and interview skills for external reviewers. *Evaluation Bimonthly*, 25, 9-12.]

蘇錦麗（1997）。**高等教育評鑑—理論與實務**。臺北市：五南。

[Su, J. L. (1997). *Evaluation system in higher education: Theory and practice*. Taipei, Taiwan: Wunan.]

Abma, T. A. (2006). The social relations of evaluation. In I. Shaw, J. C. Greene & M. M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 184-199). London, England: Sage.

American Evaluation Association (2004). *American Evaluation Association guiding principles for evaluators*. Retrieved from <http://www.eval.org/p/cm/ld/fid=51>

- Canadian Evaluation Society (2010). *Competencies for Canadian evaluation practice*. Retrieved from https://evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice.pdf
- Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. *New Directions for Evaluation*, 141, 7-23.
- Creswell, J. W., & Clark, V. L. P. (2007). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124-130.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum.
- Denzin, N. K. (1978). *Sociological methods*. New York, NY: McGraw Hill.
- Denzin, N. K., & Lincoln, Y. S. (Eds.) (2011). *The Sage handbook of qualitative research* (4th ed.). London, England: Sage.
- Greene, J. C. (2005). Evaluators as stewards of the public good. In S. Hood, R. K. Hopson & H. T. Frierson (Eds.), *The role of culture and cultural context: A mandate for inclusion, truth, and understanding in evaluation theory and practice* (pp. 7-20). Greenwich, England: Information Age Publishing.
- Greene, J. C. (2011). The construct of validity as argument. *New Directions for Evaluation*, 130, 81-91.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oak, CA: Sage.
- House, E. R. (2011). Conflict of interest and Campbellian validity. *New Directions for Evaluation*, 130, 69-80.
- Johnson, R. B., & Onwueabuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Kane, M. T. (2004). Certification testing as an illustration of argument-based validation. *Measurement*, 2(3), 135-170.
- Kane, M. T. (2006). Validation. *Educational Measurement*, 4(2), 17-64.
- Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Los Angeles, CA: Sage.
- Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *Sociological Quarterly*, 34(3), 673-693.
- Lather, P. (2001). Validity as an incitement to discourse: Qualitative research and the crisis of legitimation. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.) (pp. 241-250). Washington, DC: American Education Research Association.
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10(1), 35-57.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62(3), 279-301.

- Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, 34(5 Part 2), 1189-1208.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443-459.
- Schwandt, T. A. (2002). *Evaluation practice reconsidered*. New York, NY: Peter Lang.
- Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report*, 2(5), 36-38.
- Silverman, D. (2013). *Doing qualitative research: A practical guide* (4th ed.). London, England: Sage.
- Smith, J. K. (1992). Interpretive inquiry: A practical and moral activity. *Theory into Practice*, 31(2), 100-106.
- Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Establishing essential competencies for program evaluators. *American Journal of Evaluation*, 26(1), 43-59.
- Stufflebeam, D. L. (2001). The meta-evaluation imperative. *American Journal of Evaluation*, 22(2), 183-209.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Los Angeles, CA: Sage.
- Thomas, V. G. (2004). Building a contextually responsive evaluation framework: Lessons from working with urban school interventions. *New Directions for Evaluation*, 101, 3-23.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.