

以有效的觀察提供回饋－論教學觀察的信度與效度及高信效度觀察系統的建構

張德銳

臺北市立大學與輔仁大學退休教授

一、前言

國內《十二年國民基本教育課程綱要總綱》規定：「為持續提升教學品質與學生學習成效，形塑同儕共學的教學文化，校長及每位教師每學年應在學校或社群整體規劃下，至少公開授課一次，並進行專業回饋。」由此可見，課室的教學觀察與回饋業已成為中小學教師教學現場的常態，如何善用教學觀察這個工具，提供給授課教師有效的回饋，進而有益於中小學的「教」與「學」，是國內教育界所關心的議題。

然而教學觀察與回饋的實施有其諸多困境，諸如保守封閉的教學文化（林彩岫，2018；Lortie, 1975）、被觀察會引起教師的緊張和焦慮（Byrne, Brown, & Challen, 2010）、觀議課時間不足且增加教師負擔（巫南萱，2023；賴光真，2022），除此之外，觀察結果的信效度問題也會影響教學觀察與回饋的成效，這是一方面由於教學觀察的培訓不足或者觀察工具本身的問題，另方面由於觀察的霍桑效應（Hawthorne effect）或者觀察偏誤（observation bias），常使得教學觀察之信度與效度不高，導致不能對教師表現資料有效地蒐集、分類、詮釋、回饋給授課教師，並與授課教師共同協商討論教學改善之道（Archer et al., 2016; O’Leary, 2014）。

過去我國推動中小學教師專業發展評鑑，對於觀察工具雖然已有相當成熟之發展，但是這些工具並未進行信度與效度上的研究，在觀察員培訓上亦未強調觀察信效度的訓練，亟待在既有的教學觀察工具研發成果上，發展出一個具有高信效度的教學觀察系統，作為中小學教學觀察與回饋之用，以增進教學觀察與回饋的成效。另在國內學術界上亦罕見教學觀察信效度的論述。有鑑於此，本文茲先說明教學觀察的信度與效度概念，其次論述如何建構一個有信效度的觀察系統，而文中所述如何建構的建議主要參考Archer等人（2016）的專著《更好的回饋以獲得更好的教學：改善課堂觀察的實用指南》（Better feedback for better teaching: A practical guide to improving classroom observations）。該書的研究與出版係由比爾及梅琳達蓋茲基金會（Bill & Melinda Gates Foundation）所贊助的，其諸多觀點頗具學術與實務應用價值。

二、教學觀察的信度與效度

不管是在進行教育研究或者教學觀察時，信度（reliability）和效度（validity）

都是兩個非常重要的概念（郭生玉，2012；McMillan, 2000; O’Leary, 2014）。信度指的是教學觀察結果的一致性和穩定性。效度則指的是教學觀察工具和程序能準確測量到教學表現的程度。

信度是效度的先決條件，沒有信度必然沒有效度，是故Pianta與Hamre（2016）指出，一個有用的教學觀察工具首先必須具有高度的跨觀察者信度（inter-rater reliability）或者測試一重測信度（test-retest reliability）。跨觀察者信度指的是：在教室觀察的情境下，在相似的條件下由組內不同的觀察者進行觀察時，觀察工具和流程應該要產生一致的結果。測試一重測信度指的是：當同一觀察者在不同時間點使用同樣工具，觀察同一教室，觀察結果應該是相對一致的。

Joe等人（2013）主張，除了信度之外，一個有效的教學觀察工具更必須具備高度的效度，亦即在教室觀察的情境下，該工具應能準確評估其設計用於評估的教師的教、學生的學、以及師生互動的內容。Joe等人指出，教學觀察效度的類型有以下五種：

- （一）內容基礎效度（content-based validity）：確認觀察工具的證據表明其測量的是其設計用來測量的內容，並且代表該內容領域，例如教學表現的重要內容。
- （二）結構效度（structural validity）：評估觀察工具或試驗結構的程度，以確保其在理論上或概念上的結構是否符合預期。這包括確保觀察工具中各個維度或項目之間的關係是否與理論或概念模型相一致。
- （三）聚斂效度（convergent validity）：證實觀察工具與課堂教學實踐的其他衡量方式（例如學生對教師教學反應意見調查）得到相同的結果。
- （四）效標關聯效度（criterion-related validity）：檢查觀察工具如何預測或與其他相關效標（例如學生學習成績）相關聯的程度。
- （五）可推論力（generalizability）：確認或支持觀察結果的可推論力係基於觀察維度、場合、課程的樣本皆具有代表性。抽樣可能導致錯誤，而這種錯誤可能削弱對觀察評分等第解釋的效度論據。

以上五種觀察效度雖然皆有其理論依據與實務應用價值，但是Pianta與Hamre（2016）認為其中尤以效標關聯效度更具重要性，因為如果所觀察到的具體事實或證據能有效預測學生學習成就，則不但可以說服教師接受回饋而改善教學，而且可以使行政人員更有信心將觀察結果用於人事決定。畢竟在教育領域，學生學習是最重要的事，如果無法提升學生學習，則測量與詮釋教師教學表現將

失去意義。

三、一個有信效度觀察系統的建構

Roberson（1998）、Joe等人（2013）、Pianta與Hamre（2016）皆認為要確保信度，觀察者應接受充分的培訓，以確定他們理解觀察標準並在使用工具時保持一致性。若要確保效度，則所發展的觀察工具應與評鑑計畫的目標和內涵一致，而且將觀察工具提交專家評審，有助於確保它準確測量所期望的概念。另外，在廣泛使用觀察工具之前，應對觀察工具進行試點測試，以預先發現並解決問題。

然而要發展一個具有高信度和高效度的觀察系統是一件龐大的工程。Archer等人（2016）指出，一個值得信賴的觀察系統的組成部分主要有四：

（一）觀察量規

量規（rubrics）是一種評估工具，通常以表格、清單或評分標準的形式呈現，可用於學習評量、課程評鑑、教師評鑑等各種不同的評估場景，幫助確保評估的公平性、一致性和有效性，並促進學生或參與者的學習和發展。

Archer等人（2016）認為，一個良好的教學觀察規準必然包含有一套精心設計的「量規」（或可名之為評分表），藉以描述教學重要方面的表現之可以觀察的標的以及做為評定教學表現水準的依據。總體而言，它代表了教育社群對於有效教學強有力的聲明，它不但指引教師教學的方向，而且也使不同的觀察者能就相同的教學表現達成相同的判斷。此外，在專業回饋時，量規還提供了與授課教師進行有意義的改進的討論之共同語言。

精準設計一套合理可行的觀察量規，係發展一個高信度與高效度觀察系統的首要工作。在這一方面的活動主要有五（Archer et al., 2016）：(1)透過研究，建立一套可普遍理解的有效教學的指標及評量基準；(2)透過此一觀察量規將觀察內容限定於課程中可以預期的行為以及觀察者可以合理追蹤的行為；(3)利用簡明易懂的語言和結構來促進對觀察內涵及評量基準的理解；(4)詢問教師和觀察員對觀察量規的合理性、一致性以及對教學改進的支持程度；(5)透過研究證據，證明觀察結果對促進學生學習的程度來評估觀察工具的效標關聯效度。

據此，為建構一個有高信效度的觀察系統，本文建議如下：

1. 修正教師專業標準

我國教育部（2016）曾在2016年2月15日以臺教師(三)字第1050018281號函公布《中華民國教師專業標準指引》，該指引係依循《中華民國師資培育白皮書》揭櫫我國理想教師圖像，據以研發10個教師專業標準及29項教師專業表現指標，包含教育專業、學科教學、教學設計、教學實施、學習評量、班級經營、學生輔導、專業成長、專業責任及協作領導等面向。唯經過時間的推移，國內的教育環境已有所變化，特別是十二年國教新課綱的推行以及人工智慧在教育應用等，該指引實有必要加以修正，才能與時俱進。

2. 修訂教師評鑑規準

教師專業標準決定了教師評鑑的規準，而這些規準可以包含層面、指標及檢核重點。我國教育部為了推動教師專業發展評鑑，依據《中華民國教師專業標準指引》，於民國105年4月25日以臺教師（三）字第1050040254號函發布「高級中等以下學校教師專業發展評鑑規準（105年版）」，其主要內涵共有3個層面（課程設計與教學、班級經營與輔導、專業精進與責任）、10項指標、28個檢核重點。教師專業發展評鑑雖然已經於106學年度轉型為「教師專業發展實踐方案」，但是這個規準還是在沿用中。然隨著教育環境的改變，新的課程發展與教學策略層出不窮，該規準實有必要加以修訂，才能引領基層教師往更理想的方向精進，並作為協助教師專業發展的依據。

另外，我國的教師成績考核是目前唯一有法律依據的教師評鑑機制。現行《公立高級中等以下學校教師成績考核辦法》係2021年07月28日所修正公布的，其法源基礎係《高級中等教育法》第三十三條及《國民教育法》第十八條第二項之規定訂定之。然教師成績考核的規準亦未依據《中華民國教師專業標準指引》加以修訂，宜儘速配合調整之。

3. 修正教學觀察的量規

在教學量規的設計上，「高級中等以下學校教師專業發展評鑑規準（105年版）」，在每個檢核重點皆設有「檢證來源」與「評定等級」之說明，可做為教學觀察時可觀察的標的以及做為評定教學表現水準的依據。唯該等量規還是有三個問題亟需解決：其一，該量規缺乏專家學者與實務工作者的深入檢視，難以確定其內容效度；其二，其評量等級僅分為「推薦」、「通過」、「待改進」等三個等級，不若國外知名教師評鑑學者Danielson（2011）在《教學評鑑工具架構》（The framework for teaching evaluation instrument）將評量等級分為「傑出」（distinguished）、「精通」（proficient）、「基本」（basic）、「不滿意」（unsatisfactory）

等四個等級之精細。

4. 確定教學觀察的效標關聯效度

國內專家學者應透過一個大型的實徵研究，探討教學觀察結果對促進學生學習成效的程度來評估觀察工具的效標關聯效度。有了高度的效標關聯效度，教學觀察的結果當有高度的說服力與應用性。

(二) 觀察員培訓

Roberson（1998）指出，觀察員培訓（observer training）是教學觀察系統有效運作的一個最重要環節。觀察員培訓不能只是把一個評量工具交給觀察員，然後說「去觀察」就了事。如果這樣做，每個觀察員對觀察證據的蒐集與紀錄以及如何加以詮釋就會有多種多樣，勢將嚴重影響觀察的信度與效度。

Archer等人（2016）指出，為有效進行觀察員培訓，負責培訓的人員宜：(1) 事先蒐集適宜的教室教學影片，然後使用觀察量規來確定影片中的授課教師在每個觀察指標的基準評級和評級理由；(2) 向觀察員清楚說明觀察量規的基礎、結構、主要特徵和術語；(3) 讓觀察員意識到觀察偏誤以及如何減少這些偏誤對評量等級可能造成的影響；(4) 透過明確的指導、示範和不斷的練習來培養觀察員對觀察工具的熟練程度；(5) 說明及示範如何根據觀察結果向教師提供富有成效的回饋技巧。

據此，我國在發展出教學量規和觀察工具後，應進行嚴謹的觀察員培訓，其工作內涵如下：

1. 蒐集大量具有代表性的教室教學影片

在符合知情同意等倫理原則下，由中小學教師提供，或者在中小學大量錄製各領域、各年級的教室教學影片，以做未來觀察員練習觀察的範本。

2. 確定影片中授課教師在每個檢核重點的基準評級和評級理由

延請專家學者經由多次的實作和討論，來確定影片中授課教師在每個檢核重點的基準評級和評級理由，以作為觀察員是否通過培訓的參考標準。

3. 向觀察員清楚說明教學觀察的規準和量規

以簡明易懂的語言和結構來促進觀察員對觀察內涵及評量基準的基礎、結構、主要特徵和術語的理解。

4. 透過明確的指導、示範和不斷的練習來培養觀察員對觀察工具的熟練程度

在充分的講解與示範後，宜提供觀察員以影片作為實作練習的機會。透過不斷的練習、討論、校正，來培養觀察員對觀察工具已經能夠充分理解與掌握，並且運用自如。

5. 說明及示範如何根據觀察結果向教師提供富有成效的回饋技巧

回饋可以被視為觀察過程中最重要的階段，因為唯有透過回饋才能對授課教師的未來發展產生有利的影響。是故，必須教導觀察員在與教學者對話中，如何善用信任、融洽、引導式提問、不做價值判斷、提供資料與資源等會談技巧（丁一顧、張德銳，2009），來進行如何改善教學的討論。

（三）觀察員評估

教導了並不代表學習了；同樣的，學習了並不代表就學會了。是故觀察員評估（observer assessment）是了解與改善受培訓人員是否掌握了正確知識和技能的唯一方式。在此一方面的活動主要有三（Archer et al., 2016）：(1)創建一組衡量必要教學觀察知識和技能的績效活動；(2)設定學會教學觀察知識和技能的最低績效標準；(3)澄清當觀察員沒有表現出熟練程度時會發生什麼，包括他們將獲得哪些額外支持或處置。

據此，為有效評估觀察員的品質，有必要進行下列三個活動：

1. 設定測試觀察員觀察知能的活動

建議可以以多元的方式來測試觀察員的觀察知能，例如以紙筆測驗測試觀察員對觀察規準與量規的瞭解程度，以觀察法了解觀察員的觀察態度和投入程度，以觀察實作結果判定觀察員是否達到觀察標準。

2. 設定教學觀察知識和技能的最低績效標準

觀察員在多次練習後，組內不同的觀察員對同一教學情境進行觀察時，觀察

工具和流程應該要產生一致的結果，來確保觀察結果具有高度的跨觀察者信度。另外，更重要的是，觀察結果必須落在與既定標準可接受的誤差範圍內，才能通過培訓。

3. 提供觀察員再培訓的機會

對於無法通過測試的觀察員提供補救教學與再一次測試的機會，如仍未能通過既定標準，則提供下一期再培訓的機會。

(四) 觀察監控

Archer等人（2016）指出，觀察監控（monitoring of observations）是另一個重要的品質控制措施。觀察系統發展者需要知道實際在觀察現場發生了什麼，例如觀察員是否遵循程序，還是匆忙而草率地進行觀察？了解存在的問題意味著可以採取措施解決問題以及改善整個觀察系統。在此一方面的活動主要有三：（1）檢查觀察政策和程序是否被嚴謹地遵守；（2）確保觀察者隨著時間的推移仍保持一定的準確性；（3）確保教師能夠在回饋會談，獲得清晰、具體且可行的回饋。

據此，國內教學觀察系統的發展者可以進行下列活動：

1. 實地考察觀察員的運作情形

經由實地考察獲得第一手資料，一方面確定觀察政策和程序是否被嚴謹地遵守，另一方面發現觀察實作的缺失，以作為進一步改進觀察系統的依據。

2. 提供觀察員回流教育的機會

隨著時間推移，觀察員會有不知不覺地傾向於過度高估或低估觀察評分的現象，為解決此一問題必須每隔一段時間，要求觀察員接受回流教育，以減少觀察員的觀察偏誤。

3. 調查教師是否經由回饋獲得專業成長

教學觀察與回饋的主要目的係在提供教師教學現況的客觀回饋，然後促進其專業成長。是故有必要經由訪談與調查研究了解教師接受回饋的情形以及是否因為有效的回饋而能永續的成長。

四、結語

誠如O'Leary（2014）所言，教學觀察與回饋在歐美先進國家已經相當地普及，可以說業已融入到教師的生涯發展之中。也就是說，教師們從其職業生涯的開始到結束，皆會經由教學觀察或者被觀察，學習如何教學，進而建構個人的教學實務智慧。

「工欲善其事，必先利其器。」我國推動教學觀察與回饋已有二、三十年之久，然多年來缺乏對教學觀察信度與效度的關注；對於有信效度觀察系統的研發亦著力不深，自然成果相當有限。爾今爾後，如何在此一領域加強學術研究以及早日建構一個有高信效度的觀察系統，應是國內關心教師專業發展的教育界人士亟待努力的課題。

參考文獻

- 丁一顧、張德銳（2009）。**認知教練－理論與實務**。臺北市：五南。
- 巫南萱（2023）。國中輔導科教師公開授課歷程經驗、關注面向及轉變研究（未出版之碩士論文）。國立清華大學，新竹市。
- 林彩岫（2018）。教學文化之研究：以兩所國民小學的教學觀察為例。**教育理論與實踐學刊**，37，53-78。
- 郭生玉（2012）。**心理與教育研究法**。臺北市：精華書局。
- 教育部（2016）。**中華民國教師專業標準指引**。臺北市：作者。
- 賴光真（2022）。公開授課實施缺失之檢討評析－兼論課綱法規及學校實施之修訂調整。**臺灣教育評論月刊**，11(7)，112-121。
- Archer, J., Cantrell, S., Holtzman, S. L., Joe, J. N., Tocci, C. M., & Wood, J. (2016). *Better feedback for better teaching: A practical guide to improving classroom observations*. San Francisco, CA: Jossey-Bass.
- Byrne, J., Brown, H., & Challen, D. (2010). Peer development as an alternative to peer observation: A tool to enhance professional development. *International Journal for Academic Development*, 15(3), 215-228.

- Danielson, C. (2011). *The framework for teaching evaluation instrument*. Princeton, NJ: The Danielson Group.
- Joe, J. N., Tocci, C. M., Holtzman, S. L., & Williams, J. C. (2013). *Foundations of observation: Considerations for developing a classroom observation system that helps districts achieve consistent and accurate Scores*. Princeton, NJ: ETS.
- Lortie, D. C. (1975). *Schoolteacher: A Sociological Study*. Chicago, IL: The University of Chicago Press.
- McMillan (2000). *Educational research: Fundamentals for the consumer* (3rd ed.). Boston, MA: Addison Wesley Longman.
- O'Leary, M. (2014). *Classroom observation: A guide to the effective observation of teaching and learning*. New York, NY: Routedge.
- Pianta, R. C., & Hamre, B. K. (2016). Implementing rigorous observation of teachers: Synchronizing theory with systems for implementation and support. In J. A. Grissom, & P. Youngs (Eds.), *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 22-36). New York, NY: Teachers College Press.
- Roberson, T. J. (1998). *Classroom observation: Issues regarding validity and reliability* [paper presentation]. Annual Meeting of the Mid-South Education Research Association, New Orleans, LA, United States.

