

講倫理才有道理： 歐洲研究倫理規範的新進展

朱容萱*、黃之棟**

摘要

學術倫理、信度、效度被認為是評估研究品質的重要指標，學界一般也多主張三者的關係乃三足鼎立的鐵三角，相互之間不可化約，重要性亦不分軒輊。本文旨在挑戰此傳統理解，並提出學術倫理、信度、效度間層疊關係的看法。我們認為，衡諸當前國際學術脈動，以人權保障為出發點的倫理規範顯然最獲歐美學界重視。對此等國家而言，遵守學術倫理要求早已不再是不甚具約束力的道德勸說，而是一種強化專業素養、鞏固研究根基與確保學術品質的必要條件。此一動向亦可從各國際學術組織紛紛開始擘劃倫理規約的潮流中看出端倪。為闡明以上思潮，本文以一實證研究為例，剖析學術倫理意識的興起對傳統信度、效度觀點的衝擊。除了文本與法理的分析，文中亦參酌歐洲重要學者的訪談資料，藉以釐清倫理規範如何與專業素養連結，以及此連結建立後隨之而來的可能誤區。在例示並解析倫理規約條文之後，本文重構了學術倫理、信度、效度間的關聯，並指出在捕捉三者關係時，我們不該視它們為鞏固學術品質的金三角，而應該看到層層相疊的金字塔關係。

關鍵詞：學術倫理、信度、效度、歐洲

* 英國愛丁堡大學教育學院博士候選人

** 國立空中大學公共行政學系助理教授



壹、緒論

2010年4月的一個上午，一場英語研究的口語測驗正在某國中裡如火如荼地進行著。考官與應考學生進行了英語問候與相關對答後，考官隨即注意到受試者緊張且不住顫抖的雙手。帶著強烈道德意識的考官，深覺不應讓學生過度煎熬；在詢問學生的意向之後，考官毅然終止了這場測驗。仍緊張難平的學生離開了考場，下一位強作鎮定卻還能面帶微笑的學生則開始了他的測驗。道德考官的一念之仁，卻種下了後續的給分爭議及對於整個研究信度、效度的質疑。此處爭論的焦點在於：施測當時學生因緊張而沒能完成後續測驗，致使考官只能依照測驗初始時的基本問候對話來評斷該生的英語能力。然而，這樣的評斷究竟適不適當，就成了問題。

為此，這位道德考官與另兩位評分老師對於該不該給分、以及如何給分的問題起了相當的爭執。一方面來說，若用學生一開始的問候對話來評斷之後的表現，似乎對其他學生不太公平；但另一方面，若因棄考而給該考生零分，似乎也有邏輯上的謬誤。具體來說，這個研究當初設計時是想測得同學的「口語表現」，既然該生因緊張而棄考了，在他還沒開始表現的情況下，評斷者又怎能給予表現極差的零分評價呢？這個邏輯上的悖論，其實也與一般評斷研究品質的兩大指標，即「信度」(reliability)、「效度」(validity) 息息相關。信度是指研究成果的可複製性。可複製性愈高，信度愈高。¹從上面的「棄考個案」來看，考試之所以要有信度，是爲了確保該考試是公平的。也就是說，考官不會因爲考生身分而有差別待遇。反觀效度，其主要功能則在檢驗考試的結果是否真的反應了原本設定的欲測。²也就是說，研究者設計的測驗方法是否真的幫他測到了想測的東西。從同樣的棄考個案來看，考生因太緊張而缺考了，所以該生的口語能力沒有完全被測得。但基於公平（也就是信度）考量並權衡現實，在沒有補考機會而評分者又非得給分的情況下，零分似乎就成了不滿意但卻不得不接受的決定了。

其實，再回頭檢視本案例即會發現，真正導致信度、效度問

¹ Alan Bryman, *Social Research Methods (3rd editon)* (Oxford: Oxford University Press, 2008), p. 31.

² Alan Bryman, *op. cit.*, p. 32.



題的原因其實來自考官的詢問。正因為考官意識到了學生的不自在與恐懼，又主動詢問了是否需要中止測驗，才出現了學生棄考的決定。倘若當時考官沒詢問，學生也就不會棄考。問題是，不棄考問題就真的解決了嗎？就信度而言，若考官刻意壓抑自己的倫理關注而不提出是否中止考試的提問，雖然表面上考試流程的一致性確實得以確保，研究的信度也得以達成。但就效度而言，學生是在極度惶恐下被強迫參與研究且完成考試的，即使該考題確實能測出學生的口語能力，但我們測得的似乎也不是平時該生的口語能力，而是該生驚慌失措時的表現罷了。由此可見，一位鐵面無私、不講情面的考官，還是會招來信度與效度的疑慮。這也是為什麼各學術單位（如英國教育研究協會（British Educational Research Association））大力推動學術倫理規範建制的的原因。

本文即是在這個脈絡下，探討學術倫理、信度、效度間的關聯。不同於一般社會科學研究法教科書將三者分開討論的提法。本文強調我們不應把此三項要素視為各據一方的學術品質三角支柱，而應該將它們視為層層相疊的品質保固金字塔。為論證此點，本文首先釐清信度、效度與倫理規範的意義。接著利用歐洲各組織的倫理規約來闡述三者間的連結。這個部份，我們也將從訪談分析裡，梳理出過度注重倫理規約可能發生的危險。最後，本文將重構學術倫理、信度、效度的關係，並提出新的觀察視角作結。

貳、信度、效度與道德倫理的交錯

信度與效度是評估學術研究質量的標準。不論是傳統的量、質化研究 (quantitative/qualitative research)，還是近來崛起的混合型研究 (mixed methods research)，基本上都將信度與效度奉為圭臬。不過，由於量化、質化與混合型研究所遵循的典範(paradigm)不同，加上它們在本體論 (ontology) 與認識論 (epistemology) 上的看法亦相左，這些研究對衡量研究品質的標準當然也不盡相同。有趣的是，儘管學科脈絡的歧異且對信度與效度有著不同的認知，不同學科 (discipline) 的學者還是傾向通過「信度」與「效度」這兩個名詞來探討研究品質。質言之，各派學者常用不同的名詞來解釋相同概念，或用同一名詞來指涉不同概念。名詞的交錯使用很可能產生雞同鴨講、或是錯把馮京當馬涼的誤解。由於這個緣故，在探討學術倫理、信度、效度的關係前，本文先就量化、質化研究中信度與效度的概念，分別做一簡單的介紹與整理。



限於篇幅，本文不擬對每一名詞做全面闡釋，合先說明。

一、信度

信度所扮演的角色，如前所述是在確保研究結果的可複製性。但對於何謂可複製性與怎樣才能確保它的實現，量化與質化研究者的看法稍有不同，宜分別觀察。首先，在量化的傳統下，可複製性至少包含三個面向：(一)、結果的穩定性 (stability)。這意味著多次反覆測驗後，所得的結果都相當接近。換言之，反覆施測後所得的多次結果之間若呈現正相關 (positive correlation)，穩定性的目標即告達成；(二)、內在信度 (internal reliability)。內在信度適用於多重指標 (multiple indicator) 的研究測驗裡。比方說，以問卷為主要資料收集工具的研究，多會利用內在信度來衡量問卷中各問題與其對應概念間的契合度。³大體來說，內在信度與幅合「效度」(convergent validity) 的概念是相同的，而這兩個概念又與區辨「效度」(discriminant validity) 是對立的。具體來說，前組概念是在測試問題與概念間的契合度。區辨效度的功能則正好相反。區辨效度主要用來區分測量不同概念的多重指標群組間的差異性。⁴也就是說，它的功能在於確認當初設計用來測試不同概念的指標群組間，「沒有」強烈的正相關性；(三)、觀察者一致性 (inter-observer consistency) 及觀察者內信度 (intra-observer consistency)。前者評估不同觀察者解釋相同現象的一致性；後者檢試同一觀察者在不同情況下對相同現象闡釋的一致性。當不同的觀察者們對於所觀察現象的解釋很相近時，代表觀察者一致性很高。同樣地，當同一位觀察者在每種情形下對同一個現象都有類似的解釋時，則通過觀察者內信度的考驗。⁵對於以上信度的三個面向，皆可靠統計相關分析 (correlation analysis) 來一一檢驗。⁶值得注意的是，基於檢驗信度的各相關分析法皆有其背後的統計原理與假設，因此在決定使用何種相關分析來檢驗信度的各面向之前，研究者宜根據研究資料的屬性來做選擇。同時，對於相關係數質的解釋，也會根據不同的學科或研究樣本數的大小而有所差異。限於本文重心不在探討統計分析法，有興趣的讀者宜參考

³ 同註 1, pp. 149-151.

⁴ David A. De Vaus, *Surveys in Social Research* (3rd Edition) (London: University College London Press, 1993), pp. 29-31.

⁵ Sari Luoma, *Assessing Speaking* (Cambridge: Cambridge University Press, 2004), p. 179.

⁶ 例如 Pearson 相關係數、Cronbach Alpha 相關係數、Cohen Kappa 相關係數等。



其他相關文獻。⁷

相較於量化研究以統計推演信度，質化研究由於不以數字來背書或使用非系統式的資料收集與分析，這使得它們的信度有無飽受量化學界的質疑。⁸為此，質化研究學者 Kirk 及 Miller 呼籲研究者把資料收集與分析方式清晰地記載下來。⁹另外，Hammersley 也重新定義了信度。Hammersley 將信度視為不同觀察者劃分觀察現象的一致性，或是相同觀察者在不同情況下劃分觀察現象的一致性。簡言之，這個脈絡下的信度所關心的是，被觀察的現象是否能被劃分到相同的類別裡。¹⁰其實這項概念和量化研究中的觀察者一致性和觀察者內信度的概念是相通的，因為他們的重點都放在研究者對相同的研究現象提出相近的闡釋，並做出相似的分類。但是，這樣的觀念在質化研究卻有另一個名字，即「內在信度」(internal reliability)。呼應此處內在信度的是質化研究中的外在信度 (external reliability)。外在信度追求的是研究結果與經驗的可轉移與分享程度。其基本意涵可對應量化研究的可複製性與概化目標。¹¹縱然質化學者一再強調可複製性的觀察結果並非他們的研究重心，但這並不代表質化研究者不致力達成外在信度的標準。¹²例如，LeCompte 及 Preissle 就針對質化研究的複製性議題提出建言。他們認為若想複製質化研究結果，研究者的身分、研究參與者的選擇、社會情境、資料收集及分析法等都得相同才行。¹³

從量化與質化脈絡下的信度分類，我們很容易就發現名詞使用不一致的問題。同樣一個「內在信度」，在不同脈絡下卻有截然

⁷ 請參考 Jane Fielding and Nigel Gilbert, *Understanding Social Statistics* (2nd Edition) (London: Sage, 2006) 及 Dennis Howitt and Duncan Cramer, *An Introduction to Statistics in Psychology* (5th Edition) (Harlow: Prentice Hall, 2011).

⁸ David Silverman, *Doing Qualitative Research: A Practical Handbook* (London: Sage, 2000), pp. 8-10.

⁹ Jerome Kirk and Marc L. Miller, *Reliability and Validity in Qualitative Research* (London: Sage, 1986), p. 72.

¹⁰ Martyn Hammersley, *What's Wrong with Ethnography?* (London: Routledge, 1992), p. 67.

¹¹ 同註 1, p. 376.

¹² Louis Cohen, Lawrence Manion and Keith Morrison, *Research Methods in Education* (5th Edition) (London: RoutledgeFalmer, 2000), p. 119.

¹³ Margaret D. LeCompte and Judith Preissle, *Ethnography and Qualitative Design in Educational Research* (2nd Edition) (London: Academic Press Ltd, 1993), p. 334.



不同的意義。在量化的脈絡下，此內在信度意謂著多項子指標相互間的連結性；但在質化的脈絡下，那裡的內在信度指的卻是近似於量化研究裡的觀察者一致性及觀察者內信度。若再把量化內在信度的概念與幅合效度等一起來看，我們立刻就會發現問題的複雜。因為在信度的分析裡，竟然還交雜著「效度」的名詞。像這樣名詞交錯使用的狀況，在下面效度的分類中也看得到。

二、效度

效度是用來衡量研究方法是否合於研究目的基準，也是評估研究者的實作是否確實體現了研究概念的量尺。白話地說，效度關心的是研究者是否真的測到了他想測的東西。¹⁴我們同樣可以從量化與質化等兩個面相來觀察。在量化研究裡，至少存在著四種效度的分類。第一種效度著重衡量研究資料收集法的適切度，一般稱為測量效度 (measurement validity)。這個面向觸及研究中所有研究測量的相關活動。第二種效度為內在效度 (internal validity)。內在效度的關注點在於如何證明研究結果與原因之間的關聯。也就是說，研究者對自己推敲的前因後果有幾分的信心。第三種效度為外在效度 (external validity)，用以衡量研究的概化程度。第四種效度則在衡量研究本身跟現處社會之間的距離，即所謂的生態效度 (ecological validity)。生態效度旨在評估研究計畫對現實生活的貢獻。這點之所以重要，是因為社會科學的研究成果如果只能在特定的技術面上成立，卻無法適用在現實生活裡，那麼這種研究就成了社會科學家不具實用價值的理論奇想或藝術收藏品了。¹⁵

在上述四大效度中，測量效度在文獻中受到的矚目最多。¹⁶學界因此從中延伸出了眾多子概念。一般來說，測量效度涵括了表面效度 (face validity)、同時效度 (concurrent validity)、預測效度 (predictive validity)、構念效度 (construct validity)，以及前節在談信度時提到的幅合效度和區辨效度等。¹⁷表面效度，顧名思義是指該測量必須「看起來」就像是在測量所要檢測的概念。例如，英

¹⁴ Robert Adcock and David Collier, "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research," *American Political Science Review*, Vol. 95, No. 3, 2001, pp. 529-546.

¹⁵ 同註 1, p. 33.

¹⁶ 同註 12.

¹⁷ Dennis Howitt and Duncan Cramer, *Introduction to Research Methods in Psychology* (London: Pearson, 2011), p. 273.



語測驗就該看起來像是在測驗參加者的英語能力，因此若測驗試題參雜了其他外語的使用，就會惹人非議。同時效度是指以現有的指標來測量結果的效度。比方說，我們把工作請假的頻率當成測定人們對工作滿意度的指標。若測量後發現，常請假跟不常請假的人對工作的滿意程度相同，此時我們就該檢討是否該用請假頻率來評斷工作滿意度。換言之，這個研究沒有達到同時效度。和同時效度具有相同本質的是預測效度，只是預測效度參雜了未來時間的因子。預測效度期待某項測驗的結果能夠預測未來另一項測驗的結果。例如，利用英國雅思考試 (International English Language Testing System) 的結果來預測日後學生就讀英國研究所時的學業表現。最後是身處最核心的構念效度。達成構念效度有兩個要件，第一要件為研究學理依據。也就是研究目的與研究問題的形成應當有理論的支持。第二要件為測量工具的學理依據。研究者必須說明採用某項測驗或資料收集法跟其研究目的的連結。¹⁸比方說，在測驗「口語能力」的考試裡，測試者若選用了紙筆測驗來評斷口語能力，這就很可能招致對構念效度的批評。

另一方面，質化研究者用效度來衡量研究者對其研究現象闡釋的真實性。¹⁹質化學界對於如何解釋及如何展現研究現象真實面的問題已爭論了半世紀之久。²⁰長久以來，一直都有學者提出建言。²¹其中，又以 Lincoln 及 Guba 的指標最為人所熟知。Lincoln 及 Guba 呼籲跳脫量化效度的框架，改用可信度 (trustworthy) 來特指質化研究的效度。可信度包含了四個面向：可信性 (credibility)、可轉換性 (transferability)、可靠性 (dependability) 及驗證性 (confirmability)。²²縱然 Lincoln 及 Guba 想提倡用不同的角度來看待質化研究的效度問題，但若把 Lincoln 及 Guba 的指標與上述量化的指標相比，則依舊可發現相似之處。像是可信性與可轉換性就能在前段量化效度的敘述中找到類似的概念（對應的情

¹⁸ 同註 1, p. 152.

¹⁹ Jeasik Cho and Aleen Trent, "Validity in Qualitative Research Revisited," *Qualitative Research*, Vol. 6, No. 3, 2006, pp. 319-340.

²⁰ Paul Atkinson, Amanda Coffey and Sara Delamont, *Key Themes in Qualitative Research: Continues and Changes* (New York: Altamira Press, 2003).

²¹ 如 Joseph A. Maxwell, "Understanding and Validity in Qualitative Research," *Harvard University Review*, Vol. 62, No. 3, 1992, pp. 279-300; Anthony J. Onwuegbuzie and Nancy L. Leech, "Validity and Qualitative Research: An Oxymoron?" *Quality and Quantity*, Vol. 41, 2007, pp. 233-249.

²² Yvonna S. Lincoln and Egon G. Guba, *Naturalistic Inquiry* (California: Sage, 1985), p. 219.



況可參照下表 1)。又，第三面向用來建議質化研究者清楚說明其資料收集與分析過程的可靠性，本項則可對應量化研究中的整體信度概念。最後，唯一較為不同的當屬驗證性指標了。由於傳承社會建構論 (constructivism) 的觀點，質化研究擁護者深信量化研究強調的客觀 (objectivity) 及價值中立 (value-free) 原則在社會科學研究中是不可能實現的，故強調以驗證性來強化質化研究的價值附帶 (value-bound) 論點。

除此分類之外，質化研究中亦有內在效度 (internal validity) 與外在效度 (external validity) 的分法。前者與量化研究中的構念效度類似，探討研究設計跟理論間的契合度。簡單來說，研究者所欲觀察的現象必須吻合研究目的，並解答他的研究問題；至於後者，由於受到量化研究的影響，質化研究的外在效度通常可以對應到量化研究「概化」的概念上去。對此，質化研究的外在效度常引發討論甚至引起質疑。其原因在於質化研究都以個案研究為主，因此所使用的樣本數也比較小，如何從個案窺見問題全貌當然容易延伸出問題。²³另一方面，質化研究強調從研究對象的角度，來了解其所研究的現象。因此質化研究常自我設定其目的在回答「什麼」(what is)、「如何」(how)、跟「為什麼」(why) 的問題上。²⁴也就是說，質化研究者在意的是能否深度了解其所觀察的現象(或人)，而不在概化其研究結果。

由上面的分析，我們發現雖然所有研究者都講信度與效度，但由於量化與質化研究依循的本體論與認識論不同，連帶使得他們看待現象的角度與研究手法也產生相當的歧異。²⁵從兩造的爭辯中除了可以看出信度、效度對學術研究的重要性外，雙方的爭辯也說明了我們對信度、效度的看法與定義上，仍有商榷和深拓的空間。由於討論質化與量化的擅長與缺失並非本文的重心，故我們不再對此多加說明。有興趣的讀者可參考相關文獻。²⁶以下茲列出本文

²³ 同註 1, p. 376.

²⁴ Liz Spencer, Jane Ritchie, Jane Lewis and Lucy Dillon, "Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence," *Government Social Research Service*, 2003, <<http://www.civilservice.gov.uk/my-civil-service/networks/professional/gsr/index.aspx>>.

²⁵ Robert B. Johnson and Anthony J. Onwuegbuzie, "Mixed Methods Research: A Research Paradigm Whose Time Has Come," *Educational Researcher*, Vol. 33, No. 7, 2004, pp. 14-26.

²⁶ 請參考 John W. Creswell and Dana L. Miller, "Determining Validity in



提及的信度與效度對照關係圖（請參照表 1）。

表 1：量化與質化研究的信度與效度對應關係

信度		效度	
量化	質化	量化	質化
{可複製性}	{外在信度}	{外在效度}	{外在效度} {可轉換性}
[觀察者一致性、觀察者內信度]	[內在信度]		
		(構念效度)	(內在效度) (可信性)
<內在信度>		<幅合效度>	
穩定性		面效度 同時效度 預測效度 生態效度 內在效度	可靠性 驗證性

資料來源：

1. 筆者參考 Bryman (2008)及 Lincoln and Guba (1985)²⁷後自行整理。
2. 表中附有相同括號的名詞代表其概念相似。

三、學術倫理

現今各學術領域對研究倫理的重視，可追溯到二次世界大戰時代。由於二戰期間出現了諸多枉顧人權的醫學實驗，因此戰後湧現了檢討醫學倫理的聲浪。1947 年的紐倫堡公約(the Nuremberg Code)、1964 年的赫爾辛基宣言(Declaration of Helsinki)²⁸及 1968 年在英國成立的基層研究倫理學委員會 (Local Research Ethics Committees, LREC) 都是對這段歷史醒思的成果。自此，道德倫

Qualitative Inquiry,” *Theory Into Practice*, Vol. 39, No. 3, 2000, pp. 124-130. 以及 Robin Whitemore, Susan K. Chase and Carol L. Mandle, “Pearls, Pith, and Provocation: Validity in Qualitative Research,” *Qualitative Health Research*, Vol. 11, No. 4, 2001, pp. 522-537.

²⁷ 請參考註 1 及註 22。

²⁸ 關於這些宣言的簡要介紹可參考蔡甫昌、林芝宇與張至寧，〈研究倫理的歷史、原則與準則〉，《台灣醫學》，第 12 卷第 1 期，2008，頁 107-122。



理正式成爲學術研究的討論主題，現在更是各學術單位資助研究計畫的首要考量之一。²⁹

以當代歐洲學界爲例，歐盟執委會 (European Commission) 主管研究與教育的執委 Janez Potocnik 表示，學術倫理是現今任何研究中的首要元素。換句話說，不合乎倫理的研究就是不合格的學術研究。這種研究當然也不該被出版：

歐盟資助的學術研究一定要將學術倫理放在最高順位。倫理是學術研究中不可或缺的部份，從最初的構思一直到最後的發表過程皆是如此。學術倫理的省思應充斥在研究的每一個點上。只有確切地進行學術倫理把關，才有可能獲得優秀的研究成果。³⁰

總的來看，各學術領域對道德倫理的看法雖有不同，但都源自於前述的醫學研究倫理規範。在醫學相關的領域裡，倫理規範的訂定主要依據三大原則：(一)、參與者的權利考量原則；(二)、道德倫理考量原則；(三)、參與者參與研究的後果考量原則。簡言之，研究者必須尊重研究參與者，並確保參與者在參與研究過程中不會受到傷害。同時，研究者必須尊重參與者的參與意願 (自主 (autonomy))，且確保在研究過程中每位參與者都能得到相同的待遇 (正義 (justice))。再者，該研究本身必須是有助益 (行善 (beneficence))，且不會造成傷害 (不傷害 (non-maleficence))。³¹

然而，在把這套醫學倫理守則運用在社會科學範疇時，經常會出現不適用的情況。³²原因在於社會科學研究的道德倫理規範有其情境特殊性 (context specific)，³³這使得嚴密的醫學倫理規範會

²⁹ Rose Wiles, Sue Heath, Graham Crow and Vikki Charles, "Informed Consent in Social Research: A Literature Review," *Research Methods: An ESRC Research Programme*, 2005,

<<http://eprints.ncrm.ac.uk/85/1/MethodsReviewPaperNCRM-001.pdf>>.

³⁰ 轉引自 Eleonore Pauwels, "Ethics for Researchers: Facilitating Research Excellence in FP7," *European Commission*, 2007,

<<ftp://ftp.cordis.europa.eu/pub/fp7/docs/ethics-for-researchers.pdf>>.

³¹ Priscilla Alderson, "Ethics," in Sandy Fraser, Vicky Lewis, Sharon Ding, Mary Kellett and Chris Robinson (eds.), *Doing Research with Children and Young People* (London: Sage, 2004), pp. 97-112.

³² 戴華、甘偵蓉、鄭育萍，〈人文社會科學與研究倫理審查：執行研究倫理治理架構計畫的考察與反思〉，《人文社會科學簡訊》，第 12 卷第 1 期，2010，頁 10-18。

³³ Maurice Punch, "Politics and Ethics in Qualitative Research," in Norman K.



在社會科學研究操作上出現困難。³⁴ 因此，社會科學界也開始發展出專屬的倫理規範。如，英國社會學學會 (the British Sociological Association, BSA)、社會研究學會 (the Social Research Association, SRA)、以及美國社會學會 (American Sociological Association, ASA) 等，都有各自的學術倫理主張。此外，不同領域的社會科學學門也逐步擴展出自己獨特的學術倫理規範。比方說，英國心理學會 (British Psychological Society, BPS)、英國教育研究協會 (British Educational Research Association, BERA) 等。其中，歐洲語言測驗評量協會 (European Association for Language Testing and Assessment, EALTA) 也趕上了這股重視學術倫理的風潮。該會於 2011 年 5 月假義大利舉行的年度會議³⁵，議程中更以學術倫理為主軸邀請三位講者 (Alan Davies (本研究之受訪者)、Elana Shohamy 及 Wim van der Linden) 針對測驗倫理發表演說。

從各學科紛紛量身訂做其倫理規範的趨勢中，我們不難發現社會科學研究倫理的進展已愈趨成熟。學術倫理所涵蓋的範圍以及影響，也不再局限於人權維護這一傳統的項下。這些特徵還反應在本文作者與歐洲語言測驗評量協會顧問 Alan Davies 教授所進行的面對面的訪談³⁶裡。在本文作者與 Alan Davies 教授進行的訪談中，Davies 教授特別提到學術倫理規範的制定與公布本身就是一種專業的表現：

國際語言測驗機構就像是語言測驗學界的發言人。這些機構把語言測驗的用處收集起來，好向其他的機構或政府團體發聲。所以，我認為語言機構最主要的功能，就是要表現得像一個專業的機

Denzin and Yvonna S. Lincoln (eds.) *Handbook of Qualitative Research* (California: Sage, 1994), pp. 83-95.

³⁴ 有我國學者認為，就研究傷害而言，一般醫學研究行為產生的傷害多發生在生理層面，而社會與行為科學的傷害主要出現在心理與社會層面。由於這樣的特質，雖然後者產生的傷害也可能是「既難預測且不易平復」的，但由於其顯著性較低，研究者很容易就會忽略。請參考林正介、陳祖裕，〈社會及行為科學研究相關之倫理議題〉，《人文社會科學簡訊》，第 12 卷第 1 期，2010，頁 19-25。

³⁵ 會議資訊請參考：“Ethics in Language Testing and Assessment,” paper presented at the European Association for Language Testing and Assessment Annual Conference, 2011, <<http://ealta2011.siena.cils.unistrasi.it/articolo.asp?sez0=130andsez1=0andsez2=0andart=209>>.

³⁶ 朱容萱，當面訪談，Alan Davies，英國愛丁堡大學(英國)，2010 年 3 月 11 日。



構，好讓學者等一些在此領域工作的人知道這是一個專業的組織。關於這點，我認為特別可以從他們發展倫理規約 (code of ethics) 的企圖中看得出來。現在大部分的機構都在發展這個 [倫理規約]，有倫理規約和操作手則 (codes of practice) 等等。這些規範給了在這個領域內的人一個指引，引導他們如何對待他們的工作，以及讓他們明白這份工作對他們的期待。

從 Davies 教授的回應裡，我們發現倫理規約的精神不在硬性約束機構 (此處單指語言測驗機構) 的道德操守，而在激勵身為專業人員該有的風範與自我戒律。換言之，正因為專業，所以有所為、也有所不為。就像 Coady 提醒我們的，當一個人「被迫」當好人時，他行為裡展現出來的道德涵養，就已經被打了折扣。³⁷Coady 的這段話引出了另一個醒思。在學術倫理意識高漲的時刻，為倫理規約投入心力，當然是一個政治正確的選擇。不過，這些作為究竟是從善如流後的行禮如儀，還是發自內心的專業意識與自重，就頗耐人尋味了。道德的自我意識其實是相當重要的，Davies 教授這麼說道：

我認為發展倫理規約是有用的。...發行規約一方面來說，可以當成證明「你是專業的」的指標；另一方面，它也可以是種偽善的假動作。假裝說「你看！我們都是好人喔！」，然後什麼也不做。你懂我的意思嗎？所以，發展倫理規約是有其危險性的。我的意思是，就某層意義上來說，你不應該，你也幾乎可以說你不需要發展這個東西，因為不論如何，你的行為本來就應該合乎道德倫理才對。所以，我覺得這其實跟你的價值觀有關。因為就某層意義上來看，你也可以說：「這就是我們相信的！這就是我們想做的！而且大家都知道」。

Davies 教授的解釋說明了成立規約的「偽善」危機。更重要的是，他道破了成立規約後的困境。這個困境在於：我們無法強制研究者必須遵從倫理規約，也無法得知研究者是否真的遵循了規約。畢竟，遵守學術倫理是自發性的行為，且在沒有倫理規約審查機制的情況下，對於怎麼實踐的問題我們只能交給每位研究者來決定。只不過，當把決定權下放時，我們還是無法確認研究者是否具備了足夠的學術倫理知識與訓練來輔助他們做出相關的決定。³⁸問題不僅如此而已，規約制定時常刻意使用不明確的字

³⁷ Margaret M. Coady, "The Moral Domain of Professionals," in Margaret Coady and Sidney Bloch (eds.), *Codes of Ethics and the Professions* (Melbourne: Melbourne University Press, 1996), pp. 28-52.

³⁸ Marie Smyth and Emma Williamson, "Introduction," in Marie Smyth and Emma Williamson (eds.), *Researchers and Their 'Subjects': Ethics, Power, Knowledge*



詞，以方便研究者依個別研究情境彈性調整。因此，就操作面來看，學術倫理的品質其實是建立在研究者個人的道德基礎上。³⁹總之，缺乏操作方法的倫理規約，即使提供了大方向的道德守則，在沒有具體指引的情況下，通常也無法發揮其真正的效用。⁴⁰更不用說允許太多解讀空間可能導致刻意、或不小心的曲解。有鑑於此，某些專業學科在訂定了倫理規約後，還會陸續訂立操作守則。⁴¹國際語言測驗協會 (International Language Testing Association, ILTA) 就是這麼一個例子。

2000 年時，國際語言測驗協會在溫哥華的年會裡發表了倫理規約，緊接著也在 2005、2007、及 2010 年的年會中積極發展操作守則的撰寫與推行。在該倫理規約的首頁上，該會明確說明了此規約規範了作為一個專業人員所應具備的道德與理想；該會的操作守則也明確指出了怎麼將蘊含在上面規約裡的理想付諸實行，甚至具體指出了測驗機構該有與不該有的行為。⁴²國際語言測驗協會的倫理規約共有九項原則。每項原則皆以「語言測試機構(者)應該」(Language testers shall...) 做為開頭，說明專業測驗機構該有的考量與責任。不但如此，每項原則還附有註釋來進一步說明規約的意涵。更具體來看，該操作守則包含兩部分：第一部分涵括六大項目 (A-F)，主要規範語言測驗機構該有的考量、義務與責任；第二部份 (項目 G) 則著重在告知受試者 (test taker) 應被賦予的尊重與權益。在最後幾點說明中，該操作守則更提醒受試者應主動詢問自身權益。例如，受試者應主動詢問本身的責任及有關考試結果保密條款 (confidentiality) 的規定等。由此可見，倫理規約與操作守則想要傳達的不是單向的制約，而是希望創造

and Consent (Bristol: The Policy Press: University of Bristol, 2004), pp. 1-16.

³⁹ Robin Small, "Codes Are Not Enough: What Philosophy Can Contribute to the Ethics of Educational Research," *Journal of Philosophy of Education*, Vol. 35, No. 3, 2001, pp. 387-406.

⁴⁰ Alan Davies "Ethics, Professionalism, Rights and Codes," in Elana Shohamy and Nancy H. Hornberger (eds.), *Encyclopedia of Language and Education* (London: Springer, 2008), pp. 429-443.

⁴¹ Kenneth Boyd and Alan Davies, "Doctors' Orders for Language Testers: The Origin and Purpose of Ethical Codes," *Language Testing*, Vol. 19, No. 3, 2002, pp. 296-322.

⁴² 請參閱 International Language Testing Association (ILTA), "ILTA Code of Ethics," 2000, <http://www.iltaonline.com/index.php?option=com_contentandtask=viewandid=57andItemid=47>.



一種測驗機關與考生間相互協調與督促的氛圍，俾使所有在此領域的利害關係人，都能積極主動關心自己及他人的權益。其他的測驗機構，如歐洲語言測驗評量協會⁴³與歐洲語言測試者協會 (the Association of Language Testers of Europe, ALTE)⁴⁴也都發表了類似的操作守則，並詳實標明了實行考試與評量該有的作為。

然而不可否認的是，再詳盡的描述也無法窮盡測驗時的所有情況。像國際語言測驗協會這樣的國際組織，要規劃出一套符合各國文化又適合各種社會背景的操作守則，就更是件難上加難的任務了。為因應各國的國情與文化，Davies 教授鼓勵各國測驗機構發展出自己的操作守則，並送交國際語言測驗協會審核。⁴⁵藉由檢測各國擬定的操作守則，這種作法既解決了如何融入各國文化差異的難題，同時也可藉此機會檢試倫理規約的適用性。隨著這樣的多向互動，讓倫理規約更具意義，也更能發揮操作守則的功用。

參、學術倫理、信度、效度的關係

從社會科學研究學界對學術倫理的重視就可以發現，倫理道德早已成了研究的首要條件（學術倫理早已是信度、效度之外的重要指標）。一般探討社會科學研究的參考文獻必定會提到學術倫理、信度、效度對衡量學術研究的重要。但此三大鞏固學術品質的三巨頭、金三角，卻常常被分開來討論。⁴⁶縱使試圖串聯信度與效度的討論不少，⁴⁷但真正將學術倫理、信度、效度看成「相互影響關係」的研究卻不多。在尋傳統方式對三者一一介紹之後，以下本文將扼要藉由國際語言測驗協會的倫理規約與操作守則來探討學術倫理、信度、效度間的關係。此外，我們也將進一步說明為何一個可通過學術倫理檢驗又具有效度的學術研究，必然也可以獲得信度的保證。

⁴³ The European Association for Language Testing and Assessment (EALTA), "EALTA Guidelines for Good Practice in Language Testing and Assessment," 2006, <<http://www.ealta.eu.org/documents/archive/guidelines/English.pdf>>.

⁴⁴ The Association of Language Testers of Europe (ALTE), "The ALTE Code of Practice," *Quality Assurance*, 1994, <<http://www.alte.org/cop/index.php>>.

⁴⁵ Kenneth Boyd and Alan Davies, *op. cit.*

⁴⁶ 同註 12。

⁴⁷ 請參見許禎元，〈社會科學信度與效度的檢定及其關聯性〉，《醒吾學報》，第 27 期，2004，頁 1-23。



基於行文的方便，讓我們先從信度與效度的關係談起。基本上，有效度就確保了信度。這點不管是對量化還是質化而言都是相通的。以量化研究信度的三個面向（即，穩定性、內在信度與觀察者一致性及觀察者內信度）來看。首先，缺乏穩定性的測量結果意謂著每次測到的結果都是不一樣的。造成結果浮動的原因很多，可能是用了不夠精良的測量工具，也可能根本用錯了測量工具等。不論此處的不一致源自於何處，不一致的結果使得研究者無法確定哪次測到的結果才是真的。更重要的是，研究者也無法確定是否真的測到了他想測的東西。例如，假設我們想測的東西是某人的身高，若三次測得的結果都不相同，那麼究竟哪一次才是他真正的身高？又，研究者如何才能在這樣的情況下宣稱他真的測到了身高，就會成為問題。另一方面，對於質化研究而言，穩定性雖不是這類研究的首要目標，但是如前所述（即，質化研究者對複製性的看法），質化研究者仍認為研究結果該有某種程度上的轉移性與複製性。因此，即便質化學者一再以社會現象會隨時間不斷變化為由，⁴⁸強調質化研究難以（且也不期待）屢次取得穩定的研究結果，但從質化學者持續發展信度指標（請參閱前節對信度的討論）的行為來看，我們還是可以見到質化研究者對研究結果穩定性的企圖。

第二，缺乏量化內在信度代表著用來測量的指標同時測到了一個以上的概念。也就是說，研究者測到的東西是多種概念交織後的結果。如此一來，該研究者就無法釐清所測到的結果究竟是哪個或哪些概念。在這種情形下，研究者自然也無法利用測量結果來解釋想要預測的現象。例如，在一份探討自信心與數學考試成績關係的問卷裡，倘若其中參雜了詢問學生喜好與數學知識的問題，這就很可能會造成低內在信度的結果。同時，這樣收集到的資料也無法直接拿來和數學成績做連結。換句話說，由於這些資料裡參雜了喜好與數學知識的變因，因此這些資料無法用來解釋自信心與考試成績的關係。當然，這個結果會被認為是沒有效度的。若將此觀點放入質化研究的脈絡下，我們也會發現類似的推論結果。以訪談為主的質化研究為例，若研究者誤把喜好與知識看成自信心的子項目，而用喜好與知識的問題來解釋自信心與數學成績的關係，該研究同樣也會遭受到外界對其研究效度的質疑。

⁴⁸ Catherine Marshall and Gretchen B. Rossman, *Designing Qualitative Research* (London: Sage, 1989).



第三，低觀察者一致性或低觀察者內信度代表著研究者對所觀察到的現象沒有一致的解釋。當這種情形發生時，便會面臨解釋現象的難題，因為我們無法判斷哪位研究者，或是哪一次觀察到的結果才是真實的。在質化研究中，人們雖對此概念有不同的名稱（即，內在信度），但論者同樣也強調觀察一致性對解釋現象的重要。甚至有學者認為這項概念是評斷質化研究信度的唯一指標（請參閱前節 Hammersley 對信度的定義）。綜觀以上推論，學術研究得通過上面總總信度的檢驗，才有可能保障效度。換句話說，當研究者在宣稱其研究有效度時，他的研究也應當具備了信度才對。⁴⁹

在得知效度與信度的關係後，接下來就必須探討效度與學術倫理之間的關係。在前面介紹效度時，我們已經提到量化研究中所說的構念效度，其實就是質化研究裡的內在效度。量化的構念效度與質化的內在效度都是在評估研究者想要測量的東西究竟有沒有意義。也就是說，研究者真的測對了他應該測的東西了嗎？這裡的關鍵字是測「對」了，而不是測「到」了。換言之，研究者一開始想要探知的概念就必須是「對」的，才能符合他的研究需求。也只有「對」的概念框架下，研究者才能設計出「對」的測量方法。同樣地，唯有在決定了測量方法並實際施測後，我們才能（會）去評估設計出來的測量方法準不準確、有沒有信度。那麼，為何需要確保研究的效度？如上所言，社會科學研究目的大都隱含了「增進對社會的了解，並做出貢獻」的假設。說得更具體一點，研究是爲了社會而做、是爲了對研究現象（對象）做出貢獻而做的（請參考前節關於生態效度的討論）。這樣的目標可以和前面醫學道德規範的行善與不傷害原則相互呼應。總而言之，有效度的研究是爲了探索真實的現象，而追求真實的目的又是爲了增進研究對象的福祉。基此，研究者當然應該把被研究者的權益擺在第一位。事實上，被研究者的福祉與權益，正是學術倫理規範的標的。

對於前述這個規範的目標，國際語言測驗協會在其倫理規約 (ILTA Code of Ethics) 裡有清楚的說明。該規約的第四項原則⁵⁰ (Principle 4) 說明語言測試者不該濫用他們的專業知識與技能。在

⁴⁹ 同註 1, p. 153.

⁵⁰ 該原則原文為：Language testers shall not allow the misuse of their professional knowledge or skills, in so far as they are able.



同項第一款註釋⁵¹ (annotation) 中也提到，語言測試者不該蓄意利用專業知識或技能，來損害受試者的權益。比方說，當語言測驗者被要求擔任某專業語言測驗的考官時，他們要清楚知道自己的職責，注意自己的行為不會危害受試者的權益。⁵²此點說明了測試者在施測時，要以專業的角度出發，真實地回報擔任試考者 (trial subjects) 的心得，並進行反饋。這點其實就是在要求測試者，要用專業來評斷測驗是否具有一定的品質。這也就是在對測驗的效度與信度進行把關。

同樣的情形在該規約的第五項原則⁵³中也可以看見。該原則鼓勵語言測試者持續增進自身的專業知識，並與其他從業同仁分享這些知識。從這個規定裡，我們也可以看出規約制定者想要確保測驗效度的用心。第五項的第一款註釋⁵⁴說明了持續學習與增進個人知識，是身為專業人士的根本，無法做到這點將會對受試者造成傷害。同項第二款⁵⁵則更明確說明了測試者應當盡可能利用各種進修機會來充實自己。比如說，盡可能參加專業會議、閱讀專業出版品等，⁵⁶都是可以增進專業知識的具體管道。以上兩條註釋都強調語言測試者應當增進自己對語言測驗的認知，以確保受試者

⁵¹ 本註釋原文為：Language testers shall not knowingly use their professional knowledge or skills to advance purposes inimical to their test takers' interests. When the progress of the tester's intervention is not directly to the benefit of the test takers (for example when they are asked to act as trial subjects for a proficiency test designed for some other situation), its nature shall be made absolutely clear.

⁵² International Language Testing Association (ILTA), "ILTA Code of Ethics", 2000, <http://www.iltaonline.com/index.php?option=com_contentandtask=viewandid=57andItemid=47>.

⁵³ 本原則原文為：Language testers shall continue to develop their professional knowledge, sharing this knowledge with colleagues and other language professionals.

⁵⁴ 本註釋原文為：Continued learning and advancing one's knowledge are fundamental to the professional role; failure to do so constitutes a disservice to test takers.

⁵⁵ 本註釋原文為：Language testers shall make use of the various methods of continuing education that are available to them. These may involve participation in continuing language testing programmes and professional conferences, and the regular reading of relevant professional publications.

⁵⁶ International Language Testing Association (ILTA), "ILTA Code of Ethics," 2000, <http://www.iltaonline.com/index.php?option=com_contentandtask=viewandid=57andItemid=47>.



的權益。由於測試者是制定測驗的人（單位），因此這些人（單位）對語言測驗的認知會影響他們對語言能力的看法，進而也會影響他們出題的內容。這種把對語言能力的看法用來呼應考題的概念，其實也是前述構念效度的要旨。因此，由本規約的第五項原則裡，我們又再次印證了倫理規約與效度的高度連結。

最後，從國際語言測驗協會的操作守則 (ILTA Guidelines for Practice)⁵⁷裡，我們更能清楚看見學術倫理、信度、效度間的關係。操作守則第一部分 A 項第二款⁵⁸提到，所有考試都必須提供可用來推斷考試結果用途的足夠資訊。以商業英語溝通考試來說，因為商業英語溝通能力是一種概念，因此測驗設計者必須詳細說明這個概念裡有哪些子概念，而測驗成績又是如何對應到這些概念上的。緊接著，在同樣 A 項第三款⁵⁹中又提到，所有的測驗都必須是有信度的，測驗結果必須具有一致性，且必須有一定程度上的概化力。有關信度的規定，還可在 B 項第六款⁶⁰與第八款⁶¹中發現。此兩項條文強調測驗評分者必須熟悉他所評分的測驗活動、確保所有的受試者都經歷相同的測驗程序，並且公布評分者的前後一致性 (intra-rater reliability) 與評分者間 (inter-rater reliability) 的一致性。此處一致性的要求與前節討論信度時所提到的「觀察者一致性」與「觀察者內信度」的概念是相通的。在這個操作守則裡，像這樣與信度、效度相關的條文俯拾即是。

⁵⁷ International Language Testing Association (ILTA), "ILTA Guidelines for Practice," 2005, <http://iltaonline.com/images/pdfs/ILTA_Guidelines.pdf>.

⁵⁸ 該條條文原文為：All tests, regardless of their purpose or use, must provide information which allows valid inferences to be made. Validity refers to the accuracy of the inferences and uses that are made on the basis of the test's scores. If, for example, the test purports to be measuring the ability to use English in business communication, the inferences based on the test score are valid to the degree that the test does in fact measure that ability. However, since the ability to use English in business communication is a construct, the test developer must spell out just what that construct is or what it consists of. The test score inference or interpretation can be valid only if the test construct offers as accurate as possible a picture of the skill or ability it is supposed to measure.

⁵⁹ 該條條文原文為：All tests, regardless of their purpose or use, must be reliable. Reliability refers to the consistency of the test results, to what extent they are generalizable and therefore comparable across time and across settings.

⁶⁰ 該條條文原文為：Those doing the scoring should be trained for the task and both inter and intra-rater reliability should be calculated and published.

⁶¹ 該條條文原文為：Care must be taken to ensure that all test takers are treated in the same way in the administration of the test.



自此，我們可以確定，學術倫理、信度、效度會相互影響，而三者間也存在著代表此種相互影響關係的隱形雙向箭頭。本文認為，了解三者間的相互影響關係只是第一層次的理解而已。更重要的是，我們必須知道他們之間關係是「層層相疊的」。倫理規約必定是學術研究得以展開的第一順位要件。也就是說，它是所有研究的基石。唯有在確定研究根基的正當性後（也就是有了倫理考量後），我們才能進一步發展所欲探索的研究概念，並規劃如何運用具體的測量工具來衡量抽象的概念（或說，這時我們才開始構思構念效度與內部效度的問題）。達成了這些工作後，一個研究才具體化到如何執行這些測量（也就是信度）的考量上。當然，在實際測量的層次，一致性的執行過程是最重要的，因為這關係到測量結果的可複製性。有了可複製性的保證，我們才能確定測到的結果是有效度的，也才能進一步拓展出概化的可能性。

釐清學術倫理、信度、效度三者間的層疊關係除了有理論上的價值，也有實際操作上的需要，這點可從歐洲各學術單位歷年來積極修正倫理規約且大力推廣落實規章的行動中得知。歐洲的倫理規約風氣經歷一開始的道德勸說階段，到日見成熟的明文規定時期，直至現在已成為鞏固研究信度與效度品質的基石。由此我們可以發現落實倫理規約對學術研究的影響。其中一方面的影響來自於外界對研究者的道德觀感；另一方面則來自於學界對研究品質的看法。後者又會最終影響出版研究與發表論文的机会。基此，國內學者在進行（歐洲相關）學術研究時，宜清楚了解各學術單位的倫理規定。這個釐清的動作應在研究初期（甚至規劃研究之前）即開始進行，直至研究的最終階段。隨者研究目的、研究問題的形成，資料收集及分析方法的演變，研究者都必須將倫理規章考慮在內。研究者應當要有持續反思研究倫理的體認；同時，應當了解違犯規章也或多或少會威脅信度與效度的認識；更重要的是，吾人更應當了解違反倫理規約終將折損其研究的價值與貢獻。



肆、結論

學術倫理、信度、效度的重要性對研究者來說也許早已不是件新鮮事，我們幾乎可以說它們早就成為眾所皆知的學術常識了。即便是這麼常識性的問題，相關文獻似乎還是傾向把信度、效度放在一起討論，然後才在它們的後頭加上一小段學術倫理的討論以為點綴。這樣的論述模式雖不至直接影響研究的品質，但如此的安排很容易就會讓人（特別是讀者和新進研究者）忽略了三者間的連動關係。有鑑於此，本文以作者自身的真實研究經驗為例，引出學術倫理、信度、效度三者難以兼得的拉扯與疑慮。這個個案現出了學術倫理對信度和效度可能造成的影響。在這樣的醒思驅使下，本文接著也對前述三項概念進行了討論與重構。

經過本文抽絲剝繭的層層解析後，我們首先發現量化與質化研究文獻中討論的信度與效度概念，常有重疊與交錯使用的狀況。有時人們用同一名詞來闡述不同觀念，有時人們又用不同名詞解釋同一概念。這樣的交錯使用，使得我們在談論信度、效度時，常會陷入雞同鴨講或似是而非的誤區。在釐清了錯綜的名詞問題後，本文接著仔細說明了兩者間的關聯。即，沒有信度的測量結果，就無法成就有效度的研究。這是因為效度需要信度中的穩定性、內在信度與觀察者一致性或觀察者內信度的保證。這個關聯性在量化研究中如此，在質化研究談論內在信度與外在信度的範疇內亦是如此。

了解了上述關係後，本文又以國際語言測驗評量協會的倫理規約和操作守則為例，進一步說明學術倫理與信度、效度問題的緊密關聯。一一推敲規約與守則的內容，吾人不難發現，學術倫理、信度、效度的訴求是環環相扣且相互重疊的。它們之間的關係比較像座一層疊著一層的金字塔，而不是傳統見解下相互分離的金三角。我們認為，學術倫理可視為金字塔的底座，成為效度、信度的根基。此根基意謂著對學術研究的行善、不傷害與道德原則的確保。緊疊著這個基座的是效度，用以確保研究的結果是真實的、是符合研究目的的、是對社會有貢獻的。接著，為達成該目的還需要信度的幫忙。有了高信度的測量工具才能成就高效度的測量結果，最終也才能達致高道德的學術研究。誠如 Davies 教授所言，道德規範不只是崇高的道德情操，更是一種專業的展現。這種專業的展現，標示了當前歐洲社會科學研究的動向，也為正



在積極建立研究倫理準則的我國，指出了新的方向。



參考書目

中文

- 林正介、陳祖裕，〈社會及行爲科學研究相關之倫理議題〉。《人文社會科學簡訊》，12 卷 1 期，2010，頁 19-25。
- 許禎元，〈社會科學信度與效度的檢定及其關聯性〉。《醒吾學報》，27 期，2004，頁 1-23。
- 蔡甫昌、林芝宇、張至寧，〈研究倫理的歷史、原則與準則〉。《台灣醫學》，12 卷 1 期，2008，頁 107-122。
- 戴華、甘偵蓉、鄭育萍，〈人文社會科學與研究倫理審查：執行研究倫理治理架構計畫的考察與反思〉。《人文社會科學簡訊》，12 卷 1 期，2010，頁 10-18。

外文

- Adcock, Robert and Collier David, "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research," *American Political Science Review*, Vol. 95, No. 3, 2001, pp. 529-546.
- Alderson, Priscilla, "Ethics," in Sandy Fraser, Vicky Lewis, Sharon Ding, Mary Kellett and Chris Robinson (eds.), *Doing Research with Children and Young People*, London: Sage, 2004, pp. 97-112.
- Atkinson, Paul, Coffey, Amanda and Delamont, Sara, *Key Themes in Qualitative Research: Continues and Changes*. New York: Altamira Press, 2003.
- Boyd, Kenneth and Alan Davies, "Doctors' Orders for Language Testers: The Origin and Purpose of Ethical Codes," *Language Testing*, Vol. 19, No. 3, 2002, pp. 296-322.
- Bryman, Alan, *Social Research Methods (3rd Edition)*. Oxford: Oxford University Press, 2008.
- Cho, Jeasik and Aleen Trent, "Validity in Qualitative Research Revisited," *Qualitative Research*, Vol. 6, No. 3, 2006, pp. 319-340.
- Coady, Margaret M., "The Moral Domain of Professionals," in Margaret Coady and Sidney Bloch (eds.), *Codes of Ethics and the Professions*, Melbourne: Melbourne University Press, 1996, pp. 28-52.
- Cohen, Louis, Lawrence Manion and Morrison Keith, *Research Methods in Education (5th Edition)*. London: Routledge Falmer,



- 2011.
- Creswell, John W. and Dana L. Miller, "Determining Validity in Qualitative Inquiry," *Theory Into Practice*, Vol. 39, No. 3, 2000, pp. 124-130.
- Davies, Alan, "Ethics, Professionalism, Rights and Codes," in Elana Shohamy and Nancy H. Hornberger (eds.) *Encyclopedia of Language and Education*, London: Springer, 2008, pp. 429-443.
- De Vaus, David A., *Surveys in Social Research (3rd Edition)*. London: University College London Press, 1993.
- Fielding, Jane and Nigel Gilbert, *Understanding Social Statistics (2nd Edition)*. London: Sage, 2006.
- Hammersley, Martyn, *What's Wrong with Ethnography?* London: Routledge, 1992.
- Howitt, Dennis and Duncan Cramer, *Introduction to Research Methods in Psychology*. London: Pearson, 2011.
- Howitt, Dennis and Duncan Cramer, *An Introduction to Statistics in Psychology (5th Editin)*. Harlow: Prentice Hall, 2011.
- Johnson, Robert B. and Anthony J. Onwuegbuzie, "Mixed Methods Research: A Research Paradigm Whose Time Has Come," *Educational Researcher*, Vol. 33, No. 7, 2004, pp. 14-26.
- Kirk, Jerome and Marc L. Miller, *Reliability and Validity in Qualitative Research*. London: Sage, 1986.
- Le Compte, Margaret D. and Judith Preissle, *Ethnography and Qualitative Design in Educational Research (2nd Edition)*. London: Academic Press Ltd, 1993.
- Lincoln, Yvonna S. and Egon G. Guba, *Naturalistic Inquiry*. California: Sage, 1985.
- Luoma, Sari, *Assessing Speaking*. Cambridge: Cambridge University Press, 2004.
- Marshall, Catherine and Gretchen B. Rossman, *Designing Qualitative Research*. London: Sage, 1989.
- Maxwell, Joseph A., "Understanding and Validity in Qualitative Research," *Harvard University Review*, Vol. 62, No. 3, 1992, pp. 279-300.
- Onwuegbuzie, Anthony J. and Nancy L. Leech, "Validity and Qualitative Research: An Oxymoron?" *Quality and Quantity*, Vol. 41, 2007, pp. 233-249.
- Punch, Maurice, "Politics and Ethics in Qualitative Research," in Norman K. Denzin and Yvonna S. Lincoln (eds.) *Handbook of Qualitative Research*. California: Sage, 1994, pp. 83-95.
- Silverman, David, *Doing Qualitative Research: A Practical Handbook*. London: Sage, 2000.



- Small, Robin, "Codes Are Not Enough: What Philosophy Can Contribute to the Ethics of Educational Research," *Journal of Philosophy of Education*, Vol. 35, No. 3, 2001, pp. 387-406.
- Smyth, Marie and Williamson Emma, "Introduction," in Marie Smyth and Emma Williamson (eds.) *Researchers and Their 'Subjects': Ethics, Power, Knowledge and Consent*. Bristol: The Policy Press, University of Bristol, 2004, pp. 1-16.
- Whittemore, Robin, Susan K. Chase and Carol L. Mandle, "Pearls, Pith, and Provocation: Validity in Qualitative Research," *Qualitative Health Research*, Vol. 11, No. 4, 2001, pp. 522-537.
- International Language Testing Association (ILTA), "ILTA Code of Ethics," *International Language Testing Association*, 2000, <http://www.iltaonline.com/index.php?option=com_content&task=view&id=57&Itemid=47>.
- International Language Testing Association (ILTA), "ILTA Guidelines for Practice," *International Language Testing Association*, 2005, <http://iltaonline.com/images/pdfs/ILTA_Guidelines.pdf>.
- Pauwels, Eleonore, "Ethics for Researchers: Facilitating Research Excellence in FP7," *European Commission*, 2007, <<ftp://ftp.cordis.europa.eu/pub/fp7/docs/ethics-for-researchers.pdf>>.
- Spencer, Liz, Jane Ritchie, Jane Lewis and Lucy Dillon, "Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence," *Government Social Research Service*, 2003, <<http://www.civilservice.gov.uk/my-civil-service/networks/professional/gsr/index.aspx>>.
- The Association of Language Testers of Europe (ALTE), "The ALTE Code of Practice," *Quality Assurance*, 1994, <<http://www.alte.org/cop/index.php>>.
- The European Association for Language Testing and Assessment (EALTA), "EALTA Guidelines for Good Practice in Language Testing and Assessment," *International Language Testing Association*, 2006, <<http://www.ealta.eu.org/documents/archive/guidelines/English.pdf>>.
- The European Association for Language Testing and Assessment (EALTA) Annual Conference, *International Language Testing Association*, 2011, <<http://ealta2011siena.cils.unistrasi.it/articolo.asp?sez0=130&sez1=0&sez2=0&art=209>>.
- Wiles, Rose, Sue Heath, Graham Crow and Vikki Charles, "Informed Consent in Social Research: A Literature Review," *Research*



Methods: An ESRC Research Programme, 2005,
<<http://eprints.ncrm.ac.uk/85/1/MethodsReviewPaperNCRM-001.pdf>>.



Research as Applied Ethics: Research Ethics in the European Context

Rong-Xuan Chu^{*}, Chih-Tung Huang^{**}

Abstract

This paper presents a different viewpoint over research ethics and academic integrity. The role of ethics is undoubtedly crucial in the process of conducting social science research. Indeed, ethics has been the highest priority in all strands of social studies conducted within the European research framework. For some academics, keeping ethical issues in mind is not only the best research conduct, sometimes it is the only conduct considered to be legitimate. Yet, a renewed role of ethics in research is demanded to achieve utmost professional attitude and commitment to academic development. Beyond the traditional focus of ethics on human rights, this paper argues that there is a need to reconceptualize the role of ethics in research and to renegotiate relationships among ethics, reliability and validity.

Keywords: Ethical issues, Reliability, Validity, Europe

* Ph. D. Candidate, Moray House of Education, the University of Edinburgh

** Assistant Professor, Department of Public Administration, National Open University

