

# 期刊評價與學術評價中的CSSCI

沈固朝

---

**[提要]** 引文索引作為一種檢索工具，利用研究之間的相互引證關係為學者們提供跨越時空的豐富文獻，客觀、系統地反映出科學研究工作和成果的繼承性和發展性。在這個意義上，引文索引被科研管理部門作為學術成果質量的定量評價工具。引文統計分析結果與使用多項定量、定性評價指標的評估結果有較高的相關性，但也有其局限性。論文分析了CSSCI用於期刊評價和學術評價的利弊，討論了以影響因子來評估研究者個人業績的問題，指出了產生問題的體制性根源。作者呼籲要正確認識引文索引在期刊評價和學術成果評價中的作用。

**[關鍵詞]** 期刊評價 學術評價 CSSCI 引文索引

**[中圖分類號]** G236; G304 **[文獻標識碼]** A **[文章編號]** 0874 - 1824 (2017) 03 - 0114 - 11

---

2017 ~ 2018 年度 CSSCI 來源刊目錄的發佈再次揭開了潘多拉匣子，各種問題和評論如預料撲面而來。無論是調侃式的吐槽、無奈的歎息還是冷靜的分析，都從一個個側面重複著多年的“老問題”。“學界被影響因子綁架”、“引文成為標準的理念和體系”、“量化的學術評價標準是架在學者身上的枷鎖”，“請‘核心期刊’走下神壇”等等聲音可能有一定的代表性。以“影響因子神話”為關鍵詞搜索一下網絡，還會發現同一用詞的不同用意，一種是介紹刊物如何通過吸收優秀稿源、拓展獨創性科技論文發表的“快速通道”、培養專業編輯隊伍、與科學家直接對話等等獲取期刊的高影響因子；<sup>①</sup>另一種是鞭撻期刊影響因子的操縱行為，呼籲“去魅”、“打破神話”，或提高計算的科學性。<sup>②</sup>

國內有多種評價工具，無論哪一種，發佈其目錄都會引發學術界大小不同的“地震”，這在國際上是少見的。筆者近年利用接待來訪和期刊界大小會議，多次呼籲全面和正確地看待期刊評價工具。所謂“全面”，就是既要看到評價工具的有利一面，也要看到不足一面；在使用過程中要關注它的問題，要看到引用的不同動機及其複雜性所造成的“引文率”被扭曲的問題。所謂“正確”，就是要客觀地評價引文索引的作用，既不誇大，也不貶低。引文分析當然不是解決所有評價問題的萬能鑰匙，但研究表明引文統計分析的結果與使用單項或多項定量、定性評價指標的評估結果有很高的相關性，經過長期和廣泛的實踐檢驗，迄今還沒有更有效的工具取而代之。<sup>③</sup>將它“神化”的原因，是在使用中將高引用與高水平等同了起來，將影響力與創新力等同了起來，

將來源期刊收錄標準與學術評價標準等同了起來。量化的指標不是期刊影響力評價的全部指標，也不是期刊質量高低的唯一標準，更不是學術評價的衡量標準。但這種呼籲只是一種期望。正如編輯朋友們調侃，說得不錯，可沒有用！

那麼，出路何在？響應“舊金山宣言”，在科研評價中停止使用基於期刊的計量指標，如期刊影響因子<sup>④</sup>或者遵循“萊頓宣言”的原則，<sup>⑤</sup>“請‘核心期刊’走下神壇”？筆者懷疑，在不觸動評價體制的情況下單純取消工具，不僅無助於解決問題，還會把暴露的問題又重新掩蓋起來，因為產生問題的土壤還存在。我們的視線不能僅僅停留在表象和問題上。筆者願借助此文再談一下對這些問題的一得之見。

### 一、引文索引的本質是檢索工具

自引文索引的先驅、美國 1873 年出版的供律師查找判例的檢索工具——謝潑德引文（Shepard's Citation）問世以來，引文索引已經走過 140 多年的歷程。期間，尤金·加菲爾德（E. Garfield）於 1963 年編製的“科學引文索引”（Science Citation Index）以及於 1973 年、1978 年分別創辦的“社會科學引文索引”（SSCI）和“藝術和人文科學引文索引”（A&HCI）影響最大。這種索引利用文獻中普遍存在的參考引證現象，為人們提供了一種新的查找文獻的方法。如果到 Web of Science 的站點瀏覽一下，可以發現，絕大部分欄目都是在介紹如何利用各種檢索功能去跟蹤研究趨勢，如何查找作者、刊物和高質量的論文，如何通過文獻之間的引證關係從前人的著述中獲取所需的知識，瞭解科學研究的發展脈絡和思路。在 SSCI 板塊，讀一下“為什麼要利用 SSCI”（“Why SSCI?”），在所介紹的五大功能中，隻字未提學術評價。同樣，中國科學院文獻情報中心於 1996 年、南京大學中國社會科學研究評價中心於 1998 年先後建立的 CSCD 和 CSSCI 不僅最初是以檢索工具的面目出現，直到今天這兩大文獻數據庫的界面展現給讀者的仍然是各具特色的檢索功能。

說引文索引是檢索工具而非評價工具，不是旨在為種種問題開脫，而是說明這種工具一開始就不是為評價而設立的，無論是期刊評價還是學術評價。CSSCI 與其他評價工具的不同之處，在於它只能用自身產生的數據，如果要進行評價，還需要搜集其他數據，建立其他指標，特別是定性指標。可以說，南京大學中國社會科學研究評價中心長期以來主要是在進行數據加工，在少數情況下依據用戶要求承接過一些引文分析項目，但它尚未從事過真正意義上的人文社會科學研究評價。

在引入全文檢索之前，引文索引是標準的二次文獻。與普通的文獻檢索工具相比，引文索引庫是經過科學地篩選後確定的世界範圍內自然科學、社會科學和人文科學方面的重要期刊的綜合性、多學科、具有權威性的文摘類檢索數據庫。它既可以從篇名、關鍵詞、文摘、著者、來源期刊、機構、地址等角度檢索世界範圍的各學科重要期刊文獻信息，也可以查閱普通檢索工具難以查檢的某些交叉學科的資料。引入全文檢索之後，引文索引“檢”（線索）和“索”（取原文）的界線消失了，兩大功能在數據庫中合二為一，使學術服務工具的特點更加突出，分析的功能得到了進一步發揮。例如，若來源文獻 A 和來源文獻 B 都引用了文獻 C（A 和 B 的引文），則 C 就是文獻 A 和文獻 B 的引文耦。引文耦愈多，其相應的來源文獻之間的相關性愈高。很多人將引文索引視為評價工具，忽略了引文索引的特殊檢索功能，這實在是使用這類數據庫的一大損失。僅就期刊編輯而言，無論是選題還是組稿，甚至在審稿時借鑒相關的研究，引文索引都是非

常好的幫手。文獻之間的引證和被引證關係所揭示的研究專題之間甚至作者之間存在的某種內在聯繫，不僅指明了與讀者需求最密切的文獻線索，而且包含了相似的觀點、思路、方法，反映了科學交流活動，顯示了科研成果之間、刊載文獻的期刊之間以及文獻所屬學科之間的內在聯繫，使許多論文有機聯繫起來，構成論文網、著者網、文獻網，這是引文索引最初及至現在的最重要價值所在。以後西方在綜合性引文索引的基礎上又進一步細化出了專科引文索引，如“生物科學引文索引”（BioSciences Citation Index）、化學引文索引（ChemSciences Citation Index）、臨床醫學引文索引（Clinical Medicine Citation Index）等。ISI 還將 Web of Science 與其他數據庫連接起來，如與 ISI 化學服務數據庫（ISI Chemistry Server<sup>®</sup>）、德溫特發明索引（Derwent Innovation Index<sup>®</sup>）、美國國家生物技術中心（NCBI）基因庫（GenBank）等連接起來，從而使研究人員能夠清晰地瞭解學術研究中所涉及的專利，推動科學研究迅速轉化為生產力。

## 二、引文索引的選刊和評刊作用

當引文索引在國內推廣開來以後，期刊評價這一索引的附加功能逐步受到了重視。體現評價功能最明顯的依據是根據文獻計量學原理編製的“期刊引用報告”（Journal Citation Reports, JCR），即對每一份被收錄期刊用以下的統計數據進行分析：通過影響因子評價和比較一種期刊相對於同學科中其他期刊的影響力；通過即年指數（Immediacy Index）比較同一年內期刊中的論文被引用的頻率；通過被引半衰期（Cited Half-life）反映被引論文價值的持久性；通過引用期刊列表（Citing Journal Listing）及被引期刊列表（Cited Journal Listing）反映引用某特定期刊的頻繁程度；通過總引用次數（Total Cites）反映該刊歷年文獻在統計當年被引用的情況等等。引文分析發現，被引頻次居前 20% 的期刊包含了 SSCI 期刊總被引頻次的 70% 以上，即較少的期刊包含了大多數的被引頻次。這一被稱為 20/80 定律可為圖書館藏書建設提供選刊依據，可為學者高效率地獲取所需學術論文提供參考，亦可為建設特定專業期刊網絡提供信息源。

引文分析及其相關的研究引起了圖書館學界的重視，引發了以文獻計量學理論為指導的期刊研究熱潮。據南京大學博士生彭玉芳統計，1998 年至 2016 年 9 月 17 日，知網總共有 756 篇文獻與引文分析相關，平均每年發文量為 39.8 篇。尤其在 2008 年至 2011 年期間，引文分析作為重要的評價工具，對整個學術界、期刊界產生巨大影響。從這些文獻使用的總共 1,148 個關鍵詞來看，引文分析涉及期刊評價、機構評價、學術評價、科研評價、研究方法評價、人才評價、數據庫評價、學科評價、軟件評價、引文評價、核心作者分析、期刊影響力、學術聲望等。當然，反對和質疑聲音也開始出現，相應關鍵詞是：同行評議、自引率、引文動機、計量評價模式、引證動機等。

當然，文獻計量工作者毫不避諱用計量方法進行期刊評價的缺陷。SSCI 期刊的學科地域分佈是極不均衡的，不同的國家有著各自獨特的歷史文化背景和意識形態，社會科學研究取向、學術成果交流方式受地域性限制較大，這種不均衡現象使其作為社會科學的國際可比性指標的公正性大打折扣。<sup>⑥</sup>CSSCI 用與引證相關的影響因子指標來評價不同引證習慣、不同學科論文的綜合性刊物，也具有不可比性。<sup>⑦</sup>由於各學科專業期刊影響因子的差別，以及來源期刊被作為“核心期刊”而人為地限制了其數量，使那些引用局限於較小領域的刊物入選極為困難，有的甚至再怎樣努力都沒有可能。

即使評刊，也有基於形式的定量評價和基於內容的定性評價兩種方法。依筆者之見，引文索引作為評刊工具，到目前為止還主要是一種基於形式的量的比較（如論文的聚類度、相關度、規

範化、價值的持久性等等），而非對期刊學術價值的質的評價。之所以稱為“比”而非“評”，是因為第一，評審的對象是一系列反映被評客體實際狀況的數據，或可供相互比較的信息（可比性原則）；第二，被比較的各刊處在同一個層面之上（相容性原則），因而對“同質”的關注多於“異質”；第三，要求在評價中徹底摒棄各種“己見”或“學術偏見”，堅持一種多元或中性的立場（平等性原則）；第四，在具體操作上，要求評價人員盡可能遵循國家或國際上統一規定的統計口徑和數據處理方法，盡可能減少數據處理的主觀選擇範圍（可操作原則）。這些恰恰都是比較的特點。根據上述原則所“比”出的高與低、多與少、強與弱、優與劣等等，得出排序。顯然，這種“評價”與基於內容進行價值判斷的“評價”是有很大差距的，它甚至不要求“評價者”能夠讀懂被評對象的內容，只要求掌握數據處理的方法。評比的目的是促進競爭，評比的過程具有博弈的特點，評比的結果是將量的大小作為質的高低的參照。這種方法的優點——避開了基於內容評價所帶來的非共識性、耗時性和難操作等缺陷，因而受到科研管理機構的歡迎，尤其在分配緊張的學術資源時，它是解決矛盾的有利工具。儘管我們可約定俗成地用評價一詞涵蓋幾乎所有的評議、評判、評獎、評優、評級、鑒定、考核等活動，但區分基於量的比較和基於質的判斷，有利於瞭解各自的優點和局限，避免混用和互相取代。名為評價實為評比，就可能將一些不可比或難比比的要素排斥在外，用統一的標準來比較不同研究對象、不同研究方法、受不同傳統習慣的影響而表現出的不同形式的學術研究成果及其特點各異的表述方式，其公平性就很難掌握。

還需要指出，來源期刊的遴選並非全部都是定量方法。從編輯選稿、同行評議、主編定稿到最終由各學科代表組成的指導委員會或學術委員會審定，以及部分學科在遴選前的問卷調查都可視為基於內容的定性評價。讀者引用本身就是從眾多論文中選擇（或稱內容評價）的結果。這是來源期刊產生的前提和基礎。它之所以被戴上“量化的學術標準”的帽子，也許是因為最終決定排序的兩個指標——他引影響因子和被引總數（總被引頻次）是定量指標。但這種指標，所依據的是在大樣本數據基礎上總結出來的各類學科文獻交流規律，並非專門針對文科、理科或某一學科。引文索引統計源的規模樣本保證了它的總體數據的可靠性和公正性，十幾萬篇論文的數據量是任何抽樣調查和同行評審都做不到的。少數失範行為很難扭曲它在總體上的公正性。

實踐也證明，無論在哪一種評價工具中，絕大部分口碑較好、價值較高的學術期刊都收錄其中，只是排序稍有不同。在這個意義上，期刊影響因子在評價學術期刊、推薦學術論文方面所提供的參量，是具有統計上的借鑒意義的，但只是作為影響力大小和變動的參照，不宜作為期刊質量高低的標準。正如健康檢查，其影像或化驗的報告給出的數據僅是某些健康指標的信號，報告上的結論也只是“臨床印象”，疾病或健康與否的定性結論，只能靠醫生結合其他指標和經驗作出判斷。

### 三、被視為學術評價工具的引文索引

無論是從事學術研究的學者、攻讀學位的學子還是從事科研管理的工作人員，在論文開題、確定選題、申報項目時往往需要解決下述問題：如何在前人研究的基礎上進一步推進研究？如何防止重複研究？如何瞭解有影響的學者？怎樣知道人文社會科學研究的熱點與發展趨勢？怎樣知道某一觀點或研究的演變軌跡？我國人文社會科學研究所產生的有重要影響的學術資源在哪裡？各學術機構的學術地位及影響力如何？能否根據論文的引用文獻類型、類別等來考察各學科的研究特點、學者的閱讀和資料利用趨向、學科發展的成熟度，等等。

就上述問題，筆者將學術評價分為宏觀或中觀的評價，以及微觀的、對個人學術成果的評價。幾乎所有引文索引庫的優點都賦予了它們回答前一類評價問題的能力，即：通過各學科論文的關鍵詞數量以及關鍵詞的變化趨勢，考察各學科研究熱點和研究趨勢；通過各學科論文引用期刊的數量、期刊的影響廣度、半衰期以及期刊影響因子可評價各學科期刊的影響力；通過對機構、地區的發文和被引統計分析得知機構和地區的學術影響力；通過對機構篇均被引數據可分析機構成果的整體影響力；通過論文和著作在某一學科的被引數據統計分析，給出在這一學科學術影響最大的論文和著作的學術影響的評價；通過這些論文和著作的主題可考察學者關注的問題。Web of Science 在 2004 年提供檢索“結果分析”(Analyze Results)，分析的選項包括作者分析、國家或地區分析、文獻類型分析、機構分析、語種分析、出版年分析、期刊分析、主題分析。依據在不同的選項中的文獻量的降序排列來實現對檢索到的數據集合的量化，如前 10 位、前 25 位……前 500 位，也可以根據特定的需要進行大量的人工操作以滿足深入分析的要求。同時，通過開發數據庫的鏈接功能，研究人員可以獲得某些論文的全文。在這個意義上，引文索引數據庫可在宏觀和中觀層面上服務於學術研究，但在微觀層面上，難以對個別學術成果的學術價值進行評判，只能通過被引用情況瞭解個人的學術影響力，為內容評價提供某種借鑒。

期刊評價工具介入學術評價有它產生的環境和時代因素。它之所以能夠在上世紀 90 年代“熱”起來，越來越深、越來越快地介入了學術評價，緣於某些流於形式的專家評議使得同行評議失去了往日的地位和作用，<sup>⑩</sup>當時對單一的學術成果定性評價的抨擊不亞於今天對量化評價一統天下的責難。為解決評價中的矛盾、減少“主觀因素”的干擾而產生了量化學術評價的需求。而 SCI 的引入推動了高校和科研機構的排名，科研與利益掛鉤的激勵機制以及隨之而來的科研行政管理的強化，推動了由排名產生的“學術榜”效應，易用性又進一步強化了這種效應。

但畢竟期刊評價和學術評價是兩種性質、對象、目的和方法截然不同的評價，它的合理性往往掩蓋了它的不合理性，不妨套用“存在即合理”這句哲學名言，看一下何以期刊界在評價中“缺位”(筆者謂之“辦刊者不評價，評刊者不辦刊”)，以及期刊評價和定量分析被“越位”用於學術評價。

以學術評價的內容之一——作者研究為例，美國人很早就開始了這一探索。1926 年美國學者洛特卡(A. J. Lotka)在著名的《華盛頓科學院報》(*Journal of the Washington Academy of Sciences*)上發表題為“科學生產率的頻率分佈”(The frequency distribution of scientific productivity)的論文，旨在通過對發表論著的統計來探明科技人員的生產能力及對科技進步和社會發展所作的貢獻，也開創了用定量方法探索科學文獻數量與作者數量之間的關係。此後，洛特卡的研究經科學史學家普賴斯(D. S. Price)的補充，確定了從發文量篩選科學界核心研究力量的普賴斯定律(Price Law)。但這個定律沒能揭示科研成果深層次的“質”，即多產作者群不等於核心作者群。隨著研究的深入，加菲爾德在上世紀 50 年代不僅開創了從引文角度追蹤科學發展動態的新方法，而且與普賴斯一起在引文索引的基礎上發展了引文分析技術，並用於評價科研成果的質量和影響力，進而測定某領域的核心作者群。2005 年，物理學家喬治·赫希(Jorge Hirsch)提出 h 指數的新思路，避開了以往的指標只傾向於科研人員發表論文的期刊這一缺陷，用論文積累的指標衡量科學家個人學術成就。然而這一研究雖可以評價一個科學家的終身成就，卻無法揭示研究的最新進展，不合適於評價年輕科技人員。最重要的是，無論發文量還是被引量，只是科研活動的表現形式，只能直觀地反映學者的學術活躍度，並不能完全反映一個科學家在某

領域中的學術貢獻。<sup>⑨</sup>可見，無論是論文檢索、刊物評價還是引文分析，都可作為學術研究的輔助工具，這是引文索引在我國被視為學術評價工具的原因，也是短時間內迅速普及、得到廣泛應用的原因。

期刊評價工具被用於學術評價，還有如下原因：第一，學術期刊是承載學術成果的載體，對載體的評價和對內容的評價很難割裂開來；第二，科學家認為在一定的時期內一篇文章的重要程度可以通過被引用的程度來衡量。畢竟，引用反映了學術共同體對某一學術成果的關注和重視，通過對引文的分析，可以瞭解該研究對其他研究的影響，揭示相關學科之間、相關研究成果之間的關係，這是研究成果學術價值的體現；第三，依據文獻計量學原理編製的年度性的“期刊引用報告”為科研管理機構進行學術資源的分配、學術活動的管理提供了便利，使得複雜的學術評價變得具有可操作性；第四，量化的引文分析結果與用其他方法評定的結果有極強的相關性；<sup>⑩</sup>第五，由於學術評價是與學術機構團體利益和學者的個人利益掛鉤的，這種評價的競爭性，以及由於管理所要求的即時性等特點，使其在掌握公平性和易操作性方面具有相當的難度，以計量為主的期刊評價工具則具有相對優勢。

然而，要用基於期刊整體的評價代替基於個體的學術成果評價有其先天的缺陷。首先，所謂“核心期刊”是根據佔總量 20% 的期刊上的論文可以滿足學術論文引用需要的 80% 這一文獻計量學的定律而確定的，代表這 20% 的來源刊不能滿足散佈在 80% 期刊上的另外 20% 左右的引用需求，也就是說另外 20% 左右的高學術價值的論文未被承認；其次，論文發表後的被引時機和時間長短因不同學科、不同研究領域而有不同；同樣，科學研究中的價值從被發現到被重視、承認和廣泛接受亦有一個過程。因此，成為“熱點”而被大量引用不能與價值完全劃等號，未被引用的亦不能與“沒有價值”劃等號。科學史上已有許多例子證明這一點。再次，任何一個刊物不可能每篇論文都得到引用，而期刊評價卻是以該刊的所有論文為整體的評價，可稱之為一種“中觀”評價；學術成果評價則是對個別學術成果（在期刊中指學術論文）的學術價值的評判，是一種“微觀”評價。Thomson Reuters 在題為“保持期刊影響因子的完整性”的一段論述中，用粗體字強調，不要用期刊影響因子去評價個人的業績（Do not use the Journal Impact Factor to assess the performance of an individual researcher），期刊影響因子只能用於評價期刊。<sup>⑪</sup>最後，正如前述，目前的期刊評價大多是一種基於形式的評價（或依筆者之見是一種評比），而學術評價則是針對學術成果內容的評價。學術評價早在計量方法引入之前就已存在，它的內涵遠比文獻計量學所能測度的內容豐富得多，複雜得多，以目前的方法論和信息技術而言，還難以用數據來剛性地、單一地分析複雜的人類腦力勞動和精神產品。為此，人們只得從影響力（引用量、轉載量、閱讀量或借閱量、發行量等）、社會聲譽（獲獎數和獲獎層次）、學術地位（作者職稱或機構的權威性）、規範程度（題錄信息的著錄、引文質量、差錯率）、制度保障（有無匿名評審制度、有無學術違規行為、有無收取版面費）等等形式上的、定量的觀測結果進行判斷，只在宏觀和中觀層次上有借鑒意義，但在微觀層次上，計量方法目前還難以對個性化的學術研究結果，特別是對其內容的質量高低作出科學、合理和相對準確的價值判斷。

學術評價是針對內容的。這就要求評價者（專業編輯和同行評審者）能夠讀懂論文的內容，能夠從研究問題難易程度、專業知識掌握程度、論證水平、觀點、方法和資料駕馭水平等等尺度進行論文的綜合質量分析，在此基礎上進行科學性、價值性、前沿性、學術貢獻、學術特色、風格、專長的比較，最終得出刊物整體的創新性、研究的多樣性、爭鳴性、價值性的判斷。同時，還需

要對作為載體的學術刊物的研究導向、欄目設置、編稿質量（校對、差錯率）、學術規範（審稿、編輯、著錄、出版及學術不端檢測等）進行審查。顯然，上述評價不是引文索引的數據加工人員能夠做到的。

因此，引文索引能否用於學術評價，不是個“能”或“否”的簡單答案，要看用在哪些方面，以及如何應用。正如前述，在宏觀或中觀層面上，依據引文資料所做的學科分析是具有一定的參考價值，但如果用於個人學術成果的評價，一定要慎之又慎。論文是否收錄在某核心期刊上可以成為衡量學術水平的參考指標，但不宜作為硬性指標。憑什麼說不在核心期刊上的文章就一定不是好文章？如果作者自薦加上學術同行的審核認定，為什麼就不能得到科研管理部門的認可呢？不幸的是，當這個問題遭遇“操作程式”，“參考”就變成了“硬性”，因而一再驗證了我們常常批評學術評價的“簡單化”、“一刀切”和“行政化”，反映了科研管理體制中深層性的老問題。

#### 四、學術評價：影響力還是創新力？

科研管理部門在使用文獻管理部門的期刊評價工具進行學術評價時，以學術期刊的整體質量來“估測”某一學術論文的質量，是只利用了期刊作為學術成果載體的特點，未充分認識到，反映文獻分佈規律和老化規律的指標，包括影響因子、被引頻次、即年指標、期刊影響廣度、地域分佈、半衰期等等可量化的指標無法取代不同類型的科研與教學活動所具有的不同的價值標準。

創新力應是衡量學術論文價值的根本標準。一篇沒有創見的文章，對社會進步和科學技術發展不起作用，也無法提供科技領域新的內容。創新性大，論文價值高；反之，論文價值就低。所謂的創新性是指論文中闡述世人尚未談過的新理論、新方法、新技術或創新性的模仿，而且需要在實踐中經過反復探索、研究而形成。借助影響力指標來評價創新力面臨著技術性難題：這種始於問題的覺知、繼於心智活動的探索和方案的提出、終於問題的解決和驗證的活動，沒有統一的表現形式可以描述，沒有固定的行為模式可以遵循。此外，創新力的評價需要對相關的學術領域有深刻的瞭解和洞察，難以避免同行專家因其學識、興趣等個人因素的影響所造成的歧見。而創新的發現要求對“異質”關注多於“同質”，更側重個性而非共性的比較。一種在求異或逆向思維主導下的創見在最初未必會取得共識，成為熱點，因而在一定時間內（例如兩年引用期）不一定有較高的引用率，也不一定達到峰值。最早揭示核心期刊現象的英國著名文獻計量學家布拉福德（S. C. Bradford）將其發現公佈於《工程》（*Engineering*）雜誌上的時間是1934年，這篇題為“關於特定主題的情報源”當時沒有引起多大反響，兩年後也未產生“引用峰值”。解決科學問題本身需要時間，原創成果的傳承和傳播需要時間，創新成果的價值檢驗需要更長的時間。因此，創新力的幾個要素——創造前所未有事物的能力、解決問題的能力、在思考過程中能夠產生創新事物的能力、將可聯結的要素結合成新的關係的能力——很難用文獻計量學的語言來描述，至少到現在為止評價工具還沒有理想的定量方法和可以量化指標。

當然，強調學術評價的定性的創新力標準，並非要否定定量的影響力標準，而是想說明，當我們不得不借助影響力指標輔助學術評價時，不能忘了學術研究的根本目的在於探索真理，解決問題，提升創造力。將影響力作為借鑒，只是從易操作的角度而言的，不能忽視由此帶來的急功近利的副作用，特別是當學術評價與利益分配捆綁在一起的時候。

期刊評價目前被錯位地用於個人成果的學術評價是現行體制下的一種“無奈的選擇”。說明這一點的目的在於不要混用這兩種評價，要看到期刊評價用於學術評價的缺陷，也要看到期刊評

價行使這種作用的合理性，不能避難就易地將學術評價排除在現有的評價體系中。相反，在認識到這種“不可替代性”的基礎上要積極推進以學術創新力提升為導向的學術評價研究，因為學術評價導向不僅影響著社會科學研究的導向，也是決定社會科學研究的發展源泉。學術評價如何改進現存評價體系中的不合理部分，讓它發揮推動和引導科研活動朝著更具創新力的方向發展，是評價研究工作者、學者和管理部門共同面對的話題。

## 五、問題的根源：評價工具還是評價體制

無論哪一種工具，在用於評價時都會暴露出自身的缺陷。無論採用哪些評價指標，都可以找到反例來批判說明這些評價標準不恰當的地方。<sup>②</sup>產生上述問題是緣於我們面臨著一系列矛盾：管理人類極為複雜的思維和智力勞動不得不用相對統一的標準和簡單、易於操作的方法；創新力最活躍的因素——科研好奇心、獨立思考、探索勇氣、建樹激情、問題意識和批判精神往往被套牢在追逐利益的功利枷鎖中；缺乏對定量指標局限性的認識，走入“定量指標＝客觀評價”的誤區；統一的定量指標如何對個性化的學術研究，特別是研究成果的質量高低作出科學、合理的評價，等等。一旦評價結果形成了按刊物影響力大小排序的目錄並且在科研管理中得到應用、推廣，形成了在績效分配、職稱評聘、崗位晉級、學位頒發中的一整套考核規定和制度，這套制度又會反過來固化當前的評價方法，兩者相輔相成，甚至互為因果：刊物對作者說，請給我好稿子。作者說，你是“核刊”嗎，如果不是，給了也是白給，因為單位不算“工分”。那麼，是有了好稿子才可能有“核刊”，還是有“核刊”才能吸引好稿子？陷入這個“雞”與“蛋”孰先孰後的怪圈永遠無解。

如果始終盯著工具，而忽略了工具的使用，則是本末倒置。能否跳出這個怪圈，從使用的角度，也就是說從體制這個根本性的問題上作一番思考？這裡所說的評價體制是由評價客體、評價主體和規則制度組成。評價客體是學術論文還是作為論文載體的期刊？評價主體是學術同行還是數據加工人員或者是科研管理者？評價規則是基於創新力還是基於影響力？是基於解決問題的科研成果還是基於刊物等級的論文數量？擺放的位置不同，發揮的作用及其產生的結果自然不同。期刊評比之所以不能取代學術評價，在於影響力的評價不能代替創新力的評價，形式評價不能代替內容評價。它們之間關係的“缺位”、“越位”造成了“以形式代替內容”、“以量代質”、“以刊評文”等等的“錯位”，而造成這種錯位的恰恰在於“規則”——以科研產出為目標，以產出數量、等級與利益分配掛鉤的績效主義評價為激勵。尤其是將來源期刊作為學術榜，凡不在該類刊物上的論文不被認為是好論文，得不到學術共同體的承認，這種硬性規定將“好稿源”趕到了那一小部分“核心期刊”上，由此增添了“馬太效應”中的人為因素，不僅高質量論文的積壓造成發表時間延長，而且加劇了刊物的競爭——往往不是“互補、互識、互鑒”式的良性競爭，而是比“學術聲譽”的“老大、老二、老三”的座次爭搶，在核心刊和非核心刊之間，甚至在核心刊之間“比”出了怨言和矛盾。

更嚴重的是，“大一統”的剛性標準所面臨的評價有不同的類型（成果評價、機構評價、刊物評價、科研評價、論文評價、著作評價），不同的對象（基礎研究、應用研究、理論研究與實證研究），不同的層次（宏觀層面的學科評價、中觀層面的機構評價與科研評價、微觀層面的成果評價與科研人員評價），不同的內容（學術問題的評價、學術目標的評價、學術論證的評價、學術傳承精神的評價），不同的標準（定量與定性、學術標準與非學術標準、直接標準與間接標



準、價值標準與科學標準、人文學科標準與社會科學標準、基礎理論研究與應用實踐研究），不同的期刊定位（普及刊、工作刊、學術刊）等。要用一把鑰匙開那麼多鎖，再怎樣改進工具都無濟於事。一些人把怨氣出在工具上，試想，如果取消諸多評價工具中的一種就可以解決上述問題，那麼，取消了學位與研究生教育發展中心曾試圖推出的“A類期刊”，有沒有解決“以刊評文”這個體制性的問題呢？<sup>⑩</sup>再試想，如果取消“學術榜”式的“核心期刊”目錄，真正回歸數據庫收錄期刊的來源目錄，影響力大小的問題，無論“前100位”、“前200位”交給各類用戶根據自身的情況自行確定，原來“圈子”裡的刊物會贊成嗎？失去“可操作、易操作”工具的科研管理單位會同意嗎？讓被視為“核心”的數據庫真正回歸數據服務的角色並非是一件簡單的事，就眼前來說，它涉及評價體制的問題，就長遠而言，更涉及我國科研管理、科研政策改革這類更深層的問題。不少人說，有工具比沒有工具好，畢竟來源期刊目錄中的絕大多數是被學界認可的高質量的期刊，取消工具豈不亂套？筆者的理解，這個“套”就是我們現在習慣了的評價體制和科研管理體制。評價體制中的問題不解決，被這些問題放大了工具的缺陷是難以克服的。

當然，“規則”的改變已經開始，例如，在成果評審方面，延長評價週期，關注“代表性成果”，改評價標準“一刀切”為分類指導與分層次考核相結合；在評價主體方面，加強學術評價中的匿名評審、雙盲評審、迴避原則，細化專家庫中的“小同行”以避免“外行評內行”；在職稱評審標準方面，對科技人才不再將論文發表作為評價應用型人才的限制性條件，對工程技術人才、正高層次系列不再共用一個評價標準，對於企業、基層一線和青年人才，破除論資排輩，突出創新創業實踐能力的評價；在評價制度方面，逐步建立公示制度、監督制度、申訴制度、糾錯機制，等等。希望這種改革能夠真正觸動體制這個“套”，產生抑制學術生態失衡的實效。

## 六、“解套”的起點：走出認識上的誤區

走出困境，首先要走出認識上的種種誤區。

第一，淡化核心期刊意識。正如考試是學生的“指揮棒”，當評刊工具成為考核工具的時候，也就變成指導一切研究和教學工作的“指揮棒”。“核心”者，“中心”或“主要部分”（《新華字典》，北京：商務印書館，2001年），它必須是某一學科較為集中地刊載原創性論文的學術性期刊，它必須是少量的、具有代表性的期刊，它也必須是刊載某一學科文獻密度大、引文率及利用率相對較高、代表該學科現有水平和發展方向的期刊。<sup>⑪</sup>核心期刊是一種客觀存在，可當它成為追逐的目標和行動的自覺時，就變成了指導一切工作的“指揮棒”。去除“指揮棒”作用，首先要淡化“核心期刊意識”。科研管理上的“核心期刊意識”是將該刊目錄作為科研定量管理的硬性指標，忽視或輕視低引文率的好論文，在此思路下出台的各種評價指標和配套政策成為教學和科研的“抓手”，導致只認刊物不認論文的學術評價；刊物建設上的“核心期刊意識”是擯棄學術期刊發展自身的規律和初衷，只以期刊評價指標為“抓手”，以提高影響因子為唯一目的，這有可能導致期刊偏離了學術追求和社會關切，走向“世俗化”、“功利化”和“同質化”；作者的“核心期刊意識”是以論文能否登載上核心期刊為目標而不論是否適合發表自己的成果，導致“為‘核心’而投稿”、“為論文而論文”，一旦刊物離開了排行榜，立即撤稿改投他處，哪怕削減一半篇幅也在所不惜。這種功利性的“研究”越多，放在真正研究上的精力越少。長此以往，學術成果的價值大打折扣；學術評價機構的“核心期刊意識”是試圖掌握評刊的話語權並引領學術研究方向，而不是關注如何通過學術批判和學術爭鳴去提升創新力。

當然，產生“核心期刊意識”的真正原因並不都在刊物、作者或評價機構，要克服和抵制由此帶來的負面效應也非一刊、一人、一機構能勝任。評刊工具能做的，就是盡力淡化充當“核心期刊排行榜”的角色，回歸為學術活動提供數據服務的原位上來。就 CSSCI 而言，就是利用多年的數據積累發揮向用戶推優薦優的作用，突出檢索工具的角色，加強統計分析的功能，為作者和讀者的閱讀與投稿提供依據，為科研管理提供信息服務，為優秀學術成果的傳承和傳播提供檢索的便利，為辦刊提供有益的建言，能夠利用豐富的數據回答諸如期刊在學科中處於什麼水平、發展趨勢如何、讀者關注什麼、與同類優秀期刊相比差距在何處、該怎麼獲得優質稿源、怎麼找到學科內的優秀作者等等刊物真正關注的問題。畢竟，作為期刊評價工具，它的主要職責就是讓學術成果的價值得到後人不斷地解讀和引用，能很方便地被他人獲取並用於相關的研究中，使其在更大範圍內得到學術同行的認可。如果 CSSCI 能協助實現這一目標，就是它最大的價值體現。

第二，走出“定量＝客觀”的誤區。客觀的數據不等於用在評價上也一定是客觀的。文獻的相互引用是一種複雜的思維過程，索引加工人員是難以深入考察引用行為間的內在聯繫的。引用行為受許多主觀因素的影響，包括引用文獻的可獲得性，論文發表的時間、語種、所在學科專業、引用者的檢索能力、二次出版物報導等因素。而刊物排名和影響因子的大小也很容易受到諸如期刊載文量的急劇改變、刊名改變、關鍵詞不確切等因素的影響。文獻計量學所揭示的文獻分佈規律是期刊在百花齊放的發展過程中自然形成的，但在不良的學術生態和行政干預下，高學術價值論文向少量刊物集中的規律有可能受到人為因素的扭曲，使得刊物質量的升降與好稿源被切斷與否發生了某種聯繫。而好稿子的發文空間受到擠壓所催生出的種種弊端，如果不加以控制亦會使數據失真，在一定程度上反過來影響數據的客觀性。因此，文獻計量學家們很早就文獻計量工具有於評價時指出，只有當人們對文獻計量學指標的自身缺陷有了足夠的認識時，也只有在這類指標與其他更多的定性評價信息相融合時，文獻計量學指標才能成為科研績效評價的有效工具。

第三，不要將高引用與高質量完全等同起來。被引次數上的微小差別，不能說明論文質量上的高低。文獻被引用反映的是某種影響力，高影響力可能與高質量有關，但不能一概而論。在一些案例中，被引用越多，也許越象徵著一種否定的評價。不常被引用的文獻亦不能認為不重要。要解釋某一成果的學術影響力本身就是很複雜的，沒有同行的介入，沒有對內容的認真分析，是很難得出質的判斷的。

第四，正確對待評價後的排名。評價似乎永遠與排名、身份認同糾纏在一起。對排名的熱情是人類的心理偏好之一，排名在不同利益主體間產生競爭有積極意義，也有消極意義。<sup>⑤</sup>在現行的評價體制下，只要存在對排名的需求，對排名研究的熱潮就不會降溫，期刊“曬排行”、“秀榜單”還將長期存在。要刊物不追求轉載率、文摘量、引用率等仍是困難的，畢竟在目前情況下這關乎刊物的生命，但重要的是正確對待排名結果。所有的排行榜都是動態的，與其提心吊膽地關注“會不會掉下去”，為保持或提升排名而傾力攻關、買論文，甚至不惜冒學術不端之險“做數據”，不如把精力放到內涵的提升。“看淡”影響因子需要自信和勇氣，需要跳出“怪圈”，看穿影響力和“排座次”這個表面的東西及其所帶來的負面效果，始終如一地走“好文章——好刊物——好評價”的路線，始終扣住需要解決的社會實踐問題，依照辦刊本身的規律潛心做好組稿、審稿、選題策劃、規範化等工作，而不是反其道而行之（先把評價需要的數據“做”上去，解決“入榜”問題後再吸收好稿源）。對於文獻計量工作者和評價工具的研製者來說，要更科學地設置指標體系，盡可能將定性和定量標準科學地結合起來，建立起排名研究的質量監督機制，

引導人們科學、理性地瞭解產生排名的環境和現狀，而不是一方面抱怨評價體制的不公，一方面仍在繼續強化期刊的等級、身份和門戶。對於科研管理部門來說，重心能否放在協助破“名”、破“利”、破“身份固化”方面進行制度設計，為避免評價簡單化、功利化、行政化而盡一己之力。既然高校已經開始取消“985”、“211”院校、“一本”、“二本”等等名分，期刊能否也走出為“名”所累、為“評”所動的怪圈？能否給刊物“鬆”名次之“綁”，“減”排行之“壓”，使之無包袱地致力於科學發現、知識經驗的交流、對學科建設的支持，對學術創新的貢獻，對人才成長的支撐？

- 
- ①程磊：《中國期刊的影響因子神話：從2.16到12.41！》，《生物探索》，<http://m.biodiscover.com/news/research/176131.html>
- ②董少校：《影響因子神話亟待打破》，北京：《中國教育報》，2017年2月27日。
- ③袁培國、吳向東、馬曉軍：《論引文索引數據用作評價工具的科學性和局限性》，合肥：《學術界》，2009年第3期。
- ④舊金山科研評價宣言小組著：《〈關於科研評價的舊金山宣言〉——讓科研評價更加科學》，李宏、王建芳編譯，北京：《圖書情報工作動態》，2013年第5期。
- ⑤《用萊頓宣言進行科研評價》(The Leiden Manifesto for research metrics)，<http://www.wtoutiao.com/p/1f43SCO.html>
- ⑥姜春林、梁若愚、田文霞：《SSCI期刊分佈及其對我國社會科學評價的影響》，武漢：《科技進步與對策》，2008年第1期。
- ⑦王文軍：《分學科評價：綜合性學術期刊評價的合理路徑——以教育部“名刊工程”入選綜合性學報為例》，南京：《南京大學學報》，2011年第3期。
- ⑧⑫覃紅霞、張瑞菁：《SSCI與高校人文社會科學學術評價之反思》，武漢：《高等教育研究》，2008年第3期。
- ⑨郝若揚：《如何測度學科核心作者》，北京：《中國社會科學報》，2016年9月20日，第1版。
- ⑩袁培國：《中文社會科學引文索引的研究評價作用》，太原：《山西大學學報》，2003年第1期。
- ⑪ Thomson Reuters, Preserving the Integrity of The Journal Impact Factor Guidelines from the Scientific business of Thomson Reuters. <http://community.thomsonreuters.com/t5/Citation-Impact-Center/Preserving-the-Integrity-of-The-Journal-Impact-Factor-Guidelines/ba-p/1218>
- ⑬喬新生：《取消“A類期刊”恰逢其時》，北京：《中國科學報》，2016年5月26日。
- ⑭段青：《核心期刊評價的基礎理論研究》，長春：《情報科學》，2009年第1期。
- ⑮李俠：《排名如何更科學》，北京：《光明日報》，2016年6月24日。

作者簡介：沈固朝，南京大學信息管理學院教授，南京大學中國社會科學研究評價中心副主任。南京 210093

[責任編輯 劉澤生]