

新聞剪影

專題報導

科技短訊

科技政策

科技新知

學人及社團活動

求才專欄

合作機會

科技臺灣

歐盟科技報導

相關網站



本月專題報導



## 不同學門領域對於學術發表的態度 德國觀點系列26：引證率作為研究成果評估的基礎

作者：駐德國代表處科技組 現職：駐德國代表處科技組

文章來源：宏博基金會

發佈時間：2011.01.11

引證率作為研究成果評估的基礎 此方法的證明力與極限

原文：Forschungsbewertung auf der Basis von Zitierungen – Aussagekraft und Grenzen der Methode

載於：各學門的學術發表與研究成果評估（Publikationsverhalten in unterschiedlichen Disziplinen. Beiträge zur Beurteilung von Forschungsleistungen），132-155頁

作者：Werner Marx博士，德國斯圖加特馬克斯·普朗克固態研究所化學物理科學所資訊組主任。

本文原先由薩爾邦大學（University of Saarland）教授兼宏博基金會Feodor Lynen材料科學研究獎學金審核委員Frank Muecklich提議而撰，後併入此宏博基金會討論集。

### 摘要

採用以引證為基礎的學術成就評估指標需要目測能力和背景資訊。如果缺少這兩項能力，而採用倍受質疑的測量資料來補助研究和主導研究，那麼評估就不會達到預期效果。如果評估要避免錯誤或不公平，就必須深度認識現有資料庫與工具，同時也要清晰了解各種以不同引證為基礎的指標，以及此方法可能產生的錯誤與扭曲。本文將以現實為基礎，概括指出此方法的潛能與限制，同時

還提出引證資料在製作過程和分析解釋過程可能發生的危誤。本文所舉例子將予以匿名。針對一些重要的，但是常被誤解或濫用的指標，如「Journal Impact Factors」以及「Hirsch-Index」，本文將進一步闡釋。自然科學和其相關應用領域是本文主要討論目標。人文科學的專書引證和其使用上的受限將只大略述及。

## 引言

「研究學者或研究機構以及其研究結果是否達到優秀水準，這個問題的討論有時相當激烈。如果我們能更清晰地訂定指標的潛力及限制，消除錯誤的期待，並且注意某些應用原則，那麼所有的爭執就可以就事論事具體討論。研究學者習慣於享受自我評估的特別待遇，他們當然首先反對外界的評估。他們越懼怕失去這些特別待遇，就越強硬要求參與外界評估，這也是可以理解的。」

（見 Prof. Peter Weingart 於Bild der Wissenschaft之文，1993）[1]

目前在德國，引證數據在學術成就評估的程序中作為作品回響評價，也成為一項廣獲討論的議題[2-4]。自從引證索引（citation index）出現之後，人們就已經激烈地爭論影響力（Impact）和品質相互關連的程度，但是本文不討論這點。我們通常認為一篇作品如果越常被引用，就對該學門領域的後續發展越重要、越有利。在研究評估的架構下，我們希望透過這個引證率基礎找出特別具有創新力的研究學者或研究機構，繼而明確地針對目標提供補助。對學術成就評估而言，引證率已經成為一項附加的重要指標[5-7]。

引證（citation，或譯引文、引述）一詞表示一份作品在另一份作品之中被引用。通常進行引證的作者透過引證行動表示他對被引證的作品與其作者相當重視。但是光是引證數量並不就是真正的評價，也不是品質或作品價值的量表。尤其是後者的範疇屬另類的，根本很難予以量化。引證展現的是該作品在領域同仁之間的回響，而不是最終的評價。但是引證數量無疑地可以顯示其強點和弱點 不多些也不少些。

這個方法的基礎是一個名為「書目計量學」（Bibliometrie，或稱文獻計量學）的學門，也稱科學計量學（Scientometrics）。這些名稱也是這個半世紀以來密集發展的學門之同義名。然而採用引證率來衡量學術工作的效應只是眾多議題之一而已。最初，書目計量學的研究結果其實只是對圖書館員具有實際效用。但如今引證對研究理論和研究社會學，尤其是研究評估越來越重要：哪

位作者或是哪個實驗室目前最受重視？目前最熱門的研究題目是什麼？哪位年輕學者具有豐富潛力足以勝任研究高職？一位研究學者、一份作品和一所研究機構的引證在過去時間是如何發展的？

在學術上，一篇文章是否值得刊登，一項計畫是否夠資格獲得補助，以及甄選教職評審申請人的資格等，都是由資深學術同儕（Peers）審查評定的。這些「學術的衛士」應該透過同行評審制度的架構來保證研究水準。然而此項行動也附帶風險，評審人員可能無法辨識一項學術工作的意義，與評審領域不相干的因素可能摻入評審程序，而影響一個客觀公正的評定。況且審查小組常常只是由少數專家組成，這更加強其主觀態度。因此學界時時提出抗議，要求以公平獨立與量化指標進行學術評估。採用許多專家同仁的意見而評定的引證分析學，也就因此趨勢而興起。

毫無疑問的，研究成就永遠不能只依賴唯一的衡量數字來評審，而完全不顧及資深審查學者的意見來評定。唯有研究本身能提供評估其研究成就所需的內行專門知識。因此引證數量不應該是決定學術成就評估的單獨因素，專家同儕鑑定（Informed Peer Review）也必須參與才行。引證率絕對不能取代學者的審核，引證率只是一個附加的指標條件，可用來釐清相互矛盾之處：如果一位被公認為傑出的研究學者（或是自認如此）只具有較低的影響指數，他就應該更審慎地受到審查，其嚴謹程度與審查一位默默無聞，但是具有極高影響指數的研究學者相同。這種態度同樣也適用於審查研究小組和研究所。

由於公家資金越來越短缺，連政治界也要求一個盡量客觀易解的評估範疇，以便能有效補助那些具有潛力的研究項目。因此，德國研究界在過去幾年間越來越需面對以引證率顯示出版物銷售量的情況。德國研究評估基本上仍然依賴專家鑑定。引證資料卻已經被個案作為附加指標採用而列入下列各種審核程序：審核教授講座職位（教授職審核小組）、確保研究所水準（專業諮詢小組、管理委員會）、在國際比較展示研究情況（新聞報導）、期刊的最佳化（適宜的期刊）、頒獎（頒發學術獎或頭銜）。

使用者所期待的程序以及使用者不可缺的背景資訊闡釋仍然非常不一致。有關取用方法及個別工具的好處和壞處有時多不為人所知。因此說明方法的效能和風險，還有統一其製作與表達可說是急需之務。有意義的使用和濫用有時只有一線之差。相關的研究應該對此方法進行適當的探討，發展有意義的應用，以

期找到並訂立一個適當的標準。德國研究基金會（DFG）和科學委員會（Wissenschaftsrat）在這方面越來越多的活動可作為一種激勵。

#### A. 基礎

科學引證索引（Science Citation Index, SCI）

書目計量（Bibliometrics）和科學計量（Scientometrics）主要是經過Eugene Garfield和Derek de Solla Price兩位學者的力推而達到今天的重要性。De Solla Price於1963年發表其入門專著《Little Science – Big Science》，奠定了科學計量學為研究科學的科學[8]。目前書目計量學和科學計量學最著名的期刊為《Scientometrics》。Garfield是美國費城科學資訊研究所（Institute for Scientific Information, ISI，今名為Thomson Reuters）的創始人。1963年，他還建立了「科學引證索引」（Science Citation Index，以下簡稱SCI），成為評估與調查引證的資料基礎[9-10]。然而最初使用引證是為尋找重要作品，這可以省卻較為麻煩的關鍵字尋找法。因為一份作品若引證另一份作品，就表示這兩份作品極大可能是研究同一議題，或議題至少互相關連。

初始時期，引證是學生用來舉出其老師的作品的，到19世紀中葉才漸漸發展成今日的用法。引證可說是研究內涵的隱形知識網。引證分析的內容包括個別文章、某一研究小組、某一研究所、某一所大學、甚至某一所研究機構（如馬克斯·普朗克研究院，以下簡稱馬普研究院，MPI）的引證狀況，或是各聯邦之間在引證方面的比較。引證分析的目標主要是測量出版物、學者、研究所和期刊（見下文）的學術影響力。例如調查早期文品效用時間的發展，或是以圖表顯示引證聯繫網（如E. Garfield研發的軟體HistCite）。

今日，書目計量調查的資料基礎主要是Thomson Reuters集團（其前身為ISI）出版的引證索引（Citation Index）[11]，首推SCI，它收錄大約6000份自然科學及醫學期刊。雖然期刊數量只佔目前全球自然科學期刊的10%到20%，但是其包括的文章及其引證則可說概括全部。對核心學科，如化學、物理學、以及生物學來說，是綽綽有餘的，因為其顯示相當少的期刊佔有優勢（布萊德福定律，Bradford's Law；意指少部分的出版物可以滿足大部分讀者的需求，譯者註）。其他學門，如數學、資訊學、材料科學和材料技術、以及工程學門，SCI收錄的不那麼多，這當然有可能造成錯誤評估或過低評估。

除了SCI以外，還有「社會科學引證索引」（Social Science Citation Index，

簡稱SSCI) 以及「藝術人文科學引證索引」(Arts and Human Citation Index, 簡稱A&HCI) 列有約3000份期刊, 所以這三個索引總共收有約9000份目前刊行的期刊(以總時段計算, 則為16000份期刊)。被收入索引的範疇為編輯者的國際聲譽、其涵蓋學門的程度、是否定期出版等等因素, 其他範疇則屬主觀因素。一份未列於SCI的期刊並不一定就是不重要的期刊。目前已經另有「會議論文集引證索引-科學」(Conference Proceedings Citation Index, Science, 簡稱CPCI-S), 以及「會議論文集引證索引-社會與人文科學」(Conference Proceedings Citation Index, Social Science & Humanities, 簡稱CPCI-SSH), 兩者都是1992年開始發行的。

其他Thomson Reuters集團特別針對學術評估最佳化發行的的重要索引為[11]: Essential Science Indicators (ESI)、HighlyCited.com資料庫、具有期刊影響指數的Journal Citation Report (JCR) (見下文)、各國特有的High-Impact資料庫, 以及一些應需求而製成的個人報告(Personal Reports) 與機構報告(Institutional Reports)。

Thomson Reuters集團將這些引證索引透過自己的使用者平臺Web of Knowledge (WoK) 以Web of Science (WoS) 或是經由外界資料庫提供者, 如STN International (見Fachinformationszentrum Karlsruhe, 卡斯魯爾專門資訊中心) 提供學界使用[12]。對浩大的引證分析工程來說, STN高度的系統要求是重要的。這些引用文獻可以各按執照的涵蓋時期、單獨或是整體的需求被搜尋。在WoS搜尋SCI可回溯到1900年, 在STN搜尋則只能回溯到1974年。提供搜尋SCI的各個不同的搜尋系統(retrieval systems) 依需求不同而決定不同的功能, 例如引證統計必須時時更新, 就需要特別的功能。

目前已經有其他引證索引出現: Google Scholar、Elsevier開發的Scopus資料庫、化學摘要文獻資料庫(Cheical Abstracts)、Physical Review Online Archive (PROLA)、以及屬專門學門的CiteSeer(資訊學)等等, 它們涵蓋的期限大多只有十年, 但是在這期限內, 就各學門看來, 部分又比SCI所收錄的還完整。例如收於Scopus的出版物數量和引證數, 在應用學門如材料科學和材料技術方面常較高。然而大部分書目計量調查的資料來源仍然是Web of Science, 因為它包含有Thomson Reuters集團直接發行的引證索引。若搜尋工程浩大, 也可以購買SCI全部資料而就地發行供用(見下文)。

SCI所具資料跨學門，其所儲存的可查詢書目遠較其他索引為多，這個特點使它佔有一個極為特殊的地位：除了與傳統文獻資料庫所收的書目資料（名稱、作者、登錄期刊）與內容摘要之外，SCI還收有所有書目附帶的文獻指引（註解、參考處、引證），並以簡要形式和特定標準儲存。

這個發展的結果產生兩個性質完全不同的引證索引子單元

（Untereinheiten）：(1) 文獻資料庫（包括來源文獻（Source-Item，或譯[來源項[目]、引用文獻）、來源記錄Source-Report），此資料庫存有所有刊載於Thompson Reuters收錄的期刊文章，具備典型的文獻資料庫功能。此資料庫中的文章是引證文獻（Citing Papers）。(2) 被上述文章引用的被引證文獻資料庫（被引文章、被引項目、參考項目）。這個資料庫包含附加的文獻引證資料（註解、參考項目、引文），提供資料庫作為引證索引的功能。這兩個子單元收錄的作品絕大部分是相同的。

原則上，每一篇文章可以出現在這兩個子單元當中：如果一篇文章刊登於索引收有的核心期刊，它可作為執行引證的文章；如果核心期刊的任一作者至少被引證一次時，它就成為被引文章。從1900年（SCI於WoS下）到2008年期間，總共有四千萬來源文獻（source item）與六億參考項目（reference item），這表示一篇文章在這段期間內平均被引證15次（六億除以四千萬）。只有那些未以文獻形式被收錄，但在收錄的文章中被提及的作品，還有那些雖然被收錄，但從未被引證的文章，才各只出現在一個子單元而已。

這兩種資料庫造成兩種不同的搜尋方法：首先為一般搜尋法（General Search），可執行引證索引作為文獻資料庫的功能；另一為被引參考搜尋（Cited Reference Search），在參考文獻中瀏覽搜尋。這兩個不同的搜尋方法各具有其搜尋視窗，可在WoS網頁左上方按鍵叫出。

透過一般搜尋法視窗可以找到書目資料（文章名稱、作者、作者地址、期刊名稱、出版年等），1991年以後的出版物也可以用摘要的字詞以及相關詞語搜尋。搜尋只限收於索引的核心期刊的文章，即期刊文章（文章、書評、通訊Letters、註解、新聞、社論等），而不包括專書、書輯文章、博士論文、研究報告與會議報告（除非也登載於期刊）。這些出版物也因此不被列入學者或研究機構的出版清單。以下一為在WoS的一般搜尋模式尋找德國學者Max

Planck文章的視窗 [在此從略]；二為搜尋到的結果，即搜尋紀錄，其中包含被引證的九篇文章 [在此從略]。

Max Planck的文章之所以被搜尋到，是因為那些文章刊於《物理年鑒》（Annalen der Physik），該期刊是SCI在收錄年限內收有的約六千份期刊之一。它在此是執行引證的文章，引證其他九篇文章。兩篇被引證的文章又也是SCI的文獻。其他七篇是1900年以前發表的，因此未以文獻收錄於SCI（而以被引文獻儲存）。Max Planck本人的文章從1900年到現在（2008）又被其他23篇文章引證過，即見於文章後面附上的引用書目。他的文章成為被引文獻，而非引證文獻。下表顯示那23篇文章之一的搜尋結果 [在此從略]。

利用被引參考搜尋模式（Cited Reference Search）可以找到所有被引文獻（所有收於索引的核心期刊所有文章的引證文獻），並按照字母序列列出。但是被引文獻只以簡短形式呈現：首位作者、出版年、期刊號、以及首頁。期刊名稱則以被引寫法呈現，沒有統一的形式。參考資料被完整取用，不只限於索引收有的核心期刊文章而已。

在這個搜尋模式下可以找到被引專書、書輯文章、以及進行引證的作者所提出的引用文獻——即上文提到的1900年以來索引中的期刊收有的六億參考文獻。這個搜尋模式具有兩個重要的好處：一為可以搜尋到那些未收於來源項目的文章和專書引證，當然必須知道首位作者的名字才行。二是可以查到誤寫的引證。視窗上，來源項目用顏色標出，並且連結核心期刊的文獻。下表為Max Planck早期作品的部分搜尋結果 [在此從略]。

[搜尋結果顯示]物理年鑒中被正確引證的Planck文章用藍色標出，並且附加連結，Planck其他被引文章，包括他1900年的革命性發現（見表中被引證129次和75次的文章）只出現在引證中，不屬於SCI收錄的文獻。然而其執行引證的文章可經搜尋而標出（即SCI收錄的包括那129次加上75次的文章）。

說明：本文在此描述的是基本關連，尤其是WoS引證索引的結構（但也關係到其他資料庫），讀者最好自己從事搜尋，才能得到實際經驗！

目前透過這兩個模式搜尋出來的結果能夠予以分析。進階搜尋（Refine）的功能提供許多方法，透過不同的搜尋點將所需的資料範圍縮小，並且清除不必要的資料。分析功能則可以進行這些文獻的解析，例如出版物整體可以就其所屬

領域、出版年或期刊進行檢索（以按照年數列出）。最近才加上的引證報告（Citation Report）功能可完成出版物（研究學者、研究機構）的調查、製表、時間發展製圖等任務，同時也可尋到含有引證的作品與其引證內容（見下文）。

### 研究評估的指標

在評估學者的研究成就時，我們最喜歡看他們的發表成績。在自然科學方面，他們在SCI收錄的期刊發表的文章數已經成為其學術產量（Output）量化的標準了。但是一份學術出版物定義不足，通常也未區分著名期刊的長篇文章與短訊，光是出版數量實在不是最佳的指標。由於引證索引是檢索的基礎，對那些應用方面學科的學者就更不利了。他們的文章很可能並不刊載於SCI收錄的核心期刊上（或甚至登在某本專書中或報告中，而這些原本就不收於引證索引）。

因此，學界一直認為出版物應該就其意義和重要性來評估。最常受採用的標準是期刊影響指數（Journal Impact Factors，簡稱JIF），它決定作品的重要性。然而這也非常值得爭議，因為一份期刊的期刊影響指數對其所刊載的各篇文章的影響指數只能提供低微的資訊（見下文）。有時，高影響指數期刊（High-Impact Journals）、發行甚廣的跨領域期刊如《Science》和《Nature》、或是著名的通訊期刊如《Physical Review》等刊載的文章會被作為評比程序的品質範疇。但是這個方法也因同樣的情況而有爭議之處。既然已經運用引證法了，就應該由每篇文章的被引證數量來測量該文的影響效果。

### 引證的規模與分佈情況

個別文章出版以後的影響效果從絲毫無回響到被引證達上百次或甚至上千次都有。有史以來，被引證最多次的作品是1951年由O.H. Lowry撰寫的文章，該文到目前為止已經被引證三十萬次[13]。那篇文章的主題是蛋白質質量化鑑定方法。然而這篇文章的效用只是那一時的單一事件。繼之的是被引證一萬多次的幾十篇文章，多出自生化和分子生物學。1990年共有二十篇文章各被引證一萬多次，一千四百多篇文章各被引證一千次以上。大部分期刊只被引證幾次而已（一至十次），有更少的期刊被引證多些（一百到一千次）。SCI從1945年到1988年收錄的期刊中有50%多只被引證過一次（見下文）。所以可說大部分引證集中於為數相當少的文章。我們常常採用的平均值其實曲解了引證極為不均



的分佈狀況。

這樣的分佈情況不符合具有最大近平均值數的高氏分佈定律，分佈極度歪斜，極度不均勻，算術平均值與最常值相差甚遠。典型的研究機構的文章被引證的分佈情況如下：1%最常被引證的文章就已經達到總效果的20%，10%最常被引證的文章已經取得50%以上的效果，50%最常被引證的文章等於取得所有文章總效果的90%以上。這樣不均的分佈也見於期刊引證：WoS所包含的九千份期刊當中，有一千份期刊就涵蓋引證量的75%。這種歪斜不均的分佈也稱為「帕雷多分佈」（Pareto-Distribution）[14]。大約一百年前，意大利的社會學兼經濟學家Vilfredo Pareto調查國民財富的分佈情況，他發現20%的家族擁有80%的財富（即所謂的80/20-規則）。

### 各篇文章的影響發展

除了引證次數和其分佈之外，其影響效果在時間上的發展趨勢也具有說明力。如果我們查某一篇文章在某一時段內被引證的情況，就可取得其引證歷史（Citation History），這可視為股票市場或銷售情況的曲線圖。我們可以看到，通常學術文章在出版的同年就受到同仁的重視，其被引證次數急劇增加。大約三年以後，引證率達到最高點。那以後，這些文章通常被新出的文章排擠，其影響效果就漸漸降低。然而也有一部分文章在幾十年以後還常被引證。此現象被稱為學術文獻的老化，或用放射性術語來說，學術文獻半衰期。

至此所述僅是許多文章平均效果的時間性發展。個別文章的效果發展情況有時與此模式相差很遠。有些文章在許多年之後，甚且幾十年以後才達到其效果頂峰。有些早期出版的文章在幾十年以後才受到注意，其特別重要的益處也就這麼晚才被發現。

早期馬普研究院一年的期刊文章總體可顯示其影響效果在時間上的發展情況。圖一[在此從略]的曲線表示收於SCI的馬普研究院1975年的兩千份作品。採用這年的文章是因為其具有較長的觀察期，可保證那些文章在這段期間應該會被引證。然而我們討論馬普研究院一年之間提出的作品時，是只挑廣泛自然科學領域方面的文章，取其平均值為基準。

如果我們（就該年度作品）分別觀察各學門領域的話，就發現各發展曲線形狀類似。雖然各學門有不同的頂尖效果時間點，但是其最大效果狀況不相上下。

只是其效果消失的下落曲線陡斜程度不同：生物科學下落得最快，數學最慢，物理和化學則居其中。

長期以來，資訊激烈大幅膨脹，此導致學術界極度注重當下時效性。研究學者忽略以前的作品可見於下列情況：在某一段時期廣受引證的文章，其出版時日短，當時間陣陣過去時，那些文章就越來越少被引證。時間刻度尺很快就變成品質刻度尺，而使以前的文章或被錯誤評估為拙劣，或甚至全被遺忘。

## B. 應用

### 個別研究學者的比較

引證分析通常指的並不是單獨作品的分析，而是作品整體的調查。如同單獨作品的引證歷史（Citation History），我們可以藉引證分析來檢索某位研究學者所有作品總影響效果在時間上的發展，並且以圖表顯示出來。影響力的時間曲線（見WoS Citation Report）顯示出每年引證次數，所涵蓋期限溯至作品出版年。要注意的是：時間刻度尺以執行引證的文章出版年為準，而不是取作者出版該作品的時間。有關其作品量選取的是累積值，即不只是那引證的出版物，還包括到引證那年出版的所有作品。涉及引證量時，則不採取累積值：指數只表示出在引證的那單一年間的引證次數而已，而不是累積到那一年的總引證次數。

圖二[在此從略]顯示的是五位不具名學者的影響指數發展曲線圖。此圖包括從以前即累積的被引證次數到當下的被引證次數，其作品的影響效果也隨著學者的年歲持續直線增加。如果該學者從某一時期開始較少出版，或是其出版物的影響指數逐漸降低，或是退休，那麼其曲線就下降。由於引證行為通常是一段時間以後發生，其某一特定時段（如當下）的影響指數大多是建立在該學者以前的作品。其目前的出版物尚少有機會被引證，其被引率不值得一提。一位學者的影響指數曲線可被視為其股票市價或其銷售曲線（Sales Curve）。如果要以一個數字表達學者的影響力，可以取曲線所顯示的各數值的平均值，此法不受限於年齡，而且可以標出每年被引證次數，此引證次數高於其平均值，即提高其影響效力。

[圖二]各曲線數值的意義：以名為BLAU的曲線於2005年被引證500次為例，此表BLAU曲線代表的學者在2005年一年內總共被引證500次。引證涵蓋該學者到2005年底的所有出版物，但是執行引證的出版物只限於那些在2005年出版的

期刊文章（而不是累積到2005年的引證）。現在解釋一下各曲線的發展變化：BLAU曲線的學者在三年內持續提高其影響效果。GRUEN曲線的學者在其學術事業的前半段影響效果極端大幅增加，然後到目前為止持續下降。ROT曲線代表的學者可說是此圖中的明星級學者，他的影響效果從開始到目前一直持續增漲，途中毫無絲毫掉落。CYAN曲線的學者也是一樣，但是該學者初次出版遠較為晚。MANGENTA曲線則代表那些即使長期發表，但只達到低微影響效果的學者群。

除了整體影響效果之外，此圖還顯示出其流程（即某些文章引起的影響指數增高或掉落的時間點）以及目前的趨勢，我們可就對未來的發展作保守的預測。為了進行較適當的比較，我們可以將所有曲線移置於一個觀點而與學者的學術年齡相對比較（即從事學術活動的年數）。這樣，就見ROT曲線居最高位，次之為BLAU，繼而明顯降落到MANGENTA為最低，其間的CYAN曲線則代表年輕、充滿潛力的後進學者。

說明：無可避免的是年齡較長的學者會遭到被低估的命運，因為在前幾十年，引證活動很明顯地遠較為少。有關進一步資料（包括影響指數曲線升高的確切數學表達方式）可參閱下文「Using time dependent citation rates (sales curves) for comparing scientific impacts」：<http://arxiv.org/abs/physics/0611284>

#### 單獨學者的年期圖表

另一個有意義的層面是一位學者的創作時間點，即其被引證的文章的出版時間，而不是引證其作品的文章的出版時間（如圖二、圖三所示）。這種顯示方法猶如標示出產年的葡萄酒釀製年份圖表（Vintage Diagramm），可稱為引證年期圖表。就如酒釀製年份不同而有不同的量和質（價錢），一位學者也因學術產量（Output）的變換而達到不同的影響效果（Impact）。圖四[在此從略]顯示ROT曲線學者歷年的影響指數的柱形圖表。年代越近，不只是較新年期的柱子就越高，連較早期年代的柱子也有稍微程度的增高。

論柱子高度的意義：ROT學者1994年的出版物到現在被引證次數達700次（暗色），到2000年被引證約500次（淺色）。深色柱子表示到現在（2005年）的引證次數增高發展，淺色柱子顯示到2000年底的被引證發展。這兩個不同時間

點的數據資料充分顯示出被引證的發展有一段初始時程，時間上有些延緩，總效果影響力的絕大部分是在最近五年內產生的。此圖中顯示的數值往現在方向持續降低，這並不表示該學者的影響力降低，而通常是因為引證起始時間延緩而造成的。與圖二和圖三不同的是，圖四無法對未來發展做出任何預測。

### 各研究所的比較

以影響力曲線為基礎的引證分析可以毫無困難地從單獨研究學者擴展到研究小組和整個研究所。圖五[在此從略] 仿圖二圖三 明確顯示三個最大研究所每年的引證數量，這三個被選出的研究所具備上述各圖中五位學者的研究領域。與圖二至圖四不同的是，圖五因技術上原因未採用被引作品，而改採進行引證的出版物作為評估影響力的標準。一份進行引證的出版物可能引用一個以上的文獻，因為一個研究所的兩篇或其以上的作品同時被一份作品引證。但是在影響力比較範圍內，這個現象並不會造成什麼改變。

研究所的規模，就其學術人員數（博士生、博士後、固定聘用工作人員）而論，在過去三十年來大約沒什麼變動。如果我們將此規格化，將每位研究學者受時間因素限制的每年（被）引證次數圖表化，可以發現有另外一種上升曲線。研究所的時間發展曲線大致類似單獨學者的趨向發展曲線。

一個研究所的出版物構成一個較大作品整體，其引證曲線則顯得較不平衡。例如ROT研究所1%被引證最多次的作品就已經涵蓋所有引證的20%了，10%最常被引證的作品包含引證量的一半以上，50%最常被引證的作品甚至佔引證總數的97%。這所研究所約15000份出版物當中，只有一份被引證一千次以上。不少出版物被引證幾百次。大部分的出版物不是只被引證一次，就是只有幾次而已。三分之一的出版物到現在沒被引證過，連其作者本身也未引證自己的作品。然而我們必須指出：這些出版物中有許多是新近才出版的，要達到其被引證或甚至常被引證為時尚早。

### 一個研究所的影響力趨勢

在研究所評估方面，以下的列圖法證實有效：我們先選出一所研究所在某些年的出版物（例如該所在五年內的出版物），然後計算出在那段期間內的被引證作品數（或進行引證的作品）。此法可以追蹤到最近的發展。圖六[在此從略]顯示圖五的ROT研究所在作品出版後五年間每年的出版物影響力的平均值。圖六也可以放棄考量被引證作品，而取進行引證的作品作為製圖的基礎。

以五年為時段作為評估影響力時間，起碼可以顧及部分作品統計學上來說需要的兩三年時間（見圖一），同時也可使曲線均勻。這個顯示影響力的方法比WoS Citation Report所提供的曲線更為合適。當然除了上述的五年為比較期限之外，我們也可以採用較短或較長的時段作為計算出版物和引證的基礎，然後每次延後一年：例如每年的出版物和該年以及以後兩年的引證。這樣，我們可以將觀察的期限朝向現在推展兩年。

限制在最初幾年內的時間影響力平均值可以作為基礎，來預算以引證呈現影響力的趨勢。例如一個持續下降的趨勢應該要視為警訊，應該受到批判質疑。影響力的絕對值對趨勢沒什麼證明力，因為該值因學門領域的不同而有所差異。如果我們不採用進行引證的作品，而改用被引證次數，則會產生一項有趣的比較：圖例中的研究所在最初五年期間，每份作品平均被引證六次。與之比較：化學和物理學門在同樣時間內，每份作品平均被引證四次（見Essential Science Indicators ESI-10/2008）。

#### Hirsch-Index作為衡量指數

不久前，有一種以引證為基礎的影響指標出現，可用來評估學者個人。該指標按其發明人名為Hirsch-Index（亦名為Hirsch-Factor，h-index或h-number）[15]。Hirsch-Index有一個優點，即任何人可上網到Web of Science就可迅速又簡易地進行搜尋檢索。我們只需找出要評估的作品，將那些作品按引證率分列出來（WoS: Sort by Times Cited）。Hirsch-Index就是符合引證次數的排序數字。例如某學者的Hirsch-Index為44，表示該作者著有44篇作品，每篇最少被引證44次。Hirsch-Index不但顯示出版量（Output）（作品數量），也指出影響力（引證次數）。但是它並不表明整個影響範圍，而只是限制於較高的被引證部分（所謂的Hirsch-Index收有的核心期刊）。這樣可以防止單一的高引證率扭曲真相。雖然這樣的一個指標有些地方仍顯粗略，但它已成為衡量學術成就持續發出影響效力的指標了。

不同學門的作品具有不同的平均引證率，差異很大，這也影響到以Hirsch-Index來做比較的評估方法。在自然科學與技術領域內，平均值可能相差十倍，因此未作適當的規格化時，只能就同一領域內的學者進行比較。況且，Hirsch-Index還依著述年齡而定。因此，h-index除以學術事業年數（比如從完

成博士論文到當下)，而得到一個m-index值，這個m-index結果可以對同一領域內不同年齡的學者進行比較。到目前為止，h-index已經經過多次修改[16]。其某些版本（如g-index、h(2)-index）的目標是更加注重被引率高的作品。其他的修改版（如r-index、ar-index）則以Hirsch核心作品不同著重點的引證率為基礎。

與其他用引證率為基礎的指標比較之下，Hirsch-Index顯然較少出現搜索錯誤。但是Hirsch-Index避免錯誤的條件是作者名字必須完整，附加的地址資料能確定該位作者身份。然而許多Web of Science的使用者不知道作者全名，或是根本沒意識到欠缺這些資料可能引起的困難。再者，我們不應該相信唯一的一個影響指數就足以評價一位學者的學術成就。除了採用引證率之外，還應該顧及以下幾點：該學者的研究領域（屬主流領域或是小型領域）、其高引證率作品的類型（文章、書評、通訊）、身為首位作者的文章數量、未被引證文章所佔比例、影響力的時間曲線（見上文）、其作品以期刊和/或學門領域為準的引證率（見下文）。

#### 期刊影響指數（Journal Impact Factor）

從1975年以來，Thomson Reuters集團出版的。「期刊引證報告」（Journal Citation Reports，簡稱JCR）就將期刊按照其期刊影響指數（Journal Impact Factor，簡稱JIF）的高低排序，每年都予以更新[17]。如果談到影響指數（Impact Factor），通常都是指期刊影響指數，Thomson Reuters出版集團從未設定其他的定義。起初，期刊影響指數（JIF）只是用來評估期刊整體。它被用來排名，主要是因為一篇文章或一份期刊如果越常被引證，就會越被同仁看重[18]。「期刊引證報告」（JCR）具有兩種版本：「科學版」（Science Edition）包含的刊物為科學引證索引（Science Citation Index, SCI）收有的約6000份期刊，「社會科學版」（Social Science Edition）則備有社會科學引證索引（Social Science Citation Index, SSCI）中的1700份期刊。「期刊引證報告」（JCR）不包含所有現在發行的期刊，而只精選一部分，大約是學術期刊總量的10%到20%。

要確定期刊影響指數需先計算一份期刊在某一特定兩年時段所刊載的文章及至一年以後的引證量。其方法是，引證次數除以那兩年內可被引項目數目（citable items，即文章、概貌文章、通告短文）。例如我們要算出2007年的期刊影響指數就必須首先確定2005年和2006年可引證的文章總數。然後再調查

2007年出版物參考文獻中提及2005年和2006年可供引證文章的期刊名稱（期刊刊名）出現多少次。最後所計得的文章數和其被引證次數就可算出商數值。簡單說法：一份期刊的期刊影響指數是這份期刊的文章在出版後兩年間的每年平均被引證次數。

除了期刊影響指數之外，「期刊引證報告」（JCR）還提供三個有趣的期刊引證行為參數，Thomson Reuters將其定義如下：

- 引證半衰期（Citing Half-Life）：一份期刊從所選年度往前推算，該年引用各年度期刊數達到總引用數的一半所需的時間年數。此所處理的是一份期刊中所有文章備有的參考文獻。
- 被引證半衰期（Cited Half-Life）：一份期刊從所選年度往前推算，在該年被引用數達到全部被引用數的一半所需的年數。此所涉及的是一份期刊被引證而列入另一期刊中文章的參考文獻。

引證半衰期顯示出一份期刊諸位作者的引證活動，反映出他們在引證時如何做選擇。引證者的年齡分佈顯示被引文章的時效性（Actuality），一份期刊的文章就建立於此時效性根基上。被引用半衰期告知的是一份期刊的文章被其他文章引證的情況，反映出這份期刊的文章留在讀者記憶中的時間有多長。引證半衰期和被引證半衰期都呈現影響力降落的情況，因為同樣的作者或文章引證其他文章，也被其他作者引用。引證半衰期（就某一特定期刊的文章來看）是主動的；被引證半衰期則是被動的。

- 立即指數（Immediacy Index，簡稱II）：一份期刊某年度的文章數除以該期刊該年度可供引用的文章數。

計算某一期刊的立即指數（II），舉II-2007為例，此表示該期刊於2007年的引證總數除以2007年文章總數所得到的數值。立即指數（II）可以視為期刊影響指數（JIF）的一種特殊表達方式，它主要並不是測量整個影響效果，而是測量學界同仁的反應速度。

很遺憾的是，採用刊登學者文章的期刊之期刊影響指數來評價學者的學術成就已經廣為通行。然而引證時間點不均勻的分佈使一份期刊的期刊影響指數只建立在少數文章上，對期刊每一篇文章其實無法作出正確的評估。期刊影響指數平均值和最常被引數值常常相差甚遠。換句話說：一篇文章刊載於高期刊影響

指數的期刊上，並不代表這篇文章就具有高的影響效果。即使在高期刊影響指數的期刊，如《Nature》和《Science》，絕大多數的文章並未達到非比尋常的高影響效果。

因此，我們現在面對的是一個單向的關係：文章的影響力多多少少會提升其發表期刊的聲譽，但是高期刊影響指數的期刊登載的文章並未導致這些文章明顯地更常被引證。所以某學者文章被引率和刊載期刊的期刊影響指數之間的關係相當微弱，這也不足為奇[19-22]。期刊影響指數對評估學者文章的影響力完全不適用。雖然實際計算文章被引次數需要更多的時間，而且易犯其他的錯誤，但是要評估學者作品的影響效果仍然以算出其文章事實上被引證多少次為要。

下列僅略述採用和闡釋期刊影響指數時必須注意的事項：

1. 單獨作品的影響力和所刊載期刊的期刊影響指數之間的關係微弱！
2. 不同領域的期刊不能直接就用其期刊影響指數互相比較。
3. 充滿動力的學門領域，如生命科學，能較快累積引證數，期刊影響指數的公式使這些學門受到優惠。
4. 其他單獨學門，如數學，大部分發表於專書上，在期刊影響指數的計算上就顯不利了。
5. 不同形式的期刊（如通訊、評論）無法直接就其期刊影響指數互相比較。
6. 每年出版極少期的期刊由於其計算時段短，所以其期刊影響指數非常不穩。
7. 跨學門的期刊，如《Nature》和《Science》，得到高的評價，其實是因為它們刊載許多不符傳統典型的文章。
8. 期刊改名（期刊頭銜更改）對期刊影響指數的影響甚至還持續到改名以後許多年。
9. 最新的期刊影響指數是評估過去兩年的發表物，而不是期刊當下年度的期份。
10. 期刊影響指數不提供期刊文章長期的永續性。

結論：期刊影響指數（具有上述的限制）評估的是期刊整體，因此用它來評判個別文章，繼而評估研究成果是絕對不適宜的！

學者不應該只是一味地向期刊影響指數高的期刊投稿。期刊影響指數較低的期刊絕對也有可能成為提供資訊給學術同仁的較佳選項。只要所發表的文章被相關讀者閱讀到，也引起這些讀者的興趣，那麼該篇文章就會被引證。



## C. 闡釋

### 曲解失真的可能性 概論

引證動作的理想是記錄所有的知識借用活動。原則上，所有的出版物都應該引證從前的作品，這些有的是繼續討論該出版物，有的是成爲該出版物的基礎。但是每一位學者都經歷過，採用引證的選擇過程或引證是否完整，常常不是完全建立於專業的考量的。學術的嚴謹性雖然督促學者正確紀錄知識借用的過程，但是引證也是一項社會行動，不完全只是以專業考量為範疇，而是也附帶某種程度的隨意取決。許多年以來，引證理想可能碰到到什麼樣的偏差，一直是文獻計量學調查的議題[23-24]。以下簡短列出造成引證曲解最常見的原因：

1. 學者引證自己的文章或建立所謂的引證聯盟，即一小群學者互相引證，不為群外的人知道。
2. 學者通常在可引證對象中選出引證者時，偏愛那些具有高知名度的學者，這樣就更加提高那些學者的名望。
3. 在可引證的衆多文獻當中，那些與作者論調相同的作品較會被引證，因為這樣可更加強自己的論證力。
4. 有些被引證的文章之所以會被引證是因為它們包含錯誤，立論與結論不實（即所謂的負面引證）。
5. 參考文獻有可能是隨意取決的，也有可能是因為疏忽或缺乏資訊而顯得不完整。
6. 除了典型的引證之外，還有其他方式表示引證活動，例如在文中以括號寫出作者名字和縮寫，或是在謝詞中和私人訊息中提出。
7. 概貌文章（如書評）和方法論文章（如化學分析法和合成法）被引證的可能性不尋常的高。

這些因素毫無疑問地可以扭曲以引證為基礎的評估真相。這項曲解造成高估，也造成低估。因此在調查許多文章作品的時候，採取計算平均值；只有在某些個案中才標明錯誤的來源。許多上述的因素也常被過量化高估。在此，我們僅就自我引證（本身引證）為例進一步討論：

SCI的搜尋系統提供包含自我引證與不包含自我引證的計算方法。我們發現，這類引證通常只佔10%以下，很少超過15%到20%以上。只有發表量尚少的年輕學者自我引證的比例有時高些。由於只有很少自我引證的例子顯示該學者的

影響力產生明顯的改變，因此我們通常也就不排除自我引證，標準引證計算方法包含自我引證，但是並不會降低引證次數的證明力。

#### 依各學門領域為準的引證率

足以大幅影響引證分析結果而造成曲解的還有其他因素。首推以學門領域為準的引證率（但是若透過適當的規格化還是可以採用的）。我們觀察自然科學和技術學科領域內各出版物的引證率，可以發現達到十倍的差異！其原因是各學門具有互不相同的引證慣例，或確切地說：出版物附錄的平均參考文獻數量各有不同。

數學領域的文章平均引證十篇作品，化學和物理方面的引證作品數為二十到三十，分子生物學則有五十到六十篇的參考文獻。按照這個比較，在分子生物領域的文章被引證的機率就比其他學門高。因此，如果忽略各學門不同的引證慣例，而直接作跨學門的比較，所得結果是不可靠的。表一摘自Essential Science Indicators的Baselines所包含的過去十年時間Thomson Reuters集團認定的22個學門的平均引證率[25-26]。

#### ISI認定的學科 / 1998-2008平均引證率

所有的學科	9.7
農業科學	6.6
生物生化學	16.12
化學	9.49
臨床醫學	11.76
電腦資訊學	3.03
經濟&商業	5.02
工程師學門	3.03
環境科學/環保科學	9.5
地球學	8.48
免疫學	20.58
材料科學	5.56
數學	2.99
微生物學	14.77
分子生物學/基因學	24.75

多國學科 ( Multidisciplinary ) 4.09

腦神經科學/行爲科學 17.82

藥學&毒理學 10.74

物理學 7.99

植物動物學 6.90

心理治療/心理學 9.68

社會科學，一般 4.06

太空科學 12.91

表一：1998-2008 年間，Thomson Reuters出版集團認定的22個研究範圍 ( Research Fields ) 的平均引證率。資料來源：Essential Sciences Indicators ( ESI-10/2008 ) 的Baselines。

除了數學以外，工程師學門和技術科學也甚少被引證。材料科學和材料技術方面的平均引證率低，其原因是這個領域以應用和技術為基礎。一位在材料科學領域工作的學者，其引證活動有別於一位在傳統科學領域工作的化學家或是物理學家。即使是材料科學有極大部分牽涉到化學或物理，我們也不能將材料科學專家和化學家或物理學家直接互相比較。另外，材料學科學家的出版物也常不見於WoS的核心出版物。

由於ESI的比較資料只做粗略區分 ( 只按期刊種類區分，而不看個別文章 )，以其為基礎的跨學門比較也就理所當然地不是很精準。除此之外，當下熱門領域的文章 ( 即所謂的主流研究或流行研究 ) 平均起來多較常被引證 ( 見下文 )。這兩個因素聯合作用，就導致不同專業的研究學者或研究機構的出版物無法直接互相比較！

那些大量採用以引證為基礎的資料來評估研究的機構 ( 如歐洲荷蘭萊登大學的Centre for Science and Technology Studies，簡稱CWTS；以及德國比勒費德 ( Bielefeld ) 大學的Institut fuer Wissenschafts- und Technikforschung，簡稱IWT ) 本身就儲存有科學引證索引 ( SCI ) 資料，可以就近在自己找出的參考參數基礎上進行規格化。首先取學者發表的期刊或學者所屬機構作為參考數值。除了要調查的期刊之外，其出版年與出版類型 ( 文章或書評等 ) 也列入考量。經此得來的引證率Journal Citation Score Mean ( JCSm ) 可以被認定為期待數值，與實際的每篇文章引證率 ( Citations per Paper，簡稱C/P ) 比較，就得到一個以期刊為基準的相對引證率 ( CP/JCSm )。

這個以期刊為基準的相對引證率有一個缺點，即它不區分在高影響指數期刊和低影響指數期刊發表的文章的參考數值。換句話說：它不討論一個高的CP/JCSm值是如何產生的，是以名望高的期刊為基準，或是以名聲較低的期刊為基準，但兩者卻是有不同的要求。因此，以學門為基準的平均引證率（Field Citation Score Mean，簡稱FCSm）被認為是比較適當的參考數值，而常被採用，並且與實際引證率被定義為以學門為基準的相對引證率（CP/FCSm）。一個專門領域的定義是按照Thomson Reuters集團對該領域期刊的歸類而定，而不一定與受評估的文章總體所具備的期刊版面顯示的相同。

以期刊為基準或以學門為基準的相對平均引證率都有的主要缺點是，分類以期刊整體為基礎，而不是以期刊內個別文章的專門內容為出發點。況且，一些著名的期刊，如《Nature》和《Science》被歸類為多面學科（Multidisciplinary），這根本不適宜當作參考數值的基礎。然而由於缺乏其他方法，以學門為基準的相對平均引證率常被視為皇冠指標（Crown Indicator）而廣受運用。在此我們必須說明，一個唯一的指標很少能夠做到公平正確的歸類。文獻計量學的一項基本規則指出，唯有數個指標綜合起來才能提供充足的證明力。因此引證率絕對必須與其作為根據的作品數量值聯合起來討論才對。

研究社群規模的影響？

另一個值得思考的問題是，像化學或物理領域又分有許多專門學科，這些學科之間又有什麼不同。我們以2000年的物理期刊為例，探討其門下七個按照SCI定義的學科。由於該年度的出版數量差異為五倍，時常被提出的一個問題就是：研究範圍較廣的科學家比那些研究範圍較小，因而人數較少的學者更常被引證嗎？

較大的研究範圍確實是會產生較多能進行引證的專家同仁。但是他們提供可供引證的作品時，競爭壓力也較大。因此，引證文章和被引文章的比例並不受研究範圍規模影響，而永遠是同樣的。要注意的一點是，每篇文章的平均引證文獻量不一定與研究範圍大小有關，而是各與每個學科的引證文化有關。

表二顯示物理學各學科的每篇文章平均引證數，最高為最低的1.5倍。如果考量到對排名的影響，這個數值算不低。就出版數來看，最多產的學科為最少的

五倍多，但是與其平均引證數卻無關連：應用物理學出版量最大，但是其平均引證數卻最低。反過來，那些較小的學科，如高能物理和電漿物理，卻達到較高的平均引證數。

表二：SCI認定的七門物理學科的平均引證數，調查以2000年的出版文章以及其在2000年1月到2006年4月的引證為基準。

ISI認定的物理各學科 / 2000年文章數 / 2000-2006被引次數 / 每篇文章被引次數

物理，應用 24367 / 171611 / 7.04

物理，原子，分子&化學 11097 / 111839 / 10.08

物理，固態體 21332 / 155676 / 7.30

物理，流體&電漿 4764 / 43507 / 9.13

物理，數學 6828 / 47913 / 7.02

物理，核子 5799 / 46190 / 7.97

物理，粒子&場 7693 / 81216 / 10.56

物理學科總數 9.05\*

\* ESI ( 1996-2006 ) 平均引證量，搜尋日期：2006年4月7日

主流研究議題佔優勢嗎？

一個值得探討的問題是，當下特別受重視的研究（主流研究，有時也暱稱為流行研究）是否比那些屬於所謂的隙縫範圍的研究更具有影響力。我們在尋求這個問題的解答時，鍵入一些流行的範圍和一些較不為人矚目的範圍名稱和摘要包含的關鍵字。表三顯示主流研究比物理和化學的出版文章可高達三倍。我們必須正視這個主流議題文章比隙縫議題文章大為有利的引證優勢所發出的警訊。因為這兩類研究都有其存在價值：沒有一個國家能夠忽略目前特別受到注視的議題。一方面，這有違對現代學術的認知，同時也意味著，擱置具有協同作用的潛力而不加以利用。另一方面，基礎研究絕對需要一個特定的空間以開發新的研究範圍，沒有這項行動就沒有進步。這兩類研究不能互相對抗，互相摧殘。

SCI認定的主流學科 / 2000年文章出版數 / 2000-2006年被引證次數 / 每篇文章平均被引數

nanotube? ( 篇名，摘要 ) 1152 / 37419 / 32.48

dark energy/matter ( 篇名, 摘要 ) 797 / 19792 / 24.38

string theory ( 篇名, 摘要 ) 644 / 13843 / 21.50

nano? ( 篇名, 摘要 ) 11924 / 209179 / 17.54

global warming ( 篇名, 摘要 ) 1803 / 27954 / 15.50

fullerene? ( 篇名, 摘要 ) 1500 / 18497 / 12.33

化學, 所有學科 10.45\*

物理, 所有學科 9.05\*

supercond? ( 篇名, 摘要 ) 6292 / 50231 / 7.98

supercond? ( 只有篇名 ) 2794 / 21544 / 7.71

low temperature? ( 只有篇名 ) 1505 / 10906 / 7.25

mineral? ( 只有篇名 ) 2536 / 18019 / 7.11

mechanic? ( 只有篇名 ) 4910 / 34838 / 7.10

high pressure ( 只有篇名 ) 1063 / 7287 / 6.86

\* ESI ( 1996-2006 ) Average Citation Rates, 搜尋日期: 2006年4月7日

圖表三: 目前較流行和較不流行的研究領域的平均被引證率, 以2000年出版的文章和2000年1月到2006年4月被引證次數為基礎, 按被引證率由高向低排列。

主流研究議題很明顯地佔有被引證優勢, 這應該提醒使用者小心闡釋這類資料。如果忽略此點, 研究會很快地陷入反效果的危險中!

如果我們看一下每單年份期刊的影響指數曲線, 就可以發現那些當下受重視的議題遠比那些較不受重視的議題的研究佔優勢。圖七[在此從略]是2000年奈米科學的出版文章與超導體文章的比較。該圖顯示主流議題的文章到現在的影響力持續增加 顯然是因為此研究領域越來越受重視。反之, 超導體領域文章的影響力在最初兩年以後就漸漸降低。甚至在1990年, 該領域發展進入頂峰, 該年份期刊的強勢卻也呈現類似的發展。

說明: 從事跨領域比較而未考慮引證行為的差異, 實屬文獻計量學最大的缺陷!

### 美國與歐洲的比較

我們在評估研究機構的學術成就時常常會遇到一個問題, 即很難找出可以互相比較的機構。美國的大學及研究機構常被採用作為比較的標準, 因為這些機構常享有高聲望, 大家喜歡以它們作為模範和衡量標準。但是美國與歐洲之間的

比較涉及一個國與國之間進行比較時產生的，到今天仍然無法圓滿解答的老問題：美國的學者是否比較喜歡引證他們的美國同仁？許多年以來，我們發現這類比較常將美國機構列於上方，即使內行人也認為其學術成就與其他機構其實是同等的，我們仍將其列於較優秀的地位。

我們幾乎可以認定美國學者引證其美國同仁比歐洲學者引證歐洲同仁來得多。美國學者在其（多數為親身認識的）同仁或美國出版的文章已經可以找到足夠的引證來源。因此，美國學者的文章被引證的可能性也隨著增高。但是這樣的引證行為無可避免地會引起水準方面的影響。除此之外，那些不屬於自然科學核心學科的研究領域在歐洲期刊發表，而不被收入 Thomson Reuters 集團出版的引證索引中。像社會經濟學的出版物可能不被刊載於著重美國的期刊，另如材料科學也是如此。

美國與歐洲的比較終究是很有問題的，因為雙方的出版文化和引證文化均有所差異，若直接比較等於將梨子和蘋果互相比較（得不出一個結果）。這個問題可以顯示文獻計量學研究的必要，以提高比較調查的證明力，並達到一個真正公平的評估。基於此議題具有的實際意義和政治迫切性，我們還需進一步說明。

由 H.F. Moed 撰寫的《Citation Analysis in Research Evaluation》一書[6]特別針對這個問題作詳盡的探討，但仍未提出令人滿意的答案。書中提到：《...美國的科學家到某一程度引證其他美國文章，不是因為那些文章優秀，而好像是有其他原因，例如他們對外國的作品不熟悉，或是他們具有島國人的狹窄心態...這種趨勢，特別是當美國學者或機構與非美國學者或機構進行比較時，可以導致引證分析宣告作廢。》

### 遲來的肯定

通常一份作品在出版以後不是很明顯地被多次引證，就是罕見被引或根本不被引證。但是我們如果觀察1950年以前著名期刊最常被引證的作品時，可發現有部分作品是在出版以後幾十年再度頻繁地被引證[27-29]。二十世紀前半葉的一些作品在過去的二十年間更常被引用，比許多諾貝爾獎得主的文章都還頻繁。此現象被稱為遲來的肯定（Delayed Recognition）或「睡美人」（Sleeping Beauty）。其中有些是當時學界無法理解的文章，因為作者的思想太先進，因此被學界置之不理。雖然一篇被當時學界長期忽視的文章後來被尊為先驅文章

的可能性不高，但是研究的發展好像只受相當少數文章的影響。後來才產生影響力的作品數雖然不多，但是對學術界卻可能具有重大的意義。因此睡美人給我們的警訊是：只以幾年為限的引證率所算出的評價值絕對不是最終結的定論。

我們可以舉Gustav Mie在1908年於《物理年鑒》（Annalen der Physik）發表的一篇文章為例說明，該篇文章探討光在粒子中的散射比光波的散射為大[30]。在二十世紀的前半葉，這篇文章幾乎未被任何人引用。由於膠態體、懸浮金屬、以及星際分子的研究大有進展，這篇文章漸漸變得重要。從五十年代開始，這篇文章就常常被引用，而成為《物理年鑒》最常被引證的文章（其被引證還比愛因斯坦的文章為多）。這篇文章與目前約四千篇文章同屬於早期最常被引的物理方面的文章。圖八表明Mie文被引歷年發展曲線（共3500次）[在此從略]。

有些較晚才獲得肯定的文章由一位著名研究學者或透過被引用於常被引證的作品之後，而再度被頻繁引用。有如那位王子吻醒睡美人。那些使以前的作品重新被閱讀傳播的，多為書評（Reviews）或是如《Nature》或《Science》著名期刊所刊載的作品。

#### D. 附錄

##### 未被引證的作品

有些為數相當大的學術作品在出版之後的一段很長時間從未被引證（連作者本身也未自我引證），而消失於學術文獻庫：SCI收有的期刊從1981到1985年有20%多的文章在出版以後就從未被引用，只有約20%的文章被引用過一次以上。其他兩份引證索引的文章當中，未被引證過的則高達50%[31-33]。

德國馬普研究院（Max-Planck-Gesellschaft）從1975到1980年發表的文章因時段長而有機會被引證，可提供一個適當的文章數讓我們查驗。其結果見於圖九[在此從略]。圖九顯示馬普研究院的出版物在十年內降低80%而達到10%，維然後就維持在這個數字。從統計學觀點來說，如果一篇文章在這十年間未被引證，其後被發現而被引證的機率是很低的。

未被引證的文章所佔的比例具有兩個意義：一是我們應該思考文章未被引用（在任何文獻參考中被提及）的原因是什麼。學者撰文到底也是運用到金錢和



個人時間精力等資源，未獲得預想的成果，我們就必須考慮其原因。二是未被引證的作品比例可作為評估的指標之一。

### 專書引證

在自然科學領域的引證有90%為期刊引證。人文社會科學只有50%為期刊引證。自然科學最常被引證的作品幾乎全是期刊文章，人文社會科學最常被引證的作品則是專書或其中文章。人文社會科學平均引證年齡明顯高於自然科學。人文社會科學的議題較以國家為重，其研究結果因此常限於國內而已。

領域之間的差異也明顯反映在馬普研究院人文科學部門（Institute der geisteswissenschaftlichen Sektion，簡稱GSHS）的出版物和引證。在此僅舉出科隆馬普社會研究所的一些數字說明：1998年馬普年刊所列的出版物只有12%以該所為地址，被收入那三份引證索引當中。雖然該所的出版物期刊文章佔約30%，但是其中許多作品未發表於Thomson Reuters出版集團認定的核心期刊。未列入索引的有許多是專書文章。詢問GSHS各研究所之後，所得到的答案是上述比例數值也適用於大部分研究所。有一些研究所甚至根本不被列入Thomson Reuters集團出版的引證索引當中，要不就是只有單獨文章出現。

原則上，專書可以被其他著述或期刊引證。Web of Science儲存參考文獻並不限出版方式，因此專書引證也可搜尋到（期刊引證專書）。但是因為專書不被收列於WoS的引證索引，其引證只能透過已知書名的首位作者查到。達到此結果的方法是利用被引參考搜尋（Cited Reference Search），然而這需要一些經驗（見上文）。

專書和專書文章不以文獻形式被收入引證索引，這造成下列後果，即所有引證（專書和期刊）若被專書執行（專書引證期刊），則不被儲存於引證索引。其所佔比例在那些以專書出版為慣例的學科當然是很高的了。這麼一來，在調查人文科學方面的引證狀況時就很容易獲得殘缺不全的結果而導致錯誤的結論！

數學是一個很好的例子，它說明引證索引的收錄規則對接近自然科學的領域也可能不足。因為數學領域較少發表於傳統的期刊。其作品有一大部分又常被專書引用，就造成我們所找出的比例不完整。除了在書中提及的引證之外，還缺乏那些引證索引之外的期刊文章的引證資料。

問題：資料庫錯誤

除了上述的曲解以外，還有一大堆的錯誤，原因多是資料庫建立者的問題[34]。現在舉化學期刊為例說明：期刊《應用化學》的文章首頁（書評 Reviews）從1996年開始就以較大單元收錄於SCI文獻中[35]。資料庫建立者，例如進行引證的作者，被這個引進各篇文章的整頁顯示（預告標題）弄混了。由於這個原因，在進行引證分析時，那些佔較大比例的以文章真正首頁為基礎的引證未列入計算，而只算那些為數較小的錯誤引證。因為有些進行引證的作者引用[預告標題]首頁，就算成一次引證，而使引證次數增加一頁（次）。這種錯誤只有在評價單獨作品時才能避免。結果是應用書評（Angewandte Reviews）的引證有80%可以忽略。這個錯誤雖然在2001年中已經消除，但是根據Thomson Reuters出版集團表示，過去五年間已經無法事後改正這個缺點。對有關學者，這個錯誤可能對他們以引證率為基礎的影響效果評估會產生負面的影響。

這類錯誤和此方法可能帶來的陷阱就讓我們在進行引證分析時必須更加小心了。目前引證資料容易取得，搜尋系統所提供的搜尋平臺也較方便易用，這使我們容易誤認為外行人只要按鍵就可以又迅速又簡單地建立資料庫。但是事實上，普通使用者仍然會面對方法上的諸多困難，無法發現大部分隱藏著的資料庫問題，因此就無法達到嚴謹的評估，也無法修正錯誤的結果。以圖表和曲綫展現的搜尋結果誤導我們認為該方法是正確可靠的，但是其資料基礎缺乏完整性，並且充滿錯誤，這個方法其實是不可靠、不正確的[36]。如此造成不具批判性的闡釋和不負責的使用，同時也影響資料的散播[37]。

### 結論

一般而言，我們對一種方法的證明力應該盡量不要抱有不實際的期望，同時也不應該吹毛求疵。引證分析提供的絕對不是精準的衡量尺標，而只是一種取向。如果受調查的作者、研究小組或研究所的引證數值或其比例數字相差少許，這絕對無法說明其學術成就具有分別。因此我們不應該認為某篇文章、某一位學者、某一研究小組或某一研究所比別篇文章、別位學者、別的研究小組或別的研究機構優秀到幾個百分比或差幾個百分比，甚至說好一倍或只達到一半。即使是利用引證分析撰寫研究結果程式也是很有問題的，因為引證有部分在時間發展上是延緩的，這可能先模糊了研究結果以後才會達到的重要地位。

另一方面，引證數值也顯示某位研究學者目前是否享有比其他同仁不尋常高或

甚低的影響力。在此，我們不應該忘記，出版物影響指數高和低的差別可達三倍，影響指數高的學者比較低影響力的學者之間的效果差別可達兩倍。在評估一個研究機構時，我們可以預估其影響力發展趨勢。發展趨勢若突然極度降低，這就是警訊，我們必須再利用其他指標加以測查。這正是引證數值的功能：顯示出強點和缺陷，並且指出危險的發展趨勢！

在討論此方法的證明力時，我們常常忘記這個方法所衡量的不是研究的品質。引證次數首先是一個回響或效果的（文字）紀錄，顯示的只是一份出版物在學術同仁之間引起的反應。一份作品被視為先驅作品的條件首先是具有較強影響效果，至少能維持一段較長時期，但那絕對不是唯一的先決條件。一些方法論方面的文章出乎意料之外地常被引用，而一些大家公認為重要的作品卻反而較少被引證。因此我們在影響力/效果/反應和品質/意義之間應該做清楚的區分。後者多因興趣和價值觀而定，因此客觀性和量化測量均較困難。

對這個方法持批評態度的學者有時用一則笑話表達他們的意見：一個喝醉了的酒鬼在路燈照明下尋找他遺失了的鑰匙。有一個人剛好走過，就問他為什麼在那兒找。那酒鬼答說，因為那兒有燈。這個笑話道出的是光在有燈的地方找丟失的鑰匙是愚蠢的。批評者認為學術評估採用以引證為基礎的指標為測量方法，這種觀點是狹窄的。但是這其實就是自然科學的特點：以光線集中於盡量緊密的光波長度範圍測量分子、固態體和星球。在這個程序中，我們當然不會獲知所有的資料，而引證分析也正是如此。

研究機構的研究義務、國內與國際上的競爭、以及證明正確使用公款補助等，在在當然（在可能範圍內）都要求一個保證品質水準的機制。即使如此，對研究評估（即對文獻計量學）持有斟酌批判的態度仍然是應該的，而且也是必要的。製造者（資料庫建立者、文獻計量學者）和評估資料使用者（研究政策人員、政治家）對效果均須負責。一個因失誤而受質疑的名聲很難被平反。因此，在此以兩個批判性論點作為結束：(1) 科學歷史告知我們，科學革命（如果有的話，也是非常稀罕的）很少以計劃研究為基礎發展的，而是常常靠那些外行人和半路出家的人造成的。學術成就若能依計劃而來，就違反一個常理，即不尋常的研究成果常常是走彎道曲路而得來的，很難針對目標給他們提供補助。(2) 研究者、學者和期刊目前已經對採用引證的評估方法作出反應，他們開始蓄意繞過其操控。這麼一來，學術評估就反而有可能違背其原本旨意

了。

引用書目：

1. P. Weingart: Welche Forschung ist gut? ( 何種研究是優秀的 ? )  
Bild der Wissenschaft 5, 34-37 (1993).
2. W. Marx, H. Schier: Zitierungsraten- eine Messlatte zur Bewertung von Forschungsqualitaet? ( 引證率-研究品質的評估尺標嗎 ? )  
Physikalische Blaetter 57, 25-29 (2001).
3. U. Schmoch, A. Hullmann: Noten fuer die deutsche Forschung? ( 為德國研究打分數 ? )  
Bild der Wissenschaft 1, 42-47 (1999).
4. F. Frick: Ueber Zitate zu Zensuren und Zaster ( 經過引句而至成績分數和金錢 )  
Bild der Wissenschaft 12, 72-76 (2004).
5. L. Bornmann, H.D. Daniel: What do citation counts measure?  
A review of studies on citing behaviour.  
Journal of Documentation 64, 45-80 (2008).
6. H.F. Moed: Citation Analysis in Research Evaluation. Springer Verlag, Dordrecht (2005).
7. M. Jokic, R. Ball: Qualitaet und Quantitaet wissenschaftlicher Artikel – Bibliometrische Aspekte der Wissenschaftskommunikation ( 學術文章的質與量 學術溝通的文獻計量學層面 )、 Schriften des Forschungszentrums Juelich ( 尤利希研究中心刊物 )、 Reihe Bibliothek/Library ( 系列圖書館 )、 Band 15 (2006).
8. D.J. De Solla Price: Little Science – Big Science. Suhrkamp Verlag, Frankfurt (1974).
9. E. Garfield: Citation Indexing – Its Theory and Application in Science, Technology, and Humanities. John Wiley & Sons, New York, Chichester, Brisbane, Toronto (1979).
10. E. Garfield: Random Thoughts on Citationology – Its Theory and Practice. Scientometrics 43, 69-76 (1998).
11. Thomson Reuters: <http://scientific.thomsonreuters.com/products/>

12. STN International: <http://www.stn-international.de>
13. O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall: Protein measurement with the Folin phenol reagent.  
Journal of Biological Chemistry 193, 265-275 (1951).
14. V. Pareto: Cours d'economie poliique 2, 299-345 (1987).
15. J.E. Hirsch: An index to quantify an individual's scientific research output.  
Proc. of the Nat. Acad. of Sci. of the USA 102, 16569-16572 (2005).
16. L. Bornmann, H.D. Daniel: What do we know about the h index?  
Journal of the American Society for Information Science and Technology 58, 1381-1385 (2007).
17. Journal Citation Reports (JCR-2008), Science Edition:  
<http://scientific.thomsonreuters.com/products/jcr/>
18. E. Garfield: The history and meaning of the journal impact factor.  
JAMA – Journal of the American Medical Association 295, 90-93 (2006).
19. D. Adam: The counting house.  
Nature 415, 726-729 (2002).
20. A. Amin, M. Mabe: Impact Factors: Use and Abuse.  
Perspectives in Publishing 1 (October 2000).
21. P. Mestecy: Impact Factors-Handle with care.  
Materials Today 1, 8-12 (1998).
22. P.O. Seglen: Why the impact factor of journals should not be used for evaluating research.  
British Medical Journal 314, 498-502 (1997).
23. M.H. MacRoberts, B.R. MacRoberts: Problems of Citation Analysis: A Critical Review.  
Journal of the American Society for Information Science NY, 40, 342-349 (1989).
24. IMU Report:<http://www.mathunion.org/fileadmin/IMU/Report?CitationStatistics.pdf>.
25. Essencial Science Indicators (ESI-2008):<http://scientific.thomsonreuters.com/products/esi/>

26. ESI baselines:<http://esi.isiknowledge.com/baselinesmenu.cgi>
27. A.F.J. van Raan: Sleeping beauties in science. *Scientometrics* 59 (3), 461-466 (2004).
28. Redner: Citation Statistics from more than a Century of Physics Review. *Physics Today* 58, 49-54 (2005).
29. W. Marx: Dornroeschen und Mauerbluemchen ( 睡美人與牆邊小花 ) . *Physik in unserer Zeit* 38, 34-39 (2007).
30. G. Mie: Contributions to the optics of turbid media, especially colloidal metal solutions. *Annalen der Physik* 25, 377-445 (1908).
31. D.P. Hamilton: Publishing by – and for? – the Numbers. *Science* 250, 1331-1332 (1990).
32. D.P. Hamilton: Research Papers: Who's United Now? *Science* 251, 25 (1991).
33. D.A. Pendlebury: Science, Citation, and Funding. *Science* 251, 1410-1411 (1991).
34. Editorial: Errors in citation statistics. *Nature* 415, 101 (2002).
35. W. Marx: Angewandte Chemie ( 應用化學 ) in Light of the Science Citation Index. *Angewandte Chemie* 113, 143-148 (2001) 以及 *Angewandte Chemie – International Edition*, Vol. 40, 139-143 (2001).
36. M. Cardona , W. Marx: Verwechselt, vergessen, wieder gefunden – Referenzen, das fehlerhafte Gedächtnis der Wissenschaft(ler) ( 被搞混了 , 被忘記了 , 再度被發現 – 參考文獻 , 充滿錯誤的學術(人員) ) *Physik Journal* 3, 27-29 (2004).
37. P. Weingart: Das Ritual der Evaluierung und die Verführung der Zahlen ( 評估的儀式與數字的誘惑 )  
In: *Die Wissenschaft der Öffentlichkeit: Essays zum Verhältnis von Wissenschaft, Medien und Öffentlichkeit*. Weilerswist: Velbrueck Wissenschaft, 206, 102-122 (2005).

註解

本文只包含一些重要的資訊，以許多使用者極端微弱的背景知識與研究評估的需求為導向，仿馬普研究院內引證資料的使用者所需撰寫而成。本文有關資料庫技術與實際執行，以及有關文獻計量學闡釋與其關係的觀點絕對尚待求完整無缺。

---

[TOP](#)

Copyrightsc2003 Department of International Programs. All rights reserved. 國科會國際合作處 版權所有

[隱私權宣告](#) [版權宣告](#) [最佳瀏覽解析度1024x768](#)