Scientometrics (2012) 90:343–360
DOI 10.1007/s11192-011-0523-6

The relationship between co-authorship, currency of references and author self-citations

Wen-Yau Cathy Lin · Mu-Hsuan Huang

Received: 7 January 2011 / Published online: 22 September 2011

Abstract This paper attempts to identify the relationship between co-authorship and the currency of the references and author self-citations in the key journals of environmental engineering. The results show that the self-citation rate of co-authored articles is higher than in single-authored articles. A statistically significant correlation is identified between the numbers of co-authors, the rate of author self-citing and the author self-cited; though it was a low correlation. The value of coefficient correlation between the number of co-authors and the author self-citing rate is slightly higher than that between the number of co-authors and the author self-cited rate, which indicates that the number of co-authors hold a stronger correlation with the self-citing rate than the self-cited rate. Meanwhile, self-citing references are found to be more up-to-date than references to others. The range of publication years of self-citing references is smaller than that of references to others, indicating that researchers tend to preferentially cite their own recent works. There is no significant difference in the latest references between self-citing references and the references to others. It might result from electronic journals that provide an easy access to the most current publications.

Keywords Self-citation · Author self-citation · Co-authorship · Publication lag

Introduction

In studies of citation analysis, self-citations are often regarded as 'noise' or diversions removed from analyses (Hellsten et al. 2007). This situation has happened since the assessment of research or scientific policy-making has employed a citation analysis (Moed 2005; Pichappan and Sarasvady 2002). Scholars hold different opinions on the meaning

W.-Y. C. Lin
Department of Information and Library Science, Tamkang University, Taipei, Taiwan
e-mail: wylin@mail.tku.edu.tw

M.-H. Huang (✉)
Department of Library and Information Science, National Taiwan University,
No. 1, Roosevelt Road Sec. 4, Taipei 10617, Taiwan
e-mail: mhhuang@ntu.edu.tw

and role of self-citations in the analysis of citation. Similar disputes occur in the assessment of research or scientific policy-making. It is possible that author self-citations could be manipulated to raise the counts of citation, thus ensuring one's position in the scientific community. Numerous studies have indicated that only few research papers contain no self-citations, and many researchers tend to cite their previous own works over the publications of others (Hyland 2003; Phelan 1999; Tagliacozzo 1977; White 2001). In contrast, self-citations are considered a natural phenomenon in knowledge dissemination by bibliometric scholars. From this perspective, the lack of self-citations over a long-term period is just as pathological as a high proportion of self-citations (Glänzel et al. 2004).

In research related to author self-citations, the relationship between the number of co-authors and self-citations has received much attention. Garfield (1979) states that if author self-citation is defined as self-citations by the first author, it would account for at least 10% of total citations; the ratio would be even higher when this is extended to the second author or all co-authors. In recent work, author self-citation is usually defined as when the references cite any identical author of the original articles (Aksnes 2003; Glänzel et al. 2004; Snyder and Bonzi 1998). Following this rule, the rate of author self-citation will be higher than Garfield's estimate. Another study suggests that the citation counts of co-authored articles have shown an increase in line with the number of co-authors (Herbertz 1995). This leads to the question of whether the number of co-authors has an influence on the self-citing rate and the self-cited rate. The author self-citing rate refers to the ratio of author(s) citing their own previous work(s) as reference(s) to all the references of the article. The author self-cited rate refers to the ratio of articles cited by authors themself/selves to the total cited times. The study by Aksnes (2003) has confirmed this phonemonon and shown that papers receive an average of more than two citations with each of additional co-author. However, self-citations only partially explain the increasing citation rate. It is necessary to go through a clear and accurate investigation into the influence of co-author number on author self-citations.

Comparing the currency of the references of citing works from author him- or herself to the currency of the references to works from others cited is another potential issue. According to Bonzi and Snyder (1991), self-citations are more current than citations of others' work. It is easier to gain access to one's own previous publications than those of others, which makes sense of those latest references in articles with self-citations. However, is there a correlation between the currency of references and the author self-citations?

This paper attempts to discuss the relation between the number of co-authors and author self-citations, which also intends to build an accurate undersanding of author self-citations, espesically when the relation is employed in a citation analysis for the evaluation of reasercher. Key journals in environmental engineering are selected for analysis regarding the impact of the number of co-authors on the self-citing rate and the self-cited rate, as well as the correlation between the author self-citations and the currency of references.

Literature review

Co-authorship

Former studies have found that co-authored papers which were subject to higher self-citation rates or more citations are found coming from many disciplines and countries, e.g. Norwegian publications (Aksnes 2003), psychological research on tourism (Barrios et al. 2008), ecological articles (Leimu and Koricheva 2005), mathematics and statistics

publications (Rousseau 2001), papers in life sciences, natural sciences, engineering & material science, and interdisciplinary sciences in the UK (Katz and Hicks 1997) and management science articles (Smart and Bayer 1986). In comparison, the difference between co-authored papers and single-authored papers is not statistically signicicant in the fields of finance (Avkiran 1997), clinical psychology, and educational measurement (Smart and Bayer 1986). In theoretical physics, most papers are co-authored, and the small number of single-authored papers have received more citations than co-authored ones (Rousseau 2001). Smart and Bayer (1986) and Rousseau (2001) both support the hypothesis that co-authored papers usually receive more citations than single-authored ones, however, they also stress that this inference should not be applied to all fields of science.

Regarding to the relation between the number of co-authors and self-citations, Herbertz (1995) points out that collaborative papers from researchers from different institutes have higher citation rates than those from researchers from the same institute; nonetheless, the former usually sees higher self-citation rates. van Raan (1998) agreed on the higher self-citation rates of co-authored papers and regarded self-citations as an having an "amplifying" effect. He showed a conservative attitude toward the greater impact of international collaborative papers in which self-citations are also higher, and he stressed that self-citations in this case should be interpreted differently in the academic evaluation process. Similarly, Persson et al. (2004) have assented to van Raan's idea, the results of their study also supporting the concept that the number of references increases with the number of co-authors, especially when the co-authors come from different countries. However, they also discovered that references from other authors are higher than references to the authors themselves, and thus the suggestion of self-citations will not necessarily lead to a higher number of references in co-authored papers. Meanwhile, in the study of self-citations from Norwegian researchers during the period of 1981–1996, Aksnes (2003) has confirmed the correlation between co-author number and self-citations. He notes that every single author correlates to an increase of two cited times in the original article, and that more co-authors would lead to a higher self-cited rate, though he also points out that self-citations could only partially explain the increasing self-cited rate. Leimu and Koricheva (2005) analyzed 837 co-authored research articles in the ecological journal *Oecologia* during the period of 1998-2000, finding signs that there are more citations received in co-authored articles and the rate of self-citation is higher than the ones of single-author.

There are studies that present different viewpoints. Tagliacozzo (1977) demonstrates that the number of co-authors does not correlate with the number of self-citations, and that the seniority of co-authors is influential. If there is no-one of equivalent seniority amongst the co-authors, the senior author often contributes the majority of the citations by providing his/her previous relevant research, which makes the self-citations of co-authored articles close to single-authored ones. On the other hand, if co-authors have equivalent seniority, the mutual resistance will lead to each co-author contributing less of their previous relevant research to co-authored paper(s) than his/her individual work. However, the correlation inferred by Tagliacozzo remains unproven. Similar evidences are found in the studies of Glänzel and Thijs (2004) and Schubert et al. (2006). Glänzel and Thijs (2004) suggest that a co-authored paper will increase the self-citation rate, but the increase in the rate of self-citations is much lower than the rate of being cited. Reasons in support of this include: (1) the productivity of an author whether he/she owns co-authored papers or not is usually not directly correlated to average citation counts, (2) a hierarchical structure might exist among co-authors, and (3) multiple authorship is not the result of the work of stable terms, and the topics of research may not be persistent. The second suggestion is somewhat similar to Tagliacozzo's inference. The study results might be inconsistent to some extent, yet one thing can be guaranteed is that a

disparity in the correlation between co-author number and self-citations has been found across different disciplines. The researchers entirely admit that the results should only be applied to the discipline(s) they have focused on (Aksnes 2003; Glänzel and Thijs 2004; Herbertz 1995; Smart and Bayer 1986; Tagliacozzo 1977; van Raan 1998).

Currency of references

It is easier and more common for an author to acquire his/her own previous publications than those of others. Also, authors tend to cite their latest work while also citing older publications (Hellsten et al. 2007). Looking at the age of references, the lag in publication among references with self-citations and original articles is shorter than that of the references citing from others and original articles. This logical inference has been confirmed by numerous studies (Aksnes 2003; Al-Qallaf 2003; Bonzi and Snyder 1990, 1991; Glänzel et al. 2004; Lawani 1980; Tagliacozzo 1977).

According to a study by Tagliacozzo (1977), over 50% of the self-citing references antedate the original articles by 2 years or less, and more recent than those references authored by others. Bonzi and Snyder (1990) conducted a comparison study of self-citations in the fields of chemistry, geology, economics, and sociology. They found that the average age of self-citations in these four fields is below 4.5 years, which is smaller than 9.3 years of the citations to others. Another study by Bonzi and Snyder (1991) shows that the range of the age of self-citations is between one and 30 years, with an average of 5.3 years; the range of the age of citations to others is between one and 84 years, with an average of 11.3 years. The figures suggest that the articles self-cited by an author are indeed more recent than those citations to others. The study by Aksnes (2003) shows that the author self-cited rate is as high as 63% in the first year of publication, but drops to 40% in the second year, and ultimately drops to 9% by the 15th year.

If researchers set the citation window for paper publication and production longer, the self-cited rate will be lower (Aksnes 2003; Lawani 1980). Therefore, Aksnes (2003) has recommended that the range of citations should be enlarged to reduce the bias caused by self-citations while calculating the impact factor of journals,. Glänzel et al. (2004) have voiced a similar opinion. They note that, in a reasonable situation, the self-cited rate drops very fast after the first 3 years of publication, and the citation rates of the self and others will reach a balance. For this reason, they have advised that the citation window should be set not less than 3 years, not longer than 4 years is sufficient for reliable bibliometric analyses. Based on the study results by Aksnes (2003) and Glänzel et al. (2004), Journal Citation Reports (JCR) may consider to enlarge the current citation window of 2–3 years while evaluating the rationality of self-citation rates.

Meanwhile, since an author can more easily acquire his/her in-press or new publications, an of self-cited article could be more up-to-date. If an author cites others' latest publications, it implies that a special social network exists between the citing and cited authors. This observation might be used to discuss and identify the concept of the "invisible college."

Research method and data collection

Self-citation refers to the situation where the author of a specific source article shown in the reference list is also one of the authors of the citing paper. This study follows the definition

conceptualized by previous research efforts: under the circumstance of multiple authors, if any authors of a source article are also amongst the authors shown in the reference list, this is what is referred to as "author self-cited" or "author self-citing" (Aksnes 2003; Glänzel et al. 2004; Snyder and Bonzi 1998). The ratio of the number of times that an author cites a reference he/she has authored to the total number of references cited is the "author self-citing rate." The ratio of the number of times an author self-cites to the total number of times cited is the "self-cited rate." The following are the formulae for calculating author self-citing and author self-cited rates in this study.

Author self-citing rate of author A: $A_{SC}(A)$

$$A_{sc}(A) = \left(\frac{C_{as}}{C_{as} + C_{ao}}\right) \times 100\%$$

C_{as} is the number of articles previously published by author A in the reference list of a specific work, C_{ao} refers to the number of articles published by other authors in the reference list of a specific work.

Author self-cited rate of author A in the journals published in year Y: $A_{SD}(A,Y)$

$$A_{SD}(A, Y) = \left(\frac{\sum_{i=1}^{n} D_{AS}}{\sum_{i=1}^{n} D_{AS} + \sum_{i=1}^{n} D_{AO}}\right) \times 100\%$$

D_{AS} is the number of references previously published by author A in the year of Y and cited by author A in the period of $Y + n$ years. On the other hand, D_{AO} means the number of references published by other authors in the year of Y and cited by author A in the period of $Y + n$ years.

This paper focuses on research articles in 20 key environmental engineering journals selected by the Environmental Engineering Program of the National Science Council in Taiwan as the scope of study, shown in Table 1. At the beginning of the study, the data regarding references is directly retrieved from WOS. However, several bibliographic errors (omissions, typos, etc.) are found in the WOS data, which may affect the analytical results. Similar errors are shown in Scopus. Additionally, the data currently available in WOS and Scopus is insufficient for this study. To collect sufficient information and improve

Table 1 Journal list

Journal title	
Aerosol Science and Technology	Applied Catalysis B: Environmental
Atmospheric Environment	Environmental Health Perspectives
Environmental Science & Technology	Environmental Toxicology and Chemistry
Journal American Water Works Association	Journal of Aerosol Science
Journal of Computing In Civil Engineering	Journal of Contaminant Hydrology
Journal of Environmental Engineering, ASCE	Journal of Environmental Quality
Journal of the Air & Waste Management Association	Journal of Toxicology and Environmental Health. Part A
Journal of Water Resources Planning and Management -ASCE	Science of the Total Environment
Soil Science Society of America Journal	Water Environment Research
Water Research	Water Resources Research

accuracy, the full text data are downloaded manually instead of conducting the research electronically.

In order to get the complete references of research articles, the beginning of the period of study is defined as the late 1990s, when electronic full text articles became more prevalent. This paper chose samples of articles published between 1999 and 2008 so as to make valid samples and a sufficiently long timespan for discussion of correlation between self-citations and time. Besides research articles, journals often contain many forms of literature, including editorials, letters, corrections, errata, review articles, additions, excerpts, and discussions. In line with its focus on academic research, this paper has limited the sample articles to research articles.

Author names are occasionally recorded in different formats in different journals, and so to verify authorship, this study examines the affiliation, publication history, and journals of authors which are shown in the articles. Research subjects, co-authors, correspondence data, personal websites, and other useful information are employed to clarify possible variations of authors' names. An authority control file is established to ensure accuracy in self-citing and self-cited data.

There is a sum of 49,502 research articles in the 10-year interval of these 20 journals. Given limited manpower, time, and resources, this paper has selected 600 articles with a 95% confidence level (confidence interval <4). Since this study focuses on the correlation between co-authorship and the currency of references on author self-citations, each sample ought to contain author self-citations. A random stratified sampling method is used to select three articles each year with author self-citations from the 20 journals over the 10-year period. Sample articles without author self-citing reference(s) are abandoned and replaced by other sample articles with self-citing reference(s) collected from the same issue of the same journal. The procedure repeats until self-citing examples are found. A list is made of 600 sample articles, and the PDF files of each article are downloaded from the online full text journal database. Each sample article has its author number, number of references and publication year recorded for further analysis. The bibliography information of cited references in this study is gathered from the Web of Science database, with data retrieved in the last week of March 2009.

Results

Co-authorship

Among the 600 sample articles, not excluding the repetition of authors, there are 2,302 authors, with an average of 3.84 authors (standard deviation 2.19). The average author self-citing rate is 17.77%, and the author self-cited rate 22.47%. Additionally, the average number of references is 33.08, and the average number of times cited is 12.1.

Among the sample size of 600 articles, 4.17% of articles are single-authored, with the remaining 95.83% co-authored articles. There are 152 articles with two authors (25.33%), 139 articles with three authors (23.17%), and 122 articles with four authors (20.33%); this is a total of 413 articles with two to four authors, accounting for 68.83% of the total sample articles. The highest co-author number for an article is 17 authors. 14 articles are published by more than 10 authors, only accounting for 2.33%. The distribution of co-author number is shown in Fig. 1. As shown in Fig. 1, we found there is a phenomenon of increase/decrease between the number of co-author and the amount of articles. There are a total of 152 double-authored articles, 139 triple-authored articles, and 122 quadruple-authored

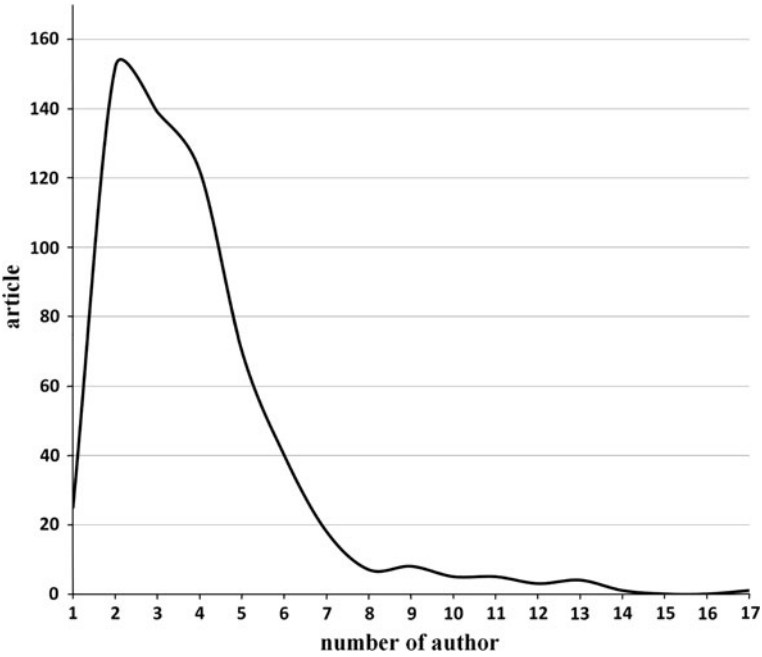

Fig. 1 The distribution of number of author

articles. As the number of co-authors increases, however, the amount of articles decreases, showing a long tail tendency.

Figure 2 shows the average number of co-author by year. It can be observed that the lowest average number of co-author is 3.38 in year 1999 and the highest is 4.37 in year 2007. A drop is shown in year 2004 in which 24 articles with two authors (40%), it's the highest ratio in all the years. The result tells that there is a decline in the average number of co-author. Overall, the trend among the period of 1999-2008 shows an increasing tendency in the average number of co-author.

Figure 3 shows the average number of references by years. The trends of the mean and the median demonstrate a slightly increasing tendency in the references of environmental engineering articles. In fact, the increasing tendencies, whether in co-author number, the number of references or the total publications, have been noticed in some prior research (Glänzel 2008; Persson et al. 2004). The results of this paper confirm this trend.

Self-citing and self-cited rates

The average author self-citing rate among a total of 600 sample articles is 17.77%, with a maximum of 77.78%. Only one article shows a self-citing rate higher than 70%; 197 articles (32.83%) a self-citing rate lower than 10%, and 505 articles (84.17%) a self-citing rate lower than 30%. There is a total of 588 articles with self-citing rates lower than 50%, accounting for 98%. This suggests that the author self-citing rate of most of the articles is below 50%.

When it comes to the rate of author self-cited, the average rate for the 600 sample articles is 22.47%. There are 222 non-self-cited articles, accounting for 37%. Based on the

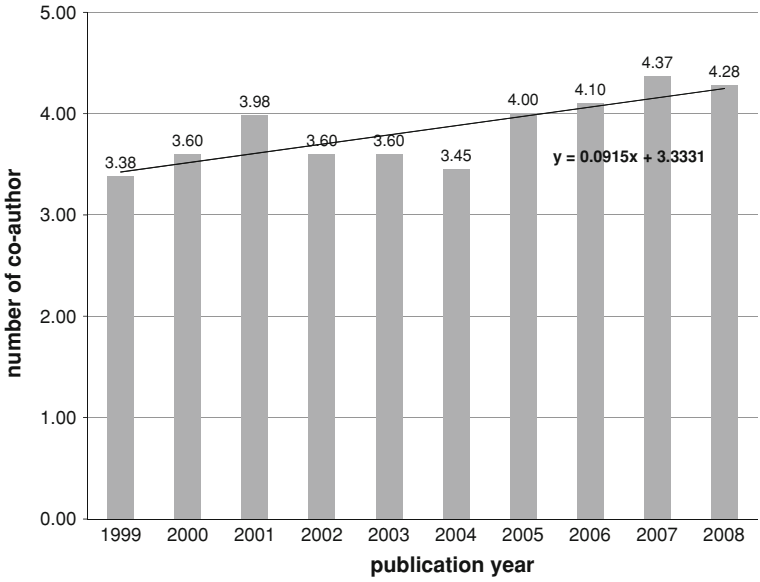

Fig. 2 The average number of co-author by year

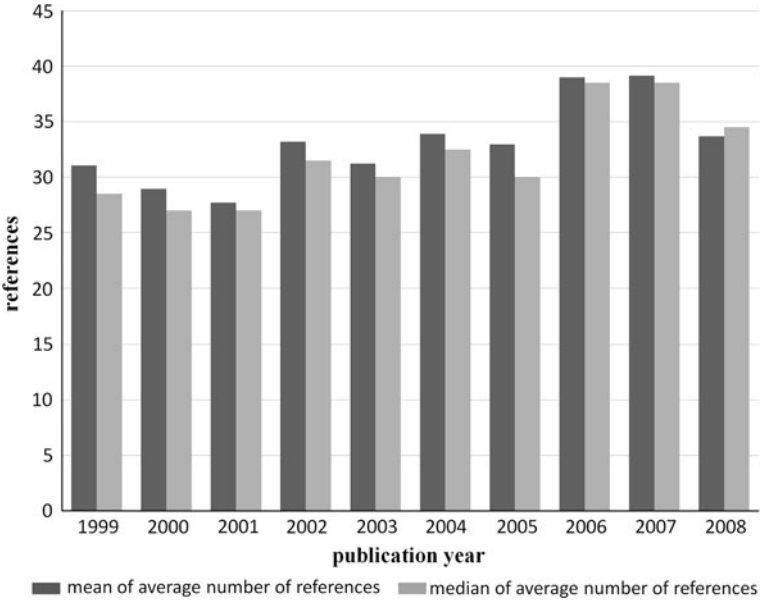

Fig. 3 The distribution of average number of references by year

cited rate of zero percent or non-zero, these 222 articles could be divided into two categories. The first group of 130 articles has been cited by others, and the second group of 92 articles has not been cited by themselves or others. In the second group, there are 44 articles published in the year of 2008. Since the accumulation of citation takes time, there

are not commonly cited. Another peak of the sample articles is revealed at a self-cited rate of 10% to 19%, with 84 articles accounting for 14%. There are 24 articles with a self-cited rate of 100%, accounting for only 4%.

The results of this paper have shown that the average author self-citing rate in journal articles on environmental engineering is 17.77%, which is close to the average of author self-citing rate of 18% of the 289 diabetes articles in the year 2000 studied by Gami et al. (2004). It is also lower than the suggestion, which is not uncommon or inordinate, of a standard 25% author self-citation rate through the biomedical literature studied by Pendlebury (2009). In addition, our study has found that there is no significant difference in the average self-citing rate between single-authored and co-authored articles. Yet the average self-cited rate in co-authored articles is higher than that in single-authored articles by 10.19%. This finding confirms the results of previous studies in the fields of tourism psychology, ecology, mathematics & statistics, life sciences, natural sciences, engineering & mathematics, and management science (Aksnes 2003; Barrios et al. 2008; Katz and Hicks 1997; Rousseau 2001; Smart and Bayer 1986). It should be noted that this study only sampled research articles with author self-citations. The exclusion of non-self-citation articles may cause a small degree of increase in the self-citation rate.

Does the self-citing rate increase over time as well as the co-authored articles and the number of references did? Since the author self-citing rate could be extremely high or low, this research also calculates the mean and the median in order to avoid any misinterpretations. The 600 sample articles are classified by publication year, and the mean and median are calculated for the annual author self-citing rate, shown in Fig. 4. There is no drastic fluctuation in the annual author self-citing rate, but in the 10-year interval from 1999 to 2008, both the mean and median show a slight decreasing tendency.

Figure 5 displays the mean and median for the author self-cited rate by year. In the general condition, the accumulation of citations takes time. In 2007, the number of articles

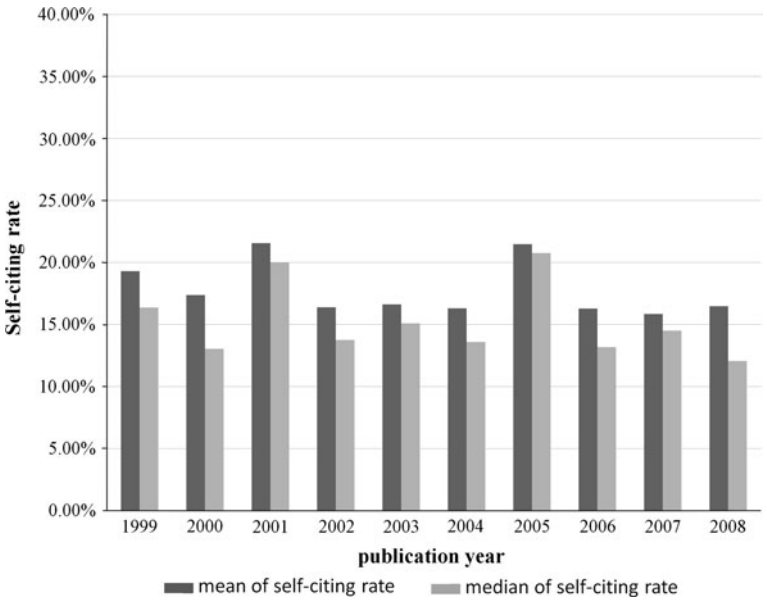

Fig. 4 The distribution of the author self-citing rate by year

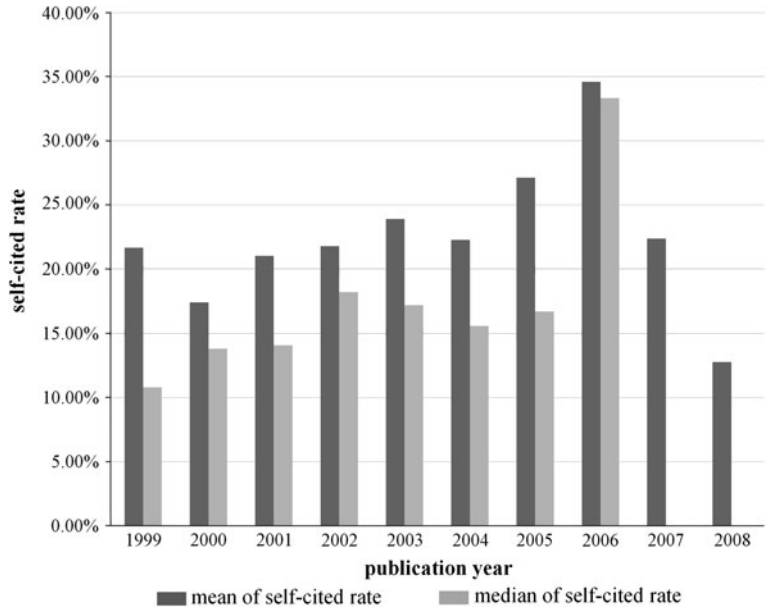

Fig. 5 The distribution of the author self-citied rate by year

with an author self-cited rate of zero percent is 35 (58.33%), and the number has reached 49 (81.67%) in 2008. These results in author self-cited rates in 2007 and 2008 with a median represented as zero.

From the above analysis, the results show a declining tendency in the annual mean and median of the author self-citing rate, but the drop is slight. In contrast, Ken Hyland claims that the author self-citing rate has grown year by year; some attribute this growth to the pumping up the numbers for citation indexes. Simultaneously, the prevalence of electronic journals might reduce this as indexing algorithms can easily detect self-citations (Anonymous 2003). In this study, the selected environmental engineering journals have been fully digitized for the time span studied; whether any reduction in the author self-citing rate may be caused by electronic journals is a matter for further research.

Co-authorship and self-citation rates

The Pearson correlation coefficient test is conducted to identify the relationship between author self-citing rate, author self-cited rate, number of co-authors, number of references, and number of cited. Since the numbers of cited in the years of 2007 and 2008 are inappropriately taken in the correlation analysis, we have removed data on 120 articles published in 2007 and 2008. Even with a reduction of sample size, a confidence level of 95% is achieved, with a confidence interval of 4.45. Table 2 indicates that the number of co-authors significantly correlates with the number of times cited, the author self-citing rate, and the author self-cited rate. This finding is consistent with prior studies (Katz and Hicks 1997; Leimu and Koricheva 2005) which suggest a positive correlation between the number of co-authors and the number of references cited. Of particular note is that the positive correlation is weak.

Table 2 Pearson's *r* of the number of co-author, number of reference/time cited and self-citing/cited rate

	No. of co-author	No. of reference	No. of times cited	Self-citing rate	Self-cited rate
No. of co-author					
Pearson's *r*	–				
Sig. (2-tailed)	–				
No. of reference					
Pearson's *r*	.087	–			
Sig. (2-tailed)	.057	–			
No. of times cited					
Pearson's *r*	.151**	.086	–		
Sig. (2-tailed)	.001	.060	–		
Self-citing rate					
Pearson's *r*	.252**	−.248**	.013	–	
Sig. (2-tailed)	.000	.000	.778	–	
Self-cited rate					
Pearson's *r*	.164**	.101*	−.075	.134**	–
Sig. (2-tailed)	.000	.027	.102	.003	–

** Significantly correlated when the significant level is set at .01 (two tails)

* Significantly correlated when the significant level is set at .05 (two tails)

Table 3 Comparison of single-authored articles and co-authored articles in author number, reference number, and self-citation rates

	No. of co-author (mean)	No. of reference	No. of times cited	Self-citing rate (%)	Self-cited rate (%)
Single-authored article	1	25.92	13.08	17.16	12.71
Co-authored article	3.96	33.39	12.06	17.79	22.90
All article	3.84	33.08	12.1	17.77	22.47

Regarding to the rate of author self-citing and author self-cited, the data of Table 2 shows that the number of co-authors is significantly correlated with the rate of author self-citing and author self-cited when the level of significance is set at .01, but the correlations are both low.

Meanwhile, the disparity between single-authored and co-authored articles deserves more attention. Based on the data shown in Table 3, the author self-citing rate of all the sample articles is 17.77%, and the author self-cited rate is 22.47%. When further dividing the sample articles into single-authored and co-authored, a comparison in Table 3 shows that there are more references in co-authored articles than single-authored ones. However, there are more references cited in the latter than the former. The rate of author self-citing is similar in both groups, but the rate of author self-cited is much higher for co-authored articles than single-authored ones.

As shown in previous analysis, co-author number, number of references, and number of times cited are statistically significantly correlated and show a positive low correlation. Why have single-authored articles received more citations than the co-authored ones? According to Rousseau (2001), single-authored articles in theoretical physics also receive

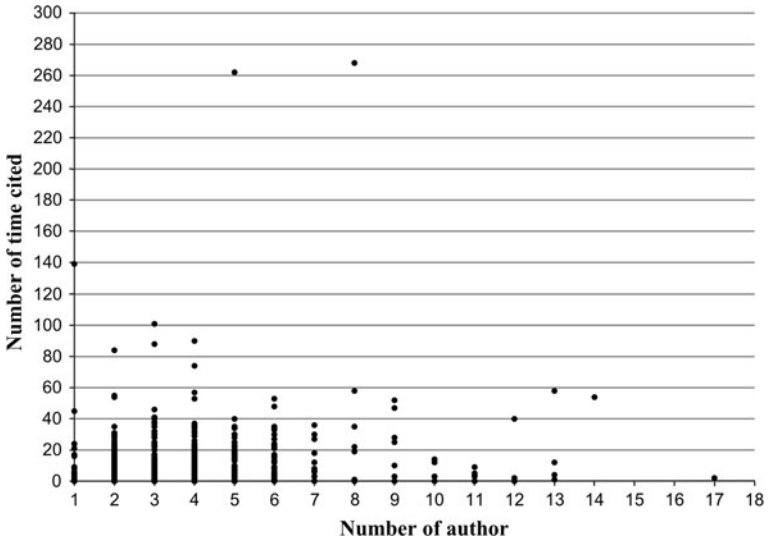

Fig. 6 The distribution of number of authors and article times cited

more citations than the co-authored ones. According to Smart and Bayer (1986), co-authored articles in management science have received more citations than the single-authored ones. However, there is no such situation present in clinical psychology or educational measurement. According to Rousseau, most theoretical physics articles are collaborative works; as a result, the situation is present in a small number of single-authored articles, which should not be used to suggest all the single-authored articles receive more citations than the co-authored ones in general. In this study, 25 of the 600 sample articles are single-authored articles, accounting for 4.17%, a rather low percentage.

Figure 6 depicts a scatter plot of the number of authors and the times cited. As can be clearly seen, one article published in *Water Resources Research* in 2000 has been cited 139 times, another article published in 1999 in *Science of the Total Environment* has been cited 45 times among the 25 single-authored articles. These two articles greatly increase the average number of times cited of single-authored articles. If these two outliers are removed, the average number of times cited of single-authored articles decreases to 6.22, lower than the value of co-authored ones. Thus it is reasonable to find a low positive correlation between co-author number, number of references, and number of times cited.

Distribution: the publication year of references

Since researchers are more familiar with their own published works than those of others, they seem more favorable toward their own works when selecting references to cite. As Narin and other scholars have suggested, an author is fond of citing recent work(s) of theirs while the research is in progress. (Hyland 2003; Narin and Olivastro 1986; Pichappan and Sarasvady 2002). One might ask whether this happens in key environmental engineering journals. A sensible approach is to investigate how the lag in publication time varies between self-citing and non-self-citing references. Furthermore, the distribution of the latest and the oldest references is worth observing when a researcher engages in self-

Table 4 Paired-samples *T*-test of the most recent/oldest references to the self or to others

	Paired samples statistics					*t*	df	Sig. (2-tailed)
	Mean	Std. deviation	Std. error mean	95% Confidence interval of the difference				
				Lower	Upper			
Self-citing & citing others (recent)	.115	2.287	.093	−.068	.30	1.232	599	.219
Self-citing & citing others(oldest)	27.06	19.06	.778	25.526	28.583	34.769	599	.000

citation or cites other work(s). In this study, the 600 sample sources pertaining to environmental engineering contained 19,849 references. A further analysis of the references cited in source articles will be presented, followed by a tentative comparison of publication years between the oldest and the latest references in two categories: self-citation references and non-self-citation references.

Among the references in the 600 sample source articles, there are 3,283 references (16.54%) with publication years and belong to references of author self-citing; 16,277 of references (82%) with publication years are without self-citation. The other 289 references are reports and webpages without publication years, or articles without available publication information for various reasons (e.g. incorrect citations, citation slips, and typographical errors). These 289 references account for 1.46% of the total references, and are excluded from the following analysis, leaving a modified number of total references of 19,560.

The references are divided by publication year. The average range between the most recent and the oldest is 6.63 years for self-citing references and 33.58 years for references to others' work. Table 4 shows a paired-samples t-test of the most recent and oldest references to self or others by articles. The results show no significant difference in the most recent references between the references to the self and to the others. However, there is a significant difference in the oldest references between the references of the self and to the others, with an average as high as 27.06 years.

As mentioned above, the results show that there is no big difference in the latest references between the references to the self and to others, indicating that the research of article authors can get the most updated literature whether from his/her own works or the works from others. This may be contributed by electronic journals. More and more in-press articles are available via academic publishers' electronic databases, which solves the problem of the long process of publication by enabling audiences to read pre-publication articles which are still in typesetting after review, or ones with typesetting complete and ready for pressing. These in-press articles help in shortening the publication time lag between citations to others and original articles, and it also reduces the currency advantage of references to the authors themselves.

Since there is no huge difference in the publication year of the latest references between references to self and to others, the disparity in the range of publication years may result from the oldest references. It is quite reasonable that the range of publication years is shorter when it comes to the citations of self-authored articles. Due to the limited time-span of the research, researchers may not have many choices when they want to cite their own works. Take the oldest article with self-citations as an example: it is published in *Water*

Research in 1999 by Frank M. Tiller, a professor of the Department of Chemical Engineering in the University of Houston, US, and the co-authors are his associate—a scholar from South Korea, in the Department of Environmental Engineering of Inje University—and a Taiwanese scholar in the Department of Chemical Engineering at National Taiwan University (University of Houston—Cullen College of Engineering 2006). In this article, one of the references is an article by first author Tiller in 1960. In 1999, Tiller was already a senior professor with more than 40 years of research experience, and surely had many prior works to his name. In addition, in the articles with self-citations, the oldest references are mainly located at the interval of 1990–1999. During this period, the authors of these sample articles have published more works for people to cite. Compared to the oldest reference to others (published in 1879) the published year of the oldest reference to one's own work is likely to be rather more recent.

Figure 7 is the distribution of the oldest references in each sample article, divided into pre-1930 and post-1930. The post-1930 period is further divided into 10-year intervals. From Fig. 6, we see that most of the references to the self are located within the interval of 1990–1999, with only 4 in the interval of 1960–1969. In contrast, the references to others are concentrated on three intervals, with 133 of them in 1960–1969, 158 of them in 1970–1979, and 118 of them in 1980–1989. This indicates that even the oldest references to the self are relatively more recent than those to others.

Another noteworthy point is that the sum of the oldest references to others before 1960 is 145, accounting for 24.17% of the total samples. Environmental engineering is part of engineering and the natural sciences, and in general, the publication dates of references in these engineering and natural sciences are more recent. However, the results of this study show that nearly a quarter of articles have cited references older than 40 years. This may be attributable to environmental engineering's newness as a sub-field of engineering and

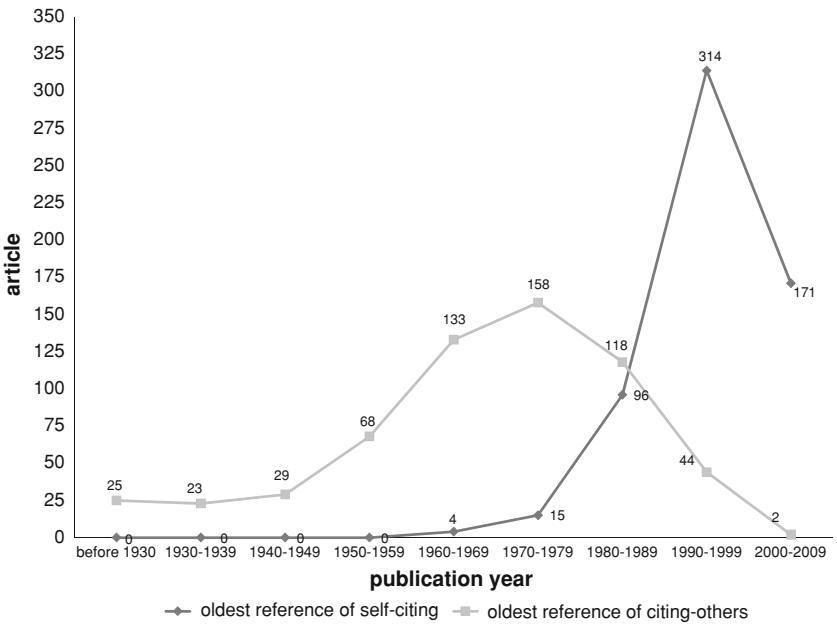

Fig. 7 The distribution of the oldest references

being higly interdisciplinary. Environmental engineering is affected by many disciplines, including civil engineering, public health, ecology, chemistry, and meteorology. It also applies the methods of physics, chemistry, biology, and medicine. As a result, researchers in environmental engineering may draw upon resources from traditional disciplines while clarifying the purpose of their studies.

Time lag: the publication year of original and reference articles

The 600 sample articles are classified by publication year, and the mean and median are calculated from self-references and non-self-references.

Table 5 shows the publication lag between original articles and the references with citations to the self/others. The average of time lag between original articles and self-citing references is from 4.18 to 5.22 years, with the median slightly lower than the mean. The average time lag between original articles and the references of citing others' is between 10.83 and 12.25 years, with the median much lower than the mean. When comparing annual changes, self-citing references are more recent than the references citing others as well. The gap between these two is from 6.16 to 7.41 years, and the median, which is easily affected by extreme values, differs from 4.37 to 5.16. From the distribution of publication lag between original articles and reference articles, it can be seen that self-citing references are indeed more recent than references citing others in each year studied. No trend can be identified in the statistics shown in Table 5.

Classified by publication year, the 600 sample articles are further analyzed by comparing publication lag between original articles and the most recent/oldest reference articles. The data from Table 6 shows that there is no major difference in self-citing and citing others in the most recent reference articles. The most recent self-citing references are more current than those citing others, except in the years of 2004, 2007, and 2008. However, for articles with the oldest references, the publication lag is more obvious between self-references/citing of others. In each year studied, the oldest non-self-citations are much older than self-citing references, with a discrepancy of more than 20 years.

Table 5 Publication lag between original articles and reference articles

Publication year of OA	SRA/OA		ORA/OA		SRA/ORA	
	Publication lag		Publication lag		Publication lag	
	Mean	Median	Mean	Median	Mean	Median
1999	5.17	5.11	12.25	9.47	7.08	4.36
2000	4.18	4.04	10.83	8.80	6.65	4.76
2001	4.65	4.33	10.87	9.08	6.22	4.75
2002	4.81	4.52	11.55	9.13	6.74	4.6
2003	4.35	3.97	11.31	9.12	6.96	5.15
2004	4.96	4.68	12.22	9.80	7.26	5.12
2005	4.68	4.19	11.18	8.67	6.5	4.48
2006	4.40	3.96	11.81	9.12	7.41	5.16
2007	5.22	4.90	11.66	9.63	6.44	4.73
2008	4.90	4.34	11.06	8.78	6.16	4.43

OA original article, *SRA* self-citing article, *ORA* citing-others article

Table 6 Publication lag between original articles and the most recent/oldest reference articles by year

Publication year of OA	PY of most current reference			PY of oldest reference		
	SRA	ORA	Difference of SRA/ORA	SRA	ORA	Difference of SRA/ORA
1999	1.85	1.9	0.05	9.02	39.52	30.50
2000	1.78	2.13	0.35	7.30	31.73	24.43
2001	1.82	2.1	0.28	8.47	30.10	21.63
2002	2.02	2.13	0.12	8.77	34.13	25.37
2003	1.88	2.27	0.38	7.92	34.85	26.93
2004	2.75	2.33	−0.42	8.18	37.80	29.62
2005	1.60	2.33	0.73	9.25	36.98	27.73
2006	1.67	2.12	0.45	8.37	37.95	29.58
2007	2.17	1.67	−0.50	9.20	37.17	27.97
2008	1.97	1.67	−0.30	9.35	36.13	26.78

The time lag between original articles and the oldest self-citing references is within a period of 10 years, and a discrepancy within more than 30 years is found between original articles and the oldest citing other's references. There is no considerable change in the time lag between original articles and the oldest references whether for reference of self-citing or citing others' annually. When compared to the oldest self-citing references, the oldest references citing others are usually much older than original articles, and their annual changes are also relatively higher. Similar to the results of much prior research (Aksnes 2003; Al-Qallaf 2003; Bonzi and Snyder 1990, 1991; Glänzel et al. 2004; Lawani 1980; Tagliacozzo 1977), this study also confirms that the publication years of self-citing references are more recent than those references citing others.

In the self-citing references, the range of publication years between most recent and the oldest references are 5.43 and 7.65 years; in the references citing others, the publication years between the most recent and the oldest references are 28 and 37.62 years. It suggests that the range of the publication years between the most recent and the oldest reference articles is four times higher for references citing others than for self-citing references.

Discussion and conclusions

Regarding to the relationship between the number of co-authors and the rate of author self-citing, the findings of this study have shown that these two are statistically correlated, but the correlation is weak. The number of co-authors and the author self-cited rate are also statistically correlated, and again, the correlation is weak. Of particular note is that the value of the correlation coefficient between the number of co-authors and the author self-citing rate is a little bit higher than that between the number of co-authors and the author self-cited rate. This infers that the number of co-authors has a slightly stronger correlation with the self-citing rate than the self-cited rate. Therefore, the omission of author self-cited times is questionable when calculating the indicators for evaluation.

Concerning the number of co-authors and the rate of self-citing/cited, there is only a low correlation between the number of co-authors and the rate of self-citing/cited. Therefore, it should not be inferred that the number of co-authors is the direct reason for any increase in

the rate of self-citing/cited. In the research conducted by Persson et al. (2004), they found that articles with more co-authors do have more references, but there is no direct evidence of a link between author self-citations and the high proportion of reference articles. With respect to the self-cited rate, it is reported that an article with more co-authors would have a higher self-cited rate, but increase is much lower than for a high citation rate (Glänzel and Thijs 2004). We can, though, be certain that the findings of this study showed co-authored articles have higher self-cited rates than single-authored ones. It might be that an author of co-authored works could base new papers on his/her prior collaborative works, and thus increase the self-cited rate for co-authored articles. In the case of single-authored articles, there is only one author in single-authored articles who could use prior works to produce new papers, and thus the self-cited rate in single-authored articles is relatively lower than that in co-authored ones. However, more co-authors does not absolutely lead to a high self-citing/cited rate for various reasons, including research directions, former research topics, or other social reasons such as seniority of authors and social hierarchy, as suggested by Tagliacozzo (1977), Glänzel and Thijs (2004), and Schubert et al. (2006).

On an average, self-citing references are more recent than the references citing others. However, the publication years of the most recent references are close between self-citing and non-self-citing references. Over several years, some most recent citings of others have even more current than self-citing references. This may result from the prevalence of electronic journals and electronic editing, which shorten the process of scholarly publication and diminish the advantage for authors of to access to their newly-published works. Consequently, the gap in publication year for the latest references between self-citing references and citing others has gradually narrowed and is currently very small. In contrast, every author is limited by their own research experience, so the oldest self-citing references are more recent than the oldest references citing others. This also reflects a smaller range of publication years for self-citing references as well as a larger range for references citing others.

Digital scholarly publications such as electronic journals and books are gaining currency. Do they have an influence on journal self-citations and author self-citations? A comparison study is suggested, focusing on different journals in a certain field over a long period of time and examining the effect of the digitalization of scholarly resources on self-citations. With the rise of open access, citation behavior has been affected by open-access journals (Mukherjee 2009). Since open-access journals are generally wanting in authoritative indexing and abstracting tools, they may easily fail to be found in traditional research approaches; in such circumstances, will this make a difference to citation behavior in terms of citations of the self and others? On the other hand, many open-access journals can be directly obtained via the Internet, which makes them more visible than subscription journals. Further discussion of those factors of may show the details of its interplay with author self-citations.

References

Aksnes, D. W. (2003). A macro study of self citation. *Scientometrics, 56*(2), 235–246.

Al-Qallaf, C. L. (2003). Citation patterns in the Kuwaiti Journal Medical Principles and Practice: The first 12 years, 1989–2000. *Scientometrics, 56*(3), 369–382.

Anonymous. (2003). Citing self. *Science, 300*(5616), 47.

Avkiran, N. K. (1997). Scientific collaboration in finance does not lead to better quality research. *Scientometrics, 39*(2), 173–184.

Barrios, M., Borrego, A., Vilaginés, A., Ollé, C., & Somoza, M. (2008). A bibliometric study of psychological research on tourism. *Scientometrics, 77*(3), 453–467.
Bonzi, S., & Snyder, H. W. (1990). *Patterns of self citation across fields of Inquiry. The 53 Annual Meeting of the American Society for Information Science* (pp. 204–207). Medford, NJ: Learned Information.
Bonzi, S., & Snyder, H. W. (1991). Motivations of citation: a comparison of self citation and citation to others. *Scientometrics, 21*(2), 245–254.
Gami, A. S., Montori, V. M., Wilczynski, N. L., & Haynes, R. B. (2004). Author self-citation in the diabetes literature. *Canadian Medical Association Journal, 170*(13), 1925–1927.
Garfield, E. (1979). *Citation Indexing: Its theory and application in science, technology, and humanities*. Philadelphia: ISI Press.
Glänzel, W. (2008). Seven myths in bibliometrics. About facts and fiction in quantitative science studies. Paper presented at the Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting. Retrieved April 29, 2009, from http://www.collnet.de/Berlin-2008/GlanzelWIS2008smb.pdf.
Glänzel, W., & Thijs, B. (2004). Does co-authorship inflate the share of self-citations? *Scientometrics, 61*(3), 395–404.
Glänzel, W., Thijs, B., & Schlemmer, B. (2004). A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics, 59*(1), 63–77.
Hellsten, I., Lambiotte, R., Scharnhorst, A., & Ausloos, M. (2007). Self-citations, co-authorships and keywords: A new approach to scientists' field mobility? *Scientometrics, 72*(3), 469–486.
Herbertz, H. (1995). Does it pay to cooperate? A bibliometric case study in molecular biology. *Scientometrics, 33*(1), 117–122.
Hyland, K. (2003). Self-citation and self-reference: credibility and promotion in academic publication. *Journal of the American Society for Information Science and Technology, 54*(3), 251–259.
Katz, J., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics, 40*(3), 541–554.
Lawani, S. M. (1980). *Quality, collaboration and citations in cancer research: a bibliometric study*. Unpublished doctoral dissertation. Michigan University, Ann Arbor, MI.
Leimu, R., & Koricheva, J. (2005). Does scientific collaboration increase the impact of ecological articles? *BioScience, 55*(5), 438–443.
Moed, H. F. (2005). *Citation analysis in research evaluation* (Vol. 9). Netherlands: Springer.
Mukherjee, B. (2009). Do open-access journals in library and information science have any scholarly impact? A bibliometric study of selected open-access journals using Google Scholar. *Journal of the American Society for Information Science and Technology, 60*(3), 581–594.
Narin, F., & Olivastro, D. (1986). National trends in physics and technology. *Czechoslovak Journal of Physics, 36*(1), 101–106.
Pendlebury, D. (2009). The use and misuse of journal metrics and other citation indicators. *Archivum Immunologiae et therapiae Experimentalis, 57*(1), 1–11.
Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics, 60*(3), 421–432.
Phelan, T. J. (1999). A compendium of issues for citation analysis. *Scientometrics, 45*(1), 117–136.
Pichappan, P., & Sarasvady, S. (2002). The other side of the coin: the intricacies of author self citations. *Scientometrics, 54*(2), 285–290.
Rousseau, R. (2001). Are multi-authored articles cited more than single-authored ones? Are collaborations with authors from other countries more cited than collaborations within the country? A case study. In F. Havemann, R. Wagner-Döbler, & H. Kretschmer (Eds.), *Proceedings of the second Berlin workshop on scientometrics and informetrics* (pp. 173–176). Berlin: Gesellschaft für Wissenschaftsforschung.
Schubert, A., Glänzel, W., & Thijs, B. (2006). The weight of author self-citations. A fractional approach to self-citation counting. *Scientometrics, 67*(3), 503–514.
Smart, J. C., & Bayer, A. E. (1986). Author collaboration and impact: A note on citation rates of single and multiple authored articles. *Scientometrics, 10*(5–6), 297–305.
Snyder, H. W., & Bonzi, S. (1998). Patterns of self-citation across disciplines (1980–1989). *Journal of Information Science, 24*(6), 431–435.
Tagliacozzo, R. (1977). Self-citations in scientific literature. *Journal of Documentation, 33*(4), 251–265.
University of Houston—Cullen College of Engineering. (2006). In Memoriam: Frank M. Tiller (1917–2006). Retrieved March 26, 2009, from http://www.egr.uh.edu/parameters/spring2006/?e=tiller.
van Raan, A. F. J. (1998). The influence of international collaboration on the impact of research results—some simple mathematical considerations concerning the role of self-citations. *Scientometrics, 42*(3), 423–428.
White, H. D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology, 52*(2), 87–108.