Citer Analysis as a Measure of Research Impact: Library and Information Science as a Case Study

Isola Ajiferuke[1] and Dietmar Wolfram[2]

[1] *iajiferu@uwo.ca*
Faculty of Information and Media Studies, University of Western Ontario, London, ON (Canada)

[2] *dwolfram@uwm.edu*
School of Information Studies, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI (U.S.A)

Abstract

The investigators studied author research impact using the number of citers an author's research is able to attract, as opposed to the more traditional measure of citations. A focus on citers provides a complementary measure of an author's reach or influence in a field, whereas citations, although possibly numerous, may not reflect this reach, particularly if many citations are received from a small number of citers. In this exploratory study, Web of Science was used to tally citer and citation-based counts for 25 highly cited researchers in information studies in the United States and 26 highly cited researchers from the United Kingdom. Outcomes of the tallies using several different measures, including an introduced ch-index, were used to determine whether differences arise in author rankings when using citer-based versus citation-based counts. The findings indicate a strong correlation between some citation and citer-based measures, but not with others. The findings of the study have implications for the way authors' research impact may be assessed.

Introduction and Literature Review

The assessment of research productivity and influence is an important component of evaluative bibliometrics. The impact of an author's research traditionally has been assessed through various measures such as numbers of refereed journal articles and citations received. Measures such as an impact factor (Garfield & Sher, 1963), and the more recent h-index (Hirsch, 2005), have been developed based on the numbers of citations a given entity (journal, author, article, department) has received. These measures have become important tools for assessing research impact.

The act of citing acknowledges the contributions made in previous works that the citing work builds on. Citation counts have long been controversial because of the varied reasons for citer motivation (Garfield, 1962; Smith, 1981; Brooks, 1985; MacRoberts & MacRoberts, 1989; White & Wang, 1997), the way citations may be counted (Cronin & Overfelt, 1993), and because citations may not address the quality of a research contribution (Lindsey, 1989; Phelan, 1989). Also, as MacRoberts and MacRoberts have noted, citing behaviour can be biased, and self-citation (i.e., when an author cites her or himself) is common (Snyder & Bonzi, 1998). In the end, the act of citing is a subjective decision made by the citing author(s). Despite these issues, a citation remains an important unit of measure because it defines a relationship between works and authors. Overall, high citation counts are assumed to be evidence of the influence or significance of a work or author associated with the work.

One may ask whether a simple count of citations serves as an unbiased indicator of the reach of an author's research given that the same work may be cited by the same people multiple times. Similarly, citations received by an author may be concentrated in a subset of an author's body of work. The h-index (Hirsch, 2005), tries to take into account this concentration by focusing on the number of publications by an author and the numbers of citations received by those publications. Still, this measure does not address the origin of the citations (i.e., citers) or their frequency.

White (2001) has noted that citation studies largely focus only on cited authors and not citers. His study revealed that the practice of reciting an author was common, with 49% or more of

Proceedings of ISSI 2009, pages 798–808.
Edited by B. Larsen and J. Leta.

citations constituting recitations for six of the eight prominent information scientists examined. The question then arises whether the reach of an author's research is more accurately determined by the number of citations received (tokens) by an author, or the number of people (types) who have cited and been influenced by a given author's work.

In this study, we propose an author research impact analysis based in part on citation counts, but more importantly on the number of unique individuals who have cited a given author. In this sense, our research is one of citer, as opposed to citation, analysis. The rationale for exploring the potential value of this perspective can be seen by example. Two prolific scholars may cite one another frequently. This can result in high citation counts for both scholars. However, do these counts represent a broad reach of each scholar's research? Similarly, is the impact of a scholar who receives one citation each from ten different authors equivalent to a scholar who receives ten citations to the same publication from a single author? We propose that the larger the number of people who have been influenced by a work--that is who have cited a work--the more influential that work is. This is comparable to Goffman's information epidemic theory model of information transmission, where a citation represents an "infection" of the author's (i.e., carrier's) work (Goffman & Newill, 1964; Goffman, 1966). Once infected, however, the citer is not re-infected by the same carrier, and therefore additional citations to the same work should not be counted. But if the carrier has contributed multiple ideas in the form of different publications that are each cited, then one can consider the citation of other publications as a re-infection with a new strain of the author's ideas.

Other than the study by White (2001), which points to the need for research with a citer focus, and an earlier study by Dieks and Chang (1976), which mathematically modelled citation processes and included the number of citers of papers, we have been unable to find any additional literature that has quantitatively analyzed citer-based data. The purpose of the present research is to explore the idea of citer analysis and compare measures we develop for citer analysis with those based on more traditional citation analysis. As an initial investigation, the focus of the present study is on highly cited authors in the field of library and information science.

Challenges with Citer-centred Measures

The investigators recognize that citer-based analysis presents challenges similar to those found in citation analysis that arise from co-authorship--a practice that is more prevalent in the sciences and some social sciences (Ajiferuke, 1991). Co-authorship can affect how cited and citing authors are evaluated and counted. Cronin and Overfelt (1993) examined citation counting using three approaches: 1) straight counts, providing credit only to the first author of the cited publication; 2) whole counts, which provide equal credit to all authors, and; 3) adjusted or fractional counts, which provide proportional credit to cited authors. Counting only the first author denies credit to co-authors who may be equal contributors to a research study. As a result, each author of a co-authored document could receive full citer credit or fractional credit.

A similar dilemma arises when counting citers based on the number of co-authors of a citing publication. A straight count implies that only the first author is a beneficiary of the cited work's ideas, which seems arbitrary. However, whole counting methods of citers will favour fields with high levels of collaboration. As an example, if author A is cited by publication P1 (single authored) and P2 (co-authored by five individuals), one could give equal credit to each citing author, or proportionately allocate credit for co-authored publications so that both publications contribute equally to the credit received. Or, to use the Goffman metaphor, are citing authors only partially infected by ideas if the citing document is co-authored? If the same document is cited multiple times, this could represent cumulative infections through

recitation. When put into practice, the calculation of cumulative reinfections or partial citer contributions can become cumbersome and abstract. For instance, using partial counts, a citing author who is one of four co-authors on a publication is considered to contribute 0.25 of a unique citer, and in a future publication is one of three citing co-authors who recite the same publication, thereby adding another 0.33 credit by a unique citer. The contribution calculations can continue upward with recitations, but should not exceed 1.0. Clearly, this type of calculation is quite artificial in that it implies that the significance of ideas is allocated proportionately across citing co-authors when each has likely benefited to the same level as a citer of a single-authored document. One could also simplify the calculations of citers by only allowing credit for one citation by a citer for the *oeuvre*, or entire body of work, of the cited author, as White (2001) did in his study. But this approach does not take into account the breadth of contributions of the cited author. For these reasons, we believe that all citing authors should contribute equally, regardless of the level of co-authorship of the citing document, and credit should be counted for different publication contributions.

Methodology

As an initial study, we have focused on highly cited researchers in the field of library and information science (LIS). Highly cited researchers were selected because they are more readily identifiable than those with low and mid-range citation counts. Researchers included in the study comprise the 25 most highly cited American LIS researchers reported by Adkins and Budd (2006) for the time period 1999-2004 and 26 British LIS researchers with the highest h-index values as reported by Oppenheim (2007). Although h-index values may not necessarily be equivalent to the highest citation counts, Cronin and Meho (2006) have found a high positive correlation between the two measures.

All publications by each author were identified using an author search in Thomson/ISI's Web of Science (WoS). Searches were limited by known address affiliations over the study period 1987-2007. This time frame was selected based on institutional database subscription limitations. All publication types were considered except book reviews, which were discarded. The number of citations for each publication represented the tally of unique publications citing the publication of interest. For each publication receiving at least one citation, the number of citers for each citing publication was tabulated using the Analyze Results feature of WoS, organized by citing author. Self-citations were removed. The definition of self-citation was extended to include citations made by publications authored by one of the co-authors of the cited publication of interest. For example, a publication written by authors A & B that is cited by another publication co-authored by authors B & C is considered a form of self-citation, even if author A is the author of interest. Data for each author were stored and tabulated in MS Excel. Citer and citation counts that take into account full and partial credit based on the number of authors and publications were calculated for each publication, along with Pearson's r and Spearman's rho correlation coefficients comparing several measures:

Citation Count – raw number of citations received by an author

Adjusted Citation Count – number of citations received by an author minus self-citations

Citations per Publication – total number of citations received by an author divided by the number of publications by the author (including publications not receiving any citations)

Adjusted Citations per Publication – total adjusted number of citations received by an author divided by the number of publications by the author

h-index – the classic definition of x publications with at least x citations (Hirsch, 2005) is used

Adjusted h-index – based on the adjusted citation count for the author

Citer Count – number of authors who have cited a publication by given author
Citers per Publication – number of citers divided by the number of publications by an author

We also propose a citer-based equivalent to the h-index, which we refer to as a citer index or ch-index:

ch-index – an author's ch-index of value x corresponds to x publications with at least x citers

[Note: We would have liked to have used c-index, but this appears to have been proposed already in a different context (Wang, Mokhtar, & Macaulay, 2008).]

The difference between an author's h-index and ch-index value will be influenced by the level of author collaboration for citing publications and the level of recitation by citing authors. Generally, an author's ch-index value will be higher than her or his adjusted h-index value, assuming that each citing publication has one or more authors. However, it is possible for the ch-index value to fall below the h-index value for low levels of author collaboration for citing publications and high levels of recitation by those authors.

Results

Tables 1 and 2 outline the citer indices for American LIS scholars and British LIS scholars respectively. In the tables, the names of the scholars have been changed to letters to maintain anonymity. The number of citers varies from 62 to 2305 for the US scholars. Apart from the two extreme values, the distribution is evenly spread with six scholars having between 300 and 500 citers each, ten having between 500 and 1000 citers each, and eight having between 1000 and 2000 citers each. The scholar with 62 citers had 20 publications included in the analysis but with the majority of the publications yet to be cited. On the other hand, scholar AA, with 2305 citers, is highly prolific with over 100 publications included in the analysis. When we looked at the number of citers per publication, the influence of being prolific is diminished, as the distribution is more evenly spread, where five scholars had fewer than 10 citers per publication, five had values between 10 and 20, another five had values between 20 and 30, six had values between 30 and 40, and three had values above 50 citers per publication. It is interesting to note that the only scholar AB, who had more than 100 citers per publication, also had only 8 publications included in the analysis, with two of these publications having been heavily cited. This also resulted in a much lower adjusted h-index value and ranking than the citer and adjusted citation counts would indicate. No authors in this group had calculated ch-index values lower than their h-index values, although three had equal h-index and ch-index values. The remaining authors had differences of 1 to 9 for these indices.

Table 1. Citer indices for American LIS scholars

Scholar	Citer Count		Adj. Citation Count		Adjusted h-Index		Citers per Publication		ch-index	
	Num.	Rank	Num.	Rank	Num.	Rank	Num.	Rank	Num.	Rank
AA	2305	1	1083	1	17	1	21.95	15	26	1
AB	1855	2	895	3	5	23	231.88	1	6	24
AC	1748	3	1014	2	15	2	30.67	9	18	2.5
AD	1197	4	742	4	13	4.5	27.84	10	15	6.5
AE	1163	5	682	5	8	16	96.92	2	8	20.5
AF	1148	6	658	6	14	3	22.96	14	16	3.5
AG	1136	7	641	7	11	8.5	39.17	4	13	10
AH	1096	8	552	12	11	8.5	17.13	17	18	2.5
AI	1003	9	590	9	10	12	23.88	13	11	15
AJ	997	10	621	8	12	6.5	31.16	8	14	8
AK	930	11	558	10	10	12	3.54	23	15	6.5
AL	906	12	418	16	7	18.5	60.40	3	10	16.5
AM	890	13	556	11	12	6.5	6.90	22	16	3.5
AN	817	14	315	19	7	18.5	34.04	6	12	13
AO	778	15	431	15	10	12	33.83	7	12	13
AP	729	16	452	14	13	4.5	14.02	20	13	10
AQ	728	17	515	13	8	16	34.67	5	10	16.5
AR	622	18	387	17	10	12	25.92	11	12	13
AS	538	19	251	20	5	23	14.54	19	9	18.5
AT	469	20	323	18	10	12	7.00	21	13	10
AU	467	21	230	21	8	16	17.96	16	8	20.5
AV	374	22	187	23.5	6	20.5	24.93	12	7	22.5
AW	359	23	187	23.5	5	23	17.10	18	7	22.5
AX	347	24	201	22	6	20.5	2.27	25	9	18.5
AY	62	25	33	25	4	25	3.10	24	5	25

In the case of the British scholars, apart from scholar BA, the distribution of citer frequency is evenly spread with four scholars having fewer than 100 citers each, eight with between 100 and 200 citers, seven with between 200 and 500 citers, and six scholars with between 500 and 1100 citers. Scholar BA is an exceptionally prolific author with about 225 publications included in the analysis, and with one of these publications having been cited more than 1000 times and six others having been cited at least 100 times. Again, the impact of the number of publications produced by a scholar is minimized when we examined the number of citers per publication. Nine scholars had at most 5 citers per publication, seven scholars had values between 5 and 10, five scholars had values between 10 and 20, four scholars had values between 20 and 50, and Scholar BA had about 50 citers per publication. Only two scholars had ch-indices equal to their adjusted h-indices, with the remaining 24 authors having differences ranging from 1 to 26.

Table 2. Citer Indices for British LIS Scholars

Scholar	Citer Count		Adj. Citation Count		Adjusted h-Index		Citers per Publication		ch-index	
	Num.	Rank	Num.	Rank	Num.	Rank	Num.	Rank	Num.	Rank
BA	15876	1	5316	1	35	1	70.56	1	61	1
BB	1025	2	636	2	13	2	31.06	2	17	4
BC	962	3	354	5	10	4	10.46	10	19	3
BD	865	4	382	3	13	3	16.96	6	21	2
BE	677	5	359	4	9	5	13.02	8	13	6
BF	632	6	264	8	7	8	5.80	15	13	6
BG	544	7	325	6	9	6	4.39	21	13	6
BH	463	8	271	7	9	7	14.47	7	11	8
BI	337	9	177	9	6	10	4.68	19	10	9.5
BJ	267	10	98	14	6	12	22.25	3	7	16.5
BK	258	11	94	15	4	19	21.50	4	4	25.5
BL	255	12	143	10	6	11	4.47	20	9	11
BM	253	13	139	11	7	9	9.37	12	10	9.5
BN	220	14	111	12	4	18	5.12	17	8	13
BO	190	15	105	13	5	13	3.33	25	8	13
BP	187	16	86	18	4	21	11.00	9	7	16.5
BQ	169	17	89	16	5	14	8.45	13	5	22
BR	153	18	72	20	5	16	7.65	14	8	13
BS	147	19	64	22	3	24	21.00	5	5	22
BT	130	20	66	21	5	17	10.00	11	7	16.5
BU	127	21	88	17	4	20	3.85	24	5	22
BV	113	22	75	19	5	15	4.35	22	7	16.5
BW	85	23	39	24	4	22	4.25	23	6	19
BX	66	24	33	26	4	23	5.50	16	5	22
BY	60	25	36	25	3	26	5.00	18	5	22
BZ	52	26	40	23	3	25	2.74	26	4	25.5

Scatter plots of the relationship between the adjusted number of citations received and the total number of citers for the American and British scholars appear in figures 1 and 2. In figure 2, the most prolific scholar has been excluded to maintain a reasonable scaling. There is a high degree of correspondence apparent between the two measures, but with more variability for more highly cited authors.

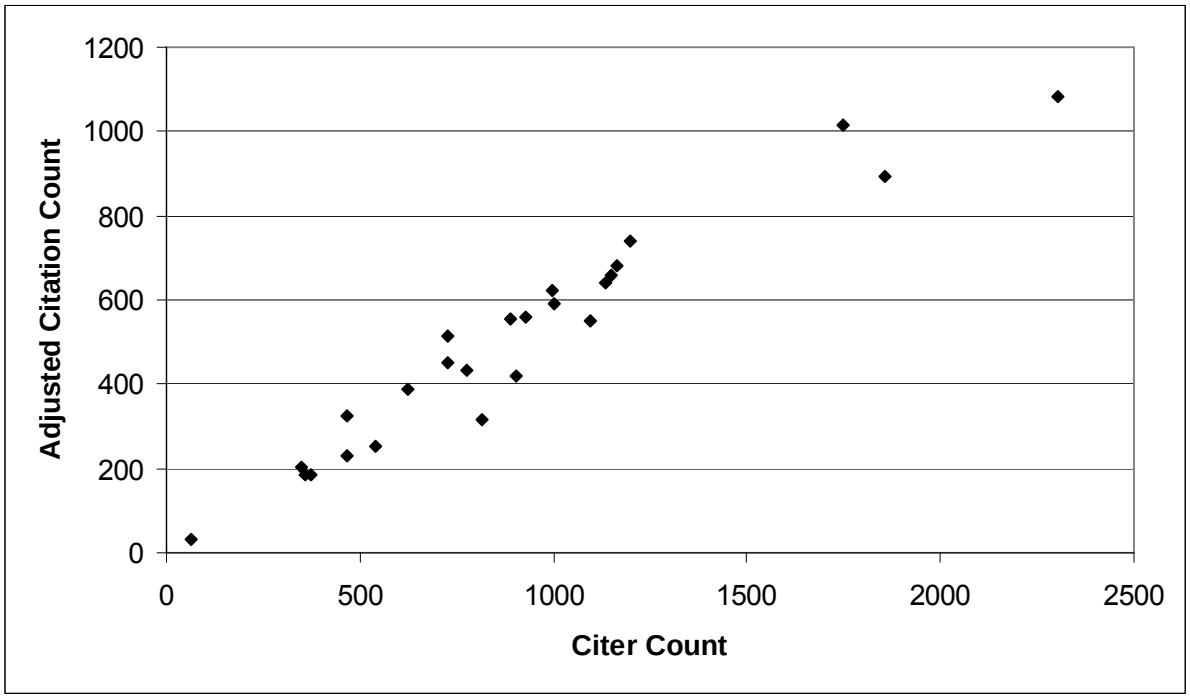

Figure 1. Scatter Plot of Adjusted Citation Count and Citer Count for American Scholars

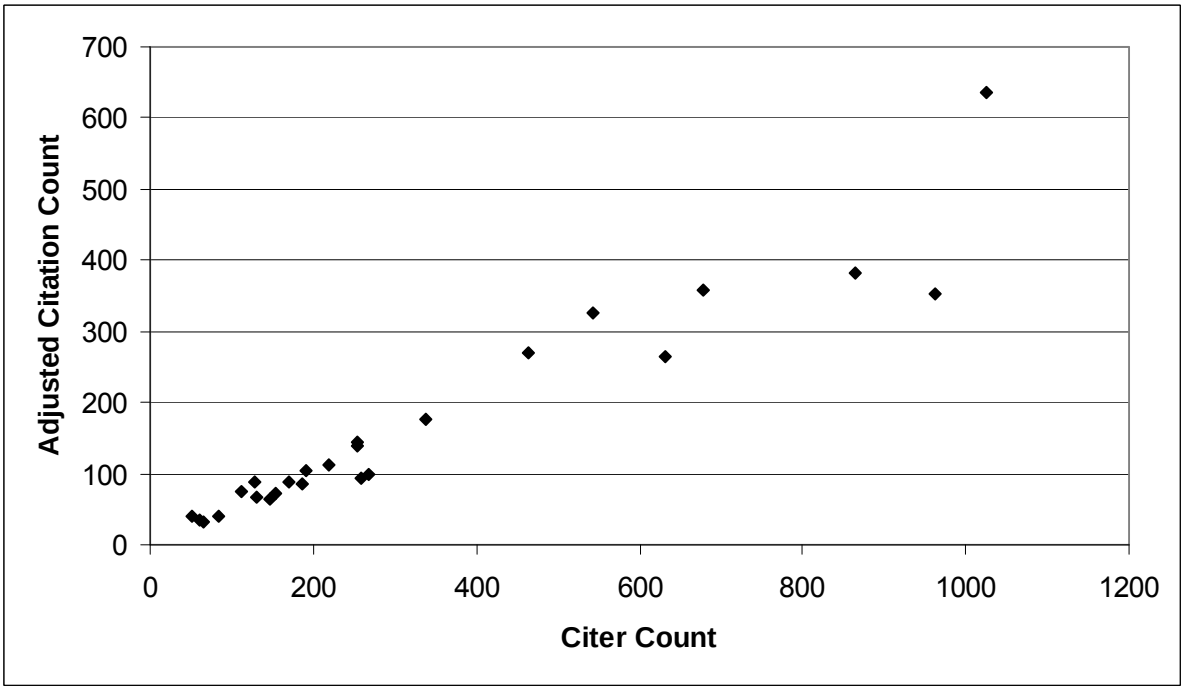

Figure 2. Scatter Plot of Adjusted Citation Count and Citer Count for British Scholars

We next examined the correlation between the three citer indices and other popular citation indices. The correlation coefficients shown in Tables 3 and 4 are mostly Spearman's rank coefficients except in a few cases where we obtained the Pearson's coefficient when none of the two distributions being correlated was skewed. For both American and British scholars, the number of citers correlated highly with the citation count and adjusted citation count, highly/moderately with the h-index and adjusted h-index, and moderately with the citations

per publication and adjusted citations per publication. The citers per publication, however, correlated highly only with the citations per publication and adjusted citations per publication. This may be due to the fact those two citation indices are rates just like the number of citers per publication.

Table 3. Correlation Coefficients between Citer Indices and other Citation Indices for American Scholars

	Citation Count	Adjusted Citation Count	Citations per Publication	Adjusted Citations per Publication	h-Index	Adjusted h-Index
Citer Count	.965 (.000)*	.965 (.000)	.573 (.003)	.574 (.005)	.638 (.001)	.624 (.001)
Citers per Publication	.395 (.050)	.445 (.026)	.978 (.000)	.968 (.000)	-.098 (.641)	.020 (.925)
ch-index	.752 (.000)	.671 (.000)	-.032 (.881)	-.064 (.762)	.937 (.000)	.893 (.000)

* Significance level in parentheses

Table 4. Correlation Coefficients between Citer Indices and other Citation Indices for British Scholars

	Citation Count	Adjusted Citation Count	Citations per Publication	Adjusted Citations per Publication	h-Index	Adjusted h-Index
Citer Count	.942 (.000)*	.960 (.000)	.574 (.002)	.431 (.028)	.838 (.000)	.883 (.000)
Citers per Publication	.410 (.037)	.430 (.028)	.922 (.000)	.923 (.000)	.270 (.182)	.440 (.024)
ch-index	.843 (.000)	.885 (.000)	.320 (.111)	.223 (.273)	.932 (.000)	0.920 (.000)

* Significance level in parentheses

Finally, given the significant correlation between the adjusted h-index and ch-index values, we wished to determine whether this relationship extended to the absolute differences between these values and the scholar citation and citer rankings. The distributions of absolute differences between calculated ch-index and adjusted h-index values for American and British LIS scholars appear in Tables 5 and 6, respectively. Although there is a significant correlation between the h-index and ch-index measures, there are sizeable differences between the absolute values they generate, which follow a Poisson-like distribution. The correlation outcomes appear in Table 7. The magnitude of the difference does not correlate significantly for American scholars, but the correlation is significant for British scholars. In both cases, the largest difference between the h-index and ch-index values was observed for the top-ranked scholar. Why this would be different for American scholars is unknown. This outcome could be influenced by differences in recitation practices or levels of collaboration for citing articles.

Table 5. Absolute differences between ch-index and adjusted h-index values for American scholars

ch-index and h-Index Difference	Number of Scholars
0	3
1	4
2	8
3	4
4	2
5	2
7	1
9	1

Table 6. Absolute differences between ch-index and adjusted h-index values for British scholars

ch-index and h-Index Difference	Number of Scholars
0	2
1	4
2	6
3	5
4	5
6	1
8	1
9	1
26	1

Table 7. Correlation comparison between ch-index and adjusted h-index differences and citer/citation counts

	Correlation Outcomes when Compared with ch-index and Adjusted h-index Differences	
	Citer Count	Adjusted Citation Count
American Scholars	0.16 (.444)*	.095 (.651)
British Scholars	0.719 (.000)	0.738 (.000)

- Significance level in parentheses

Discussion

This analysis reveals significant correlations between most measures of citers and citations, at least for the studied highly cited scholars in library and information science. The rankings of scholars based on adjusted citation counts, citer counts and adjusted h-index values show strong similarities. However, they also show the potential for large differences in rankings for

some scholars based on the concentration of citations and citers across a scholar's oeuvre. This is particularly evident with the citers per publication measure which does not correlate highly with numbers of citations or h-index values. Similarly, differences in adjusted h-index and ch-index outcomes may correlate with the citer/citation rankings of scholars. Additional investigation of this area for other fields is needed.

The results demonstrate that citer-based analysis can serve as a complement to the more traditional citation analysis. The strength of the relationships between citer and citation-based measures may vary across different fields and for scholars who have lower citation counts. Clearly, fields in which higher levels of collaboration occur will increase citer counts over citation counts. Generally, citer counts will not be lower than citation counts unless there is a low degree of author collaboration and citing authors engage in high levels of recitation. Much as it is unfair to compare raw citation counts across disciplines with very different sized research populations, it would be unfair to compare disciplines with low and high levels of collaboration based on citer counts alone. The effect of collaboration on citer-based measures warrants detailed investigation, both from the perspectives of the citing publications and the publications receiving citations.

As an exploratory study, this research has limitations. We have only investigated one field. A broader study that investigates a larger number of fields with different levels of collaboration and different citation ranks is needed to verify whether parallels are observed for the various measures. Also, all research studies are limited by the data that are available. In the case of the current study, limiting the time frame of investigation underestimates the values for prolific authors with long publication histories that pre-date the collection period.

Conclusion

In this study we have provided an initial investigation into the idea of citer analysis, where the number of citers serves as the basis of research impact assessment and not the number of citations. We believe that use of a citer (type) as opposed to a citation (token) as the unit of measure is more indicative of the reach of a scholar's research. By focusing on highly cited authors from a single field, we were able to compare and demonstrate additional measures for assessing author influence. There are strong positive correlations between citation and citer-based measures. However, this is not always the case, as the non-significant correlations between citers per publication and the more established measures of citation counts and h-index values show. The relationship between h-index and ch-index outcomes, and the influence of collaboration levels and recitation, in particular, merit further investigation across different fields.

This research serves as a foundation for undertaking future studies for the comparison of additional measures of research impact assessment on a larger scale that examines additional disciplines and not just scholars with high citation counts. Alternative assessment measures that take into account other components assessment will be undertaken. These include: 1) consideration of differential citer impact, whereby h-index or citation count values of citing authors are factored into the credit for citations received); 2) geographic reach, based on the affiliations of citing authors; 3) disciplinary reach, based on the number of journals that cite a given work within and outside an author's field, and; 4) reciter rates based on the number of authors who recite a given publication. We are currently undertaking a larger study that is investigating a range of disciplines in the sciences and social sciences.

Acknowledgements

We would like to thank Nicole Johnson for providing research assistance.

References

Adkins, D., & Budd, J. (2006). Scholarly productivity of U.S. LIS faculty. *Library and Information Science Research*, 28, 374-389.

Ajiferuke, I. (1991). A probabilistic model for the distribution of authorships. *Journal of the American Society for Information Science*, 42, 279-289.

Brooks, T.A. (1985). Private acts and public objects: An investigation of citer motivations. *Journal of the American Society for Information Science*, 36, 223-229.

Cronin, B., & Meho, L. (2006). Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57, no. 9 (July 2006): 1275-1278

Cronin, B. & Overfelt, K. (1993). Citation-based auditing of academic performance. *Journal of the American Society for Information Science*, 45, 61-72.

Dieks, D. and Chang, H. (1976). Differences in impact of scientific publications: Some indices derived from a citation analysis. *Social Studies of Science*, 6, 247-267.

Garfield, E. (1962). Can citation indexing be automated? *Essays of an Information Scientist*, 1, 84-90.

Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, 14(3), 195-201.

Goffman, W. (1966). Mathematical approach to the spread of scientific ideas—The history of mast cell research. *Nature*, 212, 449-452.

Goffman, W., & Newill, V.A. (1964). Generalization of epidemic theory: An application to the transmission of ideas. *Nature*, 204, 225-228.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102, 16569-16572.

Lindsey, D. (1989). Using citation counts as a measure of quality in science: Measuring what's measurable rather than what's valid. *Scientometrics*, 15(3-4), 189-203.

Oppenheim, C. (2007). Using the h-index to rank influential British researchers in information science and librarianship. *Journal of the American Society for Information Science and Technology*, 58, 297-301.

Phelan, T.J. (1989). A compendium of issues for citation analysis. *Scientometrics*, 45(1), 117-136.

Smith, L.C. (1981). Citation analysis. *Library Trends*, 30, 83-106.

Snyder, H., & Bonzi, S. (1998). Patterns of self-citation across disciplines (1980-1989). *Journal of the American Society for Information Science*, 48, 431-435.

Wang, W., Mokhtar, M., & Macaulay, L. (2008). C-index: Trust depth, trust breadth, and a collective trust measurement. In *WebScience'08.* Retrieved March 24, 2009 from: http://users.ecs.soton.ac.uk/dem/workshops/webscience08/papers/websci08-wang.pdf.

White, H.D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology*, 52, 87-108.

White, M.D., & Wang, P. (1997). A qualitative study of citing behavior: Contributions, criteria, and metalevel documentation concerns. *Library Quarterly*, 67, 122–154.