



Research Article

Factors Affecting Publication Impact and Citation Trends Over Time

Sandra L. De Groote

Professor and Head of Assessment and Scholarly Communications

University of Illinois Chicago, University Library

Chicago, Illinois, United States of America

Email: sgroote@uic.edu

Jung Mi Scoulas

Assistant Professor and Assessment Coordinator

University of Illinois Chicago, University Library

Chicago, Illinois, United States of America

Email: jscoul2@uic.edu

Paula R. Dempsey

Associate Professor and Head of Research Services and Resources

University of Illinois Chicago, University Library

Chicago, Illinois, United States of America

Email: dempseyp@uic.edu

Felicia Barrett

Associate Professor and Regional Head Librarian, Library of Health Science in Rockford

University of Illinois Chicago, University Library

Chicago, Illinois, United States of America

Email: fbarrett@uic.edu

Received: 17 July 2022

Accepted: 30 Jan. 2023

© 2023 De Groote, Scoulas, Dempsey, and Barrett. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

DOI: 10.18438/eblip30206

Abstract

Objective – The researchers investigated whether faculty use of the references in articles had a relationship with the later impact of the publication (measured by citation counts). The paper also reported on additional factors that may influence the later impact of publications.

Methods – This researchers analyzed data for articles published by faculty at a large public university from 1995 to 2015. Data were obtained from the Scopus abstract and citation database and analyzed using SPSS27 to conduct Pearson’s correlations and regression analysis.

Results – The number of references included in publications and the number of citations articles received each year following publication have increased over time. Publications received a greater number of citations annually in their 6th to 10th years, compared to the first 5. The number of references included in an article had a weak correlation with the number of citations an article received. Grant funded articles included more references and later received more citations than non-grant funded articles. Several variables, including number of references used in an article, the number of co-authors, and whether the article was grant funded, were shown to correlate with the later impact of a publication.

Conclusion – Based on the results, researchers should seek out grant funding and generously incorporate literature into their co-authored publications to increase their publications' potential for future impact. These factors may influence article quality, resulting in more citations over time. Further research is needed to better understand their influence and the influence of other factors.

Introduction

As noted in the Association of Research Libraries (ARL) Research Library Impact Framework, it is important to explore how the library may influence the lifecycle of research and scholarship by fostering and promoting relevant and unique research, increasing productivity, and enabling research collaborations (Association of Research Libraries, 2019). Did growing collections of e-journals over 20 years at the University of Illinois Chicago (UIC) make it easier for researchers to obtain references for their studies, leading to greater impact on their publications? And does this relationship vary across disciplines? This study explored references used in faculty publications, which are earlier works referenced in the footnotes or bibliography of publications, alongside citations, which are references to these faculty publications in later works.

A concurrent publication with a separate focus on the study data (De Groote et al., in press) showed an increase over time in the number of journals available to UIC faculty through the library’s collection, the number of publications written by faculty per year, the number of references included in the publications, and the number of publications per author. Author productivity (i.e., number of publications) was also explored in relation to the number of references used in publications. While highly prolific authors used fewer references overall in their publications, compared to productive authors, unproductive authors used the least number of references in their publications. Grant funded research included more references than unfunded research. While the number of references included in non-grant funded publications was not as high for publications that were grant funded, the number of references included in the publications

in both groups increased over time. More productive faculty members had a greater number of co-authors on their publications, and grant funded articles had more co-authors than non-grant funded articles.

For this article, researchers further analyzed data from the previous study to learn how faculty use of the literature may influence the later impact of their publications. More specifically, the researchers examined whether faculty use of references in publications has a relationship with the later impact of the publication, measured by citation counts. We also reported on additional factors that may influence the later impact of publications.

Literature Review

Faculty access to scholarly publications has changed over the past 20 years as journals moved from print to online, influencing use patterns of journals (De Groote & Dorsch, 2001; De Groote & Barrett, 2010). The number of indexing and abstracting tools increased and moved online (De Groote et al., 2007), and the number of journals available at academic institutions increased as the result of so-called big deals (Bergstrom & Bergstrom, 2004). Faculty reported that the online journals and databases allowed for easier access to more literature (Brennan et al., 2002). The density of references, measured by references per page, rose between 2001 and 2016 in articles in all disciplines except the arts and humanities, where reference density was already extremely high in 2001 and remained static (Sánchez-Gil et al., 2018).

A review of studies on factors influencing later citations focused on three categories: (1) paper-related factors (e.g., open access status of paper, number of authors, length of paper, number of references), journal-related factors (e.g., journal impact factor, language of journal, scope of journal, form of publication), and author-related factors (e.g., number of authors, funding) (Tahamtan et al., 2016). Studies on the impact of publications, as measured by their citation rates, included assessing the impact of the National Institutes of Health (NIH) public access policy on the citation rates of journal articles (National Institutes of Health, 2008). In general, NIH funded articles were cited more than non-NIH funded articles in the same journals (De Groote et al., 2015).

Some studies have concluded that the more authors were credited on an article, the more the article was cited, although there were also studies that contradicted this finding (Tahamtan et al., 2016). Bornmann et al. (2014) found that the number of authors, the number of references, and the number of pages tended to improve citation rates in a short window of time following publication. However, if multi-authored papers took longer to complete and publish, this advantage was lost. An early study reported that longer papers tended to be cited more than shorter ones, but “the larger citation rate for many-authored papers is not due to the higher citation rate for longer papers because they tend to be shorter than few-authored papers” (Abt, 1984, p. 746). Given the changing landscape for identifying and accessing the literature, it is important to also explore a potential change in citation patterns, as well as changes related to co-authorship and publication lengths.

Past studies have explored factors that potentially influenced the later impact of faculty journal articles (measured by number of citations), including the number of references included, journal prestige, co-authorship, and self-citation. Investigating the number of references included in sociology journal publications, Lovaglia (1991) found the total number of references influenced the likelihood of the articles being cited later. Researchers also observed that articles in more prestigious journals received more citations, but they were unable to conclude whether the increased citations were influenced by the prestige of the journal directly or indirectly, or whether the prestigious journals published better articles.

In contrast, a study of Malaysian review papers and 500 highly cited papers showed a positive relationship between the number of references included in the paper and the number of citations, but it was not significant (Ale Ebrahim et al., 2015). Lin and Huang (2012) examined the relationship between co-authorship and author self-citations and discovered that authors were more likely to cite their own co-authored articles compared to their sole-authored articles. Another study found that the more an author cited their own work, the more they were also cited by other scholars (Fowler & Aksnes, 2007).

Most previous studies of the number of references and later impact of the paper in terms of citations have examined articles in specific disciplines or sets of journals. However, it is critical to understand how expanding access to information, such as the increase in online journal collections through big deals, open access journals, and increased access to databases that facilitated identification to journals articles may have influenced the use of existing literature (references) in publications. Given that few studies have investigated whether the increased accessibility to online journals and databases has had an impact on the number of references included in publications and their later impact, in this study, we aimed to fill the gap by focusing on faculty from one institution over a 20-year period (1995-2015).

The primary goal of the current research project was to examine how the use of references and other variables influenced or correlated with the later impact of publications. To address these core questions, various usage statistics were collected: literature use (measured by the number of references in the publications), productivity (number of publications by faculty), publication impact (measured by the number of citations), number of co-authors, grant funding, page counts, and faculty demographic information (status and years at the institution). Our research questions were:

- In what ways do the use of references correlate with later impact of publications?
- What other variables (e.g., faculty's demographics, co-authorship, grant funding and page counts) influence later impact of publications?

Methodology

The impact of the increased access to the literature, through factors such as library big deals, open access journals, and online database on research was explored by examining the publications of UIC faculty, during the time they were at UIC, using publication data obtained from the Scopus abstract and citation database.

Types of Datasets and Procedures of Data Collection

Identification of Faculty

To explore if the publication patterns of faculty at UIC changed over time, a list of tenure system faculty members who had been at UIC for at least 5 years was requested from the Office of Institutional Research (OIR). The 5-year period is a reasonable timeframe for faculty at the university to publish and accumulate impact from their publications. Data received from the OIR included faculty rank, college, department, and number of years at UIC. Faculty located at UIC's regional locations were omitted from the study because prior to the availability of online journals, they would have had access to much smaller print collections, and thus their reference patterns prior to the online journals could have varied from those at the main university location. In addition, faculty from the arts and humanities were omitted because they typically publish in journals far less frequently than faculty in other disciplines. Also excluded were

authors in fields where publications typically involve a large number of co-authors (i.e., physics). Disciplines in the liberal arts and sciences were grouped in broader categories: chemistry, biological sciences, earth and environmental sciences, and math, statistics and computer science were recoded to *natural sciences*; sociology, economics, political science, psychology, African American (now Black) studies, communication, Latin American and Latino studies, anthropology, criminology, and law and justice, were recoded to *social sciences*. Other disciplines examined remained grouped by the main college discipline which included applied health science, business administration, medicine, dentistry, education, engineering, library, nursing, pharmacy, public health, social work, and urban planning and public affairs. We grouped faculty based on how long they had been at UIC, and we omitted older publications written by authors before they became faculty members of UIC (see Table 1).

Table 1
Publications Explored Based on Years at UIC

Years at UIC	Cut-Off for Publications Explored
5 years	No older than 2015
10 years	No older than 2010
15 years	No older than 2005
20 years	No older than 2000
25 years	No older than 1995

We searched within Scopus for each author's publications. For each publication, we recorded the author's name, number of references used in the publication, number of citations received by the article within a 5-year and 10-year timeline of publication, how many authors were involved in the publication, the length of the publication, and whether the publication was grant funded.

Procedures of Recording Publications

The list of faculty authors was divided up among the investigators. Detailed instructions were provided to each investigator and the investigators met after an initial collection of data to ensure uniformity with the process and the data. To retrieve the data from Scopus, investigators selected the *Author* tab and entered the last name and first name of the faculty member. Investigators selected the result(s) for the author if their name was a match along with the affiliation. If the investigators retrieved more than one result for an author by the same name and institution, then all were selected to obtain the full list of faculty publications. On the left-hand side of the screen, the *Year* facet was used to exclude publications outside of the date range predetermined for the faculty member. The *Document Type* facet was used to limit results to "articles." The main goal was to limit results to research articles. Limiting to "articles" did not guarantee that only research articles were included, but it did eliminate most review articles and other article types such as editorials or conference papers. Review articles were excluded because they tended to include a disproportionate number of references compared to research articles.

Next, the investigators selected and exported all publications remaining in the list. Funding details were included in addition to the bibliographic data selected by default in Scopus (authors, title, journal name, volume, issue, pages, DOI). Then, investigators copied and pasted the contents of the file into a master file. An additional column in the spreadsheet contained an assigned UIC author ID for author, so

publications by that author could be counted. Finally, investigators manually searched for each publication in Scopus and retrieved the number of references included.

Procedures of Recording Later Impact

To determine whether the number of references included in a publication had a later influence on the number of citations the article received, investigators also recorded the number of citations that an article received. This information was recorded in additional columns in the spreadsheet of publications. To obtain this information for each publication, the investigator scrolled to the top of the page for the publication and clicked on “view all metrics.” On the metrics page, self-citations were excluded to avoid the impact of authors citing their own work. To avoid issues with older articles receiving more citations because they had been around longer, we standardized the age of the citations that were counted by adjusting the date range relative to the age of the article to obtain the citations that an article had received in the last 5 years and the last 10 years (see Table 2).

Table 2
Date Range for Citations Obtained Based on the Year of Publication

Year	5 Years	10 Years
2015-2019	n/a	n/a
2014	2014 – 2018	n/a
2013	2013-2017	n/a
2012	2012-2016	n/a
2011	2011-2015	n/a
2010	2010-2014	n/a
2009	2009-2013	2009-2018
2008	2008-2012	2008-2017
2007	2007-2011	2007-2016
2006	2006-2010	2006-2015
2005	2005-2009	2005-2014
2004	2004-2008	2004-2013
2003	2003-2007	2003-2012
2002	2002-2006	2002-2011
2001	2001-2005	2001-2010
2000	2000-2004	2000-2009
1999	1999-2003	1999-2008
1998	1998-2002	1998-2007
1997	1997-2001	1997-2006
1996	1996-2000	1996-2005
1995	1995-1999	1995-2004

Data Preparation

Once the data were collected, data from each investigator were merged into two spreadsheets. The author summary spreadsheet summarized the publications of each faculty member, and the publication details spreadsheet listed the publications of each author.

The publications details spreadsheet, in addition to including author(s), title, journal name, and date, also included the author ID (assigned by UIC), the number of references included in the publications, the number of times each publication was cited in a 5- and 10-year period (for articles at least 5 years old), and the discipline of the author. The number of commas used to separate author names was tabulated using a formula in Excel plus one (because there is one fewer comma than authors in the list of authors) to indicate the number of authors in article, so that these data became available as a variable. In addition, the length of an article was determined through a formula in Excel, to subtract page start from page end, and these data became available as a variable in the spreadsheet. If some type of grant funding was acknowledged in the funding details, then in a separate column the investigators coded the article as grant funded.

The author summary spreadsheet summarized the publication information for each author. The total number of publications, total number of references, and average number of references for each author was recorded for the following time periods as appropriate: 1995-1999, 2000-2004, 2005-2009, 2010-2014, and 2015-2019. For each author, it was noted how long they were at UIC, how many publications they produced within 5-year increments, and the average number of references included in these publications. Once the spreadsheets were merged, they were further cleaned.

For the author summary spreadsheet:

- Authors who did not have a consistent publication record (i.e., there were no publications in the last 5 years of the study) were removed from the study (N=43).
- Faculty who did not have any publications were removed from the study (N=52).

For the publication spreadsheet:

- The publications of the authors removed from the author summary spreadsheet were also removed from the study.
- In some cases, Scopus did not include the number of references that an article received. These publications were dropped from the study and the author's publication number updated in the author summary spreadsheet. This occurred primarily with publications published between 1995 and 1999. A total of 390 articles were removed for this reason. Some articles had multiple authors being faculty members of UIC, which created 4361 duplicate records when we retrieved articles for each of them (total publications included = 24702). Removal of duplicate publications did not impact the author summary spreadsheet.

Data Analysis

The investigators used SPSS 27 to run several statistical tests, including Pearson correlations and regression analysis. Pearson correlations were used to examine:

1. Whether the number of references in an article correlated to the number of citations the article received (later impact).
2. The relationships between the number of references and the number of citations at the discipline level.
3. Correlations between the number of references, the number of citations, and the number of authors for articles that were grant funded and articles that were not funded.

In this article, guided by Cohen's (1988) criteria, the strength of correlation of r less than .30 was considered as small, r between .30 and .49 as medium, and r between .50 and 1.0 as large. Multiple regression analysis determined which factors predicted the citations that articles later received. Additionally, scatterplots visualized the relationships between the number of references and the number of citations for the first 10 years after publication.

Results

We examined the publications of 802 faculty from the following disciplines: applied health sciences (28), business administration (28), dentistry (35), educations (23), engineering (80), library (12), medicine (322), natural science (92), nursing (20), pharmacy (36), public health (34), social science (64), social work (8), and urban planning and public affairs (20). The average references, citations, and co-authors per article over the years are displayed in Table 3.

Table 3
Publication Demographics

	1995-1999	2000-2004	2005-2009	2010-2014	2015-2019	Average/ all Years
Total publications	1178	1972	3967	6827	10758	
No. Authors studied	118	224	376	582	802	
Aver. pub/ Author	10	8.8	10.55	11.7	13.4	10.89
Average co-authors	4.5	5	5.47	6.66	12.76	6.88
Average references	29.52	35.24	37.5	40.87	44.32	37.49
Average citations first 5 years	10.64	12.65	13.76	14.26	n/a	12.83

Between 1995 and 1999, the average number of citations articles received after 5 years was 10.64 compared to 14.26 between 2010 and 2014. Over the years, articles have received increasing numbers of citations (Figure 1), when the years of the articles published were held constant, compared to averages from the past. Articles received a greater proportion of citations between 6-10 years of age, compared to their first 5 years, which demonstrated that the impact of articles were not immediate, and the greater impact will typically be observed several years after an article was published. For articles written between 1996 and 2009, the number of references showed a weak correlation with the number of citations an article received ($r(6160) = .180, p < .001$), suggesting that the number of references included in an article may contribute to the later impact of the article.

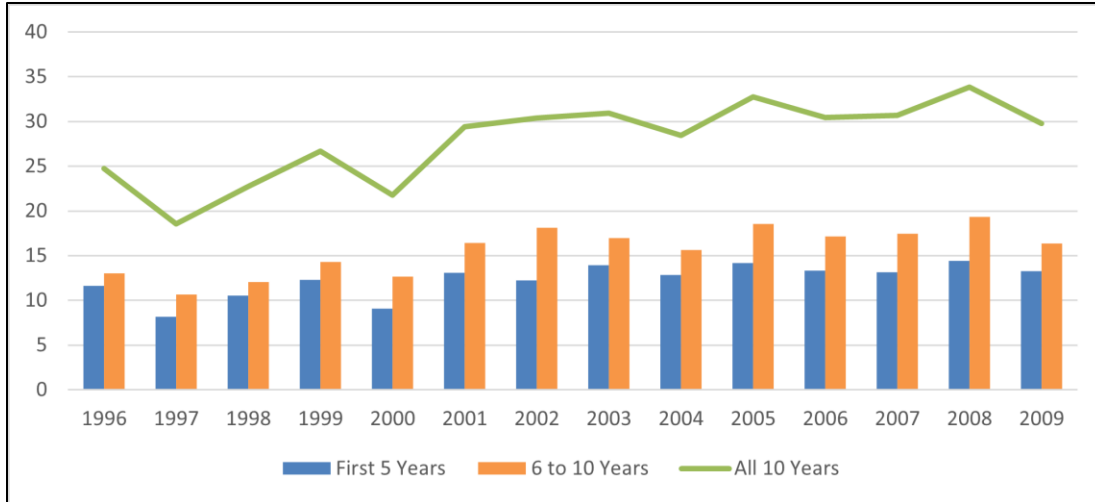


Figure 1
Average citations received by year of publication for the first 5 years, 6-10 years, and all 10 years.

Table 4
Correlations between Number of References and Number of Citations – Publications since 2000 at the Discipline Level^a

Discipline	<i>r</i>	<i>p</i>	number of publications
Applied Health Sciences	.057	.424	196
Business Administration	.553	.000**	67
College of Medicine (Chicago only)	.131	.000**	1979
Dentistry	.470	.000**	183
Education	.441	.000**	59
Engineering	.200	.000**	881
Library	.251	.386	14
Natural Sciences	.162	.000**	743
Nursing	.272	.028*	65
Pharmacy	.043	.457	304
School of Public Health	.232	.000**	284
Social Sciences	.200	.000**	362
Social Work	.354	.025*	40
Urban Planning & Public Affairs	.402	.000**	75
All Colleges	.175	.000**	5252

^a**p* < .05, ***p* < .01.

The relationship between the number of references included in an article and the subsequent citations an article received was also explored at the discipline level. There was a weak positive correlation between the number of references included in publications and the later impact of the article in all disciplines except for Applied Health Sciences, the University Library, and Pharmacy (Table 4). This suggested that in most of the disciplines, more references in a publication may have some influence on the number of citations later received by the publication.

The number of citations an article received had a weak correlation with the number of authors ($r(5252) = .136, p < .001$). This relationship between the number of authors and citations was observed despite self-citations being excluded from the data. To further explore the relationship between the number of co-authors and references included, cited references, and page count, publication data were separated between publications with one to eight authors and nine or more authors. The average references, average citations, and average page counts were obtained for the publications over time (see Table 5). Articles with one to eight authors had higher page counts compared to articles with nine or more authors. Articles with nine or more authors had more references and were cited more than articles with one to eight authors. Overall, the number of references included in publications in both groups increased over time. While the number of citations per article with one to eight authors generally increased over time, the average number of citations per article for articles with higher co-authorship (nine or more) decreased over time (1995-1999 to 2010-2014), although these articles were still cited more than articles with less authors.

The investigators also compared the number of references included in grant funded and non-grant funded publications, as well as their later impact. Only publications published in 2010 and later were explored because of concerns with underrepresentation of grant funded articles as reflected through indexing in the Scopus database (Liu, 2020). A more in-depth explanation of grant funded articles reported through Scopus is provided in the concurrent paper (De Groote et al., in press). As a result, Scopus funding data were used to explore but not confirm publication pattern differences between grant funded and unfunded publications. Grant funded articles included more references than non-grant funded articles, and the number of citations received by grant funded articles was higher, compared to non-grant funded articles (Table 6). Grant funded articles also had more authors than unfunded articles.

Table 5
Changes in Number of References, Citations (first 5 years), Page Counts Over Time by Co-Authorship Size

	1995-1999	2000-2004	2005-2009	2010-2014	Average
1 to 8 Authors					
Total Publications	1007	1599	2798	4175	2395
Average References	29.36	39.4	37.87	40.75	36.85
Average Citations	8.36	10.55	12.31	12.32	10.89
Page Count *	8.64	9.50	9.67	9.86	9.42
9 or More Authors					
Total Publications	92	157	404	820	368
Average References	31.86	36.5	40.69	45.56	38.65
Average Citations	35.6	29.19	22.54	26.26	28.4
Page Count ^a	6.57	7.36	8.27	8.96	7.79

^a Publications without page counts and duplicates were excluded from the analysis.

Table 6

Average Number of References, Average Number of Citations, and Average Number of Authors per Publication by Funding Since 2010

Funding		Number of References	Number of Citations in First 5 Years (2010-2014)	Number of Authors
Unfunded	Mean	41.2	12.99	7.36
	N ^a	7352	3536	7352
Funded	Mean	46.3	17.04	14.6
	N	6638	2039	6638

^aNumber of articles; duplicates were excluded from the analysis.

Articles published between 2010 and 2019 showed a significant but weak negative correlation between the number of authors on a paper and its page count ($r(13880) = -.022, p < .01$). A positive correlation existed between the number of references included and the page count ($r(13880) = .330, p < .01$), and also a weak positive correlation between the page count and subsequent impact of the paper as measured through citations ($r(13880) = .059, p < .001$). Articles published between 1995 and 2009 showed a weak but positive correlation between average citations and page count ($r(6305) = .038, p < .01$), and articles published between 1995 and 2014 had a weak but positive correlation between average citations and page count ($r(6305) = .023, p < .05$).

The investigators conducted a regression analysis to determine what factors predicted the citations that articles later received. Number of references, number of authors, and whether an article was grant funded were explored as predictors of later research impact, based on the number of citations an article gets after five years. The overall model with the three predictors significantly predicts the impact of an article ($R^2 = .16, R^2_{adj} = .16, F(3, 6822) = 435.67, p < .001$). Among the three predictors, all three were significant, although grant funding was right on the cusp of being insignificant (number of references: $t = 23.32, p < .001$, number of authors: $t = 25.79, p < .001$, funding received: $t = 1.96, p = .05$). The number of references, number of authors, and grant funding all contribute to the later impact of an article.

Discussion

The results of this study reflected a change in publication patterns over time. The total number of references included in articles has increased over time from 1995 to 2019. This finding was similar to that of Sánchez-Gil et al. (2018), who observed an increase in references in publications from 2001 to 2016. It is likely that several variables related to the increase in references. Authors had direct access to more articles in more journals through big deals. In addition, the increase in open access journals also increased the number of articles directly available to authors. References may also have increased due to the increase in and availability of online databases facilitating the identification of relevant literature.

Through this study, the researchers also observed that the average number of authors on an article, the total number of page numbers in an article, and the total number of citations articles received have been increasing over time. This finding prompted the question: Does a greater number of references included

in a publication result in greater impact later? This study demonstrated a weak correlation between the number of references included in a publication and the later number of citations an article received, even when self-citations were excluded, suggesting that the number of references included in a publication may contribute to the later impact of a paper. Like the findings by Lovaglia (1991), the total number of references included in a publication appeared to have a relationship with later citations. It is speculated that the greater number of references used in a paper contributed to its overall quality, and thus influenced its later impact, rather than the direct influence being the number of references themselves.

Through this study, the researchers also demonstrated that research articles were cited more on an annual basis between 6 to 10 years than in their first 5 years, which indicated that articles need time before their true impact is known. Often, the recognition of impactful scholarship has occurred a short period of time after the publication. For example, at UIC, tenure decisions are typically made during a faculty member's sixth year, but this study indicated a researcher's greatest impact would be observed after this evaluation period. Relying on citation metrics for retention decisions for relatively new publications would not have captured the full potential of impact for the publication.

The overall number of co-authors per publication increased over time. While publications with greater numbers of authors (nine co-authors or more) received greater numbers of citations, the average number of citations received by high co-authorship articles (nine or more co-authors) decreased over time. A weak correlation also existed between the number of authors and the number of citations an article received. Also observed in this study, articles with more co-authors later receive more citations. Self-citations were excluded, so it was not the situation that there were more authors to cite their own work. However, colleagues might have been more likely to cite the work of their colleagues. Further investigation could explore if coauthored publications from singular colleges or universities are cited less than multi-institutional co-authored papers. The suggested implication was that coauthored papers from multi-institution authored articles will have a greater network of non-author colleagues from multiple institutions who could cite the work, whereas single institution authored articles would be more limited to the non-author colleagues at one institution. The increase in the average number of authors on publications over time may have been related to an increased emphasis on team science both at UIC itself and in the broader research ecosystem (Cooke & Hilton, 2015), although a general increase in the overall number of faculty at UIC may also have played a role in the increased co-authorship. From 2006 to 2019, the total number of teaching faculty at UIC increased from 1163 to 2817 faculty (Association of Research Libraries, 2021). Further investigation is needed to determine to what extent the emphasis on collaboration and the increased faculty size had on increased co-authorship.

Funded publications included more references than non-funded publications, and like the findings of De Groote et al. (2015), grant funded publications received greater citations than non-grant funded publications. It is plausible that the oversight, requirements, and accountability that accompany funding may have resulted in higher quality research. This could also mean that the literature was more thoroughly explored, which led to both more references and more high-quality publications, which in turn influenced an article's later impact. Further investigation is needed to understand this observation.

The number of references, the number of authors, and grant funding were all variables contributing to the later impact of publications. Similarly, Bornmann et al. (2014) noted that the average number of references, authors, and page numbers all tended to have a positive relationship with citation rates. A regression analysis demonstrated that the number of references, the number of authors, and whether a publication was grant-funded all played a role in predicting the later impact of an article. This implied that writing grant funded publications with multiple co-authors and references altogether could increase

the later impact of the publications. It is possible that the use of more references in an article, a greater number of co-authors, and having grant funding influenced the quality of an article, thus increasing the chances that an article was cited.

Limitations

Because Scopus results were limited to the somewhat imprecise Document Type *articles*, the data may not have included only research articles, and some relevant research articles may have been omitted. Scopus is also not fully comprehensive, as only citations from journals indexed in Scopus were captured. Because self-citations were excluded from the data collection, comparisons of citation counts with and without self-citations were not possible. In this study, the researchers only explored citations 10 years out from the publication date, and as a result, it is not known when annual citations to articles would begin to decline. Are the 6th to 10th years when an article will have its greatest impact, as supported by the results here, or does the timeline extend beyond this? Additionally, in some disciplines, such as computer science, scholarship is often published through conference proceedings rather than as journal articles. This could imply that scholarly publications were underrepresented in this study for some disciplines where their primary publications were in the form of conference proceedings. Lastly, the authors would also acknowledge that collecting data was time intensive, and so future comparative research should depend on technological solutions for compiling data when possible. This would also likely reduce the potential for human error in the data collection process.

Conclusion

In this study, the researchers examined the number of references included in articles and how that may have related to the later citation impact of the publication, through faculty publication data gathered from Scopus. Variables such as page counts, number of coauthors, and grant funding were also explored as contributors to citation impact. Articles were cited more when they were 6 to 10 years old, compared to the earlier period following publication. Over time, authors have included more references in their publications, and articles were also being cited more than they were in the past. Co-authorship also increased over time. There were several variables that correlated with the later impact of a publication, as measured by citations, including number of references used, number of co-authors, and whether an article was grant funded.

Author Contributions

Sandra L. De Groote: Conceptualization, Methodology (lead), Data curation, Formal analysis, Writing – review & editing **Jung Mi Scoulas:** Methodology, Formal analysis (lead), Writing – original draft, review & editing **Paula R. Dempsey:** Writing – review & editing **Felicia Barrett:** Writing – review & editing

Acknowledgement

The authors acknowledge the support of the Association of Research Libraries (ARL) Research Library Impact Framework initiative, which was established in 2019 with a grant from the Institute of Museum and Library Services (IMLS). Special thanks to Deborah Blecic, who contributed to data collection.

References

- Abt, H. A. (1984). Citations to single and multiauthored papers. *Publications of the Astronomical Society of the Pacific*, 96(583), 746–749. <https://doi.org/10.1086/131415>
- Ale Ebrahim, N., Ebrahimian, H., Mousavi, M., & Tahriri, F. (2015). Does a long reference list guarantee more citations? Analysis of Malaysian highly cited and review papers. *The International Journal of Management Science and Business*, 1(3), 6–15. <https://ssrn.com/abstract=2572789>
- Association of Research Libraries. (2019). *Research library impact framework initiative and pilots*. <https://www.arl.org/category/our-priorities/data-analytics/research-library-impact-framework/>
- Association of Research Libraries. (2021). *ARL statistics*. <https://www.arlstatistics.org/>
- Bergstrom, C. T., & Bergstrom, T. C. (2004). The costs and benefits of library site licenses to academic journals. *Proceedings of the National Academy of Sciences*, 101(3), 897-902.
- Bornmann, L., Leydesdorff, L., & Wang, J. (2014). How to improve the prediction based on citation impact percentiles for years shortly after the publication date? *Journal of Informetrics*, 8(1), 175–180. <https://doi.org/10.1016/j.joi.2013.11.005>
- Brennan, M. J., Hurd, J. M., Bleicic, D. D., & Weller, A. C. (2002). A snapshot of early adopters of e-journals: Challenges to the library. *College & Research Libraries*, 63(6), 515–526. <https://doi.org/10.5860/crl.63.6.515>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.
- Cooke, N. J., & Hilton, M. L. (Eds). (2015). *Enhancing the effectiveness of team science*. The National Academies Press. <https://www.ncbi.nlm.nih.gov/books/NBK310394/>
- De Groote, S. L., & Barrett, F. A. (2010). Impact of online journals on citation patterns of dentistry, nursing, and pharmacy faculty. *Journal of the Medical Library Association: JMLA*, 98(4), 305–308. <https://doi.org/10.3163/1536-5050.98.4.008>
- De Groote, S. L., & Dorsch, J. L. (2001). Online journals: impact on print journal usage. *Bulletin of the Medical Library Association*, 89(4), 372-378.
- De Groote, S. L., Scoulas, J.M., Dempsey, P.R., & Barrett, F. (in press). Faculty publication patterns over 25 years at a large public university. *College & Research Libraries*.
- De Groote, S. L., Hitchcock, K., & McGowan, R. (2007). Trends in reference usage statistics in an academic health sciences library. *Journal of the Medical Library Association*, 95(1), 23-30.
- De Groote, S. L., Shultz, M., & Smalheiser, N. R. (2015). Examining the impact of the National Institutes of Health public access policy on the citation rates of journal articles. *PLoS ONE*, 10(10), e0139951. <https://doi.org/10.1371/journal.pone.0139951>

- Fowler, J., & Aksnes, D. (2007). Does self-citation pay? *Scientometrics*, 72(3), 427–437. <https://doi.org/10.1007/s11192-007-1777-2>
- Lovaglia, M. J. (1991). Predicting citations to journal articles: The ideal number of references. *The American Sociologist*, 22(1), 49–64. <https://doi.org/10.1007/BF02691867>
- Lin, W. Y. C., & Huang, M. H. (2012). The relationship between co-authorship, currency of references and author self-citations. *Scientometrics*, 90(2), 343–360. <https://doi.org/10.1007/s11192-011-0523-6>
- Liu, W. (2020). Accuracy of funding information in Scopus: A comparative case study. *Scientometrics*, 124(1), 803–811. <https://doi.org/10.1007/s11192-011-0523-6>
- National Institutes of Health. (2008). *Public access policy*. <https://publicaccess.nih.gov/>
- Sánchez-Gil, S., Gorraiz, J., & Melero-Fuentes, D. (2018). Reference density trends in the major disciplines. *Journal of Informetrics*, 12(1), 42–58. <https://doi.org/10.1016/j.joi.2017.11.003>
- Tahamtan, I., Safipour Afshar, A., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*, 107(3), 1195–1225. <https://doi.org/10.1007/s11192-016-1889-2>