Which domains do open access journals do best in? A five-year longitudinal study

Erjia Yan[1*], Kai Li[1]
[1]College of Computing and Informatics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, U.S.A. Email: ey86@drexel.edu; kl696@drexel.edu

*Corresponding author
Email: ey86@drexel.edu; Tel: (215)895-1459; Fax: (215)895-2494

Abstract

Although researchers have begun to investigate the difference in scientific impact between closed access and open access journals, studies that focus specifically on dynamic and disciplinary differences remain scarce. This study serves to fill this gap by using a large longitudinal data set to examine these differences. Using CiteScore as a proxy for journal scientific impact, we employ a series of statistical tests to identify the quartile categories and disciplinary areas in which impact trends differ notably between closed and open access journals. We find that closed access journals have a noticeable advantage in social sciences (e.g., Business and Economics), while open access journals perform well in medical and healthcare domains (e.g., Health Profession and Nursing). Moreover, we find that after controlling for a journal's rank and disciplinary differences, there are statistically more closed access journals in the top 10%, Quartile 1, and Quartile 2 categories as measured by CiteScore; in contrast, more open access journals in Quartile 4 gained scientific impact from 2011 to 2015. Considering dynamic and disciplinary trends in tandem, we find that more closed access journals in Social Sciences gained in impact, whereas in Biochemistry and Medicine, more open access journals experienced such gains.

Introduction

In his 1988 book *The Media Lab*, Stewart Brand first proposed the famous maxim: "information wants to be free; information wants to be expensive" (Brand, 1988), p. 203–205). Revealing an important paradox about the value of information, this statement—especially the first half of it—later became a famous meme in both public and academic discourses. In the late twentieth century, stronger calls to remove barriers to access to scholarly works suggest that the sentiment behind Brand's observation was widely shared by researchers. These calls are exemplified by the development of open access journals, a trend that can be traced back to the early 1990s (Laakso et al., 2011) and the launching of the first open scientific archive, arXiv.org, in 1991 (McKiernan, 2000; Straumann, 2001), the latter of which was initiated as an automated e-print archive by Paul Ginsparg (Ginsparg, 1994, 2011). These efforts were integrated into a major movement in the early 2000s, due in large part to the Budapest Open Access Initiative and its sister initiatives, such as the Bethesda Statement on Open Access Publishing (Suber et al., 2003) and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (Stratmann, 2003).

The Budapest Open Access Initiative defined two roads to open access: self-archiving and open access journals. The first approach allows authors to archive their articles in open archives that conform to open access standards, while the second approach relies upon the creation of journals committed to open access (Chan et al., 2002). In later literature, these two pathways are described as *green* and *gold* open access, respectively (Harnad et al., 2004; Mann, von Walter, Hess, & Wigand, 2009). Each of these approaches leaves room for some variation in implementation (Laakso et al., 2011). For instance, Björk and Solomon (2012) have identified three types of gold open access (OA): Direct OA (the entire journal is published open access), Delayed OA (the latest contents are only available to paid users), and Hybrid OA (authors pay to a subscription-based journal for the content to be open access). Each of these subcategories, in turn, involves further business decisions regarding publication workflow and cost coverage.

Since the advent of the open science initiative, there has been a greater awareness of self-archiving and open access journals as means of documenting scientific progress. As open access gains in popularity, there arises a need for more open access journals—a need to which publishers have been quick to respond. A quick search in Scopus' Journal Metrics shows that there were 2,481 open access journals in 2011 and 3,380 in 2015, an increase of 36%; over the same period, the share of open access journals in the database has increased from 13% to 15%. At the same time, the rise of open access brings both ethical and practical challenges for scholars and publishers: predatory open access journals are a recognized problem for researchers (Beall, 2012), and the business model of some open access journals has also been questioned (Schroter & Tite, 2006).

This study aims to evaluate the impact of open access journals. Prior research in this vein has used article- and journal-level data to reveal the impact differences between closed access and open access journals (Björk & Solomon, 2012) or between articles published in the traditional subscription-based model and those published with an open access option (Antelman, 2004; Evans & Reimer, 2009). The majority of studies seem to suggest that open access publications, both in the form of open access journals and articles with an open access option, have an advantage over closed access publications in attracting more citations (Björk & Solomon, 2012; Eysenbach, 2006). However, since most studies included a limited number of domains in their analyses, cross-disciplinary differences in impact remain unclear. Different disciplines, as is widely recognized, have varied forms of scholarly communication: they differ notably in such areas as writing, citation practice, authorship, and funding support (Demarest & Sugimoto, 2015; Gök, Rigby, & Shapira, 2016; Hyland, 1999; Leydesdorff & Rafols, 2011; Small, 2010). To account for such differences, it is necessary to examine citation impact as it varies among both access models and disciplines. To this end, we use CiteScore, an indicator that approximates a journal's scientific impact, in conjunction with a large, dynamic data set of journal-level citation data. We then analyze this data in order to address the following questions:

- What are the differences in scientific impact between closed access and open access journals in different journal ranks and across different domains?
- What are the longitudinal characteristics of scientific impact between closed access and open access journals?

- When a journal’s rank and disciplinary differences are controlled for, what are the domains in which there are a significantly higher number of open access journals whose scientific impact grows faster than the average?

Answers to these questions allow us to gain a more systematic understanding of the journal-level scientific impact of open access journals when benchmarked in relation to their closed access counterparts. This study also makes a novel contribution by revealing the dynamic characteristics of open access journals with respect to their scientific impact.

Literature review

As open access becomes an increasingly important topic in scholarly communication, it has received more attention from the perspective of scientometrics. Numerous studies have been undertaken to compare the citation rates between OA and non-OA publications. These studies have been conducted in various knowledge domains, using different samples, methods, and criteria of comparison. Appendix Table A1 is a non-exhaustive summary of the empirical evidence on this topic, based on the matrix designed by (Swan, 2010).

As echoed in the reviews by Swan (2010) and Tennant et al. (2016), most studies in Table A1 reached the conclusion that open access would increase the impact of a study, even though it is not always the case. The outcomes of OA are subject to the level of research object (journals vs. articles), the specificity of the measurements and the knowledge domains of the papers examined. Since the research designs of these studies vary in their approach to these factors, the results of the studies are highly incomparable.

Most studies measured the scientific impact on the level of papers. Overall, the outcomes are highly variant in these studies, but most affirm the impact advantages of OA publication (Antelman, 2004; Atchison & Bull, 2015; Evans & Reimer, 2009; Eysenbach, 2006; Gargouri et al., 2010; Hajjem, Harnad, & Gingras, 2006; Harnad & Brody, 2004). Within this group of studies, knowledge domain is arguably the most important factor in determining outcome; this difference is especially well presented in the studies where domain differences are part of the research questions. Antelman (2004), for example, showed that the increases in citations in mathematics, electrical engineering, political science and philosophy are 91%, 51%, 86%, and 45%, respectively. The longitudinal study of Hajjem et al. (2006) showed that OA articles enjoy a citation advantage from 36% to 172% across a broad range of knowledge domains. Xu, Liu, and Fang (2011) found that, in contrast to other broad domain categories, OA papers in the humanities are at a disadvantage. The comparability of these studies, however, is compromised by significant methodological differences, such as the adoption or non-adoption of the Altmetrics measurements and the consideration given to selection bias (authors select their best articles to publish freely online) and/or the early view effect (preprints are published earlier than peer-reviewed papers).

Despite the limited conclusions drawn from the above-mentioned research, it is obvious that disciplinary patterns of OA advantages are reflections of how open access is interwoven into the academic workflows in these fields. For example, a number of studies have confirmed that the uptake and use of OA are highly variant among knowledge domains (Björk et al., 2010; Björk, Laakso, Welling, & Paetau, 2014; Dallmeier-Tiessen et al., 2010, 2011; Fry, Spezi, Probets, & Creaser, 2016; Gargouri, Larivière, Gingras, Carr, & Harnad, 2012; Van Noorden, 2012). Based on this consideration, we would like to extend the academic pursuit of how disciplinary differences have affected the OA advantage.

Relatively few studies about this topic have been conducted on the journal level, making it difficult to acquire a comprehensive understanding of this topic. All of these studies reviewed in Table A1 (Björk & Solomon, 2012; Cheng & Ren, 2008; McCabe & Snyder, 2014; Wohlrabe & Birkmeier, 2014) seem to suggest the existence of citation advantages for OA publications across various knowledge domains. However, more evidence is required before we can make more solid claims about the effects of open access on the impact of scientific journals. The need for such evidence is a major motivation of the present study.

Data

The data set used in this study was downloaded from Journal Metrics by Scopus (https://journalmetrics.scopus.com/). Journals were included in the study if they had been assigned a CiteScore for each year between 2011 and 2015, inclusive. Journal Metrics provides several bibliometric indicators to assess a source's impact, among which we included the following in the current study: CiteScore, quartiles, subject areas, and open access status.

CiteScore: CiteScore was proposed by Elsevier in 2016 as a measure of journals' citation impact. CiteScore is analogous to Clarivate's journal impact factor with a few differences: CiteScore uses a three-year citation window instead of a two-year window and includes all document types in its calculation, whereas in the impact factor calculation, the numerator includes citations to any type of publications and the denominator includes only publications of articles and review articles. For instance, a journal's 2015 CiteScore is calculated based on the number of citations received by this journal in 2015 to its documents published in 2012, 2013 or 2014, divided by the number of documents published in 2012, 2013 and 2014. Because all document types are included in both the numerator and the denominator, a recent *Nature* news article argued that publishers that mainly publish research articles will see a relative gain in CiteScore compared with those that publish many non-research documents (Van Noorden, 2016). A preliminary analysis by Bergstrom and West (2016) showed that certain publishers including Elsevier and Emerald have seen their journals gain in CiteScore relative to others when it is used instead of impact factor. Despite these limitations, our choice of CiteScore as a proxy for scientific impact is in line with previous research (Björk & Solomon, 2012), though we are aware of the concerns inherent in relying on a single indicator in conducting research evaluations (Hicks, Wouters, Waltman, De Rijcke, & Rafols, 2015).

To understand the relationship between CiteScore and related indicators, we plotted Figure 1. In this figure, two scatter plots were produced: one is between CiteScore and

Source Normalized Impact per Paper (SNIP) (Moed, 2011) and the other between CiteScore and SCImago Impact factor (SJR) (Falagas, Kouranos, Arencibia-Jorge, & Karageorgopoulos, 2008), both contain all 18,040 journals in this data set in a log-log scale. We see in Figure 1 that both SNIP and SJR are highly correlated with CiteScore, with correlation coefficients at 0.80 and 0.87, respectively. The coefficients and the scatter plots suggest that even though they are different indicators, overall, CiteScore produces consistent results with SNIP and SJR.

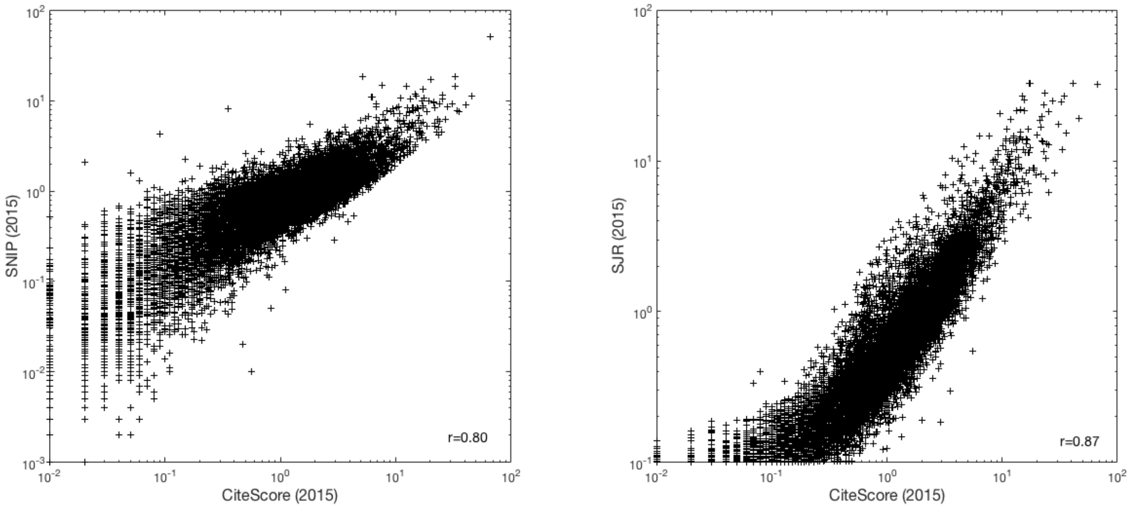

Figure 1. Scatter plots of CiteScore vs. SNIP (left) and CiteScore vs. SJR (right)

Quartiles: In ranking journals on the basis of their CiteScores, Scopus uses four quartiles plus a fifth category called “Top 10%”, which includes journals in the 99th to 90th percentile; thus, in effect, Quartile 1 includes journals between the 89th and 75th percentile[1]. A journal can be assigned to multiple quartiles if its CiteScore is positioned in different percentiles for different subject areas. For instance, *Applied Psychological Measurement* is in Quartile 2 in Psychology (miscellaneous) and Quartile 1 in Social Sciences (miscellaneous).

Subject areas: Scopus organizes journals according to a classification scheme called “All Science Classification Codes (ASJC),” which includes 27 major subject areas and 334 minor subject areas. A journal is typically assigned into one or a few minor subject areas; each minor subject area is in turn a subdivision of a major subject area. Because a journal can have multiple subject area assignments, “journal+subject” is considered as the unit of analysis for this study. For instance, we count “*Applied Psychological Measurement+Psychology (miscellaneous)*” and “*Applied Psychological Measurement+Social Sciences (miscellaneous)*” as two individual entries. This treatment, also known as multi-assignment, is consistent with prior journal-level impact assessment (Yan, 2014; Yan, Ding, Cronin, & Leydesdorff, 2013), because it avoided the arbitrariness of assigning a journal to one of its few classifications. Despite its

---

[1] Top 10% and Quartile 1 are non-overlapping in the downloadable file at https://journalmetrics.scopus.com/; however, Quartile 1 includes Top 10% journals in the searchable database.

comprehensive coverage of science, social science, and humanity domains, a few limitations of ASJC should be noted: first, biomedical journals are more heavily represented in this schema (Guerrero-Bote, Zapico-Alonso, Espinosa-Calvo, Gómez-Crisóstomo, & de Moya-Anegón, 2007); second, this schema has more detailed classification hierarchies for science domains than their social science counterparts (Yan, 2016; Yan & Zhu, 2017).

Open access status: Journal Metrics has a field to show a journal's open access status. To avoid any confounding effect, we only included journals that are consistently listed as closed access or open access from 2011 to 2015. It should be noted that it is possible that a closed access journal publishes open access articles; however, the open access status only applies to journals in this study. This operationalization is consistent with prior journal-level open access analyses, but it has limitations. One possible implication is that this treatment may lessen the impact different between closed access and open access journals: for instance, if results show that open access journals have an advantage, then in effect, this advantage may be even more noticeable, because open access papers in hybrid journals (treated as closed access journals) may actually mitigate the gap between the impact of open access and closed access journals.

It is generally believed that open access and closed access journals differ in size: open access journals—particularly those mega journals such as *PLoS ONE* and *Scientific Reports*—publish large numbers of papers than closed access journals. In this study's data set, collectively, we did not observe substantial size differences between open access and closed access journals. Open access journals published an average of 329 articles between 2012 and 2014 with a median of 136 and the numbers are 355 and 149 for closed access journals. We use Figure 2 to illustrate the cumulative distribution function (CDF) and log-logistic distribution of journal size between open access and closed access journals of this data set. Log-logistic was selected because it fits the distribution of journal size. Figure 2 shows that measured by both CDF and log-logistic distribution, the size difference between open access and closed access journals is small. Nonetheless, it should be reminded that the distribution of journal size in both groups is highly skewed, as the log-logistic distribution shows. The skewness may potentially be a confounding factor when we compare the impact of journals of various sizes—smaller journals may have different business models for financing the publishing operations, acceptance rates, and delivery methods than larger journals (Björk, 2015; Wakeling et al., 2016). Another limitation is that although we controlled the profile of journals by only including those journals that are consistently indexed as open access or closed access between 2011 and 2015, it is possible that some open access journals changed their status a few years before our observation time window (e.g., in 2009 or 2010). Such changes in open access adoption may indicate a change in journal operations, for instance, from largely volunteer work by a few scholars to a different business model that involves paid staff and dedicated communication and promotion teams. These changes will have a long-lasting impact on journal characteristics that inevitably falls into the time window of this study. These limitations in turn may affect journals' dynamic citation profiles that we aim to examine in this study.

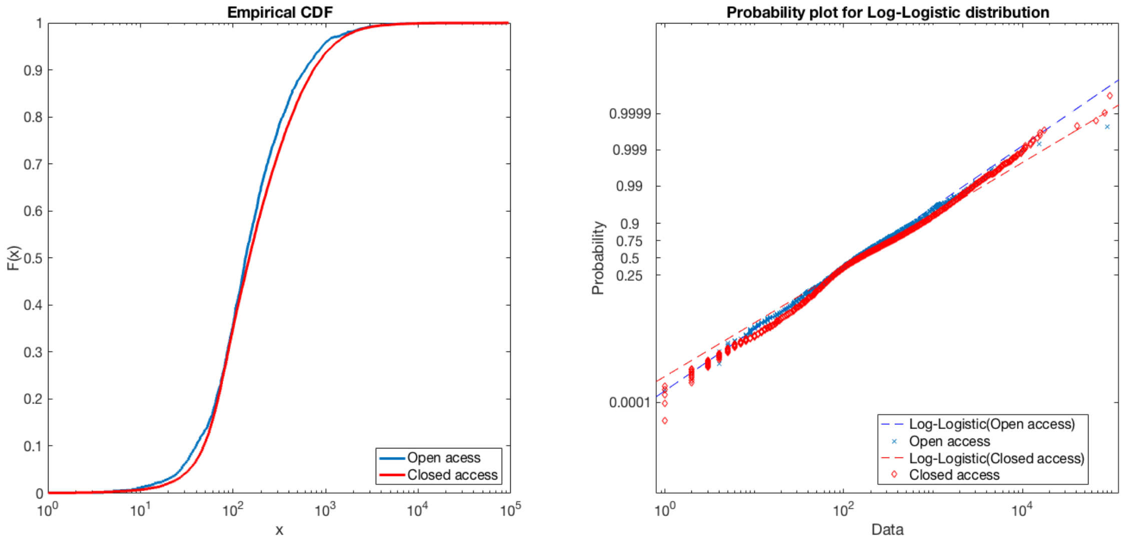

Figure 2. Distribution of journal size for open access and closed access journals

Journal Metrics includes 18,508 journals published in 2011 and 22,256 in 2015[2]. Among these, 18,040 form the basis of our study: those consistently included in Journal Metrics for each year between 2011 and 2015. Within the 18,040 journals, there are 38,941 "journal+subject" combinations, the final unit of analysis in this study. These include 34,638 closed access journal combinations (15,855 unique journals) and 4,303 open access journal combinations (2,585 unique journals). For ease of presentation, journal combinations are simply referred to as *journals* in the following paragraphs.

We use Figure 3 to show the characteristics of the data: the subgraph to the left illustrates the distribution of sources over minor subject areas and the subgraph to the right shows the distribution of CiteScores for all sources. Both distributions exhibit a power-law pattern in that a small number of subjects contain larger numbers of sources while the majority of subjects have smaller number of sources. The same pattern is true for the distribution of CiteScores in that most sources have low CiteScores while a small number of sources have high CiteScores. We also see that for both distributions, the skewness is more noticeable as the rank reached beyond 200th for subjects and about 10,000th for sources.

---

[2] Journals in this study include academic journals, book series, conference proceedings, and trade journals by Scopus. While the first three source types are traditionally seen as scholarly sources, there are also peer-reviewed sources in the trade journal type. Thus, all four types of sources were included and no distinctions was made among the source types in the current study. It should also be noted that journals of the first three source types take more than 99% of the total sources in the data set and there are less than 1% of sources that belong to trade journals.

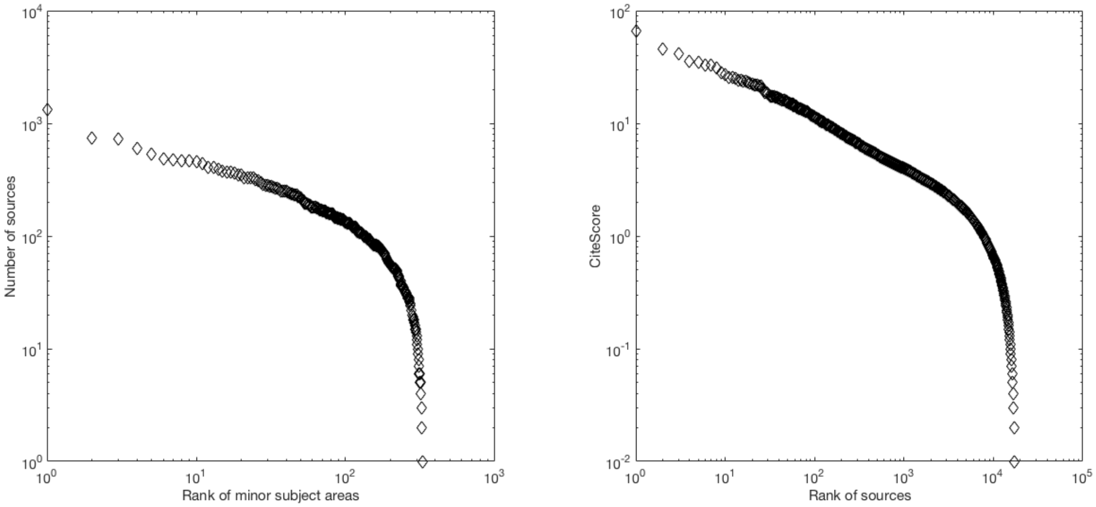

Figure 3. Distributions of sources and CiteScores

Procedures

Two sets of procedures are employed in this study. One set uses a series of visualizations and descriptive statistics to illustrate the differences in average CiteScore (i.e., mean CiteScore) between closed access and open access journals in three categories: an overall comparison that includes all journals; a quartile-level comparison that includes four quartiles and a top 10% group; and a subject-level comparison that includes 27 major subject areas. For each comparison, the mean and standard deviation are visualized for closed access and open access journals in each year between 2011 and 2015.

The second set of procedures use slopes to show the differing trends in CiteScore between closed access and open access journals. Slopes are calculated based on $y = ax + b + \epsilon$ where a is the slope, x is the time variable [1,5], and y is the CiteScore of a journal. Based on a journal's slope, journals are assigned into two groups: journals whose slopes are below the mean and journals whose slopes are above the mean. In the associated tables, *slope below mean* denotes the number of journals whose slopes are smaller than the average slope of journals in the same minor subject and in the same quartile; *slope above mean* denotes the number of journals whose slopes are larger than the average slope of journals in the same minor subject and in the same quartile.

Grouping the journals by quartile and subject allows us to control for large differences in journal rank as well as disciplinary differences in citation practices, comparing the journals only to their most similar neighbors. The specific procedures are:

- first, identify all the combinations of "minor subject area+quartile" in the data set. There are in total 1,617 combinations;
- second, for each combination, calculate the mean slope for all journals that are assigned into this combination: $slope_{i,j} = \frac{\sum_{s_x \in i,j}^{n} slope_{s_x}}{n}$, for area i and quartile j and s_x is a source in area i and quartile j. There should be 1,617 mean slopes, one representing one combination;

- third, for a source in a combination $s_x \in i, j$, use the mean slope $slope_{i,j}$ as a benchmark to see if this journal's CiteScore has increased in relation to all journals that are in the same minor subject area *i* and the same quartile designation *j*;
- last, aggregate the number of journals whose slopes are above or below the mean based on their open access status.

A series of chi-square analyses are then implemented to test whether any quartile or subject includes a significantly higher number of journals whose slopes are above or below the mean.

Results

This section first reports the trend of average CiteScore for all closed access journals and open access journals. It then shows findings on average CiteScore for closed access journals and open access journals in each quartile and each major subject area. Finally, we report a series of chi-square tests to identify quartiles and subjects in which closed access journals and open access journals performed differently, using the mean slope as a benchmark.

Figure 4 illustrates the average and standard deviation of CiteScore for 34,638 closed access and 4,303 open access journals.

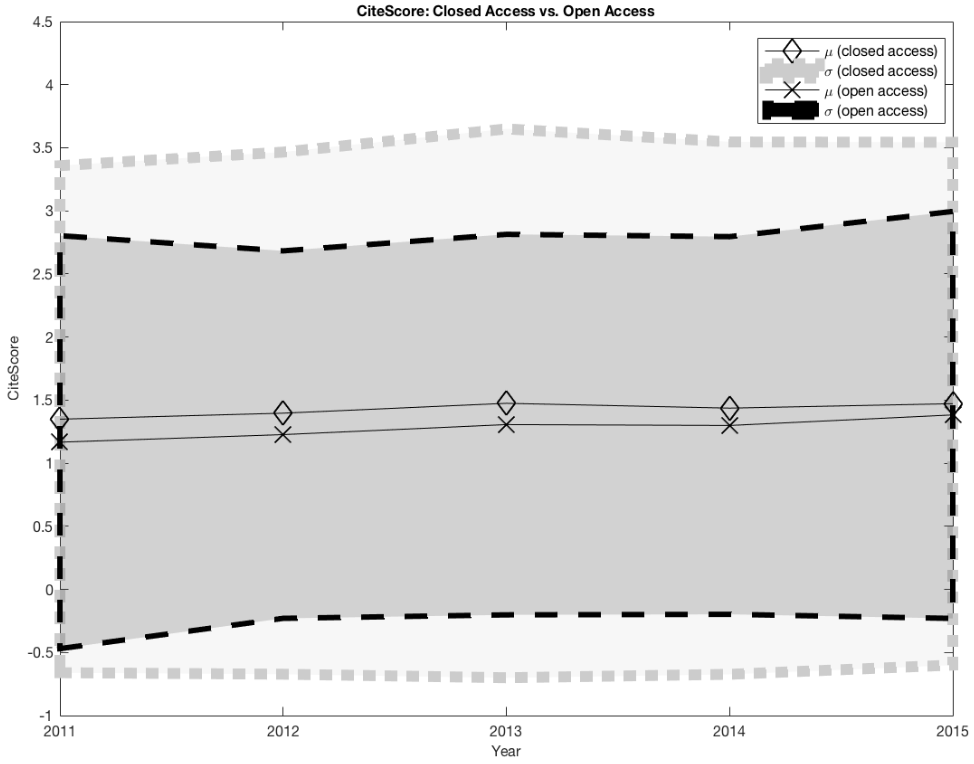

Figure 4. Average and standard deviation of CiteScore for closed access (n=34,638) and open access (n=4,303) sources between 2011 and 2015

This graph shows that, for the years 2011-2015, closed access journals had a higher average CiteScore than open access journals. The gap between these averages is, however, narrowing monotonically—from 0.18 in 2011 to 0.08 in 2015. The results demonstrate that the average scientific impact of open access journals, operationalized as CiteScore, is slightly lower than that of closed access journals, though the difference is becoming less noticeable in recent years. Because there are more closed access journals, the standard deviation of CiteScores for closed access journals is higher than that for open access journals.

Journals were then divided into four quartiles based on the CiteScore of journals in their respective minor subject area. Among the journals consistently indexed in Scopus between 2011 and 2015, there are a total of 3,972 journals in Top 10%, 9,446 in Quartile 1, 9,672 in Quartile 2, 9,873 in Quartile 3, and 5,978 in Quartile 4. The lowest share of open access journals is 7.02% in top 10; the highest is 13.95% in Quartile 3.

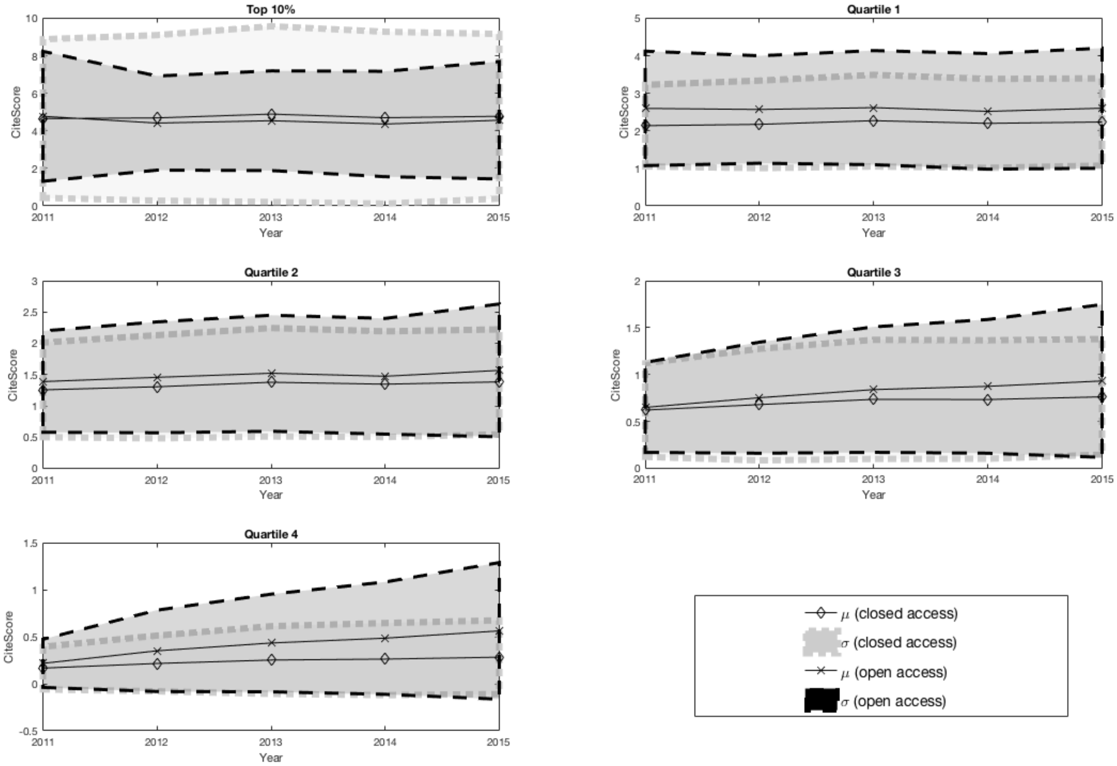

Figure 5. Average and standard deviation of CiteScores for closed access and open access sources in each quartile and Top 10% based on CiteScore between 2011 and 2015

The five subgraphs in Figure 5 shows that the average CiteScore of open access journals in all quartiles is consistently higher than that of closed access journals between 2011 and 2015. For journals in the top 10%, the average CiteScore of open access journals is higher than that of closed access journals in 2011, but closed access journals attained a higher average CiteScore from 2012 to 2015. In 2015, the average CiteScore of open access journals is 4.57; for closed access journals, the average is 4.78. Turning to dynamic trends, we see that the average CiteScore for open access and closed access journals in Quartiles 1 and 2 remained steady, but the average CiteScore of open access journals increased noticeably in Quartiles 3 and 4. In Quartile 3, the advantage of open access journals measured by average CiteScore has increased from 0.02 to 0.17; in Quartile 4, the advantage of open access journals (by the same metric) has increased from 0.05 to 0.28. The results suggest that closed access journals tend to have higher CiteScores in top percentiles, while open access journals tend to perform better in lower quartiles.

The number of open access journals in each major subject area ranges from 15 (General) to 1,126 (Medicine); the number of closed access journals in these areas ranges from 64 (General) to 5,960 (Medicine). The share of open access journals in each area varies from 4.02% (Business) to 19.15% (Neuroscience); the average share is 11%.

Table 1. Mean CiteScore for open access (OA) and closed access (CA) journals in each subject area

Subjects	2011		2012		2013		2014		2015	
	CA	OA	CA	OA	CA	OA	CA	OA	CA	OA
General	0.86	0.84	0.87	0.91	0.93	0.91	0.92	0.88	0.97	0.82
Agricultural	1.39	1.00	1.46	1.04	1.52	1.10	1.47	1.10	1.46	1.17
Arts and Humanities	0.39	0.18	0.41	0.17	0.44	0.21	0.43	0.20	0.49	0.22
Biochemistry	2.93	2.17	3.03	2.31	3.07	2.40	2.90	2.33	2.87	2.51
Business	1.07	0.38	1.10	0.38	1.21	0.44	1.22	0.38	1.37	0.43
Chemical Engineering	1.88	1.06	1.89	1.15	2.05	1.26	2.04	1.30	2.06	1.46
Chemistry	2.09	1.20	2.07	1.31	2.20	1.44	2.16	1.48	2.20	1.54
Computer Science	1.61	1.24	1.62	1.18	1.75	1.28	1.66	1.26	1.71	1.28
Decision Sciences	1.44	1.11	1.46	1.22	1.56	1.54	1.56	1.09	1.62	1.18
Earth and Planetary Sciences	1.27	0.99	1.34	1.08	1.49	1.14	1.49	1.18	1.52	1.24
Economics	0.98	0.31	0.99	0.32	1.10	0.42	1.09	0.44	1.15	0.43
Energy	1.24	0.99	1.29	1.13	1.43	1.25	1.45	1.25	1.52	1.27
Engineering	1.05	0.73	1.11	0.82	1.23	0.92	1.23	0.93	1.27	1.03
Environmental Science	1.44	1.52	1.50	1.43	1.63	1.49	1.61	1.44	1.68	1.59
Immunology and Microbiology	2.76	2.19	2.83	2.31	2.85	2.49	2.67	2.42	2.66	2.43
Materials Science	1.48	0.86	1.49	0.94	1.63	1.09	1.66	1.13	1.64	1.13
Mathematics	1.05	0.99	1.11	1.07	1.16	1.18	1.14	1.09	1.11	1.05
Medicine	1.54	1.27	1.62	1.36	1.67	1.45	1.61	1.46	1.61	1.56
Neuroscience	2.73	2.43	2.84	2.63	2.90	2.51	2.76	2.47	2.70	2.52
Nursing	0.89	0.92	0.95	1.06	1.03	1.24	0.97	1.22	1.02	1.26
Pharmacology	2.11	1.24	2.14	1.46	2.15	1.51	2.04	1.48	2.06	1.76
Physics and Astronomy	1.71	1.37	1.71	1.46	1.79	1.57	1.77	1.58	1.75	1.57
Psychology	1.57	0.60	1.63	0.71	1.71	0.71	1.68	0.75	1.72	0.75
Social Sciences	0.70	0.52	0.73	0.48	0.80	0.53	0.79	0.55	0.92	0.63

Veterinary	0.94	0.68	1.00	0.69	0.97	0.69	0.94	0.71	0.90	0.70
Dentistry	1.27	0.65	1.39	0.80	1.38	0.82	1.36	0.84	1.37	0.92
Health Professions	1.09	2.15	1.15	1.44	1.14	1.55	1.15	1.60	1.16	2.02

Three groups of subjects can be identified. The first group includes seven subjects whose open access and closed access journals have similar average CiteScores: General, Energy, Environmental Sciences, Immunology, Mathematics, Medicine, and Physics. The second group includes two subjects whose open access journals have higher average CiteScores than closed access journals (Nursing and Health Profession); the last group includes the remaining 18 subjects, whose closed access journals have higher average CiteScores than open access journals. Examples of this category include Agricultural Science, Arts & Humanities, Chemistry, and Psychology.

Within the lattermost group, the difference in scientific impact is often substantial: in 2015, closed journals in Business had an average CiteScore 220% higher than their open counterparts. For the same year, the CiteScore disparities in Economics (170%), Psychology (129%), and Arts & Humanities (121%) are also large. By way of contrast, open access journals had an average CiteScore 73% higher than closed journals in Health Profession and 22% higher in Nursing (both figures for 2015). Meanwhile, in Immunology, Medicine, and Neuroscience, the gap between the average CiteScores of open access journals and closed access journals is less than 10%. The results indicate that, in terms of CiteScore, closed access journals enjoy an advantage in social sciences, while open access journals prevail in biomedical and healthcare domains. Diachronically, the gain in scientific impact for open access journals is noticeable in Biochemistry, Immunology, and Pharmacology; the impact gap between open access and closed access journals in these domains has also become narrower.

In the remaining section, we compare open access journals and closed access journals in terms of their CiteScore trends as measured by slopes. Table 2 shows the total number of closed access and open access journals whose slopes are above or below the mean, when journals' rank and subject differences are controlled for.

Table 2. Slope comparisons between closed access and open access sources

	Slope below mean	**Slope above mean**	**Total**
Closed Access	20696 (57.75%)	13942 (40.25%)	34638
Open Access	2284 (53.08%)	2019 (46.92%)	4303
Total	**22980** (59.01%)	**15961** (40.99%)	**38941**

Among all 38,941 journals, 40.99% have a slope above the mean and 59.01% below the mean; we call these the *global statistics*. Open access journals seem to perform better in that up to 47% of open access journals have a slope above mean, compared to 40.25% for closed access journals.

We then compare the slope differences among quartiles and test whether the differences are significant using chi-square with reference to global statistics.

Table 3. Slope comparisons among different quartiles

Quartiles	**Slope below mean**	**Slope above mean**	**Total**
Top 10%*	2124 (53.47%)	1848 (46.53%)	3972
Quartile 1*	3185 (53.28%)	2793 (46.72%)	5978
Quartile 2*	5567 (56.39%)	4306 (43.61%)	9873
Quartile 3	5735 (59.29%)	3937 (40.71%)	9672
Quartile 4*	6369 (67.43%)	3077 (32.57%)	9446
Total	**22980** (59.01%)	**15961** (40.99%)	**38941**

* statistically significant at the 0.05 level (chi-square).

The overall chi-square that includes all quartiles and Top 10% is statistically significant at the 0.001 level (chi-square=436.50, df=4). With respect to individual quartiles, there is a statistical significance in Top 10%, Quartile 1, Quartile 2, and Quartile 4: compared to the global statistics, there are significantly fewer journals whose slopes are below the mean in Top 10%, Quartile 1, and Quartile 2, but significantly more journals in Quartile 4.

Our next chi-square analysis compares the differences between open access and close access journals across quartiles. Statistics for closed access and open access journals are separately used as the reference in the chi-square test.

Table 4. Slope comparisons between closed access and open access sources in different quartiles

Row Labels	**Slope below mean**	**Slope above mean**	**Total**
Closed Access	**20696** (59.75%)	**13942** (40.25%)	**34638**
Top 10%*	1966 (53.24%)	1727 (46.76%)	3693
Quartile 1*	2935 (53.16%)	2586 (46.84%)	5521
Quartile 2*	5036 (56.59%)	3863 (43.41%)	8899
Quartile 3	5018 (60.29%)	3305 (39.71%)	8323
Quartile 4*	5741 (70.00%)	2461 (30.00%)	8202
Open Access	**2284** (53.08%)	**2019** (46.92%)	**4303**
Top 10%	158 (56.63%)	121 (43.37%)	279
Quartile 1	250 (54.70%)	207 (45.30%)	457
Quartile 2	531 (54.52%)	443 (45.48%)	974
Quartile 3	717 (53.15%)	632 (46.85%)	1349
Quartile 4*	628 (50.48%)	616 (49.52%)	1244
Total	**22980** (59.01%)	**15961** (40.99%)	**38941**

* statistically significant at the 0.05 level (chi-square).

Table 4 shows that for closed access journals, there are significantly more journals whose slopes are below the mean in Quartile 4 and significantly fewer journals in Top 10%, Quartile 1, and Quartile 2. For open access journals, there are significantly fewer journals whose slopes are below the mean in Quartile 4. The results suggest that while closed access journals grow more noticeably in scientific impact in the upper percentiles, open access journals have a distinctive gain in Quartile 4.

We next examine the disciplinary differences among journals' CiteScore trends using chi-square with reference to global statistics. Subjects listed in Table 5 are indicated by the chi-square test as statistically significant. Test results for all subjects are included in the appendix table.

Table 5. Slope comparisons among different major subjects

Subjects	**Slope below mean**	**Slope above mean**	**Total**
Computer Science	1034 (56.35%)	801 (43.65%)	1835
Mathematics	840 (55.01%)	687 (44.99%)	1527
Medicine	4453 (62.84%)	2633 (37.16%)	7086
Physics	587 (55.75%)	466 (44.25%)	1053
Social Sciences	3011 (57.58%)	2218 (42.42%)	5229

The overall chi-square that includes all 27 subjects is statistically significant at the 0.001 level (chi-square=83.92, df=26). Five individual subjects are shown to be significant: there are statistically more journals whose CiteScore slopes are above the mean in Computer Science, Mathematics, Physics, and Social Sciences, but fewer in Medicine.

We use Table 6 to understand the difference in CiteScore trends between open access and closed access journals in each subject. Statistics for closed access and open access journals are separately used as the reference in the chi-square test. Subjects listed in Table 6 are statistically significant according to the chi-square test; all subjects are included in the appendix table.

Table 6. Slope comparisons between closed access and open access sources among different major subjects

	Slope below mean	**Slope above mean**	**Total**
Closed Access			
Computer Science	949 (56.76%)	723 (43.24%)	1672
Mathematics	757 (54.97%)	620 (45.03%)	1377
Medicine	3887 (65.22%)	2073 (34.78%)	5960
Pharmacology	381 (64.14%)	213 (35.86%)	594
Physics	522 (54.89%)	429 (45.11%)	951
Social Sciences	2757 (57.41%)	2045 (42.59%)	4802
Open Access			
Biochemistry	185 (47.93%)	201 (52.07%)	386
Engineering	84 (45.41%)	101 (54.59%)	185
Medicine	566 (50.27%)	560 (49.73%)	1126
Physics	65 (63.73%)	37 (36.34%)	102
Social Sciences	254 (59.48%)	173 (40.52%)	427

Six subjects are included in the closed access category and five are included in the open access category. In the closed access category, there are statistically more journals whose slopes are higher than the mean in Computer Science, Mathematics, Physics, and Social

Sciences, but fewer in Medicine and Pharmacology. In the open access category, there are statistically more journals whose slopes are higher than the mean in Biochemistry, Engineering, and Medicine, but fewer in Physics and Social Sciences. The results suggest that from 2011 to 2015, a significant number of open access journals in Biochemistry, Engineering, and Medicine have gained in scientific impact, while a significant number of closed access journals in Computer Science, Mathematics, Physics, and Social Sciences have gained in impact.

Discussion

Knowledge domains: closed access vs. open access journals

Based on the average CiteScore of closed access and open access journals in 27 subjects, we found that closed access journals have a notable scientific impact advantage in social sciences (e.g., Business, Economics, and Psychology) while open access journals have a corresponding advantage in medicine and healthcare (e.g., Health Profession and Nursing). Meanwhile, there is no apparent advantage for either closed access or open access journals in certain science domains, such as Chemistry and Energy. Dynamically, after a journal's rank and disciplinary differences are controlled for, we found that there are statistically more closed access journals that gained in scientific impact in Social Sciences, but fewer in Medicine and Pharmacology; contrariwise, there are statistically more open access journals that gained in scientific impact in Biochemistry and Medicine, but fewer in Social Sciences.

The different positions open access journals have across these disciplines may be attributed to differences in scholarly communication patterns. Scholars have found that social science disciplines tend to depend more on their own knowledge and are more self-reliant, while the disciplinary boundaries of natural science domains tend to more permeable, allowing for more interdisciplinary knowledge exchange (Yan et al., 2013). As a result, in social science domains, there may be a widely shared understanding of what journals are deemed reputable—for instance, the top journal designation in business research (Wu, Hill, & Yan, 2017); once established, these understandings are rather resistant to change. Because closed access journals form the foundation of scholarly communication in social sciences, open access journals have not had a chance to challenge their position. On the other hand, the interdisciplinary nature of many science domains has helped open access journals attain recognition as an effective scholarly communication channel.

Another possible explanation, as argued by Björk and Solomon (2012), is that in biomedical and healthcare domains, articles are more likely to be supported by grants that help cover the cost of article processing charges (APCs). Thus, researchers in these domains tend not to bear the burden of self-financing APCs; consequently, they are more likely to submit their grant-supported articles to open access journals than are social scientists, whose research often lacks funding of this type. Likewise, Evans and Reimer (2009) found that open access publication conferred no citation advantage in social science journals; personal preprint archiving, they suggested, has cancelled out the open access effect in this field.

Difference in impact trends between closed access and open access journals

When all journals are aggregated, closed access journals have a demonstrable CiteScore advantage, but this advantage is becoming less notable in recent years: the margin of average CiteScore of closed access journals over that of open access journals has decreased from 14% in 2011 to merely 6% 2015. When journals are grouped by quartile, closed access journals have an impact advantage in the top 10% category; whereas the opposite is true in all four quartiles. Dynamically, Quartiles 3 and 4 show a noticeable increase in the average CiteScore of open access journals: in Quartile 3, the advantage margin of open access journals has increased from 3% in 2011 to 18% in 2015, while in Quartile 4, the advantage margin of open access journals has increased from 24% in 2011 to 49% in 2015. The results suggest that while closed access journals tend to be more successful in top percentiles, open access journals tend to perform better than their counterparts in lower quartiles—as the outcome of this research indicates (Table 4). Dynamic analyses confirm the advantage of closed access and open access journals in different quartiles: after a journal's rank and disciplinary differences are controlled for, we find that there are significantly more closed access journals that gained in scientific impact in Top 10%, Quartile 1, and Quartile 2, but more open access journals making a notable impact gain in Quartile 4.

Conclusion

This study used a large longitudinal data set to examine the scientific impact differences between closed access and open access journals. It included 34,638 closed access journal units (15,855 unique journals) and 4,303 open access journal units (2,585 unique journals). Using CiteScore as a proxy for scientific impact, two sets of methods were designed. The first set used a series of visualizations and descriptive statistics to reveal the scientific impact differences between closed access and open access journals in five quartile categories and 27 disciplinary subjects; the second set used statistics to pinpoint the quartile categories and disciplinary subjects in which significantly more open access journals have had an impact gain, once a journal's rank and disciplinary differences are controlled for.

Using the designed methods, we found that in terms of average CiteScore, closed access journals have an advantage in top percentiles, while open access journals perform better in lower quartiles—a trend increasingly noticeable over the years included in the data set. This study also uncovered some disciplinary differences between closed access and open access journals: closed access journals had a notable advantage in social sciences (e.g., Business and Economics), whereas open access journals performed better in medical and healthcare domains (e.g., Health Profession and Nursing). Dynamically, when all journals are aggregated, closed access journals had a slightly higher average CiteScore, but this advantage is becoming less notable in recent years.

Controlling for a journal's rank and disciplinary differences, we found that over the years 2011-2015, there are statistically more closed access journals in Top 10%, Quartile 1, and Quartile 2 categories that gained in scientific impact, while more open access journals in Quartile 4 gained in impact by the same metric. At the same time, there are statistically more closed access journals in Social Sciences and more open access journals in

Biochemistry and Medicine that gained scientific impact. The results may be attributed to differences in scholarly communication practices between the social sciences and the natural sciences.

By taking a journal-level view of the differences in scientific impact between closed access and open access publications, we were able to gain a more comprehensive view of the dynamic characteristics of journal impact across different ranks and disciplines. However, journal-level analyses may overgeneralize article-level features: for instance, prior studies have used months after publication to measure the time effect on article impact, but such a feature is not available for journal-level studies. Moreover, the fact that some closed access journals publish articles with an open access option may blur the line between the two categories, which were treated as distinct in this study. In the future, a multi-level analysis—one that includes both articles and their associated journals – may offer deeper insight into the effects of open access publication.

Acknowledgement

This project was made possible in part by the Institute of Museum and Library Services (Grant Award Number: RE-07-15-0060-15), for the project titled “Building an entity-based research framework to enhance digital services on knowledge discovery and delivery”.

References

Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries, 65*(5), 372-382.

Atchison, A., & Bull, J. (2015). Will open access get me cited? An analysis of the efficacy of open access publishing in political science. *PS: Political Science & Politics, 48*(01), 129-137.

Beall, J. (2012). Predatory publishers are corrupting open access. *Nature, 489*(7415), 179.

Bergstrom, C., & West, J. (2016). Comparing Impact Factor and Scopus CiteScore.

Björk, B.-C. (2015). Have the “mega-journals” reached the limits to growth? *PeerJ, 3*, e981.

Björk, B.-C., & Solomon, D. (2012). Open access versus subscription journals: a comparison of scientific impact. *BMC medicine, 10*(1), 73.

Brand, S. (1988). *The media lab*: Penguin Books.

Chan, L., Cuplinskas, D., Eisen, M., Friend, F., Genova, Y., Guédon, J.-C., . . . Kupryte, R. (2002). Budapest open access initiative.

Cheng, W., & Ren, S. (2008). Evolution of open access publishing in Chinese scientific journals. *Learned Publishing, 21*(2), 140-152.

Demarest, B., & Sugimoto, C. R. (2015). Argue, observe, assess: Measuring disciplinary identities and differences through socio‐epistemic discourse. *Journal of the Association for Information Science and Technology, 66*(7), 1374-1387.

Evans, J. A., & Reimer, J. (2009). Open access and global participation in science. *Science, 323*(5917), 1025-1025.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS biology, 4*(5), e157.

Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., & Karageorgopoulos, D. E. (2008). Comparison of SCImago journal rank indicator with journal impact factor. *The FASEB journal, 22*(8), 2623-2628.

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PloS one, 5*(10), e13636.

Gök, A., Rigby, J., & Shapira, P. (2016). The impact of research funding on scientific outputs: Evidence from six smaller European countries. *Journal of the Association for Information Science and Technology, 67*(3), 715-730.

Guerrero-Bote, V., Zapico-Alonso, F., Espinosa-Calvo, M., Gómez-Crisóstomo, R., & de Moya-Anegón, F. (2007). Import-export of knowledge between scientific subject categories: The iceberg hypothesis. *Scientometrics, 71*(3), 423-441.

Hajjem, C., Harnad, S., & Gingras, Y. (2006). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *arXiv preprint cs/0606079*.

Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine, 10*(6).

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., . . . Hilf, E. R. (2004). The access/impact problem and the green and gold roads to open access. *Serials review, 30*(4), 310-314.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429-431.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied linguistics, 20*(3), 341-367.

Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B.-C., & Hedlund, T. (2011). The development of open access journal publishing from 1993 to 2009. *PloS one, 6*(6), e20961.

Leydesdorff, L., & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics, 5*(1), 87-100.

Mann, F., von Walter, B., Hess, T., & Wigand, R. T. (2009). Open access publishing in science. *Communications of the ACM, 52*(3), 135-139.

McCabe, M., & Snyder, C. M. (2014). Identifying the effect of open access on citations using a panel of science journals. *Economic Inquiry, 52*(4), 1284-1300.

McKiernan, G. (2000). arXiv. org: the Los Alamos National Laboratory e-print server. *International Journal on Grey Literature, 1*(3), 127-138.

Moed, H. F. (2011). The source normalized impact per paper is a valid and sophisticated indicator of journal citation impact. *Journal of the American Society for Information Science and Technology, 62*(1), 211-213.

Schroter, S., & Tite, L. (2006). Open access publishing and author-pays business models: a survey of authors' knowledge and perceptions. *Journal of the Royal Society of Medicine, 99*(3), 141-148.

Small, H. (2010). Maps of science as interdisciplinary discourse: co-citation contexts and the role of analogy. *Scientometrics, 83*(3), 835-849.

Stratmann, M. (2003). Berlin declaration on open access to knowledge in the sciences and humanities. Retrieved from https://openaccess.mpg.de/Berlin-Declaration

Straumann, T. (2001). Open source real time operating systems overview. *arXiv preprint cs/0111035*.

Suber, P., Brown, P. O., Cabell, D., Chakravarti, A., Cohen, B., Delamothe, T., . . . Hawley, R. S. (2003). Bethesda statement on open access publishing.

Swan, A. (2010). *The Open Access citation advantage: Studies and results to date*. Retrieved from https://eprints.soton.ac.uk/268516/

Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, C. H. (2016). The academic, economic and societal impacts of Open Access: an evidence-based review. *F1000Research, 5*.

Van Noorden, R. (2016). Controversial impact factor gets a heavyweight rival. *Nature, 540*(7633), 325-326.

Wakeling, S., Willett, P., Creaser, C., Fry, J., Pinfield, S., & Spezi, V. (2016). Open-access mega-journals: A bibliometric profile. *PloS one, 11*(11), e0165359.

Wohlrabe, K., & Birkmeier, D. (2014). Do open access articles in economics have a citation advantage? Retrieved from http://mpra.ub.uni-muenchen.de/56842/

Wu, C., Hill, C., & Yan, E. (2017). Disciplinary knowledge diffusion in business research. *Journal of Informetrics, 11*(2), 655-668.

Xu, L., Liu, J., & Fang, Q. (2011). *Analysis on open access citation advantage: an empirical study based on Oxford open journals.* Paper presented at the Proceedings of the 2011 iConference.

Yan, E. (2014). Finding knowledge paths among scientific disciplines. *Journal of the Association for Information Science and Technology, 65*(11), 2331-2347.

Yan, E. (2016). Disciplinary knowledge production and diffusion in science. *Journal of the Association for Information Science and Technology, 67*(9), 2223-2245.

Yan, E., Ding, Y., Cronin, B., & Leydesdorff, L. (2013). A bird's-eye view of scientific trading: Dependency relations among fields of science. *Journal of Informetrics, 7*(2), 249-264.

Yan, E., & Zhu, Y. (2017). Adding the dimension of knowledge trading to source impact assessment: Approaches, indicators, and implications. *Journal of the Association for Information Science and Technology, 68*(5), 1090-1104.