

Investigating disciplinary differences in the relationships between citations and downloads

Liwen Vaughan¹  · Juan Tang² · Rongbin Yang²

Received: 4 August 2016 / Published online: 21 February 2017
© Akadémiai Kiadó, Budapest, Hungary 2017

Abstract Several studies have examined the relationships between citation and download data. Some have also analyzed disciplinary differences in the relationships by comparing a few subject areas or a few journals. To gain a deeper understanding of the disciplinary differences, we carried out a comprehensive study investigating the issue in five disciplines of science, engineering, medicine, social sciences, and humanities. We used a systematic method to select fields and journals ensuring a very broad spectrum and balanced representation of various academic fields. A total of 69 fields and 150 journals were included. We collected citation and download data for these journals from China Academic Journal Network Publishing Database, the largest Chinese academic journal database in the world. We manually filtered out non-research papers such as book reviews and editorials. We analyzed the relationships both at the journal and the paper level. The study found that social sciences and humanities are different from science, engineering, and medicine and that the pattern of differences are consistent across all measures studied. Social sciences and humanities have higher correlations between citations and downloads, higher correlations between downloads per paper and Journal Impact Factor, and higher download-to-citation ratios. The disciplinary differences mean that the accuracy or utility of download data in measuring the impact are higher in social sciences and humanities and that download data in those disciplines reflect or measure a broader impact, much more than the impact in citing authors.

✉ Liwen Vaughan
lvaughan@uwo.ca

Juan Tang
jtang@libnet.sh.cn

Rongbin Yang
rbyang@libnet.sh.cn

¹ Faculty of Information and Media Studies, University of Western Ontario, London, ON N6A 5B7, Canada

² Consulting Group, Shanghai Library (Institute of Scientific and Technical Information of Shanghai), Shanghai 200031, China

Keywords Citations · Downloads · Disciplinary differences · Relationships between citations and downloads

Introduction

Decades of research in academic citations have contributed to our knowledge on the value and limitations of citation data. Less is known on the usage of scholarly publications. Although usage metrics are older than citation metrics (Glänzel and Gorraiz 2015), research on the former lagged behind that of the latter. The increasing availability of electronic usage data in the last decade or so prompted a growing amount of research in this area. While usage can be measured by various metrics such as library circulation (Tague and Ajiferuke 1987), library journal re-shelving data (Duy and Vaughan 2006), and bookmarks on social media (Thelwall and Wilson 2016), the most commonly studied usage data so far is downloads from online databases (papers cited below).

It is very important to distinguish different types of download data and the various ways of analyzing them so that findings from different studies can be interpreted in the right context. Download data can be collected for a particular institution (Duy and Vaughan 2003) or for all users of a particular database; the latter being more common in published research in this area. Studies have analyzed download data for a specific journal (Moed 2005), a specific discipline (Liu et al. 2011), or multiple disciplines (Gorraiz et al. 2014; Moed and Halevi 2015; Wan et al. 2010). The time frame is another variable. Some focused on a short time span after the publication of papers (e.g. Brody et al. (2006) used earlier download data to predict medium-term citation impact) while others gave a longer time span. Still others studied temporal patterns of download data (Shuai et al. 2012; Yan and Gerstein 2011).

Prior studies examined the relationships between citations and downloads but the fields studied varied greatly. The relationship is typically analyzed by correlation tests but the correlation can be calculated at the paper level or the journal level. Some studies analyzed data at both levels and compared the results from the two (Guerrero-Bote and Moya-Anegón 2014; Moed and Halevi 2016; Schlögl and Gorraiz 2010). The analysis can be done using synchronous or diachronous approaches (Moed 2005).

All these studies provided a wealth of information on the nature and characteristics of download data as well as the relationships between downloads and citations. However, findings from different studies are sometimes inconsistent and difficult to compare, mainly due to differences in data collection and data analysis as described above. We need studies that examine specific factors to gain a deeper understanding. Towards this end, we carried out a study to determine if and how the factor of discipline affects the relationships between downloads and citations. Although some previous studies compared results from different disciplines, the selection of disciplines or fields is often not comprehensive (e.g. only compared a few fields). Furthermore, the selection of journals from the disciplines or fields is often not systematic but ad hoc.

Our study covered the following five academic disciplines: science, engineering, medicine, social sciences, and humanities. We selected a broad range of academic fields (69 in total) to represent the five disciplines based on a classification scheme (details in Methodology section). For example, the discipline of science is represented by 12 fields including mathematics, physics, chemistry, astronomy, geophysics, geology, and marine science. We then selected journals to represent the fields using a systematic method (details in Methodology section). To the best of our knowledge, no prior study used a systematic

method to select such a broad spectrum of fields to examine the issue of disciplinary differences.

Shepherd (2007) proposed the idea of Usage Factor while Rowlands and Nicholas (2007) defined Usage Impact Factor parallel to Journal Impact Factor. Schlögl and Gorraiz (2010) found a moderate correlation between Usage Impact Factor and Journal Impact Factor for the 50 oncology journals in their study. We examined the correlation between downloads per paper and Journal Impact Factor for all five disciplines to determine if download data are correlated with Journal Impact Factor and if there are disciplinary differences in this regard. Earlier studies (Guerrero-Bote and Moya-Anegón 2014; Schlögl and Gorraiz 2010) found that the correlations between downloads and citations are higher when they are calculated at the journal level than at the paper level. We attempted to determine whether this is true for all five disciplines in our study, so we calculated the correlation in both ways and compared the results. Our research questions are as follows:

1. What are the differences, if any, among different disciplines in the relationships between citations and downloads?
2. Are correlations calculated at the journal level higher than that at the paper level and what is the implication of the difference?
3. What are the differences, if any, among different disciplines in the download-to-citation ratio?
4. Are downloads per paper data correlated with Journal Impact Factor data? Are there disciplinary differences in this regard?

Methodology

We used Annual Report for Chinese Academic Journal Impact Factors (Natural Science) and Annual Report for Chinese Academic Journal Impact Factors (Social Science) for the year 2014 to select journals and collect data on Journal Impact Factor (U-JIF in the Reports). Of the five disciplines that we studied, science, engineering, and medicine are included in the Natural Science volume while social sciences and humanities are included in the Social Science volume. These two annual reports classified journals into specific fields such as physics and chemistry. We used the annual reports' classification and made minor adjustments when necessary. We excluded journals that appear in more than one discipline. For example, we excluded journal *Petroleum Exploration and Development* because it appears under both science and engineering.

“Appendix” shows the disciplines and the fields in the study. It is clear that a broad spectrum of fields from the five disciplines are represented. This is important for a study that aims to compare disciplines. We selected 30 journals from each discipline. The number of journals for each field, also shown in “Appendix”, was decided based on the total number of journals in that field in the annual reports, i.e. the number of journals in the study is proportional to the total number of journals in the field.

For each field, we selected the journal(s) with the highest Journal Impact Factor scores. We chose top journals rather than randomly selected journals for two reasons. First, this makes the journals from different disciplines comparable, serving the purpose of comparing disciplines. Second, very low ranking journals may not have enough citation and download data (e.g. citations are mostly zeros) to analyze the relationships between the two variables.

For each journal in the study, we selected 50 papers published in 2011 and collected citation and download data for the papers. The year 2011 was chosen because when we

collected data in June 2016, these papers had about 5 years from publication to attract citations and downloads. Furthermore, the Journal Impact Factor data of the 2014 annual reports were calculated based on citations in 2013 to papers published in 2011 and 2012. Among the 150 journals originally selected, three did not have 50 articles in the year 2011. We replaced these journals by other journals. After filtering out non-research papers (details of filtering below), another journal did not have 50 articles and had to be replaced.

We collected citation and download data from China Academic Journal Network Publishing Database (CAJD). This is the largest full-text Chinese academic journal database in the world. The database is continuously updated covering sciences, engineering, agriculture, philosophy, medicine, humanities, social sciences etc. It is one of the databases of China National Knowledge Infrastructure (CNKI). CAJD covers over 8000 journals (CNKI 2016). Earlier studies that used CNKI databases to collect citation and download data include Liu et al. (2011) and Wan et al. (2010).

CAJD provides free searching of the database and the search results show bibliographic data such as author, title, abstract etc. The search result also displays citation and download numbers if such data are available for the item retrieved. Viewing full-text articles is available only for paid subscribers. Universities, research institutes and large public libraries in China typically have a paid subscription to the database that allows their users to have free and full access to the database.

We used the advanced search function of the database, searching by journal title and limiting the search to the year 2011. This retrieved all articles published by the journal in 2011. The default search result seemed to be sorted in descending order by the citation and download counts (i.e. papers with higher citation and download counts were presented higher in the search result), although the database did not explicitly indicate how the result was sorted. To avoid a possible bias from including papers with higher citation and download counts, we sorted the search result by publication dates. We then selected the first and last 50 articles in the search result and recorded their citation and download numbers. From this preliminary data set, we manually filtered out articles that are not research papers, which are inappropriate for the study, keeping 50 papers for each journal. The following types of articles were considered inappropriate for the study and filtered out: call for papers, editorials, editor's notes, news, book reviews, erratum, technical standards, conference announcements, agreed views on a medical issue by experts, and reports of disease statistics.

To sum up, we collected citation and download data for 7500 papers from 150 journals (50 articles from each journal). The 150 journals were from five disciplines with 30 journals from each discipline. As download data can change daily, it is important that data collection takes place in a short time period so that data for different journals are comparable. We collected all citation and download data on a single day, Saturday June 18, 2016. Filtering of articles to be excluded took place in the following days.

Results

Descriptive statistics of citation and download data

Various descriptive statistics for the citation and download data are reported in Table 1. We present the statistics for each discipline as well the statistics with the five disciplines combined. Table 1 shows that social sciences has the highest average citations per paper

Table 1 Descriptive statistics of citations and downloads

| Discipline | Citation | | | Download | | |
|---------------------------|----------|--------|------|----------|--------|-------|
| | Mean | Median | SD | Mean | Median | SD |
| Science | 11.3 | 9.7 | 8.5 | 353.6 | 283 | 339.1 |
| Engineering | 11.1 | 9.9 | 6.2 | 367.9 | 278.9 | 232.2 |
| Medicine | 8.8 | 7.9 | 4.8 | 250.5 | 211.9 | 139.6 |
| Social sciences | 20.8 | 16.2 | 15.3 | 1302.2 | 1029.6 | 955.8 |
| Humanities | 8.6 | 5.1 | 8.6 | 687 | 500.5 | 487.1 |
| Five disciplines combined | 12.1 | 9.3 | 10.3 | 592.2 | 385.5 | 639.6 |

(20.8) while that of the humanities is the lowest (8.6). Social science also has the highest average downloads per paper at 1302.2, a sharp contrast to 250.5 of medicine, the lowest.

It should be noted that the average citation and download rates reported in Table 1 are not representative of ALL journals in each discipline. This is because journals that had the highest Journal Impact Factor scores were selected for the study, as reported in the Methodology section above. The average citation and download rates would be lower if a random sample of journals were selected for the study.

Correlation between downloads and citations

For each discipline in the study, we selected 30 journals and 50 papers from each journal. The correlation between citations and downloads for each discipline can be calculated at two different levels: the paper level and the journal level. For the journal level calculation, we added up the total download and total citation counts for each journal, resulting in 30 download and citation numbers, i.e. each journal has a download and citation number. The correlation for the discipline is the correlation between these 30 download and citation numbers.

The Spearman correlation coefficients for the five disciplines are shown in the second column of Table 2. All correlations at the journal level are statistically significant ($p < 0.05$). It should be noted that although the frequency distributions of downloads and citations are not highly skewed, we used the Spearman correlation so that the comparison

Table 2 Correlation between downloads and citations

| Discipline | Journal level | Paper level | |
|---|---------------|-------------|----------|
| | | Method 1 | Method 2 |
| Science | 0.76 | 0.66 | 0.70 |
| Engineering | 0.88 | 0.71 | 0.72 |
| Medicine | 0.57 | 0.59 | 0.60 |
| Social sciences | 0.82 | 0.81 | 0.83 |
| Humanities | 0.88 | 0.67 | 0.75 |
| Average of science, engineering, medicine | 0.74 | 0.65 | 0.67 |
| Average of social sciences and humanities | 0.85 | 0.74 | 0.79 |
| Average of five disciplines | 0.78 | 0.69 | 0.72 |

between the journal level correlations and the paper level correlations, which used the Spearman correlations as reported below, is fairer.

For the paper level calculation, the unit of the analysis is a paper, i.e. each paper has a download and citation number. We calculated the correlation using two methods. In method 1, we first calculated correlation coefficients for each journal. This resulted in 30 correlation coefficients for a discipline, one for each journal. The correlation coefficient for the discipline is the average of these 30 correlation coefficients. In method 2, we still have a download and a citation number for each paper but the correlation is calculated with all 1500 papers of a discipline together, resulting in a single correlation coefficient for the discipline. In both methods of doing the paper level calculation, the frequency distributions of downloads and citations are typically skewed, so we used the Spearman correlation. The last two columns of Table 2 show the correlation coefficients of the two methods. All correlations at the paper level are statistically significant ($p < 0.05$).

It is clear from Table 2 that on average the journal level correlations are higher than that of the paper level correlations. Earlier studies (Guerrero-Bote and Moya-Anegón 2014; Schlögl and Gorraiz 2010) that analyzed specific fields reported that the journal level correlation is higher than that of the paper level. The current study extended the finding to a broad range of fields.

A more important finding is the disciplinary difference. Table 2 shows that for all three ways of calculating the correlations (represented by the three columns of data), the average correlations for social sciences and humanities are higher than the averages for science, engineering and medicine. These two groups of disciplines also differ in other ways as reported below.

Correlation between downloads per paper and Journal Impact Factor

We calculated the number of downloads per paper for each journal and correlated this number with the journal's Impact Factor. The correlation was done for each discipline. We used Pearson correlation coefficient because the frequency distributions of both variables are not skewed enough to warrant the use of Spearman correlation. The result is shown in Table 3. All five correlations are statistically significant ($p < 0.05$). The average of the five disciplines is 0.70, a strong correlation indeed. The finding that the correlations are stronger for the social sciences and humanities than for science, engineering and medicine is consistent with the disciplinary differences reported earlier regarding the correlations between citations and downloads. It should be noted that what we studied is the correlation between the Journal Impact Factor and downloads per paper, not the correlation between the Journal Impact Factor and the Usage Impact Factor as was reported in some previous studies such as Bollen and Van de Sompel (2008) and Schlögl and Gorraiz (2010). The Usage Impact Factor applies to papers published in a two-year period while papers in our study were published in a single year 2011.

Table 3 Correlation between Journal Impact Factor and downloads

| Discipline | Correlation coefficient |
|-----------------------------|-------------------------|
| Science | 0.69 |
| Engineering | 0.57 |
| Medicine | 0.42 |
| Social sciences | 0.90 |
| Humanities | 0.92 |
| Average of five disciplines | 0.70 |

Download-to-citation ratio

We further examined disciplinary differences in the download and citation relationship by analyzing the download-to-citation ratio. For each journal, we calculated the download-to-citation ratio defined as the total number of downloads divided by the total number of citations. A higher download-to-citation ratio means more downloads relative to citations. For example, a ratio of 10 means that for each citation, there are 10 downloads.

We carried out an analysis of variance test to compare the average download-to-citation ratios of the five disciplines. The test found a highly significant difference ($p < 0.01$). A Tukey’s HSD test further showed that the pattern of difference is: There is no significant difference among the three disciplines of science, engineering, and medicine; the average ratio of social sciences is significantly higher than that of these three disciplines; the average ratio of the humanities is significantly higher than that of social sciences. This pattern is clearly visible from Fig. 1 which shows the average (the mean) ratio of each discipline. Again, the disciplinary difference here between social sciences and humanities vs. science, engineering, and medicine is consistent with what has been reported above (i.e. the correlation between downloads and citations; the correlation between downloads per paper and Journal Impact Factor). We explored the possibility that the higher download to citation ratios in social sciences and humanities is the result of lower citation counts in these disciplines. We found that the average citation counts per paper are actually higher for social sciences (20.8) but slightly lower for humanities (8.6) than that of science (11.3), engineering (11.1), and medicine (8.8), as shown in Table 1. So the much higher download to citation ratios of social sciences and humanities cannot really be attributed to lower citation counts. However, the higher download to citation ratio of humanities compared with that of social sciences can be the attributed the lower citation counts of humanities.

Figure 1 also shows the standard deviations of the download-to-citation ratio, which is a measure of the variability of the ratios within each discipline. It should be noted that both the means and the standard deviations are higher for the social sciences and humanities. This suggests that although these two disciplines have higher average ratios, there are also greater variations within each discipline. Schlögl and Gorraiz (2010) reported that the download-to-citation ratio strongly depends on the publication year of the papers. Papers

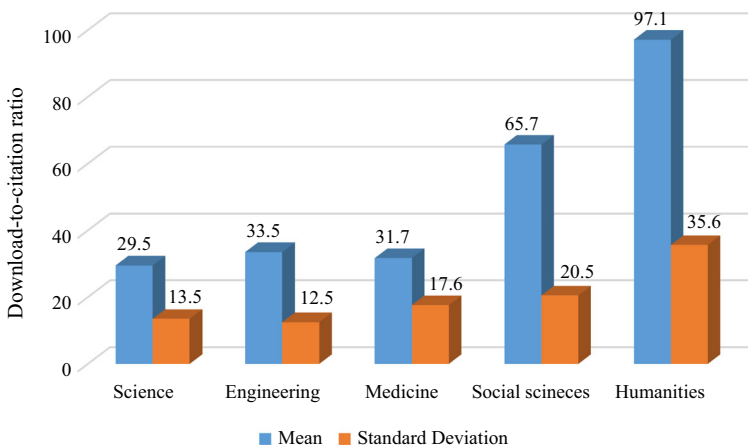


Fig. 1 Download-to-citation ratio

published more recently will have higher ratios because download counts peak earlier than citation counts. Papers in our study were published 5 years ago on average. If we studied papers published 2 or 3 years ago, the ratios would be higher. Therefore, the ratios reported in this paper should be viewed in a relative rather than an absolute sense.

Examination of outliers

To further understand the relationships between citations and downloads, we examined journals that are outliers in various measures. Among all 150 journals in the study, the two that have the highest download-to-citation ratios are *Contemporary Cinema* and *Modern Chinese History Studies*; ratios of 180.6 and 179.5 respectively. Both are in the discipline of humanities. In contrast, the two journals with the lowest ratios are *Quaternary Sciences* and *Chinese Journal of Geophysics*; ratios of 0.5 and 2.6 respectively. Both are in the discipline of science. It is clear that the two humanities journals are accessible to anyone without specific knowledge in the subject and they can appeal to a very large audience. On the other hand, papers in the two science journals are understandable only by people with highly specialized knowledge, notably researchers in those fields.

Another noticeable pattern is that journals that were published in English all have very low download-to-citation ratios. Of the seven journals that have the lowest download-to-citation ratios, four are this kind of journal. Given that there are 150 journals in the study and only five were published in English, this is a very strong pattern. In fact, the other English journal also has a below average ratio of 20.6. This suggests that relatively fewer people read or tried to read these journals, which can be explained by the fact that the English language is a deterrent. A comparison between two journals on the same subject makes the case clearer. *Chinese Journal of Rare Metals* was published in Chinese while *Rare Metals* was published in English. The download-to-citation ratios are 36.2 and 9.5 respectively; the former is almost four times of the latter. Note that the two journals were published by two different organizations and they were not the same journal in two languages. We further observed that none of the five English journals in the study are in social sciences or humanities. They are in science, engineering and medicine; 2, 2, and 1 respectively. This could be a contributing factor for the lower download-to-citation ratio for these three disciplines.

Discussion and conclusions

The study found disciplinary differences in the relationships between downloads and citations. Of particular importance is that the differences are consistent across all three measures examined: the correlation between citations and downloads, the correlation between downloads per paper and Journal Impact Factor, and the download-to-citation ratios. The disciplines of social sciences and humanities scored higher on all these measures than the disciplines of science, engineering and medicine, suggesting that the disciplinary differences found are not coincidental. Furthermore, the study used a systematic method to select fields and journals to represent each discipline, which makes the disciplinary differences found more convincing.

The disciplinary differences found in the study mean that the accuracy or utility of download data in predicting citation data, or measuring impact in general, are very different for different disciplines. This has important implications on how download data should be interpreted and used. The higher correlations between downloads and citation for social sciences and humanities suggest that using download data to predict citation data is

more reliable in these disciplines. Furthermore, the higher download-to-citation ratios for social sciences and humanities mean that papers in those journals attracted a larger readership, likely because they are accessible to more people. The lower download-to-citation ratios for science, engineering and medicine suggest that they attracted fewer readers, readers with very specialized knowledge such as citing authors. Therefore it is safe to conclude that download data are more useful for social sciences and humanities because they reflect or measure a broader impact rather than just the impact in citing authors.

The study found strong correlations (average 0.70 for the five disciplines) between the Journal Impact Factor and a journal's download per paper count. This suggests that download data can be used to estimate Journal Impact Factor and the estimation will be more accurate in social sciences and humanities where the correlations are very high (0.92 and 0.9 respectively). While Journal Impact Factor data are produced once a year and typically months after the year has ended, download data can be compiled more frequently. The ubiquity of electronic journal usage data makes this data even more attractive. Libraries have been using Journal Impact Factor to aid in decisions about journal subscriptions and cancellations. It is conceivable that download data will play an increasingly important role for this purpose. The possibility of having download data for a specific institution makes this kind of data more appealing because they are more sensitive to the institution's particular needs than the Journal Impact Factor. The broader impact that the download data measure, particularly in social sciences and humanities, is another advantage that this data has over Journal Impact Factor data, which are solely based on citation data measuring the impact on citing authors only.

The study found that the correlation between citations and downloads is higher at the journal level than that at the paper level, echoing findings from earlier studies (Guerrero-Bote and Moya-Anegón 2014; Schlögl and Gorraiz 2010). An important feature of the journal level calculation is that the frequency distributions of citation and download data are not skewed. This un-skewed distribution is not by chance but by the nature of the journal level calculation. When the download and citation numbers for individual papers are added up for a journal to get the journal level numbers, the extremely high/low numbers that would result in a skewed distribution were evened out. This means that the journal level result is less susceptible to manipulation by individual people, a concern that has been expressed over the validity of download data. For these reasons, the use of the journal level download data (e.g. for the purpose of library journal subscription) is more reliable than the use of the paper level download data (e.g. for the purpose of estimating the impact of individual papers).

One of the reasons that the study found high correlations between citations and downloads is that we calculated correlations separately for each discipline. If we merge all disciplines together and calculate a single correlation, it is 0.67 (Spearman), much lower than the average of 0.8 reported earlier. This not only points to the need to analyze disciplines separately in this kind of research but also reinforces the notion that disciplinary differences should be taken into account in the use of download and citation data. We also manually filtered out non-research articles (e.g. editorials, news, announcements, book reviews) from the study, which may have also contributed to the relatively high correlations found.

It is not clear why the discipline of medicine has the lowest correlation between downloads per paper and Journal Impact Factor, and between downloads and citations. One possible reason is that many medical journals in the study are not the top ones in the field. As was explained in the methodology section of the paper, for each field, we selected journals that were at the top of the list ranked by Journal Impact Factor. If a selected journal was not in CAJD (the database used for data collection), we replaced it with the next journal or the one after until we reached a journal that was in the database. For the

discipline of medicine, 14 out of the 30 journals used in the study were replacement journals. The reason that many medical journals were not covered by CAJD is that the Chinese Medical Association had agreements with Wanfang Data, another major Chinese database company, that gave Wanfang Data the exclusive rights for web delivery of journals published by Chinese Medical Association (Wanfang Med Online 2013). This explanation is speculative as we do not know whether the result for the discipline of medicine would be different if the top journals were used for the study. Thus findings for the discipline of medicine need to be viewed in this context.

Limitations of the study need to be discussed to place the findings in the appropriate context. As reported earlier, for each field in the study, we selected journals with the highest Journal Impact Factor scores that were in CAJD. Therefore, findings from the study may not be generalizable to all journals, particularly those that have very low Journal Impact Factor. Another limitation is that the study focused on Chinese journals, which may not represent citation and download patterns of journals in other contexts (e.g. other countries). Findings from a pilot study (Glänzel and Heeffer 2014) comparing citation and download data across countries suggest that there could be country variations.

Despite these limitations, the study is the first that examined the disciplinary differences in the relationships between downloads and citations by systematically selecting a broad spectrum of academic fields and journals from each discipline. The finding that there are disciplinary differences in the Chinese journals suggests that such differences are worth exploring in non-Chinese contexts. Another contribution of the study is the systematic comparison of different methods of calculating the correlations between citations and downloads and their effects on the results. The discrepancy in the correlation coefficients resulting from different calculation methods points to the need to report the calculation method explicitly and interpret the correlations appropriately. Finally, the finding that a non-native language (in this study the English) is a deterrent for downloading papers has implications for databases whose users have diverse languages. The download numbers so affected need to be interpreted appropriately. All these point to directions for future research, for which the current study serves as a starting point.

Appendix: Disciplines and fields in the study

| Discipline | Field | Number of journals in the study |
|------------|-----------------------|---------------------------------|
| Science | Mathematics | 4 |
| | Mechanics | 1 |
| | Physics | 3 |
| | Chemistry | 4 |
| | Astronomy | 1 |
| | Surveying and mapping | 2 |
| | Geophysics | 2 |
| | Atmospheric science | 2 |
| | Geology | 7 |
| | Marine science | 2 |
| | Physical geography | 1 |
| | Software | 1 |

| Discipline | Field | Number of journals in the study |
|--------------------|---|---------------------------------|
| Engineering | Basic engineering and technology | 1 |
| | Material science | 1 |
| | Mining engineering technology | 1 |
| | Oil and gas industry | 2 |
| | Metallurgical engineering | 2 |
| | Metallography and metals technology | 2 |
| | Mechanical engineering | 2 |
| | Arms industry and military technology | 1 |
| | Energy and power | 1 |
| | Nuclear science and technology | 1 |
| | Electrical engineering | 2 |
| | Wireless technology and telecommunications | 3 |
| | Automation and computer technology | 2 |
| | Chemical engineering | 4 |
| | Textile science and technology | 1 |
| | Civil engineering | 3 |
| | Hydraulic engineering | 1 |
| Medicine | Preventive medicine and hygiene | 3 |
| | Traditional Chinese medicine and pharmacology | 4 |
| | Basic medicine | 3 |
| | Clinical medicine | 4 |
| | Nursing | 1 |
| | Internal medicine | 3 |
| | Surgery | 3 |
| | Obstetrics/gynecology and pediatrics | 1 |
| | Oncology | 1 |
| | Neurology and psychiatry | 1 |
| | Dermatology and venereology | 1 |
| | Otorhinolaryngology and ophthalmology | 1 |
| | Oral medicine | 1 |
| | Military medicine and special medicine | 1 |
| Pharmacy | 2 | |
| Social sciences | Psychology | 1 |
| | Sociology | 1 |
| | Chinese politics | 8 |
| | World politics | 2 |
| | General economics | 3 |
| | Economic theory | 1 |
| | World economy | 1 |
| | China's economy | 2 |
| | Economic planning and management | 1 |
| | Accounting, auditing | 1 |
| Enterprise economy | 1 | |

| Discipline | Field | Number of journals in the study | |
|------------|-------------------------------------|---------------------------------|---|
| | Agricultural economy | 2 | |
| | Industrial economy | 1 | |
| | Trade economy | 1 | |
| | Logistics | 1 | |
| | Finance | 1 | |
| | Currency/finance, banking/insurance | 2 | |
| | Humanities | Philosophy | 2 |
| | Religion | 1 | |
| | Culture and museology | 2 | |
| | Language and writing | 6 | |
| Literature | 6 | | |
| Arts | 6 | | |
| History | 4 | | |
| Archeology | 3 | | |

References

- Bollen, J., & Van de Sompel, H. (2008). Usage impact factor: The effects of sample characteristics on usage-based impact metrics. *Journal of the American Society for Information Science and Technology*, 59(1), 136–149.
- Brody, T., Harnad, S., & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Society for Information Science and Technology*, 57(8), 1060–1072.
- CNKI (2016). *Introduction to the database: China academic journal network publishing database*. Retrieved November 30, 2016, from <http://epub.cnki.net/kns/brief/result.aspx?dbprefix=CJFQ>.
- Duy, J., & Vaughan, L. (2003). Usage data for electronic resources: A comparison between locally-collected and vendor-provided statistics. *The Journal of Academic Librarianship*, 29(1), 16–22.
- Duy, J., & Vaughan, L. (2006). Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination. *The Journal of Academic Librarianship*, 32(5), 512–517.
- Glänzel, W., & Gorraiz, J. (2015). Usage metrics versus altmetrics: Confusing terminology? *Scientometrics*, 102(3), 2161–2164.
- Glänzel, W. & Heeffer, S. (2014). Cross-national preferences and similarities in downloads and citations of scientific articles: A pilot study. In *Proceedings of the science and technology indicators conference*, pp. 207–215, Leiden, the Netherlands.
- Gorraiz, J., Gumpenberger, C., & Schlögl, C. (2014). Usage versus citation behaviours in four subject areas. *Scientometrics*, 101(2), 1077–1095.
- Guerrero-Bote, V. P., & Moya-Anegón, F. (2014). Relationship between downloads and citations at journal and paper levels, and the influence of language. *Scientometrics*, 101(2), 1043–1065.
- Liu, X., Fang, H., & Wang, M. (2011). Correlation between download and citation and download-citation deviation phenomenon for some papers in Chinese medical journals. *Serials Review*, 37(3), 157–161.
- Moed, H. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology*, 56(10), 1088–1097.
- Moed, H. F., & Halevi, G. (2015). Multidimensional assessment of scholarly research impact. *Journal of the Association for Information Science and Technology*, 66(10), 1988–2002.
- Moed, H. F., & Halevi, G. (2016). On full text download and citation distributions in scientific-scholarly journals. *Journal of the Association for Information Science and Technology*, 67(2), 412–431.

- Rowlands, I., & Nicholas, D. (2007). The missing link: Journal usage metrics. *Aslib Proceedings*, 59(3), 222–228.
- Schlögl, C., & Gorraiz, J. (2010). Comparison of citation and usage indicators: The case of oncology journals. *Scientometrics*, 82(3), 567–580.
- Shepherd, P. (2007). The feasibility of developing and implementing journal usage factors: A research project sponsored by UKSG. *Serials*, 20(2), 117–123.
- Shuai, X., Pepe, A., & Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: Article downloads, Twitter mentions, and citations. *PLoS ONE*, 7(11), e47523.
- Tague, J., & Ajiferuke, I. (1987). The Markov and the mixed Poisson models of library circulation compared. *Journal of Documentation*, 43(3), 212–235.
- Thelwall, M., & Wilson, P. (2016). Mendeley readership altmetrics for medical articles: An analysis of 45 fields. *Journal of the Association for Information Science and Technology*, 67(8), 1962–1972.
- Wan, J. K., Hua, P. H., Rousseau, R., & Sun, X. K. (2010). The journal download immediacy index (DII): Experiences using a Chinese full-text database. *Scientometrics*, 82(3), 555–556.
- Wanfang Med Online (2013). *Congratulations to the successful renew of agreements for the exclusive collaboration between Wanfang Data and the Chinese Medical Association*. Accessed June 10, 2016, from <http://med.wanfangdata.com.cn/meeting/hz13/detail.html>.
- Yan, K.-K., & Gerstein, M. (2011). The spread of scientific information: Insights from the web usage statistics in PLoS article-level metrics. *PLoS ONE*, 6(5), e19917.