Research Evaluation, 25(3), 2016, 264–270
doi: 10.1093/reseval/rvv049
Advance Access Publication Date: 2 February 2016
Article

OXFORD

Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data

Ad A.M. Prins,[1,*] Rodrigo Costas,[2] Thed N. van Leeuwen[2] and Paul F. Wouters[2]

[1]ResearchManagement, Pausgang 4, Groningen 9712 EZ, The Netherlands and [2]CTWS Centre for Science and Technology Studies, Leiden University, Wassenaarseweg 62A, Leiden 2333 AL, The Netherlands

*Corresponding author. Email: info@adprins.nl.

Abstract

In this paper, we report on the application of Google Scholar (GS)-based metrics in the formal assessment of research programs. Involved were programs in the fields of Education, Pedagogical Sciences, and Anthropology in The Netherlands. Also, a comparative analysis has been conducted of the results based on GS and Web of Science (WoS). Studies critical of GS point at its reliability of data. We show how the reliability of the GS data for the bibliometric analysis of the assessment can be improved by excluding non-verifiable citing sources from the full second-order GS citing data. The study of the background of these second-order sources demonstrates a broadening of the citing sources. The comparison of GS with WoS citations for the publications of the programs shows that it is promising to use GS for fields with lower degrees of coverage in WoS, in particular for fields that produce more diverse types of output than just research articles. Restrictions to the use of GS are the intensive manual data handling and cleaning, necessary for a feasible and proper data collection. We discuss wider implications of the findings for bibliometric analysis and for the practices and policies in research evaluation.

Key words: Google Scholar; research evaluation; humanities; social sciences; web of science; data reliability.

Introduction

In The Netherlands, research evaluation is organized under the combined responsibility of the Association of Universities (VSNU), the National Research Council (NWO), and the Royal Academy of Arts and Sciences (KNAW). Their combined responsibility results in a Standard Evaluation Protocol (SEP) that describes in detail the organization of research assessment, the various aspects taken care of during research assessments, and the indicators that should be part of the reporting by the committee. Peer review is the guiding principle in the assessment cycles, that cover a six-year time span and include an international assessment as well as an internal mid-term review.

In the SEP, the usage of quantitative measures such as bibliometric indicators is not compulsory. However, in many assessment cycles, bibliometric indicators are introduced to support the work of the review committee, in particular in the natural, life, and medical sciences. Owing to the lower degree of coverage of the systems that form the basis for bibliometric analyses (van Leeuwen 2013), in most of the Dutch assessments in the Social Sciences & Humanities (SSH) and Law domains, bibliometrics was not applied, with the exception of psychology, economics, and business management (Nederhof 2006). These fields stand out among the SSH and Law domains, as the communication among scholars in these domains has shifted more and more toward journal publications. Strong concerns about the design and quality criteria in research assessment were expressed in the report 'Judging research on its' merits' (KNAW 2005), initiating further thinking among scholars about these issues. Two advisory councils published reports on the humanities (KNAW 2012) and on the social sciences (KNAW 2013). These reports have influenced the new SEP, which is valid since 2015. An important shift in this new SEP is less focus on scientific productivity and more attention to societal impact of scholarly activities.

Publication cultures differ in the social sciences and humanities from most natural sciences (Hicks 2004; Nederhof et al. 2010; Van

Leeuwen et al. 2015, Archambault and Larivière, 2010). As a result, their output is not well-covered in the regular journal-based databases normally used for bibliometric studies (e.g. Web of Science [WoS] or Scopus). This is in particular the case for non-English journals, which are underrepresented (Larivière and Macaluso 2011), and for conference papers, books, and volumes (Meho and Yang 2007). An alternative to the more traditional journal-based systems of WoS and Scopus is Google Scholar (hereafter referred as GS). Although its size is not properly to ascertain (Orduña-Malea et al. 2014) and it presents a strong growth in the most recent years (De Winter, Zadpoor and Dodou 2014; Harzing 2014), a considerable number of studies have pointed to the comprehensiveness of GS in comparison to WoS or Scopus from the early appearance of GS in 2004 (Bakalbassi et al. 2006; Harzing and Van der Wal 2007) . Comparisons of outcomes of Scholar citation analyses with WoS and Scopus have been performed in relation to diverse scientific disciplines such as biomedical sciences (Falagas et al. 2008), chemistry (Bornmann et al 2009), medicine (Kulkarni et al. 2009), computer sciences (Franceschet 2010), informetrics and scientometrics (Bar-Ilan 2010), and business sciences (Mingers and Lipitakis 2010). Also, comparisons based on various fields such as chemistry physics, biology, and computing (Kousha and Thewall 2008) point to much more comprehensive results for GS.

The indexation by GS of a broader typology of publications, including books, conference papers, and also publications in other languages, has lead some authors to suggest the usefulness of the data for evaluation purposes, particularly in fields with a high diversity in publication outputs such as the social sciences and humanities (Harzing and Van der Wal 2008; Kousha and Thelwall 2008; Kousha, Thelwall and Rezaie 2011). However, this view is not uncontested, in particular because of the lack of internal quality control of GS data or the lack of information on updates (Falagas et al. 2008; Aguillo 2012). As several authors have argued, the GS database can be tampered with by feeding false publications and references (Jascó 2005, 2006; Delgado López-Cózar, Robinson-Garcia and Torres Salinas 2010, 2012, 2014). As a consequence, to use GS for bibliometric analysis, attention to data quality is pivotal.

In this paper, we report on the practical application of GS-based metric in the formal assessment of research programs, which to the best of our knowledge is a first in a real-life assessment. The fields include Education and Pedagogical Sciences, and Anthropology in The Netherlands. We address several issues on data quality, and provide a comparison of the results based on GS with WoS[1]. We conclude by discussing methodological issues in the context of assessments.

Data and methods

Education and Pedagogical Sciences (hereafter referred to as Ed/Ped) comprised 13 programs of six universities over the evaluation period 2006–11. [2] Anthropology comprised five programs of an equal number of universities, over another evaluation period, the years 2004–12. The selection of publications differs slightly for both cases. In the case of Ed/Ped, each program was asked to send in 10 publications per year (60 publications per program). Program directors were asked to send in highly valued or highly cited publications, possibly including also books. In the case of Anthropology, the selection was based on 5–10 most cited publications for each year to be evaluated, related to the size of the program. This variation in size is based on the fact that small programs are less likely than large programs to produce equal numbers of highly cited publications. Checking for publications that were listed in more than one program as double entries, the net resulting numbers of publications were 774 for Ed/Ped (with six double entries) and 328 for Anthropology (four double entries). In the comparisons of programs, the citations for double entries were counted for each program listing such publication(s). The entire publication lists for Ed/Ped and for Anthropology were included in the report about the bibliometric study and made available to the review committee. To enhance transparency, all GS downloads were stored in both HTML copies of the results pages and both these copies and the databases built from this information were made available to the program directors.

Searching the citations of the publications was done first by identifying the intended publications in GS. As a publication might have multiple entries in GS, several searches were done based on keywords of title and author, allowing for various spellings. The searches were done in the GS search engine directly, downloading the information in databases. To allow a further quality control and data cleaning process, information was retrieved for the full second-order GS citing data (this entailed the downloading of the entire list of citing publications that appear when clicked on the 'Cited by' link). The workload for downloading this information was 34 hours for Ed/Ped and 22 hours for Anthropology, the latter workload being constrained by a change in the search facility of GS restricting the results to 20 citing items per search.

The selection criterion for quality control and data cleaning was that the citing source should be traceable (verifiable) in terms of a proper working URL of Web sites of journals, publishers, or other specifiable locations. Other citing sources, in particular those without proper URL, might still be valid if checked individually, but were nevertheless excluded from the data set. Sources with defective data such as improper year of reference in comparison to the publication date of the cited reference were also removed from the data set. The net-certified citations were 22,887 (89.8% of the gross total of references found in GS for Ed/Ped), and 8,092 (89.7% of gross total for Anthropology). The workload for data cleaning was less than one hour each for Ed/Ped and Anthropology.

As part of the analysis conducted after completing the GS analysis for the assessment, a further analysis of data quality has been performed. This analysis was based on the specific URL provided by GS. It is important to note that this URL points to the indexed location of each of the citing publications, which may be the publication itself or other locations. Information provided by GS is based on the indexes produced by crawling Internet sources such as electronic academic journals, books (Google Books, Worldcat), Web sites of academic publishers, and Internet repositories such as www.jstor.org, www.cairn.info, or http://papers.ssrn.com. GS indexes also university libraries, as well as academic societies, governments, and other sources (Torres-Salinas, Ruiz-Pérez and Delgado-López-Cózar 2009). The majority of these sources contain verifiable metadata, either the proper (post print) academic publication itself, or the metadata of pre-prints or the version of record available from repositories or university libraries. However, as is also noted by Jacsó (2005), some sources such as university libraries may also contain other referring publications such as theses by PhDs or master theses, and repositories might also include conference papers and reports by research institutes. These citing sources can therefore be very diverse. The metadata were classified based on the available URL, thus revealing characteristics of the citing source, such as the

publisher or university. This classification was possible for the majority of the URLs. The classification scheme of the sources of the second-order data resulted in: (1) verifiable academic sources (including academic journals, academic publishers of volumes, and academic books), (2) university libraries, (3) repositories not identifiable as university libraries, and (4) other sources, identified as academic societies, government sites, blogs, and personal Web pages of researchers listing pre-prints of publications. In a number of cases, the available URL occurs only once in the database and did not share common characteristics. Although the data linked to these URLs might still be linked to citing academic publications, for efficiency reasons, they are classified as 'unknown'.

The analysis after the assessment also includes a comparison with citation data from WoS. Publications have been searched in the Centre for Science and Technology Studies (CWTS) in-house version of WoS source index on the basis of the available bibliographic information of the specific publication (author name as in family name and initials, title of the publication, journal title, publication year, volume, and page information). If the publication could not be found using our advanced search methods for WoS-covered publications, the publication was counted as not covered by WoS. In our in-house version of the WoS database, citations are linked to the publications directly, so citation counts are relatively easy to aggregate, after which aggregate citation scores per publication can be summed up to the level of programs, and then compared with the GS citation counts.

Results

Table 1 shows that in both fields, the academic citing sources account for more than half of the total volume of citations in both data sets. Also, the volume of 'unknown' sources, which may contain both 'exceptional' and customary citing sources, is fairly low (11% and 7%, respectively). The great majority of citing publications in GS in the data sets are retrieved from verifiable sources with strong or at very least plausible linkage to the academic publication system.

With regard to the cited output, the results for the fields of Ed/Ped and Anthropology differ in the document types. Whereas in Ed/Ped, the share of journal articles in the total set of publications is almost 90%, in the case of Anthropology, the share of journal articles is 58%, with higher percentages for books, volumes, and chapters. Differences in publication cultures are even more apparent in the volume of citations per publication types. Whereas in the case of Ed/Ped, journal articles on average are the most cited publication type, in Anthropology, books are the more cited type (Table 2).

These differences are not owing to a single or a few outliers: as Figs 1 and 2 show, the median volume of citations per document type is higher for books in Anthropology and higher for journal articles in Ed/Ped (outliers, formed by highly cited journal articles in Ed/Ped or lowly cited types of publications for Anthropology, are indicated individually with a O in Figs 1 and 2). Even though books also receive considerable attention in Ed/Ped, for the Anthropology programs, books are the most important forms of output as well as means for generating citation impact (in terms of citation counts), which, however, is not to be taken to represent or be equated with scientific impact, i.e. the influence of a publication on subsequent research (Waltman, Van Eck and Wouters 2013).

Comparing GS with WoS

Comparing the coverage of the publications in GS with their coverage in WoS, it is possible to see large differences in the coverage between the fields. Whereas over 80% of the publications in most programs in Ed/Ped were published in journals covered by WoS, this percentage fell to an average of 37.5% for the programs in Anthropology. Lower coverage was also noted for the two programs in Ed/Ped on theory & history and history & philosophy, with distinctly different publication cultures than other programs in Ed/Ped. These programs are represented as outliers marked as O's for programs in Education and in Pedagogical Sciences in Fig. 3.

A further comparison of citations in GS with those in WoS indicates higher values of citation counts provided in GS for both fields. This is also true if only GS citations from identifiably academic sources, such as academic journals, publishers, and books, are considered (Table 3).

Although there is statistical correlation in the citations per publication for GS with WoS in Ed/Ped and Anthropology, the correlation for Anthropology is based on a strongly reduced set, as only 37.5% of the publications were covered in WoS. In Figs 4 and 5, the individual programs are correlated and the differences observed among them show that programs vary with regard to their citations in GS or WoS. Whereas most programs in Ed/Ped have high degrees of correlation between GS and WoS, some programs, in particular the more theoretically or historically oriented, not only have lower degrees of correlation but show also higher visibility in GS than in WoS, which is visible in the three lower regression lines in Fig 5. In Anthropology, most programs have also lower degrees of correlation, similar to the theoretically or

Table 1. Sources for citations in GS for Anthropology and Ed/Ped

Source	Anthrop #	Anthrop %	Ed/Ped #	Ed/Ped %
Academic sources	4,573	56.5%	14,470	63.2%
University Libraries	1,677	20.7%	4,573	20.0%
Repository other than ULs	616	7.6%	1,236	5.4%
Other sources	327	4.0%	1,085	4.7%
Unknown	899	11.1%	1,523	6.7%
Total	8,092	100.0%	22,887	100.0%

Table 2. Citations per cited document type

	Citations Anthr #	Pubs Anthr #	% Anthr Citations	Citations Ed/Ped #	Pubs Ed/Ped #	% Ed/Ped citations
Book	1,818	44	41.3	700	28	25.0
Chapter	613	38	16.1	788	42	18.8
Journal	3,885	187	20.8	21,256	695	30.6
Other	357	19	18.8	109	7	15.6
Volume	1,419	39	36.4	34	2	17.0
Total	8,092	327	24.7	22,887	774	29.6

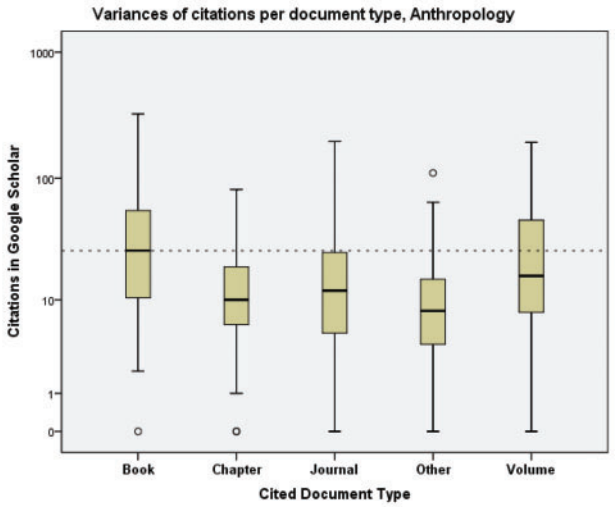

Figure 1. Distribution of citations in GS (log) per document type in Anthropology (log-scaled, dotted line is median value of highest distribution).

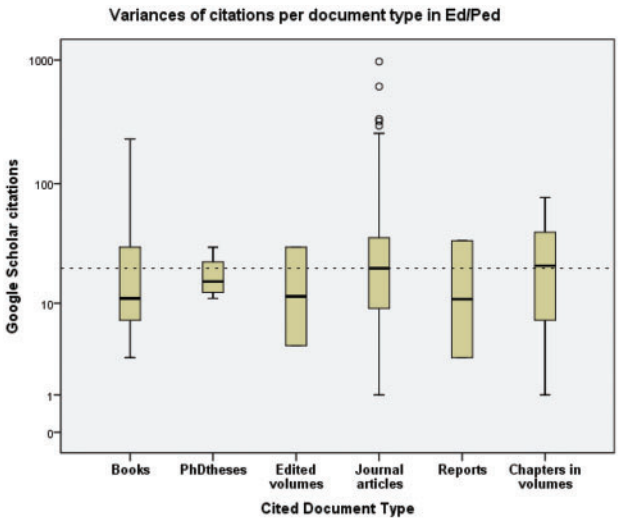

Figure 2. Distribution of citations in GS per document type in Education and Pedagogical Sciences (log-scaled, dotted line is median value of highest distribution).

historically oriented Ed/Ped programs, except for one program in Medical Anthropology (UvA).

Discussion

Our results indicate that it is feasible to perform bibliometric studies for evaluation purposes using GS, both with regards to data collection and data reliability, although still intensive manual data handling and cleaning is necessary for a feasible and proper data collection.

The comparison of GS results with WoS results indicates that it is promising to use GS for fields with lower degrees of coverage in WoS (Van Leeuwen 2013), in particular fields that produce more diverse types of output than articles in journals included in WoS. Other types of publications are important means of communication

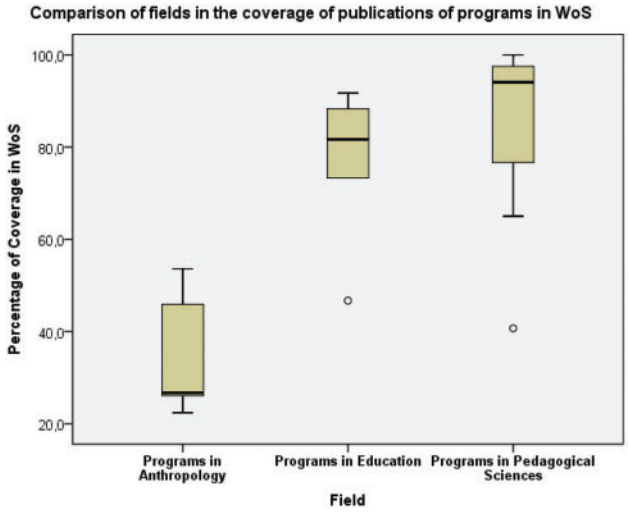

Figure 3. WoS coverage of publications per program in three fields.

Table 3. Total citations in **GS** and **WoS** for two fields

Field	Citations in GS	Citations in GS, Academic sources only	Citations in WoS
Anthropology	8,092	4,573	1,097
Education & Pedagogical Science	22,887	13,370	8,870

in Ed/Ped and even more in Anthropology, receiving considerable impact according to GS that is not covered by WoS.

There have been claims by critics of GS that the results based on this source are unreliable (Jacsó 2005, 2006, 2012; Falagas et al. 2008; Aguillo 2012; Delgado López-Cózar et al. 2012, 2014), particularly when considering raw GS data. In this study, we have observed that the information provided by GS, once retrieved on the basis of existing publication data and an important cleaning of second-order data has been performed, exhibits acceptable levels of trustworthiness in terms of the sources from which citation data are recorded. Also, the density of citation information that can be retrieved for Anthropology increases considerably to levels comparable with the results for programs in Education and Pedagogical Sciences. There are, however, important methodological and practical issues regarding how GS (see also De Winter, Zadpoor and Dodou 2014) results to be considered in the context of assessments.

Workload and data limitations

In contrast to WoS, retrieval of GS data requires considerable effort, in particular because the analysis needs to include second-order data. These second-order data are essential in establishing the traceability of citations and the source of the citation[3] as well as ensuring higher-quality data (e.g. removing duplicates, wrong links, false-positives, etc.). Recent limitations of search results to 20 per results page (which applies both for direct searches and via Publish or Perish), imposed by the GS engine, contributes to this substantial workload, combined with the impossibility of bulk processing of citation data by third-party applications (De Winter, Zadpoor and

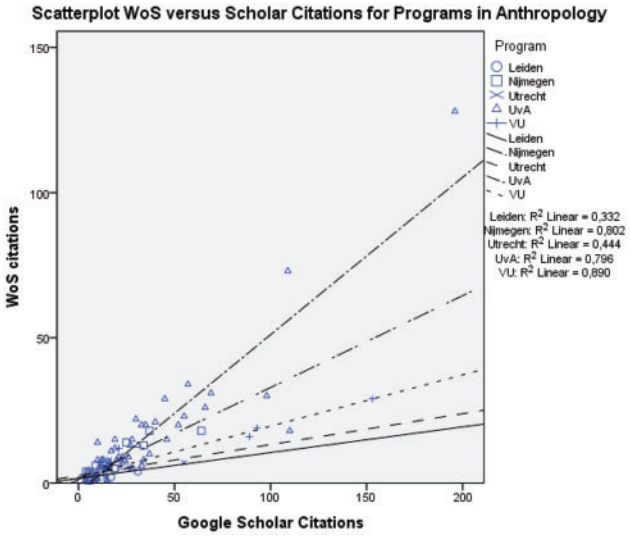

Figure 4. Scatterplot WoS and GS citations for publications of programs in Anthropology.

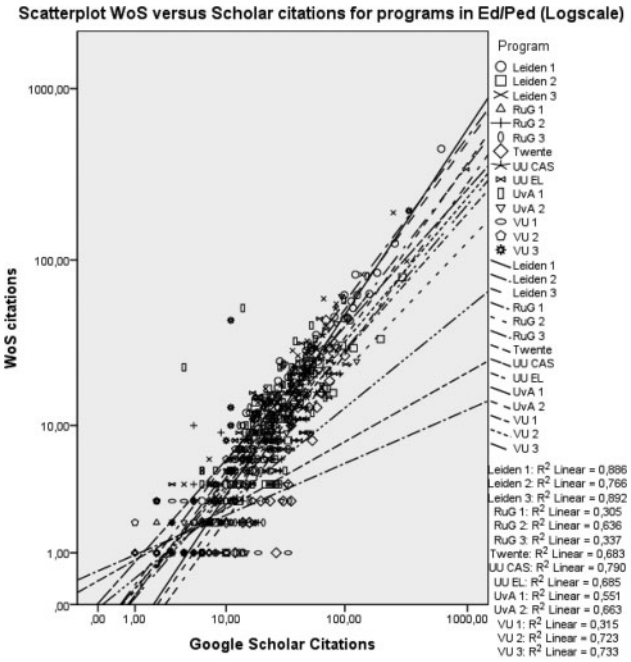

Figure 5. Scatterplot WoS and GS citations for publications of programs in Ed/Ped.

Dodou 2014, Van Noorden, 2014). This imposes limitations to how many publications can be investigated, and influences the design of large-scale and extensive bibliometric analysis, such as with analyzing big research organizations or large sets of publications. Also, although GS indexes are based on the available metadata of publishers and repositories (thus also including page numbers, issues, author lists, and journal titles), few of these metadata elements are provided directly to the end-user of GS, thus making the correct identification of the publications harder (although still possible). WoS is in this respect a more precise source, albeit that its coverage is not really relevant for fields such as Anthropology.

Possibilities for field normalization

One important limitation (partly derived from the impossibility of systematic access to GS data)[4] is that GS provides very limited opportunities for field-normalized indicators (Wouters and Costas 2012). For this study, attempts have been made for GS-based field normalization in part using the software PoP (Publish or Perish) to collect GS data for journals (Harzing and Van der Wal 2008). Although technically feasible, these attempts were rather unsatisfactory because, on the one hand, the PoP data only allowed to obtain averages of citations per paper per year for journals as recorded by GS, whereas the selection base comprised highly cited papers, and, on the other hand, because the check and cleaning of the traceability of second-order data was not possible for the reference data set (i.e. the citation data for the journals was not possible to clean). Even though the percentage of non-traceable citations was small, the comparison might still be biased. Therefore, new approaches such as those based on stratified sets of publications per journal are currently studied. In the case for Ed/Ped Sciences, attempts have also been made to include information of other sources such as *Scimago Journal Rank* (SJR) (SCImago 2007), and although possible, it leads to complicated procedures and methodological issues, which need to be studied further.

The definition of citations

Using GS implies a shift in the definition of what counts as a citation. For example, GS does not calculate or identify self-citations, which can have a relatively important effect in the citation counts. Self-citations in GS can be identified only for the first few authorships, which results only in a partial identification. GS poses restrictions on author listings in its search results, even though full identification of the authors is possible but only to the cost of highly increasing workloads. Also, whereas the citations in WoS are based on references in academic journals (although increasingly also in other academic sources, e.g. proceedings or books), the criterion in GS is the academic nature of these forms of publications as established in our data set. As a result, citations also may include references from scientific reports, PhD theses, student assignments, or student theses. Whether this shift will be accepted in view of changing perspectives on assessment standards remains to be seen.

The selection base

The selection base for publications to be analyzed is obviously relevant to the results and to the methods to be used (Abramo, D'Angelo and Di Costa 2014). In the Ed/Ped case, the selection was performed by program leaders. This led to a selection biased toward highly cited papers. However, some program leaders apparently selected publications that were deemed very relevant to the program but appeared not to be highly cited or even cited at all. As an indication, 4.5% of the selected publications were not cited at all in GS. More importantly, although technical issues—such as workload—impose limitations in selecting higher volumes of publications, the selection base is crucially related to the questions to be addressed in the assessment.

Questions to address in the assessment

The shifting ideas about assessment goals for research programs may lead to more variegated bibliometric questions, which in turn may require different research designs such as a focus on specific publications typical for the mission of institutes, or contextual bibliometrics, focusing for instance on the affiliation of citing

sources. As the precision and transparency of data show limitations and the workload is also high, the use of GS will impose significant restrictions to the possibilities to answer the desired assessment questions, but also opens up new possibilities.

Conclusions

One of the crucial reasons to not apply bibliometrics in the social sciences is the coverage of output in WoS, which is low for fields such as Anthropology and just moderate for Education or Pedagogical Sciences. There is reasonable evidence to consider GS as a valuable source for the analysis of certain fields of science, particularly in the Social Sciences and perhaps also in the Humanities (Abrizah and Thelwall 2014), as it provides more citation information based on a broader set of publication types (Kousha and Thelwall 2008; Kousha, Thelwall and Rezaie 2011; Shema, Bar-Ilan and Thelwall 2012). However, as the reviewed literature shows, in order to be able to use GS data, considerable attention should be given to data quality and reliability. Several strategies to improve the quality and reliability of the data have been shown, which are feasible also in terms of workload. However, a main limitation is the efficiency of data retrieval, restricting the use of GS as a generic tool for research evaluation. In order to be able to use GS in the context of evaluation, various complex ways for benchmarking or perform field normalization comparisons have to be worked out, for instance on the basis of available journal data, thus being able to address the issues that are critical in research assessments. It is important to bear in mind that these are not just technical problems, but these are issues also related to the questions raised in assessments, which frequently require the determination of benchmarks and contextualization frameworks that essentially need transparent, objective, robust, and clean data. Thus, although GS provides valuable and useful information for bibliometric analysis in research evaluation, the scale of its application is constrained to smaller research units or data sets. Also, when it comes to journal-based benchmarks, important additional limitations can occur in fields that rely on even higher volumes of non-journals sources than in the case of the current programs studied in this paper.

Notes

1. The first part of the actual bibliometric analysis has been performed by Ad Prins. The second part of the meta analysis has been performed by all authors. The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

2. The assignments by deans of the participating faculties to use GS in an evaluative bibliometric context were proposed due to concerns about the representation of SSH outputs in Web of Science (hereafter WoS).

It is important to mention here the involvement of the research directors of the faculties as stakeholders within their field(s) of expertise in choosing a selection base for the publications to be analyzed.

3. This data check is critical because as reported by de Winter et al., 2014, GS includes false-positive citations, duplicates, and multiple other data errors.

4. The restriction in the access to its data imposed by GS (de Winter, 2014; Van Noorden, 2014) due to arrangements with publishers strongly limits the use of GS for extensive and more robust analytical purposes.

References

Abrizah, A. and Thelwall, M. (2014) 'Can the Impact of Non-Western Academic Books be Measured? An Investigation of Google Books and Google Scholar for Malaysia', *Journal of the Association for Information Science and Technology*, 65/12: 2498–508, DOI: 10.1002/asi.23145.

Abramo, G., D'Angelo, C. A. and Di Costa, F. (2014) 'Inefficiency in Selecting Products for Submission to National Research Exercises', *Scientometrics*, 98: 2069–86.

Aguillo, I. F. (2012) 'Is Google Scholar Useful for Bibliometrics? A Webometric Analysis', *Scientometrics*, 91/2: 343–51.

Archambault, É. and Larivière, V. (2010) 'The Limits of Bibliometrics for the Analysis of the Social Sciences and Humanities Literature', in *World Social Science Report*, International Social Science Council, World Social Sciences Report: Knowledge Divides, UNESCO, París, pp. 251–54.

Bakalbassi, N., Bauer, K., Glover, J. and Wang, L. (2006) 'Three Options for Citation Tracking: Google Scholar, Scopus and Web of Science', *Biomedical Research Libraries*, 3: 7. DOI: 10.1186/1742-5581-3-7.

Bar-Ilan, J. (2010) 'Citations to the "Introduction to Informetrics" Indexed by WOS, Scopus and Google Scholar', *Scientometrics*, 82/3: 495–506.

Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A. and Daniel, H. D. (2009) 'Convergent Validity of Bibliometric Google Scholar Data in the Field of Chemistry—Citation Counts for Papers that were Accepted by Angewandte Chemie International Edition or Rejected but Published Elsewhere, Using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts', *Journal of Informetrics*, 3/1: 27–35.

De Winter, J. C. F., Zadpoor, A. A. and Dodou, D. (2014) 'The Expansion of Google Scholar Versus Web of Science: A Longitudinal Study', *Scientometrics*, 98, 1547–65, DOI: 10.1007/s11192-013-1089-2.

Delgado Lopez-Cozar, E., Robinson-García, N. and Torres-Salinas, D. (2012) 'Manipulating Google Scholar Citations and Google Scholar Metrics: Simple, Easy and Tempting', *arXiv preprint arXiv:1212.0638*, EC3 Working Papers 6: 29 May, 2012.

Delgado López-Cózar, E., Robinson-García, N. and Torres-Salinas, D. (2014) 'The Google Scholar Experiment: How to Index False Papers and Manipulate Bibliometric Indicators', *Journal of the Association for Information Science and Technology*, 65/3: 446–54.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A. and Pappas, G. (2008) 'Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and Weaknesses', *The FASEB Journal*, 22/2: 338–42.

Franceschet, M. (2010) 'A Comparison of Bibliometric Indicators for Computer Science Scholars and Journals on Web of Science and Google Scholar', *Scientometrics*, 83/1: 243–58.

Harzing, A. W. and Van der Wal, R. (2007) 'Google Scholar: The Democratization of Citation Analysis', *Ethics in Science and Environmental Politics*, 8/1: 61–73.

Harzing, A. and Van der Wal, R. (2008) 'Google Scholar as a New Source for Citation Analysis', *Ethics in Science and Environmental Politics*, 8: 61–73.

Harzing, A. W. (2014) 'A Longitudinal Study of Google Scholar Coverage between 2012 and 2013', *Scientometrics*, 98/1: 565–75.

Hicks, D. (2005) 'The Four Literatures of Social Science', in Moed H. F., Glanzel W. and Schmoch U. (eds) *Handbook of Quantitative Science and Technology Research*. Netherlands: Springer, pp. 473–96.

Jacsó, P. (2005) 'Google Scholar: The Pros and the Cons', *Online Information Review*, 29/2: 208–14.

Jacsó, P. (2006) 'Deflated, Inflated and Phantom Citation Counts', *Online Information Review*, 30/3: 297–309.

Jacsó, P. (2010) 'Metadata Mega Mess in Google Scholar', *Online Information Review*, 34/1: 175–91.

KNAW (2005) *Judging Research on its' Merits*. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

KNAW (2012) *Quality Indicators in the Humanities*. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

KNAW (2013) *Towards a Framework for the Quality Assessment of Social Science Research*. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Kousha, K. and Thelwall, M. (2008) 'Sources of Google Scholar Citations Outside the Citation Index: A Comparison between Four Science Disciplines', *Scientometrics*, 74/2: 273–94.

Kousha, K., Thelwall, M. and Rezaie, S. (2011) 'Assessing the Citation Impact of Books: The Role of Google Books, Google Scholar, and Scopus', *Journal of the American Society for Information Science and Technology*, 62/11: 2147–64.

Kulkarni, A. V., Aziz, B., Shams, I. and Busse, J. W. (2009) 'Comparisons of Citations in Web of Science, Scopus, and Google Scholar for Articles Published in General Medical Journals', *Journal of the American Medical Association*, 302/10: 1092–6.

Larivière, V. and Macaluso, B. (2011) 'Improving the Coverage of Social Science and Humanities Researchers' Output: The Case of the Érudit Journal Platform', *Journal of the American Society for Information Science and Technology*, 62/12: 2437–42.

Meho, L. I. and Yang, K. (2007) 'Impact of Data Sources on Citation Counts and Rankings of LIS faculty: Web of Science Versus Scopus and Google Scholar', *Journal of the American Society for Information Science and Technology*, 58/13: 2105–25.

Mingers, J. and Lipitakis, E. (2010) 'Counting the Citations: A Comparison of Web of Science and Google Scholar in the Field of Business and Management', *Scientometrics*, 85/2: 613–25.

Nederhof, A.J. (2006) 'Bibliometric Monitoring of Research Performance in the Social Sciences and the Humanities: A Review', *Scientometrics*, 66/1: 81–100.

Nederhof, A. J., van Leeuwen, T. N. and van Raan, A. F. (2010) 'Highly Cited Non-journal Publications in Political Science, Economics and Psychology: A First Exploration', *Scientometrics*, 83/2: 363–74.

Orduña-Malea, E., Ayllón, J. M., Martín-Martín, A. and López-cózar, E. D. (2015) 'About the Size of Google Scholar: Playing the Numbers', *Scientometrics*, 104/3: 931–949.

SCImago. (2007). 'SJR — SCImago Journal & Country Rank', <http://www.scimagojr.com> accessed 10 September and 10 October 2012.

Shema, H., Bar-Ilan, J. and Thelwall, M. (2012) 'Research Blogs and the Discussion of Scholarly Information', *PLoS One*, 7/5: e35869.

Torres-Salinas, D., Ruiz-Pérez, R. and Delgado-López-Cózar, E. (2009) 'Google Scholar Como Herramienta Para La Evaluación Científica', El Profesional de la *Inf*ormacion, 18/5: 501–10, DOI: 10.3145/epi.2009.sep.03.

Van Leeuwen, T.N. (2013) 'Bibliometric Research Evaluations, Web of Science and the Social Sciences and Humanities: A Problematic Relationship?', Bibliometrie – Praxis & Forschung. <http://www.bibliometriepf.de/article/view/173>

Van Leeuwen, T. N., Van Wijk, E. and Wouters, P. F. (2015) 'Bibliometric Analysis of Output and Impact Based on CRIS Data: A Case Study on the Registered Output of a Dutch University. *Scientometrics*, 1061: 1–16.

Van Noorden, R. (2014) 'Google Scholar Pioneer on Search Engine's Future', *Nature News and Comment*, 1–5, DOI: 10.1038/nature.2014.16269.

Wouters, P. and Costas, R. (2012) 'Users, Narcissism and Control – Tracking the Impact of Scholarly Publications in the 21st Century', in Van Berchum M. and Russell K. (eds) <http://sticonference.org/Proceedings/Vol02503/Wouters_Users_847.pdf>

Waltman, L, Van Eck, N. J. and Wouters, P (2013) 'Counting Publications and Citations: Is More Always Better? *Journal of Informetrics*, 7/3: 635–41.