

# Using Google Scholar and LexisNexis to compile citation profiles for South African journals in legal research: An exploratory study

Nelius Boshoff<sup>1</sup>

<sup>1</sup> *scb@sun.ac.za*

Centre for Research on Evaluation, Science and Technology (CREST) and the DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP), Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa

## Brief study overview

The aptness of bibliometric methods for research evaluation in the social sciences and humanities (SSH) has received much attention in the academic literature. A key argument against the use of bibliometrics relates to the insufficient coverage of SSH research outputs in Scopus and Web of Science, the two major citation databases. However, this argument is increasingly weakened by the constant growth of the Google Scholar (GS) web search engine. Despite its comprehensive coverage of publication and citation contents in the SSH, GS remains far from perfect. Not only is GS limited in terms of user functionality but also the quality of some of its data is suspect. Recently, Prins et al. (2016) developed procedures to enhance the quality of data extracted from GS, all of which require manual intervention based on small datasets and lower levels of analysis. The procedures entail the verification, capture and classification of the web addresses of the citing sources of publications in GS.

As not all citations are equal and for that reason require normalization (Ioannidis, Boyack & Wouters, 2016), citations extracted from GS should also be normalized in terms of factors like field, age of publication and document type. Currently the journal normalized citation score (JNCS) seems to be the indicator of choice, simply because it is what the journal publication data in GS allows for at present (Bornmann, 2016; Mingers & Meyer, 2017). However, the JNCS makes sense only when citations are reported for analytical units such as individuals or organisations. When the reporting unit is the journal itself, the mean JNCS for any journal will always equal one. Thus, to compile citation profiles for journals based on normalization in GS, another indicator such as the mean field normalized citation score (MNCS) would need to be reported. This raises the question as to how to demarcate fields in GS, given that GS incorporates no field classification system.

The current study circumvented the challenge of field classification by focusing on one field only, namely that of legal research in South Africa. The study conceptualized legal research as comprising all publications in any of the 30 South African journals in the field of law. As a result, the MNCS for a journal in this study does not indicate whether publications in a journal are cited more or less than the average legal research publication in the world. Instead, the relevant MNCS indicates whether a journal's publications are cited more or less than the average legal research publication in South Africa. In that way the benchmarking of a journal publication is confined to its national setting. This is an important consideration as legal research typically displays a strong national character (De Jong et al., 2011).

Apart from its national orientation, various other peculiarities set legal research apart from other fields of research (Peruginelli, 2015; Schmied, Byland & Lienhard, 2018). Among these is the dual interpretation of 'research'. On the one hand, legal research can be humanities oriented, meaning doctrinal research that focuses on the law and legal concepts and which

rests on an analysis of legal sources (e.g. court cases). On the other hand, legal research can also be social sciences oriented, relying on systematic empirical observations (De Jong et al., 2011). Equally relevant is the fact that research in legal journals can be cited in court judgements (Ambro, 2006), thereby providing an indication of the relevance of legal research to legal decision making. For that reason, the current study also investigated the extent of judicial citing of the 30 South African legal journals. The electronically accessible collection of law reports in LexisNexis South Africa was used for this purpose.

Two broad research question (with sub-questions) guided the study:

- In terms of research evaluation, what can be concluded from a bibliometric analysis of GS citations to South African legal journals? (Sub-questions: How can the issue of GS data quality be clarified? What insights emerge from the normalization of GS citation data in terms of document type and year of publication? How and why do the journals differ as to their GS citation profiles?)
- In terms of research evaluation, what can be concluded from a bibliometric analysis of judicial citations to South African legal journals? (Sub-questions: What is the average time from publication to judicial citation? How and why do the journals differ as to their judicial citation profiles? How do these profiles compare to the GS citation profiles? Is there any value in compiling and reporting on judicial citation profiles?)

Note: The study is still on-going but nearing completion. For the purpose of this conference abstract, findings for only two South African journals are reported – *Comparative and International Law Journal of Southern Africa* (CILSA) and the *Industrial Law Journal* (ILJ).

## **Methods and first results**

### *Analysis of GS citations*

The 30 South African legal journals were established over a period of 129 years, with the oldest being the *South African Law Journal* (established in 1884). The two most recent additions are the *African Disability Rights Yearbook* and the *South African Intellectual Property Law Journal* (both established in 2013). As 2013 is the first year for which publication data is available for all 30 journals, a 3-year window (2013–2015) was chosen as publication period and a 6-year window (2013–2018) as citation period. The relevant details of legal publications were obtained from different digital collections and captured in a Microsoft access database together with an indication of the document type (e.g. article or case note). The titles of publications were systematically searched for in GS. Datasets of citing publications were downloaded through the available GS user functionality. All citing publications that lack a year of publication were excluded as well as citing publications whose publication year predated that of the cited publication. For the two journals concerned, the final dataset comprised 59 articles and 2 case notes from CILSA, and 50 articles and 25 case notes from ILJ (Table 1). By 2018, none of the case notes had received any citations in GS.

**Table 1. GS citations (2013–2018) to publications (2013–2015) in two South African legal journals.**

<i>Journals</i>	<i>Year</i>	<i>Number of publications</i>	<i>GS citations, 2013–2018</i>				
			<i>Total number of citations</i>	<i>Average number of citations</i>	<i>Standard deviation</i>	<i>Lowest citation</i>	<i>Highest citation</i>
<b>Articles</b>							
CILSA	2013	20	26	1.30	1.30	0	4
CILSA	2014	19	19	1	1	0	3
CILSA	2015	20	13	0.65	0.93	0	3
ILJ	2013	17	17	1	3	0	12
ILJ	2014	18	10	0.56	1.25	0	5
ILJ	2015	15	7	0.47	0.92	0	3
<b>Case notes</b>							
CILSA	2014	2	0	0	0	0	0
ILJ	2013	5	0	0	0	0	0
ILJ	2014	7	0	0	0	0	0
ILJ	2015	13	0	0	0	0	0

Table 2 assumes that the field of legal research in South Africa comprises two journals only (CILSA and ILJ). Articles are the only document type reflected. For instance, in 2013, 37 articles appeared in both journals (Table 1, counts of 20 and 17). The 37 articles generated 43 GS citations (Table 1, counts of 26 and 17). On average, the 37 articles (Table 2) produced 1.16 citations per publication. This average served as the expected value for normalization and for computing a MNCS value. For instance, an article that was published in CILSA in 2013 and which had 2 GS citations at the end of 2018, would get a field-normalized citation score (NCS) of 1.72 (i.e. 2 divided by 1.16). The article received 1.72 times more citations than the average legal research article in South Africa did (assuming that the field of legal research comprises two journals only). If the same calculation is done for all individual CILSA articles in 2013, and similar NCS values created for all CILSA articles in 2014 and for all CILSA articles in 2015, and the mean of all those NCS values computed, the MNCS for CILSA would be 1.18. This value appears in the last column in Table 2. It means that CILSA's articles are cited 18% above the national average (controlling for the year of publication of a specific document type, which is an article). Similarly, the MNCS of 0.79 for ILJ indicates that the journal is cited 21% below the national average. Note that the 'true' MNCS values can only be finalized once all 30 journals are included in the analysis.

**Table 2. Mean normalized citation scores calculated for a hypothetical field comprising articles in two legal journals.**

<i>Year</i>	<i>Total number of articles in two journals</i>	<i>Total number of GS citations</i>	<i>Average number of GS citations</i>		<i>Journals</i>	<i>Mean normalized citation scores (MNCS)</i>
2013	37	43	1.16		CILSA	1.18
2014	37	29	0.78		ILJ	0.79
2015	35	20	0.57		Both	1.00

In order to assess the quality of citation data in GS, the web addresses (URL) of all 92 citations (Table 1, fourth column) had to be captured manually. Each URL contained information that was assigned to one of five publication sources: academic publishers (e.g.

Springer), non-academic publishers (e.g. Law Society of South Africa), university library repositories (e.g. SUNScholar at Stellenbosch University), non-university repositories and digital collections (e.g. JSTOR) and other sources (e.g. US-China Economic and Security Review Commission). By verifying the URLs of citation sources, information about the publication types (e.g. book chapter or journal article) of the citing material could also be captured.

Table 3 below shows the cross-tabulation between the publication types and publication sources that constitute the 92 GS citations. As can be seen, 39% of the 92 GS citations originated from postgraduate theses in university library repositories. A further 39% were journal articles of mainly two types (those sourced by GS from the journal publisher [15%] and those sourced by GS from non-university repositories and digital collections [24%]).

**Table 3. The 92 citations in GS classified in terms of publication types and publication sources.**

<i>Publication types</i>	<i>Publication sources</i>				
	<i>Academic publishers</i>	<i>Non-academic publishers</i>	<i>University library repositories</i>	<i>Non-university repositories and digital collections</i>	<i>Other sources</i>
Book	0%	0%	0%	3%	0%
Book chapter	4%	0%	0%	3%	0%
Conference presentation	0%	1%	0%	0%	0%
Conference proceeding/paper	0%	0%	0%	1%	1%
Hearing testimony	0%	0%	0%	0%	1%
Journal article	15%	0%	2%	24%	0%
Occasional/series paper	0%	0%	0%	3%	0%
Research assignment (Honours)	0%	0%	0%	1%	0%
Thesis (Doctoral or Masters)	0%	0%	39%	0%	0%

### *Analysis of judicial citations*

The study extracted from the law reports in LexisNexis all references made to South African legal journals in judicial decisions. The law reports go back as far as 1947. Different spelling variants and acronyms for journal titles were used to extract the relevant information, and the results were captured in Microsoft Excel. Duplicates had to be removed as LexisNexis incorporates law reports from different sources and also because a legal case could serve before more than one South African court. PDF copies of the cited journal publications were sourced from relevant digital collections of journal publications.

In Table 4, the time span of the judicial citations is different for each journal as the time span of the source publications was not restricted to the period 2013–2015. By removing time restrictions, Table 4 brought important insights to light, such as that 51% of judicial citations to publications in CILSA occurred at least 10 years after publication. In the case of ILJ, the time to citation period was found to be much shorter – 61% of judicial citations occurred within the first 5 years after publication. Table 5, on the other hand, does align with the GS analysis by specifying 2013–2015 as the publication period and 2013–2018 as the citation period. The significantly lower shares of judicial citations compared to GS citations are evident.

**Table 4. Judicial citations to publications in two South African legal journals, based on data in LexisNexis.**

<i>Journals</i>	<i>Time span of cited publications</i>	<i>Time span of judicial citations (legal decisions)</i>	<i>Total number of publications cited in legal decisions by 2018</i>	<i>Total number of judicial citations received by 2018</i>	<i>Average number of judicial citations per publication by 2018</i>	<i>Average time from publication to judicial citation</i>	<i>Minimum time from publication to judicial citation</i>	<i>Maximum time from publication to judicial citation</i>	<i>% of publications cited in legal decisions, ≤5 years after publication</i>	<i>% of publications cited in legal decisions, 10+ years after publication</i>
CILSA	1968 to 2007	1969 to 2018	41	51	1.24	12.5 years	0 years	48 years	29%	51%
ILJ	1983 to 2013	1986 to 2016	23	29	1.26	5.8 years	0 years	16 years	61%	26%

**Table 5. Comparison between GS citations (2013–2018) and judicial citations (2013–2018) to articles (2013–2015) in two South African legal journals.**

Journals	Total number of articles (2013–2015)	GS citations		Judicial citations	
		Number of articles (2013–2015) cited in the period 2013–2018	% of articles (2013–2015) cited in the period 2013–2018	Number of articles (2013–2015) cited in the period 2013–2018	% of articles (2013–2015) cited in the period 2013–2018
CILSA	59	32	54%	0	0%
ILJ	50	12	24%	2	4%

### Concluding comment

The analysis is currently being completed for all 30 South African legal research journals. In order to draw conclusions that could appropriately inform the bibliometric assessment of South African legal research, additional classifications might be needed in the analysis. Examples include specific branches of law, a classification of multidisciplinary versus specialized journals, and a classification of academic versus professional journals based on the addresses of article authors in the period 2013–2015.

### References

- Ambro, T.L. (2006). Citing legal articles in judicial opinions: A sympathetic antipathy. *American Bankruptcy Law Journal*, 80, 547–551.
- Bornmann, L., Thor, A., Marx, W. & Schier, H. (2016). The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google Scholar data for the publications of a research institute. *Journal of the Association for Information Science and Technology*, 67, 2778–2789.
- De Jong, S P.L., Van Arensbergen, P., Daemen, F., Van der Meulen, B. & Van den Besselaar, P. (2011). Evaluation of research in context: An approach and two cases. *Research Evaluation*, 20, 61–72.
- Ioannidis, J.P.A., Boyack, K. & Wouters, P.F. (2016). Citation metrics: A primer on how (not) to normalize. *PLoS Biology*, 14, 1–7.
- Mingers, J. & Meyer, M. (2017). Normalizing Google Scholar data for use in research evaluation. *Scientometrics*, 112, 1111–1121.
- Peruginelli, G. (2015). Evaluating research: The case of legal scholarly outputs. *Legal Information Management*, 15, 50–56.
- Prins, A.A.M., Costas, R., Van Leeuwen, T.N. & Wouters, P.F. (2016). Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data. *Research Evaluation*, 25, 264–270.
- Schmied, M., Byland, K. & Lienhard, A. (2018). Procedures and criteria for evaluating academic legal publications: Results of a survey in Switzerland. *Research Evaluation*, 27, 335–346.