Accepted for publication in the
Journal of the Association for Information Science and Technology

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

Lutz Bornmann*, Andreas Thor**, Werner Marx***, and Hermann Schier***

*Division for Science and Innovation Studies
Administrative Headquarters of the Max Planck Society
Hofgartenstr. 8,
80539 Munich, Germany.
Email: bornmann@gv.mpg.de

** University of Applied Sciences for Telecommunications Leipzig
Gustav-Freytag-Str. 43-45,
04277 Leipzig, Germany.
Email: thor@hft-leipzig.de

***Max Planck Institute for Solid State Research
Information Service
Heisenbergstrasse 1,
70506 Stuttgart, Germany.
Email: w.marx@fkf.mpg.de, h.schier@fkf.mpg.de

Abstract

In the humanities and social sciences, bibliometric methods for the assessment of research performance are (so far) less common. The current study takes a concrete example in an attempt to evaluate a research institute from the area of social sciences and humanities with the help of data from Google Scholar (GS). In order to use GS for a bibliometric study, we have developed procedures for the normalisation of citation impact, building on the procedures of classical bibliometrics. In order to test the convergent validity of the normalized citation impact scores, we have calculated normalized scores for a subset of the publications based on data from the WoS or Scopus. Even if scores calculated with the help of GS and WoS/Scopus are not identical for the different publication types (considered here), they are so similar that they result in the same assessment of the institute investigated in this study: For example, the institute's papers whose journals are covered in WoS are cited at about an average rate (compared with the other papers in the journals).

Keywords

Google Scholar; Bibliometrics; Humanities; Social Sciences

1 Introduction

In the classical core areas of natural and life sciences (hard sciences), quantitative methods have meanwhile become an integral part of research evaluation (Moed, 2005). In the humanities and social sciences (soft sciences) quantitative methods for the evaluation of research performance are (still) not so widespread. However, in times of limited research funding, the evaluation pressure is also rising in these disciplines, but the methodical preconditions for the application of quantitative methods are (still) not very developed.

In the natural and life sciences, bibliometrics in particular has established itself as a standard procedure for quantitative research evaluation. With respect to the selection of suitable data sources and indicators, as well as the realization of a bibliometric study, standards have been developed in the meantime and also applied (Bornmann et al., 2014; Bornmann & Marx, 2014). The most used databases are Web of Science (WoS) from Thomson Reuters and Scopus from Elsevier. WoS currently contains a core set of around 11000 journals (WoS source journals); Scopus covers more than 20000 journals. However, WoS and Scopus are multidisciplinary databases which are biased towards natural and life sciences.

1.1 Problems of bibliometrics in the humanities and social sciences

Bibliometrics on the basis of WoS and Scopus is unsuitable for use in the humanities and social sciences, chiefly for the following two reasons:

(1) A higher proportion of journals which are not included in the database: Research topics in the humanities and social sciences are often nationally or regionally orientated. Thus the corresponding publications appear in the relevant language and not in the (international) journals included in WoS or Scopus (source journals) (Butler & Visser, 2006; Frandsen & Nicolaisen, 2008; Moed, 2005; Nederhof, 2006). The problem of insufficient coverage,

particularly for WoS, does not seem to be reducing, at least in the case of the social sciences, but rather to be growing (Larsen & von Ins, 2010).

(2) A larger share of book contributions and monographs: In the natural and life sciences, research results are mainly published as classical articles (papers) in specialist journals which are largely covered by WoS and Scopus. However, this requirement is already insufficiently met in some disciplines of natural science, such as computer science and materials science. In the area of humanities and social sciences, publication tends predominantly to be in the form of books or monographs, which are essentially excluded as database documents (source items) for WoS or Scopus. Thus typical publications in the humanities and social sciences are only insufficiently captured by these databases (Marx & Bornmann, 2015). Database providers are already including proceedings and monographs, although their coverage is still poor (Gorraiz, Purnell, & Glänzel, 2013; Torres-Salinas, Robinson-Garcia, Campanario, & López-Cózar, 2014).

Since bibliometrics based on WoS and Scopus can hardly be applied to the social sciences and humanities, there is already a range of projects with the objective of introducing other indicators for evaluation in these disciplines. For example, the project "Development and Testing Research Quality Criteria in the Humanities, with an emphasis on Literature Studies and Art History" of the universities of Zürich and Basel, has the objective of developing quality criteria for research in selected subjects of the humanities (http://www.psh.ethz.ch/crus/index). But the indicators suggested in these projects are generally less practical than the indicators which are used in bibliometrics (Hug, Ochsner, & Daniel, 2013, in press; Ochsner, Hug, & Daniel, 2012a; Ochsner, Hug, & Daniel, 2012b).

The meaningfulness of bibliometric data for research evaluation ultimately depends on the coverage of the publications in the databases selected (Chi, 2013). What is not covered by the databases can also not be evaluated. The coverage of specialist literature in databases refers primarily to the publications which are recorded as database documents (source items)

and made searchable; 'non-source' items are not considered (Butler & Visser, 2006; Chi, 2014). The different level of coverage of humanities and social sciences in relation to the natural and life sciences is reflected in the different share of references (citations) of these publications which are recorded as database documents (i.e. as searchable publications in WoS or Scopus) and correspondingly linked. The difference is especially marked in the social sciences and particularly in the humanities: Although publications in the social sciences contain, on average, even more references than natural science publications, only a third of these are recorded in the WoS as database documents (Marx & Bornmann, 2015). In the case of the humanities, the share of publications recorded in the WoS is lower still by far.

1.2 The use of Google Scholar in bibliometrics

Publications represent an important form of distribution of research results in most of the humanities and social sciences. In these publications results are usually produced or discussed against the background of the research results of other scholars (i.e. citations are mandatory). Thus, the use of bibliometrics for research evaluation seems appropriate in these disciplines as well. Because of the fundamental limitations associated with WoS and Scopus, Google Scholar (GS) has been proposed in the past as an alternative (or supplement). In comparison with other existing databases (such as Chemical Abstracts, http://www.cas.org/) the use of GS has the decisive advantage of the broad coverage of the literature (Prins, Costas, van Leeuwen, & Wouters, 2014). The limitations to a core set of scientific journals mentioned in connection with WoS and Scopus disappear. This not only results in a more comprehensive coverage of publications to be evaluated, but also of citations by publications that have not appeared in core journals (Kousha & Thelwall, 2007, 2008; Kousha, Thelwall, & Rezaie, 2011). For disciplines such as computer science, GS, compared with WoS, provides a much more comprehensive and mostly more favourable picture (Franceschet, 2010; Kousha, Thelwall, & Rezaie, 2010).

However, a range of publications has pointed to many weak points and deficiencies of GS, which must be taken into account in its use (Jacso, 2005, 2009, 2012). Some years ago, GS had the problem that certain publishers denied GS access (such as the American Chemical Society, ACS), which led to very incomplete results in the corresponding specialties such as chemistry, and made the use of GS fundamentally questionable (Bornmann et al., 2009). But the situation has changed since then: The ACS publications are now also covered by GS. New studies show that GS now covers scientific publications across the specialties so well that citation analyses now appear possible in disciplines beyond the natural and life sciences: "Finally, we argue that Google Scholar might provide a less biased comparison across disciplines than the Web of Science. The use of Google Scholar might therefore redress the traditionally disadvantaged position of the Social Sciences in citation analysis" (Harzing, 2013, p. 1057). In addition, GS seems to be growing continually (parallel to the increasing output of publications) and thus to be sufficiently stable over time: "Our data suggest that – after a period of significant expansion for Chemistry and Physics – Google Scholar coverage is now increasing at a stable rate" (Harzing, 2014, p. 565).

However, certain fundamental problems still remain: GS does not supply any information on data sources, document types and time ranges or update frequencies. The citations continue to include questionable sources, such as research applications and presentations which should really not be regarded as citing documents (Meho & Yang, 2007). However, the main problem with the bibliometric use of GS is the identification and elimination of duplicates, both on the publication side as well as the side of the citations of these publications. The cause is the automatic generation of the data sets from the sources available in the internet, which leads to heterogeneous bibliographic information on one and the same publication. The names of authors, journals and title words may appear in a range of variants which have to be combined (Jacso, 2009). This combination can never be performed

satisfactorily in a purely automatic way and requires (manual) post-processing (Köpcke, Thor, & Rahm, 2010; Thor & Rahm, 2007).

The current study takes a concrete example in an attempt to evaluate a research institute from the area of social sciences and humanities with the help of data from GS. Here we follow the example of Prins, et al. (2014), by using GS in a real life assessment procedure. For this study we have consciously chosen an institute (researching into the foundations of language) which also publishes a large part of its output in journals which are evaluated for WoS or for Scopus. Our intention is to test the convergent validity of the GS results by comparing them with those based on WoS and Scopus. If the convergent validity is established (and if we arrived at similar results with GS and WoS/Scopus), we would see that as support for the use of GS for research evaluation in the social sciences and humanities. For the first time in bibliometrics, this study undertakes a normalization of citation impact on the basis of GS data. This involves a comparison of the impact of publications appearing in journals, conference proceedings, and anthologies with the impact of a reference set compiled correspondingly (Pudovkin & Garfield, 2009). The special difficulties in calculating normalized indicators on the basis of GS are indicated in Prins, et al. (2014).

2 Methods

2.1 Data set

The current study includes the publications of a research institute from the year 2009. The institute published a total of 212 publications in this year. Somewhat less than half of the publications (40%) are journal papers (see Table 1). All publication types – apart from the PhD dissertations – are included in the citation analysis of the current study.

Table 1
Publication output of the institute in 2009

Publication type	Number
Papers published in journals which are covered in WoS	56
Papers published in journals which are not covered in WoS	29
Book chapters	71
Conference papers	39 (4 full papers)
Books	10
PhD Dissertations	7
Total	212

2.2 Normalization of citation impact: journal normalized citation scores

In order to be able to compare the citation impact of papers published in different publication years and subject categories with each other, a normalization of citation counts of papers is performed in bibliometrics (Vinkler, 2010). One possibility for normalization consists in calculating the so-called journal normalized citation score (JNCS) for a unit (here: an institute), as follows: "The number of citations to each of the unit's publications is normalized by dividing it with the world average of citations to publications of the same document type, published the same year in the same journal. The indicator is the mean value of all the normalized citation counts for the unit's publications" (Rehn, Kronman, & Wadskog, 2007, p. 22). A JNCS of 1 means that the citation impact of the institute's papers corresponds to the average citation impact in the journals which published them. A score of more (less) than 1 means that the citation impact of the institute's papers lies above (below) the average in the journal.

2.3 Normalization of the citation impact of conference proceedings and book chapters

Since calculating a normalized impact is not only desirable for journal papers, but also for conference proceedings and book chapters, in this study we would like to propose a suitable normalization procedure for these publication types (Torres-Salinas, et al., 2014): (1) The citation impact of a contribution to a conference should be measured in relation to the citation impact of the other contributions to the same conference. In other words: The citation

impact of a contribution should be divided by the average citation impact of the other contributions to the same conference. In the following sections we refer to a score calculated in this way as a Conference Proceedings Normalized Citation Score (CPNCS). Since meeting abstracts are generally not included in bibliometric analyses, the normalization procedure only includes contributions which are published as full papers in the corresponding proceedings volumes. (2) The citation impact of a book chapter should be measured relative to the citation impact of the other book chapters in the book concerned. In other words: The citation impact of a certain chapter should be divided by the average citation impact of the other chapters in the same book. In the following, we refer to a score normalized in this way as a Book Chapter Normalized Citation Score (BCNCS).

2.4 Searching for publications in GS

To search for publications in GS (those from an institute or for construction of a reference set), the corresponding queries to GS were performed as follows: Firstly, each publication was searched for by title in GS and (up to) 20 results recorded. Subsequently, a query was performed and up to 1000 results recorded for each journal, conference and book by name or title. The procedure described ensures a high probability that all relevant hits can be determined in GS, even if data errors exist in GS for certain publications (such as typos in the title). We extracted all hits in GS with their own (GS internal) ID, since only these hits are have an unambiguous reference. The ID allows us to perform comparative investigations in future in which the changes in citation numbers with time could be understood (for the same publication set).

The GS hits obtained in this way would be aligned with the publications sought, i.e. the similarity of the title would be determined between the publications and the GS hits. For this, the so-called trigram similarity (ASIM) was calculated, which determines the relative agreement of trigrams (i.e. three successive characters in the title). Our experience in the past

has shown that an ASIM>.8 indicates with a high probability that the hit in GS corresponds to the publication originally sought (Thor & Rahm, 2007). It can additionally be checked whether the GS hit has the same publication year as the publication (here: 2009). We have also manually checked a range of publications to see whether it really was the publication concerned from the journal, the proceedings volume or the book. Here we concentrated on the typical problem cases where, for instance, a publication has a lot of GS hits (e.g. because it has a general title like "Editorial") or several publications have the same GS hits (e.g. because they have very similar titles). The procedure described is a heuristic proven over many years, which allows a very good assignment of GS hits to publications despite possible data quality problems. However, complete agreement between the publications sought and the hits can only be guaranteed by manual checking of every single GS hit, which is not practical with a large number of publications.

The citation window for the impact scores in GS in this study covers a period from publication date to 2014.

3 Results

3.1 The citation impact of the journal papers of the institute which are covered in the Web of Science

For the calculation of the normalized citation impact, the corresponding reference set must be compiled for every article of the institute (n=56). For this, searches are performed for all the articles in the journals in which the institute has published (n=15983). The search in GS produced an entry for a total of 15691 articles. In other words: For the 15983 articles which were sought in GS, the rate of hits was 98%. Table 2 shows the distribution of the articles sought and hit in GS across the various journals. If an article published in one of these journals could not be found in GS, it was excluded from the calculation of the citation impact for the reference set.

Table 2
Number of articles sought in GS, and number of hits in GS

Journal	Number of articles for which data was sought in GS	Number of articles with at least one hit in GS	Total number of hits for the articles in GS
1	743	736	1233
2	265	265	463
3	1026	1023	1410
4	134	134	187
5	401	399	535
6	20	20	32
7	33	33	52
8	82	81	125
9	31	31	41
10	3759	3631	4938
11	345	344	536
12	159	159	295
13	843	842	1416
14	276	276	401
15	39	7	9
16	101	100	141
17	31	31	38
18	131	130	183
19	17	17	28
20	164	164	415
21	9	9	11
22	14	14	19
23	92	92	175
24	54	54	65
25	16	16	24
26	11	11	17
27	120	120	163
28	13	10	13
29	23	23	30
31	135	130	169
32	5	1	6
33	59	59	85
34	175	175	237
35	9	9	12
36	7	7	12
37	49	30	57
38	234	233	290
39	327	327	414
40	680	680	977
41	33	29	42
42	103	103	140
43	244	243	409

44	36	36	53
45	94	94	118
46	195	195	267
47	48	48	75
48	171	171	252
49	32	32	54
50	4395	4317	5588
Total	15983	15691	22252

Besides the number of articles for which data in GS was sought, and the number of articles with at least one hit in GS, Table 2 provides the total number of hits for the articles in GS: For many articles, not just one corresponding entry is found in GS, but several. As Table 3 shows, there was a hit in GS for 11859 articles (53%). For the remaining articles, there were between 2 (n=2442) and 20 (n=1) hits. The comparable figures from Martín-Martín, Orduña-Malea, Ayllón, and Delgado López-Cózar (2014) show that the search strategy in this study (see section 2.4) allowed a reduction in the number of possible hits per publication: "83% of the documents in our sample have more than one version, whereas 40% have 6 or more versions, 19% have 10 or more versions, and 200 documents have more than 100 versions (0.1%)" (p.35).

Table 3
Total number of hits for articles in GS

Number of hits in GS	Number of articles
1	11859
2	2442
3	802
4	283
5	130
6	80
7	38
8	21
9	9
10	11
11	6
12	2
13	1
14	2

15	1
16	2
18	1
20	1

Since several entries in GS were found for around half of the articles, the question arises whether all entries, or which fraction of the entries, should be used for the calculation of the reference values. Thus, for example, around 90% of the hits in GS relate to the year 2009 (i.e. the year from which the publications of the institute come). About 10% of the hits relate to other years. We can assume with high probability that we do not need to take into account the other hits for calculation of the reference values.

Figure 1 shows the average number of citations (arithmetic mean) from GS for articles (or their article hits in GS) which were published in 49 different journals. Also shown are the average number of citations for all article hits in a journal, only for article hits from 2009 with an ASIM>.8, as well as for articles from 2009 with an ASIM>.8 and a manual correction of the data. The figure is intended to clarify which restrictions in the subgroups lead to small or large changes in the citation rates. The figure shows that the average values derived from all the articles hits differ markedly from the average values for the subgroups. In the derivation of reference values on the basis of journals, this indicates that the publication year of the hits should be taken into account. Consideration of further limitations, like the ASIM or the manual correction, hardly changes the average citation frequency at all: Across all journals, the citation rates differ on average by about one citation. However, the other limitations – besides the publication year – are still taken into account in the compilation of the reference values, so as to have the highest accuracy possible for the citation impact values.

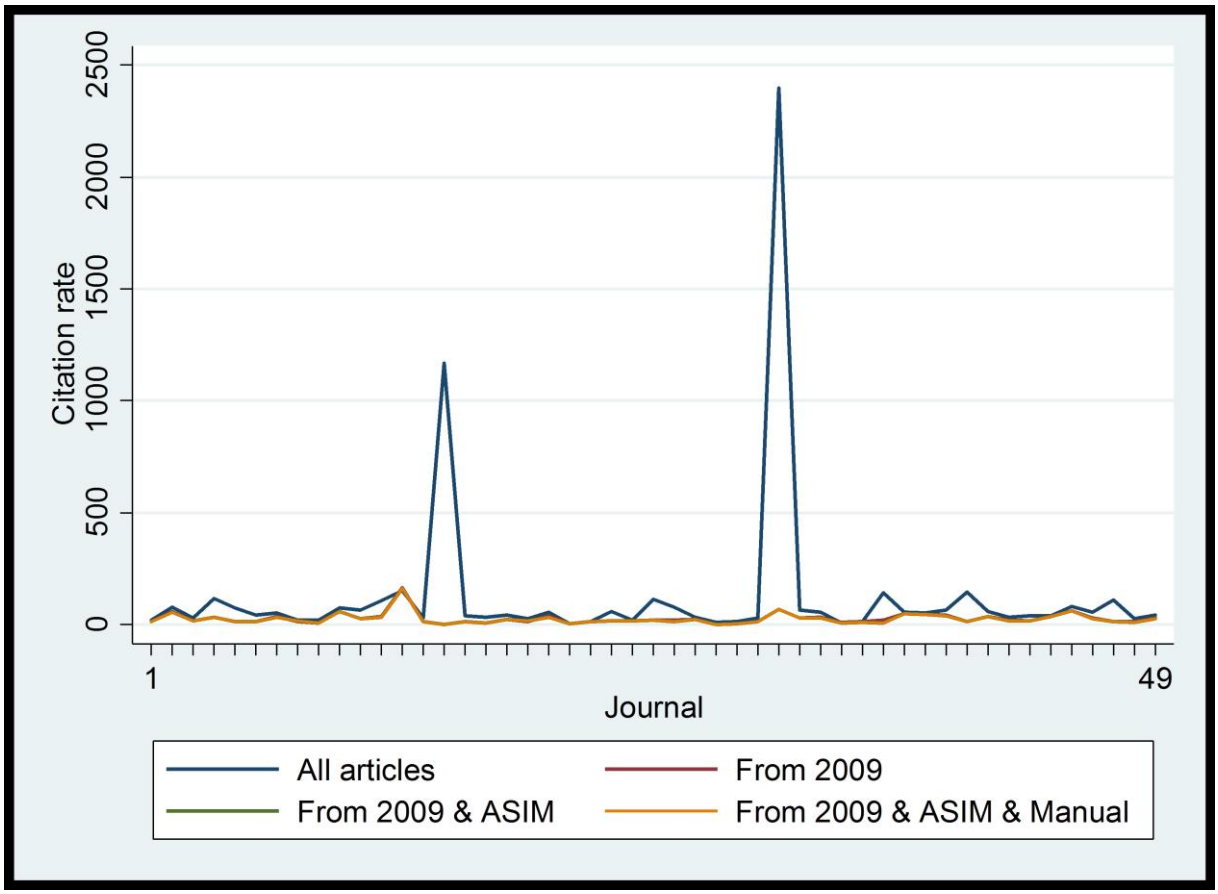

Figure 1
Average number of citations (arithmetic mean) from GS for articles published in 49 different journals. Shown here are the average number of citations for all article hits in a journal, only for article hits from 2009 with an ASIM>.8, as well as for articles from 2009 with an ASIM>0.8 and a manual correction of the data.

For all 56 of the institute's articles, citations could be searched for in GS. Table 4 gives the number of hits for these articles in GS: For a total of 56 articles there were 80 hits. However, the number of hits could be reduced to 56 when only articles from the year 2009, with an ASIM>.8 and a manual correction are taken account of.

Table 4
Total number of hits for the institute's articles in GS

Number of hits in GS	Number of articles
1	45
2	8
3	1
5	1
11	1

On the basis of the citations searched for in GS for the journals in which the staff of the institute have published their articles, we calculated the JNCS for each (based on GS). In addition, we have researched these scores in the in-house Max Planck Society (MPG) database, which is run by the Max Planck Digital Library (MPDL). This database contains the JNCSs on the basis of WoS. Whereas the citation window for the GS scores related to a period from 2009 to 2014, the citation window for the WoS scores is from 2009 to 2013. If a comparison of the scores calculated with data from the two databases indicated a similarity for the scores for the institute's articles, we could conclude that GS may be used for the bibliometric research evaluation in the area of the humanities and social sciences. A convergent validity of the results would indicate that GS comes to similar conclusions as the WoS – that is, as the database which is applied as standard to research evaluation in the sciences.

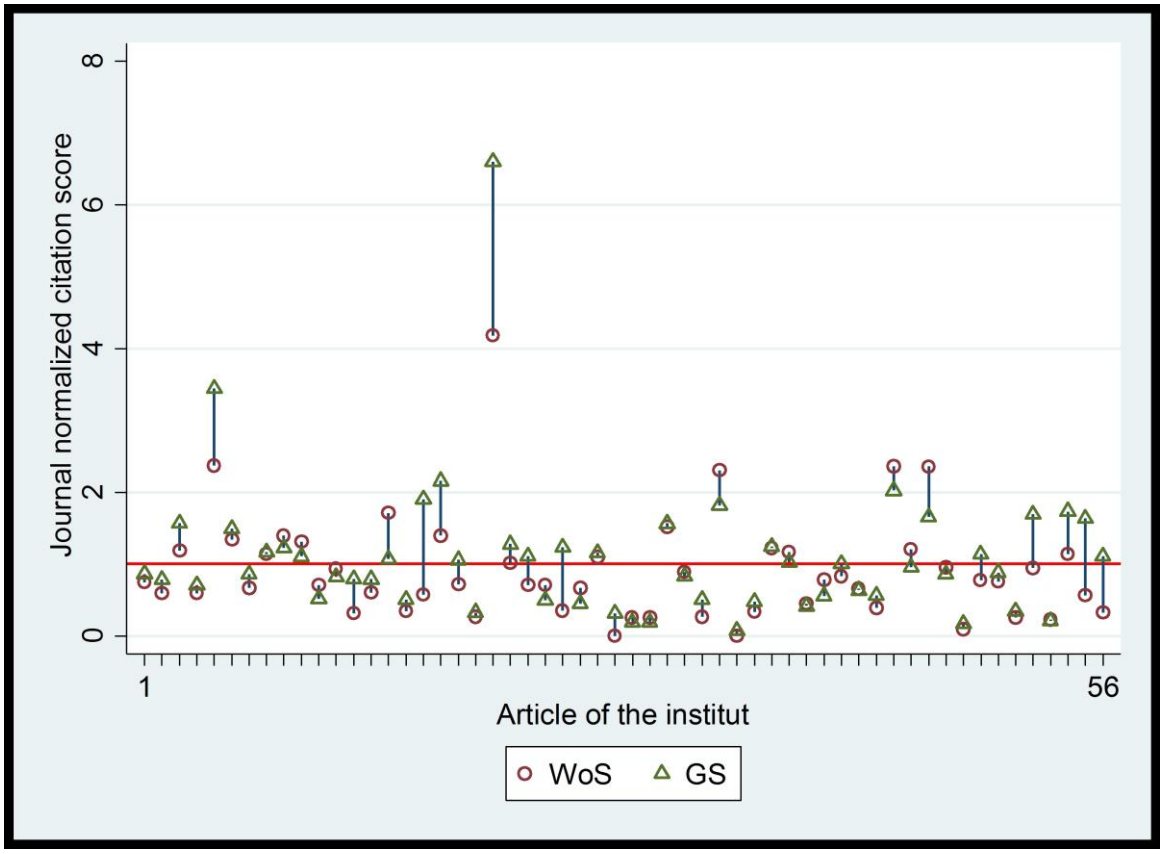

Figure 2
JNCSs for the institute's articles (n=56). The reference values calculated on the basis of GS are derived from articles which were published in 2009, have an ASIM>.8 and have been manually corrected. Whereas the reference values which were calculated on the basis of GS only took into account the subset of the articles which were found in GS (see above), the reference values calculated on the basis of WoS involve all the articles of a journal. The red line marks the citation impact of an article from the institute which corresponds to the average in the journal.

Figure 2 shows the JNCSs for the institute's 56 articles. The red line on the JNCS=1 in Figure 2 marks the citation impact of an article from the institute which corresponds to the average in the journal. As the results show, the scores differ more or less clearly. However, for most articles, the two scores agree on whether they were cited above or below the average rate.

Table 5
Citations and JNCSs for the institute's articles, derived on the basis of the WoS and GS

Statistics	Number of citations in GS	Number of citations in WoS	JNCSs GS	JNCSs WoS
Arithmetic mean	24.57	11.64	1.09	.91
Median	19	8	.92	.73
Minimum	1	0	.07	0
Maximum	126	61	6.60	4.18
N	56	56	56	56

Table 5 shows for all of the institute's articles the average citations and JNCSs, derived on the basis of WoS and GS. Whereas the average citation frequencies clearly differ between the WoS and GS, the JNCSs are similar. Thus the JNCSs are convergent valid: They agree in indicating that the citation impact of the articles roughly corresponds to the average for a journal.

3.2 The citation impact of the journal papers of the institute which are not covered in the Web of Science (but in Scopus)

A total of 29 of the institute's papers were published in journals which are not covered by the WoS (but partly covered in Scopus). Analogously to the procedure in section 3.1, JNCSs based on GS are also calculated for these papers. For this calculation we searched for the citations in GS not only for the 29 of the institute's papers, but also for all other papers in the journals in which the 29 papers appeared. As the figures in Table 6 show, a total of 2628 papers in GS were processed, of which at least one entry in GS was found for 2327.

Table 6
Number of papers sought in GS, and number of hits in GS

Journal	Number of papers for which data was sought in GS	Number of papers with at least one hit in GS	Total number of hits for the papers in GS
1	18	18	25
2	17	13	34
3	96	72	93
4	100	98	116
5	8	7	9
6	54	54	72
7	25	2	2
8	16	16	22
9	26	26	44
10	49	45	53
11	26	23	36
12	33	33	42
13	138	138	197
14	125	125	167
15	819	801	1270
16	32	31	50
17	31	30	31
18	91	76	226
19	45	43	78
20	92	90	151
21	12	9	10
22	17	17	26
23	17	16	24
24	375	203	427
25	300	296	433
26	42	33	33

27	24	12	31
Total	2628	2327	3702

Table 7 shows the number of hits for the papers sought in GS. The number of hits ranges between 1 (n=1593) and 11 (n=1).

Table 7
Number of hits for the papers in GS

Number of hits in GS	Number of papers
1	1593
2	480
3	176
4	66
5	40
6	29
7	11
8	2
9	6
10	1
11	2

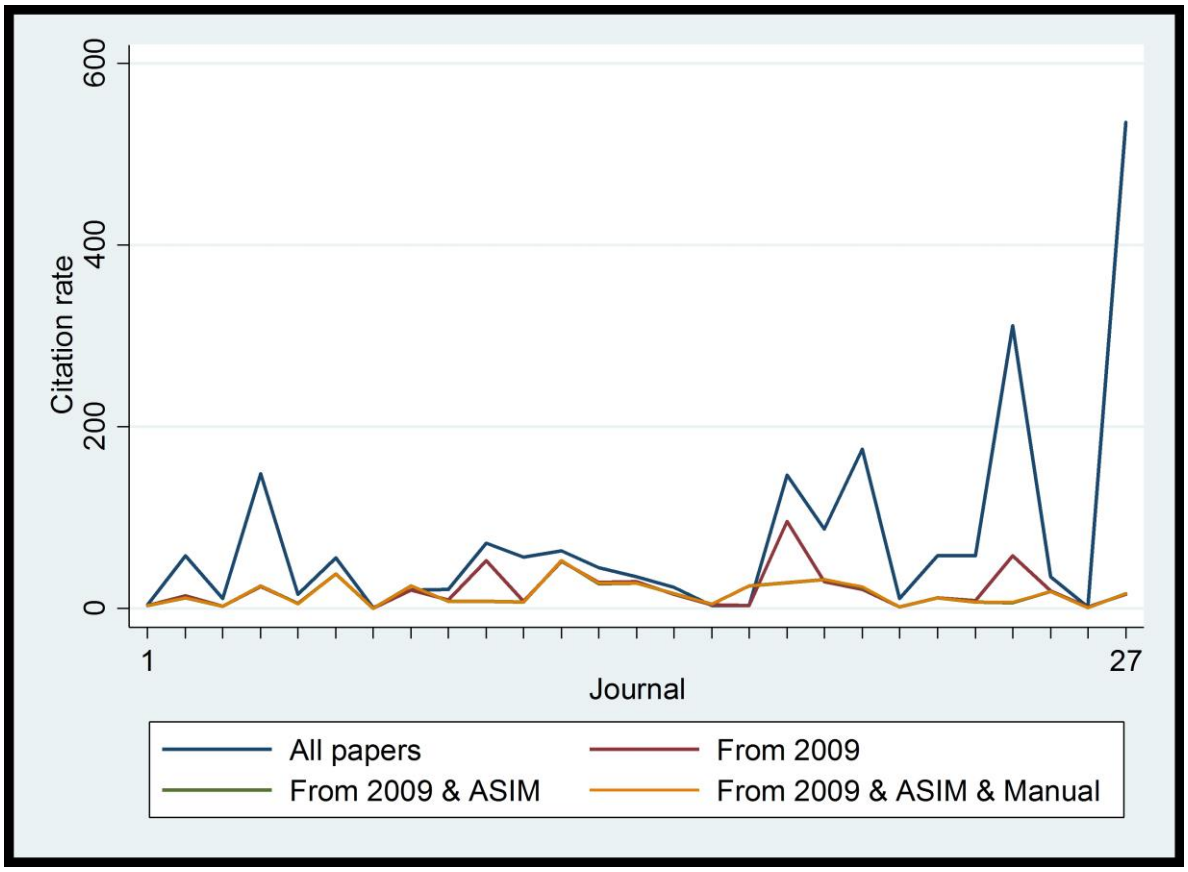

Figure 3
Average number of citations (arithmetic mean) from GS for papers published in 27 different journals. Shown here are the average number of citations for all paper hits in a journal, only for paper hits from 2009 with an ASIM>.8, as well as for papers from 2009 with an ASIM>.8 and a manual correction of the data.

Figure 3 shows the average number of GS citations for the papers published in the 27 journals in which the institute's papers have appeared. Shown here are the average number of citations for all paper hits in a journal, only for paper hits from 2009 with an ASIM>.8, as well as for papers from 2009 with an ASIM>.8 and a manual correction of the data. In agreement with the results reported in section 3.1 it is very clear that the arithmetic means, in particular, which are derived from all the papers hits, deviate from the other means. Because of the deviations in the data we only included papers for the calculation of the reference values (exactly how in section 3.1) which were published in 2009, have an ASIM>.8 and were manually corrected.

A total of 25 of the institute's papers have a hit in GS; 21 of these papers were published in 2009, have an ASIM>.8 auf and were manually corrected. Table 8 shows, for these papers, the average number of citations in GS and the average JNCS GS.

Table 8
Citations and journal normalized citation scores (JNCSs) for the institute's papers which were derived on the basis of GS and Scopus

Statistics	Number of citations in GS	Number of citations in Scopus	JNCSs GS	JNCSs Scopus
All papers				
Arithmetic mean	25.38	22.56	1.25	1.42
Median	8	11	.86	.92
Minimum	0	1	0	.40
Maximum	226	111	5.10	3.89
N	21	9	21	5
Only Papers with JNCSs Scopus				
Arithmetic mean	55	27.4	1.34	1.42
Median	10	7	.81	.92
Minimum	6	1	.35	.40
Maximum	226	111	4.26	3.89
N	5	5	5	5

For some of the institute's papers (n=9), besides the citations in GS the citations in Scopus (Elsevier) could also be searched for (see Table 8). In addition, for five of these nine papers a JNCS could be calculated. For four of the nine papers, the citations for all papers in the particular reference set were incomplete. Since in this study the normalized scores constructed on the basis of WoS or Scopus are regarded as reference values which reflect the "true" normalized impact, attention was paid to the completeness of the publications in the reference set. For this reason no JNCS was calculated for the four of the institute's papers. As a comparison of the two JNCSs (GS and Scopus) in Table 8 shows, the scores are similar and differ from each other by about 0.2. The values become even more similar (Scopus=1.42 und GS=1.34), if the calculation of the mean JNCS GS only includes those papers (n=5), which were also included in the mean JNCS Scopus.

3.3 The citation impact of the institute's contributions in conference proceedings

Of the total of 39 contributions from the institute in conference proceedings, only four appeared in proceedings volumes which included full papers. The rest were published in volumes with abstracts. Because of the limited scope of abstracts (and the correspondingly lowered expected citation rates) abstracts (meeting abstracts) are generally excluded from bibliometric analyses (Moed, 2005). For the analysis in this study, there are thus only four contributions available for normalization. There are also citation counts for two contributions from the WoS.

The reference set for the four contributions consists in each case of the other contributions published in the proceedings of the same conference. We investigated a total of 100 contributions to the four conferences in GS (of which four were published by authors from the institute). As Table 9 shows, citations in GS could be found for 65 contributions.

Table 9
Number of conference contributions sought in GS, and number of hits in GS

Conference	Number of papers for which data was sought in GS	Number of papers with at least one hit in GS	Total number of hits for the papers in GS
1	14	13	21
2	55	31	40
3	23	13	16
4	8	8	9
Total	100	65	86

From the figures in Table 9 it is clear that more than one hit was found in GS for a series of papers. As Table 10 shows, there were up to three hits for one and the same conference contribution.

Table 10
Number of hits for the conference contributions in GS

Number of hits in GS	Number of papers

1	49
2	11
3	5

As with the institute's papers which have appeared in journals (see sections 3.1 and 3.2), the question also arises for conference contributions which hits for a paper in GS should be included in the calculation of the citation rate for the reference set of a conference.

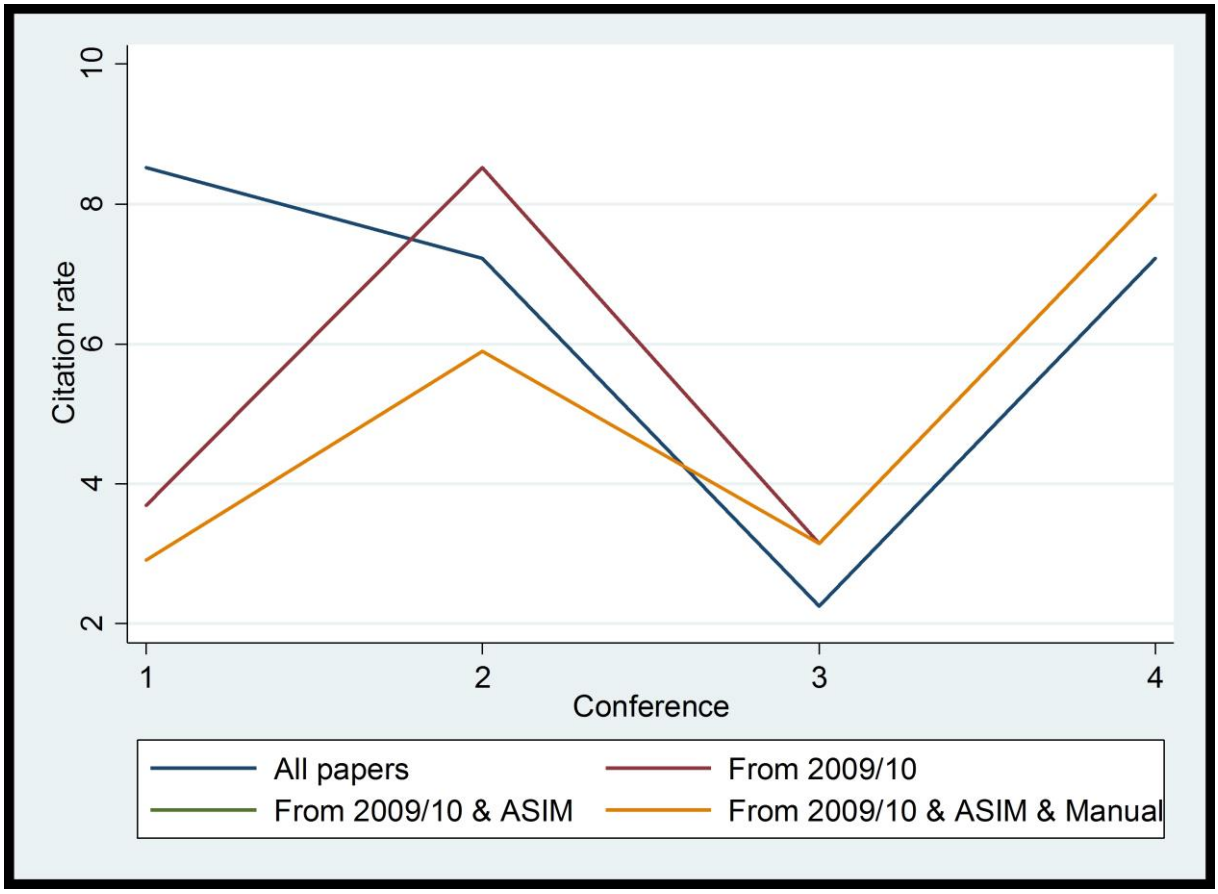

Figure 4
Average number of citations (arithmetic mean) from GS for papers published in conference proceedings. Shown here are the average number of citations for all paper hits in a conference proceedings, only for paper hits from 2009 with an ASIM>.8, as well as for papers from 2009 with an ASIM>.8 and a manual correction of the data.

Figure 4 shows the average number of citations for all paper hits for a conference, only for paper hits from 2009/10, only for paper hits from 2009/10 with an ASIM>.8, as well as for papers from 2009/10 with an ASIM>.8 and a manual correction of the data. Since the papers

from a conference (which took place in 2009) were not published in 2009, but in 2010, both years were taken account of in the evaluation. A greater deviation from the other hit groups was particularly noticeable for "All papers". The results for the three other groups are similar or largely identical.

Table 11
Citations and conference proceedings normalized citations scores (CPNCS) for the institute's papers, based on the WoS and GS

Statistics	Number of citations in GS	Number of citations in WoS	CNCS GS	CNCS WoS
1	6		2.06	
2		0		0
3	0		0	
4	16	5	1.97	2.22

Of the institute's four conference papers, three could be found with one hit each in GS. The corresponding citation counts are shown in Table 11. Whereas one paper had no impact at all, the two other publications were cited 6 times and 16 times respectively. These citation counts were used to calculate the CPNCS GS for the three papers. For this, the citations were each divided by the mean number of citations for the conference papers in the reference set. The reference set used – following the procedure in sections 3.1 and 3.2 – the respective citations from the paper hits from 2009/10 with an ASIM>.8 and manual correction of the data. As the normalized scores in Table 11 show, the two papers which were able to produce citation impact, have much higher scores than the mean value of 1. Since all the papers from one conference whose GS numbers are in the table can also be investigated in the WoS, a comparison with the impact achieved there was able to be made for one paper. With scores of 1.97 (GS) and 2.22 (WoS) the paper has similar normalized values, which indicate about twice as great an impact as for the average conference paper.

3.4 The citation impact of the institute's book chapters

The analysis of the citation impact of the book chapters includes 71 of the institute's publications, which were published in a total of 40 books. As Table 12 shows, a hit in GS could be achieved for only about half of the chapters. Thus, for example, the 17 chapters in book 1 included only one chapter with at least one hit. The chapters in one book could not be investigated at all in GS (book no. 40).

Table 12
Number of book chapters sought in GS, and number of hits in GS

Book	Number of chapters for which data was sought in GS	Number of chapters with at least one hit in GS	Total number of hits for the papers in GS
1	17	2	4
2	22	17	22
3	36	6	6
4	18	16	24
5	31	17	26
6	30	12	16
7	12	12	18
8	44	15	18
9	15	14	18
10	23	18	46
11	299	129	220
12	12	2	3
13	11	8	13
14	18	14	22
15	13	10	15
16	46	39	43
17	23	15	16
18	7	1	2
19	20	15	23
20	12	10	11
21	45	15	18
22	13	12	13
23	22	22	35
24	12	10	11
25	14	12	15
26	28	10	29
27	16	6	7
28	16	2	3
29	13	5	10
30	20	12	21

31	9	9	11
32	19	12	18
33	99	28	44
34	9	9	31
35	27	27	243
36	58	42	88
37	13	12	15
38	19	4	4
39	24	10	11
40	19	0	0
Total	1204	631	1193

From the results shown in Table 13, it is also evident that many of the book chapters found have achieved not one, but several hits in GS. This means that not only were relatively few chapters found in GS; the chapters found often had more than one hit (the latter indicates few accurate search results).

Table 13
Total number of hits for book chapters in GS

Number of hits in GS	Number of chapters
1	457
2	154
3	49
4	22
5	7
6	4
7	3
8	1
9	1
10	1
13	4
15	2
16	4
18	18
19	2
20	1
22	1

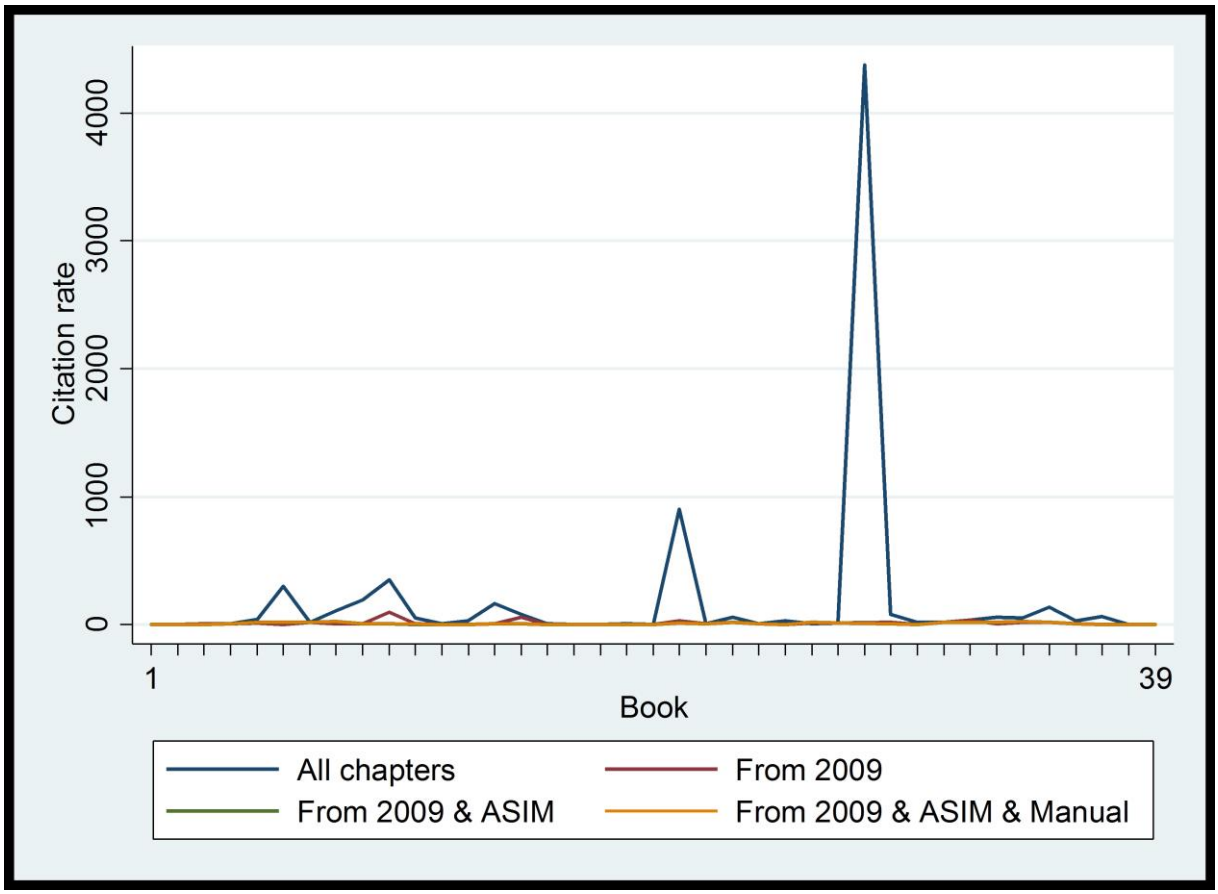

Figure 5
Average number of citations (arithmetic mean) from GS for chapters published in 39 books. Shown here are the average number of citations for all chapter hits in a book, only for chapter hits from 2009, as well as only for chapter hits from 2009 with an ASIM>.8, as well as only for chapter hits from 2009 with an ASIM>.8 and a manual correction of the data.

Figure 5 shows the average number of citations from GS for chapters published in 39 books. Shown here are the average number of citations for all chapter hits in a book, only for chapter hits from 2009, only for chapter hits from 2009 with an ASIM>.8, as well as only for chapter hits from 2009 with an ASIM>.8 and a manual correction of the data. The results are largely in agreement with the results reported in the other sections: Above all, the results relating to all chapter hits differ from those results generated on the basis of hits selected in other ways.

It was possible to find data in GS for 55 of the total of 71 of the institute's book chapters; for 48 GS also contains citation information. Many chapters had only one hit in GS, as Table 14 shows.

Table 14
Number of hits for the institute's book chapters in GS

Number of hits in GS	Number of chapters
1	33
2	12
3	4
4	2
5	2
15	1
18	1

Table 15
Citations and book chapters normalized citation scores (BCNCS) for the institute's chapters, produced on the basis of GS

Statistics	Number of citations in GS	BCNCS
Arithmetic mean	14.06	1.2
Median	5	.96
Minimum	0	0
Maximum	65	4.29
N	34	34

For a total of 34 of the institute's chapters we were able to calculate a normalized citation score (BCNCS) for which only chapter hits for 2009, with an ASIM>.8 and a manual correction of the data were included in the evaluation. As the score in Table 15 shows, the institute's book chapters were cited about 20 percent points more often than the other chapters in the books (BCNCS=1.2).

3.5 The citation impact of the institute's books

The institute published a total of 10 books in 2009. Of these, citations for 8 books could be found in GS (only two could be found in the Book Citation Index, BCI, of the WoS). The number of citations ranges between 0 and 72. For books, it is unfortunately neither

possible to investigate them in the WoS (the coverage in the BCI is too limited), nor to calculate normalized values. Since Torres-Salinas, et al. (2014) have already proposed methods to calculate normalized citation impact values based on the BCI, these methods could be used in coming years, when the coverage of the BCI has improved. Furthermore, one could try to transfer these methods from the BCI to GS.

4 Discussion

Evaluation of research based on bibliometrics has one decisive advantage: In almost all disciplines, one focuses on the primary outcome of research (i.e. publications) and their usefulness for further research (i.e. citations). Since the application of the two most important bibliometric databases WoS and Scopus is limited mainly to the natural and life sciences, we have presented in this study an instrumentarium with which GS data can be applied to the evaluation in the social sciences and humanities. As the list of Martín-Martín, et al. (2014) shows, the most important sources for publications and their citations have now been evaluated by Google: "Google Scholar's crawlers sweep the entire academic web: the most well-known scholarly publishers (such as Elsevier, Springer, Sage, Willey, Taylor & Francis, IEEE, ACS, ACM, Macmillan, Wiley, Oxford University Press); their digital hosts/facilitators (such as HighWire Press, MetaPress, Ingenta); societies and other scholarly organizations (such as the American Physical Society, American Chemical Society, ACM), government agencies (National Institute of Health, National Oceanic and Atmospheric Administration, U.S. Geological Survey), databases (Pubmed, ERIC), disciplinary repositories (such as arXiv.org, Astrophysics Data System, RePEc, SSRN, CiteBase), institutional repositories from universities or research centres, library catalogues (Dialnet), as well as personal web pages from researchers, professors, research groups, departments, faculties… hosted inside the servers of the university or research centre they belong to" (p.41).

According to estimates by Orduña-Malea, Ayllón, Martín-Martín, and Delgado López-Cózar (2014), GS today includes about 160 million documents. The results of a survey by Van Noorden (2014) show that GS and its derivatives are the most used products by scientists. More and more institutions and people are recommending that one put the URL for the GS Citations page in one's CV and on one's personal Web site. In snowball metrics – a global standard enabling cross-institutional comparisons which have been defined and agreed by higher education institutions (and Elsevier) (Colledge, 2014) – the use of GS as the primary data source for bibliometric analyses is recommended (besides WoS and Scopus). One argument for the GS use is that people can easily evaluate departments and institutions if the GS Citations pages of the faculty are easily available. Particularly when universities are evaluated, which generally cover a broad range of disciplines (Bornmann, de Moya Anegón, & Mutz, 2013), GS data could be used: According to the results of Martín-Martín, et al. (2014) on GS publications, around half of the highly-cited documents cannot be found in the WoS, and almost 20% of the highly-cited documents are books. In addition the number of books, which are hardly evaluated for the WoS, has continually increased in recent years, and "become the most frequent document type in the last five years (2009-2013)" (p. 18).

However, GS today is not without disadvantages: (1) The ease with which GS indicators can be manipulated (Delgado López-Cózar, Robinson-García, & Torres-Salinas, 2014) and the transience of the results and measures (in many cases difficult to replicate stably): A comparison of two samples of 64,000 highly cited documents (May and October, 2014) showed that "14.7% of the 64,000 documents in the most recent sample were not also present in our earlier sample. Moreover, most of these new documents are placed in pretty low positions in Google Scholar's ranking of results" (Martín-Martín, et al., 2014, p. 16).

In order to be able to use GS in the evaluation of research in the humanities and social sciences as well, we have presented in this study procedures for normalization of citation impact which are derived from the procedures of classic bibliometrics. With these suggestions

we are following recommendations as they have been formulated by e.g. Prins, et al. (2014): "To use GS in the context of evaluation, various ways for benchmarking or field normalization have to be worked out, for instance on the basis of available journal data, to address the issues of research assessments" (p. 442). The normalization of citation impact proposed in this study can lead not only to a reduction of errors in the GS data on impact measurement (errors average each other out and the statistical accuracy of prediction is mainly determined), but also relativize the generally higher citation counts of GS in comparison with WoS and Scopus: "In our sample, 91.6% of the documents have received more citations in GS than in WoS. Only 3,079 documents (9.4%) have more citations according to WoS than in GS" (Martín-Martín, et al., 2014, p. 33).

Even if we could not find citations in GS for all papers in a reference set (journals, conference proceedings, and edited books), the comparison of GS normalized citation scores with the WoS or Scopus normalized scores shows that the reference sets based on GS data are still suitable for normalization. Even if scores calculated with the help of GS and WoS/Scopus are not identical for the different publication types, they are so similar that they result in the same assessment of the institute investigated in this study: The papers of the institutes whose journals are also covered in WoS are cited at about an average rate (compared with the other papers in the journals). Whereas the papers whose journals are not covered in WoS, and the book chapters, are cited about 20 to 40% above the average, the conference papers are cited twice as often as one would expect for the papers from a conference. In the interpretation of the result for the conference papers it should be considered that it is based on only four papers which appeared in proceedings volumes.

Finally, we would like to mention a limitation of our study which future studies should address: Normalization on the basis of single journals is seldom undertaken in bibliometrics. An important reason is that this kind of normalization is disadvantageous for papers which have appeared in reputable (highly-cited) journals. With these journals, the high citation level

results in a paper published there having a worse normalized impact score than a paper appearing in a journal with a rather lower citation level. Instead of journal-based normalization, the recommendation today is normalization on the basis of the papers of a research field, and this is also general practice (Vinkler, 2012). However, we have applied journal-based normalization in this study, as it means less effort in the search for publications and citations in GS. For a research field, considerably more papers would have had to be searched for a reference set. Future research on the normalization of citation impact based on GS data should therefore concentrate on the use of the papers of a research field for the construction of a reference value.

References

Bornmann, L., Bowman, B. F., Bauer, J., Marx, W., Schier, H., & Palzenberger, M. (2014). Standards for using bibliometrics in the evaluation of research institutes. In B. Cronin & C. Sugimoto (Eds.), *Next generation metrics* (pp. 201-223). Cambridge, MA, USA: MIT Press.

Bornmann, L., de Moya Anegón, F., & Mutz, R. (2013). Do universities or research institutions with a specific subject profile have an advantage or a disadvantage in institutional rankings? A latent class analysis with data from the SCImago ranking. *Journal of the American Society for Information Science and Technology, 64*(11), 2310-2316.

Bornmann, L., & Marx, W. (2014). How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations. *Scientometrics, 98*(1), 487-509. doi: 10.1007/s11192-013-1161-y.

Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A., & 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. doi: 10.1016/j.joi.2008.11.001.

Butler, L., & Visser, M. S. (2006). Extending citation analysis to non-source items. *Scientometrics, 66*(2), 327-343. doi: DOI 10.1007/s11192-006-0024-1.

Chi, P. S. (2013). *Do non-source items make a difference in the social sciences?* Paper presented at the Proceedings of ISSI 2013 - 14th International Society of Scientometrics and Informetrics Conference.

Chi, P. S. (2014). Which role do non-source items play in the social sciences? A case study in political science in Germany. *Scientometrics, 101*(2), 1195-1213. doi: DOI 10.1007/s11192-014-1433-1.

Colledge, L. (2014). *Snowball Metrics Recipe Book*. Amsterdam, the Netherlands: Snowball Metrics program partners.

Delgado López-Cózar, E., Robinson-García, N., & 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), 446-454. doi: 10.1002/asi.23056.

Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics, 83*(1), 243-258.

Frandsen, T. F., & Nicolaisen, J. (2008). Intradisciplinary differences in database coverage and the consequences for bibliometric research. *Journal of the American Society for Information Science and Technology, 59*(10), 1570-1581. doi: 10.1002/asi.20817.

Gorraiz, J., Purnell, P. J., & Glänzel, W. (2013). Opportunities for and limitations of the Book Citation Index. *Journal of the American Society for Information Science and Technology, 64*(7), 1388-1398. doi: 10.1002/asi.22875.

Harzing, A.-W. (2013). A preliminary test of Google Scholar as a source for citation data: a longitudinal study of Nobel prize winners. *Scientometrics, 94*(3), 1057-1075. doi: 10.1007/s11192-012-0777-7.

Harzing, A.-W. (2014). A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics, 98*(1), 565-575. doi: 10.1007/s11192-013-0975-y.

Hug, S. E., Ochsner, M., & Daniel, H.-D. (2013). Criteria for Assessing Research Quality in the Humanities – A Delphi Study among Scholars of English Literature, German Literature and Art History. *Research Evaluation, 22*(5), 369-383.

Hug, S. E., Ochsner, M., & Daniel, H.-D. (in press). A framework to explore and develop criteria for assessing research quality in the humanities. *International Journal for Education Law and Policy*.

Jacso, P. (2005). Google Scholar: the pros and the cons. *Online Information Review, 29*(2), 208-214.

Jacso, P. (2009). Google Scholar's ghost authors. *Library Journal, 134*(18), 26-27.

Jacso, P. (2012). Google Scholar metrics for publications: the software and content features of a new open access bibliometric service. *Online Information Review, 36*(4), 604-619. doi: 10.1108/14684521211254121.

Köpcke, H., Thor, A., & Rahm, E. (2010). *Evaluation of entity resolution approaches on real-world match problems.* Paper presented at the 36th Intl. Conference on Very Large Databases (VLDB) / Proceedings of the VLDB Endowment.

Kousha, K., & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: a multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology, 58*(7), 1055-1065.

Kousha, K., & Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: a comparison between four science disciplines. *Scientometrics, 74*(2), 273-294.

Kousha, K., Thelwall, M., & Rezaie, S. (2010). Using the Web for research evaluation: The Integrated Online Impact indicator. *Journal of Informetrics, 4*(1), 124-135. doi: DOI 10.1016/j.joi.2009.10.003.

Kousha, K., Thelwall, M., & 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), 2147-2164. doi: 10.1002/asi.21608.

Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics, 84*(3), 575-603. doi: DOI 10.1007/s11192-010-0202-z.

Martín-Martín, A., Orduña-Malea, E., Ayllón, J. M., & Delgado López-Cózar, E. (2014). *Does Google Scholar contain all highly cited documents (1950-2013)?* Granada: EC3 Working Papers No. 19.

Marx, W., & Bornmann, L. (2015). On the causes of subject-specific citation rates in Web of Science. *Scientometrics, 102*(2), 1823-1827.

Meho, L. I., & 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), 2105-2125.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht, The Netherlands: Springer.

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: a review. *Scientometrics, 66*(1), 81-100. doi: 10.1007/s11192-006-0007-2.

Ochsner, M., Hug, S. E., & Daniel, H.-D. (2012a). Four types of research in the humanities: Setting the stage for research quality criteria in the humanities. *Research Evaluation*. doi: 10.1093/reseval/rvs039.

Ochsner, M., Hug, S. E., & Daniel, H. D. (2012b). Indicators for Research Quality in the Humanities: Opportunities and Limitations. *Bibliometrie ‐ Praxis und Forschung, 1*(4).

Orduña-Malea, E., Ayllón, J. M., Martín-Martín, A., & Delgado López-Cózar, E. (2014). *About the size of Google Scholar: playing the numbers*. Granada, Spain.

Prins, A., Costas, R., van Leeuwen, T., & Wouters, P. (2014). Using Google Scholar in research evaluation of social science programs, with a comparison with Web of Science data. In E. Noyons (Ed.), *Proceedings of the science and technology indicators conference 2014 Leiden "Context Counts: Pathways to Master Big and Little Data"* (pp. 434-443). Leiden, the Netherlands: Centre for Science and Technology Studies.

Pudovkin, A. I., & Garfield, E. (2009). Percentile Rank and Author Superiority Indexes for Evaluating Individual Journal Articles and the Author's Overall Citation Performance. *COLLNET Journal of Scientometrics and Information Management, 3*(2), 3-10.

Rehn, C., Kronman, U., & Wadskog, D. (2007). *Bibliometric indicators – definitions and usage at Karolinska Institutet*. Stickholm, Sweden: Karolinska Institutet University Library.

Thor, A., & Rahm, E. (2007). MOMA - a Mapping-based Object Matching System. In CIDR (Ed.), *Third Biennial Conference on Innovative Data Systems Research* (pp. 247-258). Asilomar, CA, USA: www.cidrdb.org.

Torres-Salinas, D., Robinson-Garcia, N., Campanario, J. M., & López-Cózar, E. D. (2014). Coverage, field specialisation and the impact of scientific publishers indexed in the Book Citation Index. *Online Information Review, 38*(1), 24-42. doi: 10.1108/OIR-10-2012-0169.

Van Noorden, R. (2014). Scientists and the Social Network. *Nature, 512*(7513), 126-129.

Vinkler, P. (2010). *The evaluation of research by scientometric indicators*. Oxford, UK: Chandos Publishing.

Vinkler, P. (2012). The case of scientometricians with the "absolute relative" impact indicator. *Journal of Informetrics, 6*(2), 254-264. doi: 10.1016/j.joi.2011.12.004.