

Anne-Wil Harzing

March 16, 2017

<http://blogs.lse.ac.uk/impactofsocialsciences/2017/03/16/google-scholar-is-a-serious-alternative-to-web-of-science/>

Google Scholar is a serious alternative to Web of Science

Many bibliometricians and university administrators are fairly conservative in their approach to citation analysis. It is not unusual to see them prefer the Web of Science (ISI for short) as “the gold standard” and discard Google Scholar out of hand, simply because they have heard some wild-west stories about its “overly generous” coverage. These stories are typically based one or more of the following misconceptions, which I will dispute below.

- First, the impression that everything “on the web” citing an academic’s work counts as a citation.
- Second, the assumption that any publication that is not listed in the Web of Science is not worth considering at all.
- Third, a general impression that citation counts in Google Scholar are completely unreliable.

Not everything published on the internet counts in Google Scholar

Some academics are under the misplaced impression that anything posted on the internet that includes references will be counted in Google Scholar. This might also be the source behind the misconception that one can simply put phantom papers online to improve one’s citation count. However, Google Scholar only indexes scholarly publications. As its website indicates: “we work with publishers of scholarly information to index peer-reviewed papers, theses, preprints, abstracts, and technical reports from all disciplines of research.”

Some non-scholarly citations, such as student handbooks, library guides or editorial notes slip through. However, incidental problems in this regard are unlikely to distort citation metrics, especially robust ones such as the h-index. Hence, although there might be some overestimation of the number of non-scholarly citations in Google Scholar, for many disciplines this is preferable to the very significant and systematic underestimation of scholarly citations in ISI or Scopus. Moreover, as long as one compares like with like, i.e. compares citation records for the *same* data source, this should not be a problem at all.

Non-ISI listed publications can be high-quality publications

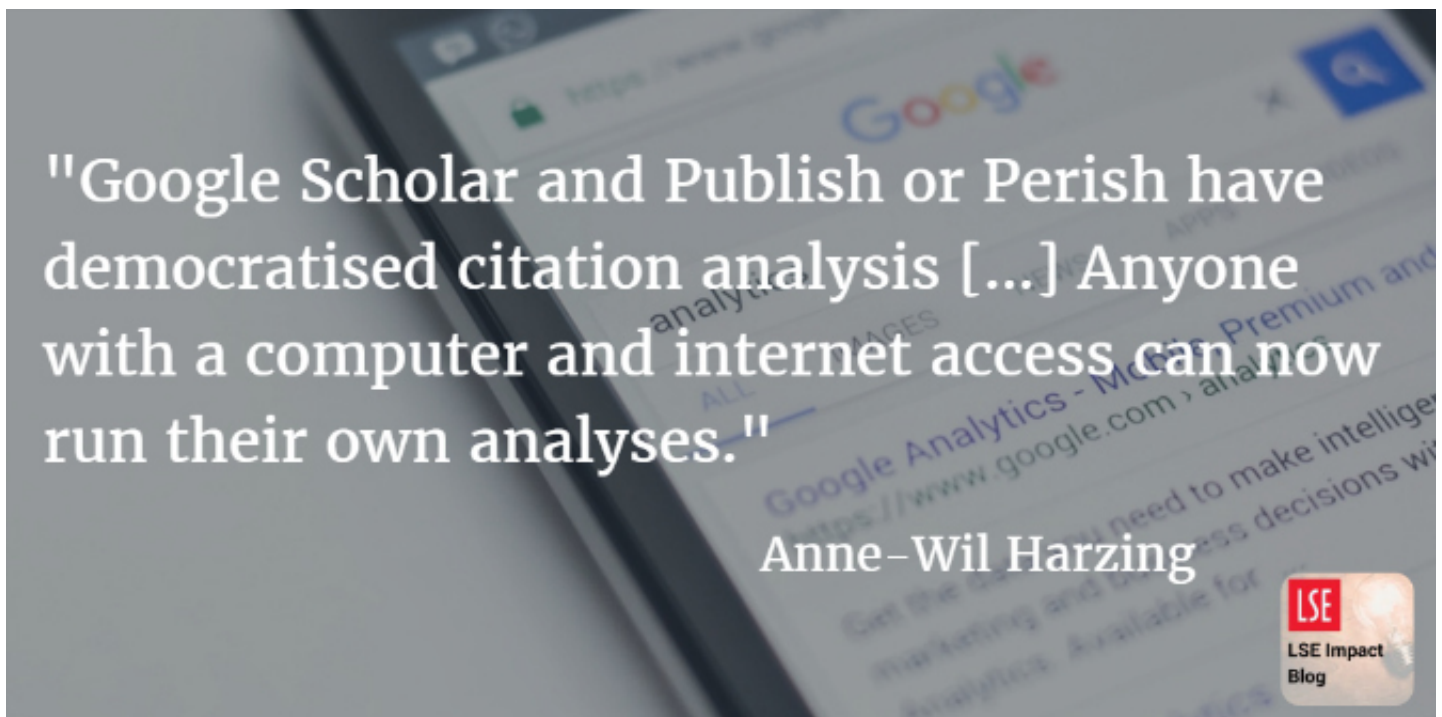
Meaning: Just because a publication isn't listed in Web of Science (also known as ISI), doesn't mean that the publication isn't high-quality.

There is also a frequent assumption amongst research administrators that ISI listing is a stamp of quality and that hence one should ignore non-ISI listed publications and citations. **There are two problems with this assumption.** First, ISI has a bias towards science, **English-language and North American journals.** Second, **ISI completely ignores a vast majority of publications in the social sciences and humanities.**

- **ISI journal listing is very incomplete in the social sciences and humanities:** ISI's listing of journals is much more comprehensive in the sciences than in the **social sciences and humanities.** Butler (2006) analysed **the distribution of publication output by field for Australian universities between 1999 and 2001.** She found that whereas for the chemical, biological, physical and medical/health sciences between 69.3% and 84.6% of the publications were published in ISI listed journals, this was the case for only 4.4%-18.7% of the publications in the social sciences such as management, history education and arts. Many high-quality journals in the field of economics and business are not ISI listed. Only 30%-40% of the journals in accounting, marketing and general management and strategy listed on my **Journal Quality List** (already a pretty selective list) are ISI listed. There is no doubt that – on average – journals that are ISI listed are perceived to be of higher quality. However, there is a very substantial number of non-ISI indexed journals that have a higher than average h-index.
- **ISI has very limited coverage of non-journal publications:** second, even in the cited reference search, ISI only includes citations in ISI listed journals. In the general search function it completely ignores any publications that are not in ISI-listed journals. As a result **a vast majority of publications and citations in the social sciences and humanities, as well as in engineering and computer science, are ignored.** In the social sciences and humanities this is mainly caused by a complete neglect of books, book chapters, publications in languages other than English, and publications in non-ISI listed journals. In engineering and computer science, this is mostly caused by a neglect of conference proceedings. ISI has recently introduced conference proceedings in its database. However, it does not provide any details of which conferences are covered beyond listing some disciplines that are covered. I was unable to find any of my own publications in conference proceedings. As a result ISI **very seriously underestimates both the number of publications and the number of citations** for academics in the social sciences and humanities and in engineering and computer science.

Google Scholar's flaws have been played up far too much

- *Citation metrics are robust and insensitive to occasional errors:* most of the metrics used in Publish or Perish are fairly robust and insensitive to occasional errors as they will not generally change the **h-index or g-index** and will only have a minor impact on the number of citations per paper. There is no doubt that Google Scholar's automatic parsing occasionally provides us with nonsensical results. However, these errors do not appear to be as frequent or as important as implied by Jacsó's articles. They also do not generally impact the results of author or journal queries much, if at all.
- *Google Scholar parsing has improved significantly:* Google Scholar has also **significantly improved** its parsing since the errors were pointed out to them. However, many academics are still referring to Jacsó's 2006 articles as convincing arguments against **any** use of Google Scholar. I would argue this is inappropriate. As academics, we are only all too well aware that **all** of our research results include a certain error margin. We cannot expect citation data to be any different.
- *Google Scholar errors are random rather than systematic:* what is most important is that errors are **random** rather than **systematic**. I have no reason to believe that the Google Scholar errors identified in Jacsó's articles are anything else than random. Hence they will not normally advantage or disadvantage individual academics or journals.
- *ISI and Scopus have systematic errors of coverage:* in contrast, commercial databases such as ISI and Scopus have **systematic** errors as they do not include many journals in the social sciences and humanities, nor have good coverage of conferences proceedings, books or book chapters. Therefore, although it is always a good idea to use multiple data sources, rejecting Google Scholar out of hand because of presumed parsing errors is not rational. Nor is **presuming ISI is error-free** simply because it charges high subscription fees.



Conclusion

As I have **argued in the past**, Google Scholar and **Publish or Perish** have **democratised citation**

analysis. Rather than leaving it in the hands of those with access to commercial databases with high subscription fees, anyone with a computer and internet access can now run their own analyses. If you'd like to know more about this, [have a look at this presentation](#).