

Jun 3rd, 12:00 AM

## **The Effect of Open Access on Citation Rates of Self-Archived Articles at Chalmers**

Lars Kullman  
*Chalmers University of Technology*, [lars.kullman@chalmers.se](mailto:lars.kullman@chalmers.se)

---

Lars Kullman, "The Effect of Open Access on Citation Rates of Self-Archived Articles at Chalmers."  
*Proceedings of the IATUL Conferences*. Paper 2.  
<https://docs.lib.purdue.edu/iatul/2014/bibliometrics/2>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries.  
Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

# THE EFFECT OF OPEN ACCESS ON CITATION RATES OF SELF-ARCHIVED ARTICLES AT CHALMERS

Lars Kullman

Scholarly Communication, Chalmers Library, Chalmers University of Technology,  
Göteborg, Sweden

Correspondence to [lars.kullman@chalmers.se](mailto:lars.kullman@chalmers.se)

## Abstract

Open Access (OA) proponents argue that OA increases the visibility and accessibility of research articles, and therefore increases the citation rate of these works. During the last decade numerous studies have been made on the possible citation advantage of OA on scholarly publications. At Chalmers University of Technology (Göteborg, Sweden) an OA policy was adopted in 2010, mandating all of its publications to be self-archived in the university repository Chalmers Publication Library (CPL). One of the arguments of the then vice chancellor was that OA would increase citations. In this study, a possible OA citation advantage of articles self-archived in CPL is examined. A total of 3424 original articles published 2010-2012 were included, 899 of which were published in full text in CPL, and 2571 that were only registered with bibliographical data. Mean normalized citation scores (MNCS) were calculated using Web of Science citation data processed by the Centre for Science and Technology Studies (CWTS) of Leiden University. Results show that self-archived articles have a 22% higher citation rate than articles that were not self-archived, and that the difference is statistically significant. The limitations and biases of the study are also discussed.

**Keywords:** Open Access, Open Access mandate, Citation advantage, Bibliometrics

## Introduction

Citation impact is a frequently used indicator in research performance assessments as well as in university rankings and other evaluations of research. A positive correlation between Open Access (OA) and citations would therefore be of great interest for the research community as a whole. University libraries have by tradition been an active part in implementing OA, both as supplier of a technical infrastructure and as a source of support in eg. legal issues. So, from a library perspective, such empirical evidence would of course serve as a great driver in the work with promoting OA.

Extensive research on whether OA articles receive more citations than non OA articles has been ongoing for more than a decade, since Lawrence (2001) first published a paper indicating an OA citation advantage in the field of Computer Science. Swan (2010), Wagner (2010) and (OpCit) offer comprehensive summaries of studies carried out on the possible OA citation advantage. A majority of this previous research suggests a positive correlation between OA publishing and citation rate. Some studies have, however, questioned a causal relationship, and pointed to confounding factors behind the observed increase in citations of OA publications.

Swan (2010) summarizes the possible explanations of an OA citation advantage as: (1) A general OA advantage: more scholars have access to papers and these therefore receive more citations. (2) An early advantage: the earlier a paper is made available, the earlier it can start accumulating citations. (3) A selection bias / quality advantage: authors choose to self-archive their best papers, and better papers attract more citations.

Existing research on a possible OA citation advantage has used various different data sources and methodological approaches. Most studies have, however, compared citations to OA and non-OA papers published in the same journal or in a set of journals within a specific research field. This has been argued to be necessary due to differences in citation practice between scientific disciplines (Harnad & Brody, 2004). An alternative approach could be to use citation-based bibliometric indicators that normalize for such differences and thus allow meaningful

cross disciplinary comparisons of citation impact (Schubert & Braun, 1996). Studies on a possible OA citation advantage utilizing field normalized citation data seem to be lacking, but could make an important contribution to this research as they are not limited to comparing likes with likes.

In this study field normalized citation scores is combined with data on self-archiving from the university repository, Chalmers Publication Library (CPL), allowing for cross-field citation comparisons between OA and non-OA articles from Chalmers research publication output.

Research at Chalmers University of Technology (Chalmers) is performed within 17 departments of varying size. About 1200 peer-reviewed articles are published yearly, 80 % of which are indexed by Thomson Reuters Web of Science (WoS) database. Much of the research conducted at Chalmers is applied research, however, and does not generate publications traditionally indexed by WoS. Of the peer-reviewed articles about 25 % are self-archived in CPL. The share of OA varies to a great extent between the departments, from a few percent up to 80 percent.

To assess whether there is a cross-field OA citation advantage for articles published by Chalmers researchers, this paper compares field normalized citation scores of self-archived and non self-archived, peer-reviewed articles in Chalmers Publication Library (CPL).

## Materials and Methods

In this study, 'self-archived paper' is used as a synonym to 'OA article', here defined as a full-text version of a paper freely available in CPL. No distinction is made between published articles (copies edited by the publisher) or final, i.e. accepted manuscripts.

In order to calculate mean normalized citation scores (MNCS), bibliographical data from CPL were matched with field normalized citation data from the Centre for Science and Technology Studies (CWTS) of Leiden University. The analysis from CWTS is based on the WoS data. In total, 3470 articles, published 2010-2012, were matched and out of those 899 were OA. For comparison, mean citation scores (MCS) based on raw citation counts were also calculated.

Abbreviation	Description
MCS	Mean citation score counted on raw citations (not field normalized)
MNCS	Mean (field) normalized citation score. Citations per paper (article, review, letter) compared to world average for publications from the same year and subject field. Publication window: 3 years, citation window: $\leq 3$ years. Self-citations excluded.

Table 1: Descriptions and definitions of man citation score and mean normalized citations score.

MNCS was calculated as:

$$MNCS = \frac{1}{n} \sum_{i=1}^n \frac{c_i}{e_i}$$

$c_i$  = number of citations of publication  $i$

$e_i$  = expected number of citations of publication  $i$  given the field and year in which publication  $i$  has been published

Given the highly skewed distributions of citations within both groups, the Mann-Whitney test was used to determine whether the citation score of OA articles is significantly different from that of non-OA articles. Mann-Whitney is a rank-sum test and the non-parametric analogue of the independent samples *t-test* (Field, 2009).

## Results and discussion

Results show that OA articles are cited 23% above world average, and have a 22% higher mean normalized citation rate than non-OA articles (1,23 compared to 1,01; table 3). The (non-normalized) mean citation score is also higher for OA articles compared to non-OA articles.

Category	Publications	MCS	MNCS
Total	3470	4,43	1,08
OA	899	4,66	1,23
Non-OA	2571	4,27	1,01

Table 2. Number of articles examined and the results from the citation scores compared.

The Mann-Whitney test shows that the citation rate is significantly higher for OA articles than for non-OA publications ( $U=1085091$ ,  $z=-2.756$ ,  $p=.01$ ,  $r=-0.05$ ,  $\bar{x} \text{ rank}=1814$  vs. 1708.05). In contrast, the test shows no significant difference between non-normalized citation counts for the same groups ( $U=1131566,5$ ,  $z=-0,945$ ,  $p=.34$ ,  $r=-0.02$ ,  $\bar{x} \text{ rank}=1762.31$  vs. 1726.13).

This study set out to investigate whether there is a possible OA citation advantage across all fields covered by articles published by Chalmers researchers. The results agree with previous studies indicating such an advantage. There was a significant difference between the two groups when using field normalized data but not when using raw citations, which illustrates the importance of using field normalized citation data in this case.

Many of the previous studies have been based on metadata retrieved from arXiv, and a logical assumption would be that papers published ahead of print have a longer window to gather citations and therefore will be cited more than papers not published as pre-prints. This early bias is also suggested by Moed (2007) to be the explanation to the OA citation advantage. The results from this study, with a high share of OA articles in the field of Physics, points to the direction that these papers might also be published in arXiv as pre-prints. An investigation of this is beyond the scope of this paper, but an interesting topic for future studies.

The observed increase in citation rate for OA papers could arguably be caused by a self-selection bias, i.e. that authors choose to self-archive their best papers, rather than the OA availability *per se*. Gargouri et al. (2010) found no evidence that mandated OA has a smaller citation advantage than self-selected OA papers, however, although it should be added that as long as the compliance level is below 100 %, the possibility of a self-selection bias cannot be dismissed. Analogously, Chalmers has an OA mandate, but as the compliance level is only 25 %, a self-selection bias cannot be ruled out.

Comparing OA frequencies of papers in high- and low-ranked journals, respectively, Xia and Nakanishi (2012) found no evidence that authors selectively make their best papers available online. This would be interesting to investigate in a future cross-disciplinary study, however. Specifically, it would be interesting to see whether there is a correlation between the citation rate of OA articles and field-normalized journal citation scores.

This study confirms the results from previous studies regarding interdisciplinary differences in the share of OA publishing (Björk et al., 2010), were disciplines with a self-archiving culture like Physics and Astronomy come out on top (see table 5.). The appearance of Telecommunications as the subject category with the third largest share of OA at Chalmers is most probably an effect of a publication project where the library helped the department of Signals and Systems upload their articles in CPL. When the project was ended the share of OA articles from the department had risen to about 85 %. This can also be seen as an example on how actions outside author behavior and OA-policies can have great effects on the share of OA at a university (Björk, Laakso, Welling, & Paetau, 2014). As a comparison to the distribution of OA articles, Table 5 shows the disciplines that Chalmers researchers are active within, the total output 2010-2012, categorized according to WoS subject categories.

Web of Science Categories	Share of OA articles
ASTRONOMY ASTROPHYSICS	16,4 %
ENGINEERING ELECTRICAL ELECTRONIC	15,4 %
TELECOMMUNICATIONS	9,9 %
PHYSICS CONDENSED MATTER	9,4 %
OPTICS	6,9 %
METEOROLOGY ATMOSPHERIC SCIENCES	5,9 %
PHYSICS MULTIDISCIPLINARY	5,3 %
PHYSICS APPLIED	4,4 %
CHEMISTRY MULTIDISCIPLINARY	4,0 %
MATERIALS SCIENCE MULTIDISCIPLINARY	3,7 %

Table 3. Distribution of OA frequencies of articles over WoS subject categories, and their share of the total OA article output at Chalmers.

Web of Science Categories	Share of total articles
ENGINEERING ELECTRICAL ELECTRONIC	15,6%
MATERIALS SCIENCE MULTIDISCIPLINARY	9,2%
CHEMISTRY PHYSICAL	8,9%
PHYSICS APPLIED	7,7%
TELECOMMUNICATIONS	5,6%
CHEMISTRY MULTIDISCIPLINARY	5,3%
PHYSICS CONDENSED MATTER	5,3%
ASTRONOMY ASTROPHYSICS	5,2%
ENERGY FUELS	4,6%
ENGINEERING MECHANICAL	4,4%

Table 5. Distribution of articles over the ten most frequent WoS subject categories (among Chalmers papers), and their share of the total output

This study has its limitations. The citation window for articles published in 2012 is just one full year. It is not taken into consideration if the OA articles are published in an OA journal or a subscription based journal. Embargo periods are not taken into consideration, nor if the papers is self-archived elsewhere (e.g. arXiv).

## Conclusion

This study confirms the results from many previous studies that there is an OA citation advantage. The OA articles studied in this paper have a 22% higher field normalized citation rate than the non-OA articles, and the difference is statistically significant.

Whilst this study has focused on the publications from just one university, a second theoretical contribution is that this study gives an example how make between field comparisons on the possible OA citation advantage using field-normalized citation data.

## Acknowledgements

The author acknowledges and thanks David Minguillo and Maria Prager for or their helpful comments and suggestions when reviewing an earlier draft of this paper. The author also thanks David for doing the Mann-Whitney tests.

## References

- Björk, B. C., Laakso, M., Welling, P., & Paetau, P. (2014). Anatomy of green open access. *Journal of the American Society for Information Science and Technology*, 65(2), 237-250.
- Björk, B. C., Welling, P., Laakso, M., Majlender, P., Hedlund, T., & Gudnason, G. (2010). Open Access To The Scientific Journal Literature: Situation 2009. *Plos One*, 5(6).
- Field, A. (2009). *Discovering Statistics Using SPSS*: SAGE Publications Ltd.
- Gargouri, Y., Hajjem, C., Lariviere, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *Plos One*, 5(10). doi: 10.1371/journal.pone.0013636
- Harnad, S., & Brody, T. (2004). Comparing the impact of Open Access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*, 10(6).
- Lawrence, S. (2001). Free online availability substantially increases a paper's impact [1]. *Nature*, 411(6837), 521.
- Moed, H. F. (2007). The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology*, 58(13), 2047-2054.
- OpCit. The Open Citation (OPCIT) Project. 2014, from <http://opcit.eprints.org>
- Schubert, A., & Braun, T. (1996). Cross-field normalization of scientometric indicators. *Scientometrics*, 36(3), 311-324. doi: 10.1007/BF02129597
- Swan, A. (2010). The open access citation advantage: Studies and results to date. *Key Perspectives Report*.
- Wagner, A. B. (2010). Open access citation advantage: an annotated bibliography. *Issues Sci Technol Librariansh*.
- Xia, J. F., & Nakanishi, K. (2012). Self-selection and the citation advantage of open access articles. *Online Information Review*, 36(1), 40-51.