Do Open Access Working Papers Attract more Citations Compared to Printed Journal Articles from the same Research Unit?

Peter Ingwersen[1,2,3] and Anita Elleby[4]

[1] *pi@iva.dk*
Royal School of LIS, Birketinget 6, DK 2300 Copenhagen S (Denmark)
[2] Oslo University College, Pb 4 St. Olavs plass, 0130 Oslo (Norway)
[3] Åbo Akademi University, Fänriksgatan 3 B, FI 20500 Turku (Finland)

[4] *anita.elleby@gmail.com*
The National Museum, I. C. Modewegs Vej, Brede, DK 2800 Kongens Lyngby (Denmark)

Abstract

This paper presents the results of an empirical case study of the characteristics of citations received by 10 open access non-peer reviewed working papers published by a prestigious multidisciplinary, but basically social science research institute, compared to 10 printed peer reviewed journal articles published in the same year (2004) by the same institute and predominantly by the same authors. The study analyzes the total amount of citations and citation impact observed in Web of Science (WoS) and Google Scholar (GS) received during the five-year period 2004-09 (February) by the two publication types, the citation distributions over the individual sample publications and observed years as well as over external, institutional and personal self-citations. The institute concerned is the Danish Institute for International Studies (DIIS), Copenhagen. The results demonstrate that the open access working papers publicly accessible through the DIIS e-archive became far less cited than the corresponding sample of DIIS journal articles published in printed form. However, highly cited working papers have higher impact than the average of the lower half of cited articles. Citation time series show identical distinct patterns for the articles in WoS and GS and working papers in GS, more than doubling the amount of citations received through the latter source.

Introduction

It has been argued that open access (OA) full text journal articles become more cited than more traditionally published and accessible articles (Harnad & Brody, 2004). In a later and more methodologically rigorous study by Moed (2007) OA articles (posted in arXiv) indeed seemed to receive more citations for the same period of time and set of journals than did non-OA. However, the difference was not great and seems to depend also on the observation window length. This picture may turn more fuzzy as printed journals increasingly also publish an online version (posting) or/and allow pre-print online access (early view) through tables of contents made public on their websites.

A profound review of OA citation research characteristics including discussions of methodological issues and central influencing factors, such as 'early view' and 'selection bias' on the research results is provided by Craig et al. (2007). The latter concept signifies when an article (also) is posted on a highly visible e-repository like the arXiv. According to the review a substantial number of OA citation studies have based their data gathering on that archive. The review also discusses the potential effect of other factors than OA on the citation patterns, such as funding or number of authors and their status, that may influence the patterns mixed with the OA effects.

Another dimension concerns the data source used for citation extraction. The amount of citations to the same publications for the same period of time is found constantly to be substantially larger through Google Scholar (GS) than via Web of Science (WoS), but more cumbersome to validate in the former (Jacso, 2008).

Our case study explores the effect of posting non-peer reviewed *working papers* (WPs) on an institutional e-repository compared to the citation fate of peer reviewed journal articles, predominantly published by the same authors from that institution in the same year as the

former and observed over the same fixed time span. The journal articles were published in traditional printed form, with no early view involved, whilst the WPs formed part of a recognized institutional series. The research institution concerned is the Danish Institute for International Studies (DIIS).

The motivation behind the investigation lies in the assumption that open access working papers in full text from a prestigious research institution, as a quite visible knowledge source through its e-archive, on average may obtain the same amount of citations through the citation indexes as the traditionally peer reviewed but printed journal articles on the same subject matter from the same institution. A sub-hypothesis is based on the additional assumption that this equality in impact appears through Google Scholar in particular. A larger ratio of internal institutional and personal self citations given to the working papers are considered important reasons underlying these assumptions.

Previously Frandsen studied citation patterns for WPs compared to selected journals in economics (2009). She examined a 10-year period (1996–2005) and her study showed "[that] working papers are increasingly becoming visible in the field specific databases. The impact of working papers is relatively low; however, high impact working paper series have citation rate levels similar to the low impact journals in the field. There is no tendency to an increase in impact during the 10 years which is the case for the high impact journals." (Frandsen, 2009: 124). The result of the study did not provide evidence of an open access advantage for WPs in economics.

Our present case study also incorporates economics, but observes a broader range of academic fields. It looks into the following aspects of OA working paper citations: The analysis of the total amount of citations and citation impact observed in Web of Science (WoS) and Google Scholar (GS) received during the five-year period 2004-08 by DIIS journal articles and WPs (carried out in February, 2009); the citation distributions over the individual sample publications and observed years; the average impact for WPs compared to selected quantiles of the articles by impact; and the citation distribution over external, institutional and personal self-citations.

The remaining of the paper describes briefly the Danish Institute for International Studies (DIIS) and outlines the data extraction and analysis methods used. This is followed by the case study findings and a concluding discussion.

The Danish Institute for International Studies - DIIS

The Danish Institute for International Studies (DIIS) was founded January 1, 2003 as a fusion of four research institutions, thus creating a highly interdisciplinary independent research unit with a full time staff of approx. 55 researchers during the analysis period. Humanistic and social science fields are predominant, but also Agricultural, Environmental and Geo-fields are represented.

The present analysis covers ten working papers and ten international peer reviewed journal articles published in 2004 by DIIS authors and cited 2004-08 (February 2009). The DIIS publication volume varies slightly during the citation analysis period, but a combined citation and publication analysis of DIIS production based on its 2006 publications demonstrates the following quite typical distribution across document types (Elleby & Ingwersen, 2010): journal articles: 119 (22 internationally peer reviewed and 80 of popular nature); book chapters/conference papers: 64 (27 peer reviewed and 30 editor reviewed); monographs: 23 (15 in non-Danish); and non-peer reviewed working papers: 23. Our case study hence samples approx. 50 % of the peer reviewed journal articles and of the non-peer reviewed WPs, respectively.

Data Collection and Analysis Methods

10 WPs published in 2004 by DIIS authors were drawn stratified across the subject field distribution; a corresponding number of 10 peer reviewed international journal articles by predominantly the same authors and covering the same subject fields as the WPs were drawn from the pool of international journal articles published in 2004. Not all authors produce WPs as well as peer reviewed journal articles during that same year. A few articles were consequently by other but similarly prestigious authors as the WP authors, as far as possible covering the same subject fields. WPs and articles were searched one-by-one in the Thomson-Reuters WoS data system in order to establish the number of received citations during the five-year period 2004-08. Similarly, GS was searched during February 2009 to include open access citations given to the same 2 x 10 publications covering the same time period. This provided some additional citations found for January-February 2009 in both citation indexes to both publication types.

The citation impact analysis calculated the distribution of citations, including citedness, and impact across the two publication types and the two citation indexes as well as over external, internal and personal self citations and demonstrated the distribution across the individual documents. Chi^2 tests were applied to the ratio between external and internal/self citation distributions.

The annual distribution of citations covered the five-year period 2004-08, omitting the few additional 2009 citations as well as citations with no traceable publication year. With reference to our hypotheses we compared the mean WP citation impact with the upper and lower half of the cited journal articles' average impact applying the median point as separator.

Case study findings

Citation impact and citation origin

Table 1 shows the overall citation impact of the DIIS articles and WPs as well as distribution over origin (external, internal or personal self citations) and data source (WoS or GS).

Table 1. Citations received 2004-2009 (Feb.) to publications published 2004 and their distribution over origin and citation index; * signifies statistical significance.

Publication Type	Publ.	WoS	GS	WoS Impact	GS Impact	WoS Cit %	GS Cit %	WoS+GS Cit (%)
Printed Articles	10	34	126	3.4	12.6	100	100	160 (100)
External citations		26	114	2.6	11.4	76	90*	140 (87*)
Internal citations		3	6	0.3	0.6	9	5	9 (6)
Self citations		5	6	0.5	0.6	15	5	11 (7)
OA Work Papers	10	8	37	0.8	3.7	100	100	45 (100)
External citations		6	28	0.6	2.8	75	76	34 (76)
Internal citations		2	5	0.2	0.5	25	13	7 (16)
Self citations		0	4	0	0.4	0	11	4 (8)

The difference in ratio of external citations over internal and personal self citations was statistically significant only for the *printed journal articles* across both citation indexes (87 % external vs. 6 % and 7 %) against the WPs (76 % external vs. 16 % and 8 %). This is owing to the more differentiated distribution found through GS (90 % vs. 5 and 5 %) for the articles. Within the open access working papers a substantial difference of ratio indeed exits, but it is not statistically significant. For WoS alone the differences are not significant for any of the two document types.

For articles as well as WPs the volume of citations (and impact) found through GS was approx. *four times* that found in WoS for the same items. Nevertheless, we observe Figure 1 an outlier with respect to the GS citations to articles (article *b*). Most of the citations given that article are external ones. If the analysis omits the outlier document, the GS impact for the

9 articles is reduced to 5.9 on average. Although still more than *doubling* the amount found on average through WoS (2.6) and 1.5 times that found for WPs by GS (3.7), without the outlier *no statistical significant difference* exists between the article and WP ratios of external vs. internal/self citations for any of the citation indexes (i.e. approx. 75 % external vs. 25 % internal/self-citations).

Citedness

The diagrams, Figure 1, demonstrates the volume of citations received during the analysis period per article (left) and WP (right). The diagrams include the mentioned outlier (article *b*).

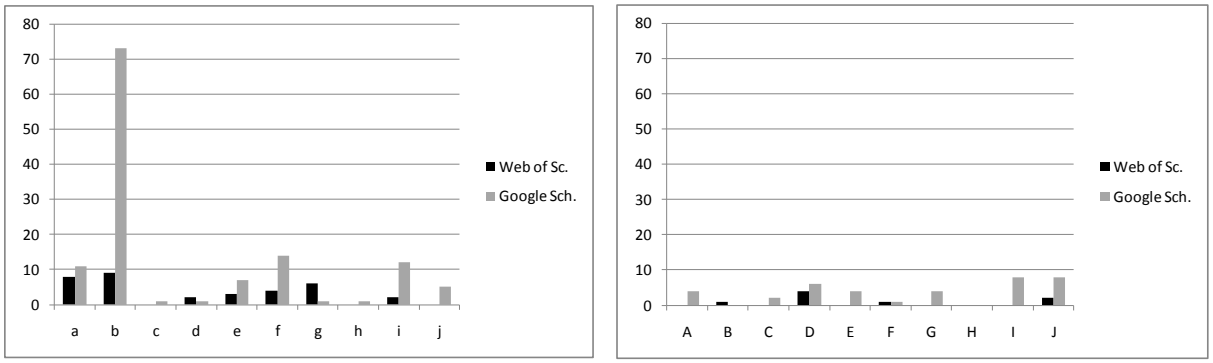

Figure 1. Distribution of citations over the analysed 10 journal articles (left) and 10 working papers (right) and across the two data sources, Web of Science and Google Scholar.

The citedness over the five-year period for articles in WoS is 7/10 = 70 %; in GS this ratio increases to 100 %. For WPs the citedness through WoS is 4/10 = 40 % whilst in GS the WP citedness reaches 8/10 = 80 %.

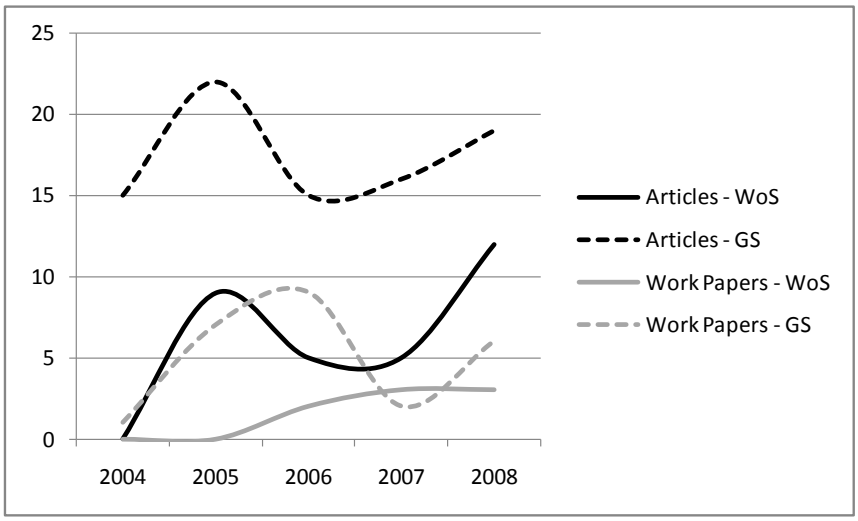

Figure 2. Annual distribution of citations based on Web of Science and Google Scholar across 10 selected DIIS articles and working papers.

Annual distributions of citations

Figure 2 displays the developments of the citations from WoS and GS for the two document types. One should note that the analysis *does not* contain the 46 citations found in GS not attributed to any publication year among their metadata. 25 of these 46 citations are given to the outlier article (b), Figure 1, 12 citations are distributed across the other articles and 9 over the WPs. Hence, Figure 2 represents a fairly consistent distribution with no particular outliers.

One observes that the WP curve based on GS demonstrates the same form as those for the articles, regardless the citation index, but is delayed approximately one year.

Average impact for working papers vs. median article impact

Tables 2a and 2b display how the median cut-off point – with or without the outlier – makes comparisons between citation impacts of the upper and lower halves possible across the two document types. Obviously, any comparison GS and WoS in between is not the issue in this analysis but rather between the publication types within each citation source.

Table 2a-b. Citation sorted DIIS articles and working papers (a: left) with calculations of average impacts (b: right) covering all items, upper and lower median-defined halves (*). Bold figures show WP impact superiority on articles through the same indexes in gray areas.

Articles	GS	Articles	WoS	WPs	GS	WPs	WoS
b	73	b	9	I	8	D	4
f	14	a	8	J	8	J	2
i	12	g	6	D	6	B	1
a	11	f	4	E	5	F	1
e	7	e	3	A	4	A	0
j	5	d	2	G	4	C	0
c	1	i	2	C	1	E	0
d	1	c	0	F	1	G	0
g	1	h	0	B	0	H	0
h	1	j	0	H	0	I	0
Total	***126***		***34***		***37***		***8***

Citation impact	**GS Articles**	**GS WPs**	**WoS Articles**	**WoS WPs**
Mean, all 10 items	12.6	3.7	3.4	0.8
Mean, upper half*	6.3	**7.6**	8.5	**4.0**
Mean, lower half*	5.9	2.4	2.1	0.4
*Outlier **not** included*				
Mean, nine items	5.9	//	2.8	//
Mean, upper half*	13.3	**7.6**	6.2	**4.0**
Mean, lower half*	3.7	2.4	1.9	0.4

Without the outlier Table 2b (right, lower rows) demonstrates: 1) That the *best cited* (upper half) *working papers* in GS (7.6) as well as in WoS (4.0) have *higher impact* than the average journal article (GS: 5.9; WoS: 2.8) *and* their lower half (GS: 3.7; WoS: 1.9); and 2) That the average citation impact of the 10 WPs (GS: 3.7) is *on par* with the average impact of the lower half of the articles, with the outlier disregarded. This is not observed through WoS.

Concluding Discussion

Without the outlier the average *impact for DIIS articles* over the five-year analysis window through GS (5.9) is more than the double of that through WoS (2.8). This is in line with the findings in (Elleby & Ingwersen, 2010) which showed a 3.5 times increase for the same article items in GS, but for 3 citation years of DIIS articles published in 2006. A look at the annual citation distribution, Figure 2, demonstrates that after 3 years (2004-06) the volume of citations in GS to the articles is 4 times as large as that found in WoS.

The citation *impact for WPs* is lower than for articles published by the same institution. Their *citedness* is also correspondingly *lower*, regardless the citation index used. The former finding is in line with that of Frandsen (2009) in her study of WPs. With respect to citedness Elleby & Ingwersen (2010) found that after three years the DIIS articles showed a citedness of 50 % (WoS) and 72 % (GS), largely corresponding to the present findings of citedness covering a five-year window.

Similar to findings by Frandsen (2009) we observe that the best cited working papers *above the median value* may reach and indeed *supersede* the average impact of the lower half (by median cut-off point) of the cited journal articles from the same institution. This observation concerns in particular the case of GS, but similar indications are found for the more citation poor WoS, Table 2. In fact we observe that even the *mean impact* through Google Scholar of the WPs is *on par* with this lower half of cited articles from DIIS covering the same analysis period in GS.

These findings indicate that our initial assumption on WP impact being comparable or in line with that of journal articles from the same prestigious research institution is *not supported*. Only the best cited WPs may compare in impact to some but not all corresponding journal articles.

It was further hypothesized that WPs would attract a higher ratio of internal institutional and personal self citations – because they are easy at hand through the local institutional e-archive. This is *not the case*, if the outlier is disregarded. When included the analysis still show that the ratio of external citations to journal articles cited in WoS is not different from that calculated for WPs.

The *annual citation developments*, Figure 2, with the outlier neutralized because the volume of GS citations lacking publication year have been omitted, demonstrate a *distinctive pattern* for the journal articles in GS and WoS as well as for the WPs through GS. Although the publication and citation volumes in the present case study are only a fraction of those used in Moed (2007) and Frandsen (2009), the *shapes correspond* to those found by Moed for citation impact of OA vs. non-OA journal articles and discussed by Craig et al. (2007: 247). They also show some similarity to the more flattened curves displayed by Frandsen (2009: 128) for WPs versus journal articles in economics.

The *citation delay* observed Figure 2 for the WP citations in WoS and the delayed citation peak in GS may indicate how important it probably is not only to post the items on known e-archives, but also to make use of 'early view' in OA, prior to actually publishing the materials. This was not done in 2004 by DIIS for their working paper series nor for their journal articles.

In conclusion, the traditional peer reviewed and printed articles seem (still), on average and annually, to have an advantage in terms of citation impact over non-peer reviewed but open access scientific material – published by the same institution – provided that no early view or biased postings are applied.

References

Craig, I., Plume, A.,McVeigh, M., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact? A critical review of the literature. *Journal of Informetrics*, 1(3), 239–248.

Elleby, A. & Ingwersen, P. (2010). Publication Point Indicators: A Comparative Case Study of two Publication Point Systems and Citation Impact in an Interdisciplinary Context. *Journal of Informetrics*, 4(4), 512-523.

Frandsen, T.F. (2009). The effects of open access on un-published documents: A case study of economics working papers. *Journal of Informetrics*, 3, 124-133.

Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*, 10

Jacso, P. (2008). Google Scholar revisited. *Online Information Review*, 32(1), 102-114. Retrieved December 2010 from: http://www.emeraldinsight.com/Insight/ViewContentServlet?Filename=Published/EmeraldFullTextArticle/Articles/2640320108.html

Moed, H. F. (2007). The effect of 'Open Access' upon citation impact: An analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology*, *58*(13), 2047–2054.