



LUND UNIVERSITY

EERQI and Web of Science: Internal Report

Åström, Fredrik

2010

[Link to publication](#)

Citation for published version (APA):

Åström, F. (2010). *EERQI and Web of Science: Internal Report*. (European Educational Research Quality Indicators (EERQI)). European Educational Research Quality Indicators (EERQI).

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Characteristics of citations to EERQI content base documents

Fredrik Åström
Lund University Libraries
Fredrik.astrom@lub.lu.se

Introduction

In the EERQI project, we have decided to try using citation analyses as an extrinsic indicator of impact of research in educational research. Using citations as an indicator of the use and impact of a document is a well established – albeit debated – practice. It is, however, important not to use statistics on citation frequencies and distributions without contextualizing the data and the analyses. This is important, not the least in cases such as the European educational sciences, where citation data primarily is available through Google Scholar. In the Web of Science databases, we know that the citations to a document are made by other research articles, whereas in Google Scholar, we need to investigate from where the citations come from. To investigate this, as well as some other structural characteristics of citing documents and educational research citation structures in Google Scholar was analyzed. The analyses were performed on two different sets of data collected from Google Scholar.

Yearly distribution of citations to EERQI content base documents

To get an idea of the potential ‘life span’ of educational research documents, the distribution of citations per year was analyzed. To retrieve data, the EERQI content base was searched for documents published in the years 2000-2003. Subsequently, these documents were also searched for in Google Scholar, and in the cases the documents had been cited, the publication year of the citing documents was retrieved (Table 1).

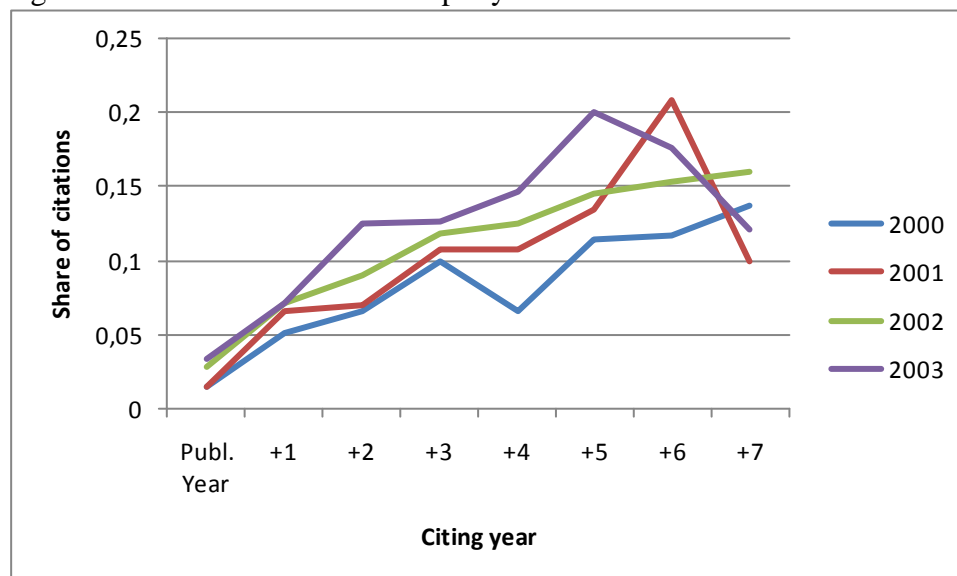
Table 1. EERQI documents and number of citations.

Publication year	No. publications	No. citations
2000	47	334
2001	78	485
2002	120	600
2003	165	927

The selection of years analyzed is limited to available data in the content base. Before 2000, there are only a few number of documents per year (e.g. from 1998 and 1999 respectively, there are only three documents indexed in the content base); and I also decided to restrict the analysis to documents that had at least eight years to be cited. A longer time period for documents to ‘gather’ citations gives us a better chance of identifying robust tendencies; and the selection of an eight year citation window is based on a previous analysis of the average ‘Ci-

tation half-life' of 8.3 years for 20 educational research journals indexed in the Web of Science databases (Åström, 2008). The average number of citations per document is relatively stable over the four years analyzed, ranging between five and seven citations per document. The distribution of citations over time shows a tendency towards a steady increase of citations, although citations to documents published in 2001 and 2003 appears to be decreasing at the end of the citation window (Figure 1).

Figure 1. Distribution of citations per year.



In the case of documents published in 2001 with a relatively steep increase in citations in 2007, followed by a decreasing number of citations in the following year can to a large extent be explained by a few articles on student participation in school reform published in 2001 receiving a very high amount of citations in 2007. The decrease of citations to documents published in 2003 is probably partly depending on the last for receiving citations is 2010, which means that there is still time for articles published in 2003 to gather more citations.

The general tendency is a relatively steady increase in citations over the eight year time period, but with the limited amount of years possible to analyze, we cannot say anything about to what extent this is analogous to the 8.3 year citation half-life identified for the educational research journals in the Web of Science databases. It should also be pointed out that the analyses builds on relatively small frequencies: e.g. for publication year 2000, we have 47 cited documents being cited 334 times and the yearly citation rates ranges from a minimum of five to a maximum of 46 citations; and when removing the highest and lowest figures, the range is between 17 and 32 citations.

Characteristics of documents citing EERQI content base articles

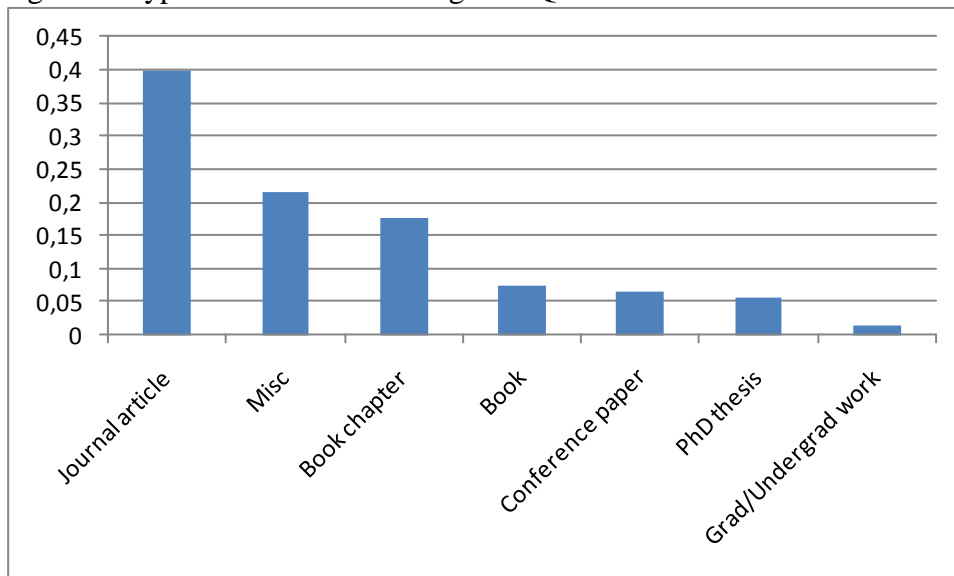
To look at different characteristics of the documents citing EERQI articles, searches were done in Google Scholar on the 100 articles being analyzed in the intrinsic and extrinsic indicator tests, retrieving information on the citing documents. In total, the 100 articles have received 347 citations, i.e. an average of 3.47 citations per paper (CPP). Apart from the analyses on the whole document set, I also split the document set to perform separate analyses on the English, the French and German articles to investigate if any differences could be detected. In terms of general citation patterns, French articles stands out with a low CPP of 1.48 and the most highly cited article attracting only ten citations, whereas the German and English equivalents received around 40 citations each.

The distribution of citations to papers was also analyzed. Generally, this distribution tends to be skewed: the Matthew effect describes a distribution of around 20% of the authors in a field receiving 80% of the citations (Merton, 1968). The distribution of citations to the EERQI material is roughly the same: 28/80, but there are some interesting variations to be noted when comparing articles in different languages: for the English language articles the distribution is 37/80 whereas for the French articles, the numbers are 20/84. And when looking at the number of articles without any citations, the figures are 32% for the whole material, whereas it ranges from 10% for the English to 64% for the French.

By matching already available information on author names for the cited documents and retrieved information on author names for citing documents, it was possible to investigate the level of self-citations; here defined as citation links between documents where at least one author is the same in both the citing and cited document. The variations between languages were very small, and the share of identified self-citations is 20% of the total amount of citations.

One issue often occurring when using citation data from Google Scholar is the lack of control over where the citations are coming from. To investigate this, the citing documents were classified according to document type (Figure 2).

Figure 2. Types of documents citing EERQI content base articles.



Most citations to the ‘EERQI articles’ comes from journal articles. The ‘Misc’ category – being the second most frequently occurring document type – contains documents such as project reports, internal discussion or seminar papers being available online and in one instance, papers/articles in publication archives/repositories where no identification on source of publication other than the archive itself could be identified and a few powerpoint presentations where the forum for the presentation could not be identified. Without having done a structured analysis of the different types of publications gathered in the ‘Misc’ category, an educated guess after having gone through all citing publications when gathering and structuring the data, is that the majority of the papers are from publication archives: publications that may also be published as e.g. journal articles or conference papers without there being information on where that would have taken place.

Also here, we can identify some differences between citing publications in different languages: whereas the English language articles are cited in other journal articles to almost 50%, the figures for the French and German articles ranges between 22-27%. And citations

from book chapters make up 32% of the German citations, while the figures for citations to English articles in book chapters are 14% and for the French only 3%. Citations to the French articles on the other hand are to more than 50% coming from the Misc category, primarily in form of publications in archives such as HAL (<http://hal.archives-ouvertes.fr>), while ‘Misc’ citations to English and German articles make up 17-20%.

In addition to document type, the language of the citing document was also investigated (Table 2), showing a large dominance of citations coming from publications in the same language as the cited document.

Table 2. Language of citing documents

Language of citing publ.	All (N=347)	German (N=100)	French (N=37)	English (N=210)
ENG	59%	10%	0%	93%
FRE	10%	0%	92%	0%
GER	28%	90%	0%	3%
OTHER	3%	0%	8%	4%

References

- Åström, F. (2008). Bibliometrics and Educational Research: a tentative analysis. Presentation at the EERQI open session, ECER 2008 in Gothenburg, Sweden.
- Merton, R.K. (1968). The Matthew effect in science. *Science*, 159(3810), 56-63.
-