LUND UNIVERSITY

Framing educational research through bibliometrics and sociology of science theories: Internal Report

Åström, Fredrik

2008

[Link to publication](#)

Total number of authors:
1

EERQI

Framing educational research through bibliometrics and sociology of science theories

Fredrik Åström, Lund University Libraries
2008-11-03

In this small paper, I intend to elaborate on some of the theoretical issues I approached in my Gothenburg presentation, as well as relating those issues to the results found in my tentative analyses on educational research; and based on this, discuss some traits I find important in terms of how educational research is organized and how that affect how we interpret quantitative information and use this for developing quality indicators.

Empirical results
To get a general idea of the structure of the scholarly communication within educational research, I downloaded records from 20 Educational Research journals (as categorized in the ISI Journal Citation Reports) indexed in Social Science Citation Index over the time period 1998-2007. The results that emerged gave us some indications on traits in educational research. When looking at the knowledge base of the field through a co-citation analysis, it seems to be of a quite general nature, stretching over a long period of time and with a substantial contribution of intellectual influence from other research fields. This tendency is further emphasized when looking at citations to current literature within the field, where the main trend is few citations in-between current educational research articles, mainly gathered within one sub-area of educational research, focusing on the use of computers and multimedia for educational purposes. To say something about the structure of the research front, based on to what extent the articles are share references in-between them, is hard without an assessment of the results from scholars in educational research.

In addition to the analyses presented in Gothenburg, I also made a journal co-citation analysis, i.e. an analysis drawing on what journals are appearing together in the reference lists of the educational research articles from ISI, resulting in the following map (Figure 1).
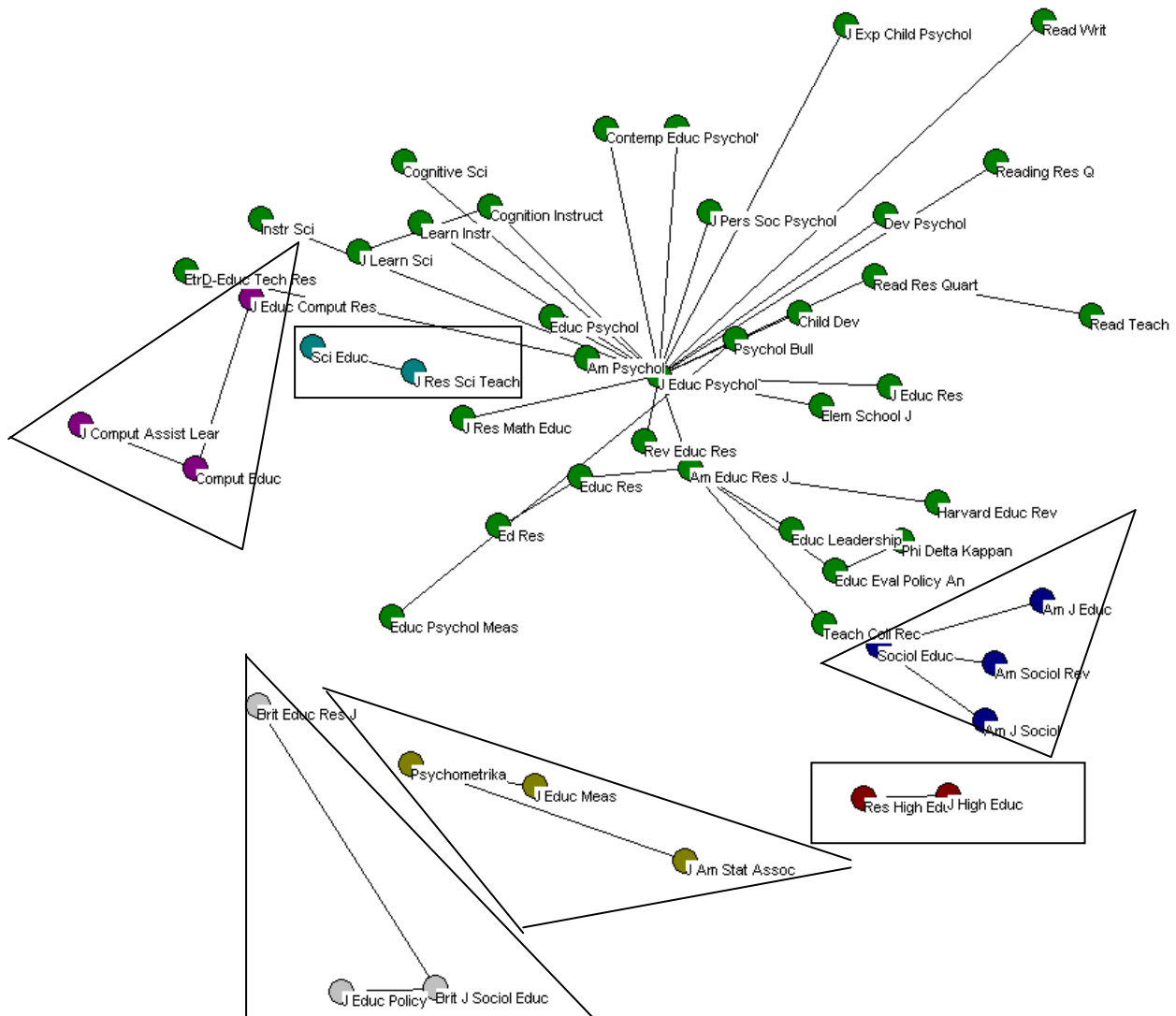
Stress=0.24855

Figure 1. Co-citation map of the 50 journals attracting 300 citations or more in educational research articles indexed in the ISI databases.

The original co-citation analysis is quite muddled (indicated e.g. by the high stress value), however, when adding a clustering analysis emphasizing the strongest de facto citation links (co-cited pairs with at least one common unit), a pattern emerges where we find: a large cluster oriented towards psychology and psychological aspects of education, one cluster oriented towards sociology/sociological aspects, one metrics oriented cluster, one computer assisted learning cluster and so on.

Apart from the different kinds of citation analyses, some analyses were also performed on keywords derived out of the DE-field in the ISI records, i.e. author added keywords. The results of these analyses can be interpreted in two different, although not necessarily unrelated, ways. The results per se show little structures in terms of different research orientations in educational research and with many similar or synonymous concepts. From one point of view, this can be explained by the nature of author added keywords, showing a low level of homogeneity and little structure in terms of dealing with various kinds of related terms: thus, without standardization of data, singular or plural forms of the same concept will show as different concepts in the analysis. How ever, this could also be interpreted as a reflection of the extent

to which terminology in educational research is formalized and generally accepted as standard, or whether there is a heterogeneous terminology, where phenomena can be described in various ways.

These traits are in no way limited to educational research: both in terms of citation and concept structures, we can find many similarities in other research fields in the humanities and the social sciences (e.g. Astrom, 2006; Hellqvist, 2008; Landstrom & Astrom, forthcoming). The wide temporal range of the research base is present in many other fields, as are the many influences from outside the own field (management research, comparative literature and library and information science, to name a few); and we see a wide range of the extent to which scholars in one field are citing contemporary colleagues in the same field. The heterogeneous terminology is also a common trait in the humanities and social sciences, where phenomena being analysed can be understood from various social, psychological and historical points of views (to name a few) and are related to a set of different contextual aspects, e.g. legal, professional, cultural and political, that varies between different geographical and cultural settings, meaning we might be describing the same basic phenomena, but we interpret them for different points of view and with different vocabularies.

Theoretical issues
The traits identified in the tentative analysis are not just related to various modes of communication in different research fields, but also to how research is organized, performed and evaluated. To take citations as an example, they do not only refer to the research you are building on, but are also an important indicator on to what extent you need to relate to current research in your own field to demonstrate how your research contributes to the intellectual goals of the field; whether if it is by building on generally acknowledged methods and techniques to discover a new strand of DNA, by using a new set of theories from another research field to re-interpret a particular social phenomenon or by showing how the analyses being performed are done so on a whole new material, previously unknown (Whitley, 2000). The ways in which you display your contributions to the field, are of major importance in terms of to what extent you refer to your contemporary colleagues in the field; and also, how these references can be interpreted as e.g. signs of impact or quality of research.

The view of citations as an indicator of research quality is very much based on a view on the sciences as being cumulative and to a large extent also being organized on a disciplinary basis. This view is not limited to how to look at citations, but has been the ideal (and very much normative) model of research organization in general throughout the 20th century, resulting e.g. in descriptions of research fields not adhering to the model as being preparadigmatic or immature. However, when studying different fields of research, as well as more recent literature on the organization of the sciences, it becomes evident that this is just one of many models of research organization and scholarly communication. Furthermore, it is not only a matter of different research fields being organized in different ways within their disciplines, especially since 1945, there has been a large increase in research being organized interdisciplinary, with a larger emphasis on applications oriented research and not the least, with research being performed in collaboration with, funded and also to some extent being evaluated by actors outside academia (e.g. Etzkowitz & Leydesdorff, 2000; Gibbons et al, 1994).

Apart from a research organization that is principally cumulative, where we are building upon the previous work of our colleagues to discover other phenomena or ways of explaining them, we also have research that is more of a negotiating nature; we use different means of explaining the same phenomena to try to negotiate a better or alternative understanding. We also find research that is essentially of a distinctive nature with scholars pursue their own tracks of research without any necessary relation to other colleagues. These different ways of

organizing research and communication implies wholly different ways of interpreting citations. In a cumulative setting, we cite to build onwards on the (high quality) works of our colleagues – and we cite them to show how we relate to them and the field in general – while e.g. 'negative' citations show that there is something wrong with the research cited (and thus of low quality). However, in a more negotiations based setting, many of the citations have a more debating function, where we are not only relating to our colleagues to build on their previous work, we both build on colleagues in our own field, but not necessarily our contemporaries: and also, on scholars from other fields. In terms of 'negative' citations, this model of communication does not necessarily imply that previous work cited is wrong or of low quality, research commented upon and debated might be of high quality and have had a significant impact, but is now contested by suggesting a different interpretation of a phenomenon. And in the case of distinctively organized research, there are few, if any, references to scholars within the same field. To a large extent, if there is related research already done, we run the risk of the topic already being covered and thus closed for further investigation. One example is e.g. annotated editions of medieval texts, where the practice is not to do two editions of the same text, but it is of course also applicable in terms of e.g. the discovery of DNA strands; you don't need to discover the same strand twice. From this point of view, the only citations going to colleagues within the field, is to some extent negative citations: we only refer to them to distinguish ourselves from our colleagues by pointing out the fundamentally unique aspect of our research: e.g. by analyzing Shakespeare from a new theoretical point of view.

Implications

To start with a disclaimer, there are still questions related to e.g. if the empirical analyses are reflecting educational research as a whole or the specific line of educational research that is indexed in the ISI databases; and wether the results of the keyword analyses are the result of indexing standards or the actual structure of the terminology of educational research. From my perspective, this is hard to say anything about until we start performing citation and semantic analyses on 'our' content base. When looking at the author addresses in the ISI articles, we see a substantial majority of Anglo-American authors in the articles indexed in the ISI databases, but will we see the same structural patterns in terms of citations and keywords when looking at educational research from other countries? The question of the representativity of the ISI databases is of major importance, not only in terms of what structures can be identified in the educational research field, but also in terms of how well the ISI databases are suited for research evaluation purposes in the humanities and social sciences. That they are poorly equipped for that, we already know, but not the least from a research policy perspective, I think it is important to be able to show how representative the ISI data actually is.

So, what then, are our possibilities of finding quality indicators through analyses of educational research literature? The seemingly heterogeneous structure of both citations and keywords/terminology, as well as how these traits seem to relate to aspects of the general organization of educational research, might seem discouraging. As I see it, two different perspectives come into play. On one hand, to as exactly as possible show the complexities of a research field like educational research – and the implications of these complexities on what kinds of quantitative analyses can be done on the literature, as well as how the results can be interpreted – is of vital importance in a time when the research policy climate is increasingly focusing on quantitative measures for the assessment of research and the distribution of funds. On the other hand, there is also the issue of the indicators that actually can be indentified, how they can be used and in what context they are used. In the recent email discussion, there seem to be three lines of evaluation contexts that will/can be discussed within the EERQI project: relevance judgements for information retrieval purposes, information extraction and refinement to support peer review processes and finally, 'research policy' type assessment of productivity and quality of educational research. From my personal point of view, I would also

like to add the perspective of science studies, to gain a wider understanding of fields of research not necessarily adapting to the model of cumulative research so often assumed in e.g. bibliometrics and science studies; and not the least, in the use of citation analysis for research evaluation; something that is of great interest in general, but also, with important implications for the three contexts of quality indicators.

As of yet and as of soon
To conclude this paper, I'd like to briefly introduce where my thoughts are right now, in terms of work to be done from my horizon. As mentioned in Ágnes Sándor's response on David Briges' and Ingrid Gogolin's "Peer review action", Ágnes Sándor and I are planning a study to type and contextualize references/citations. The basic idea is to investigate wether it is possible to match the types of citations identified in the preliminary experiments made by Ágnes Sándor in preparation for the Leuven workshop and the tentative models of research organization/scholarly communication suggested by me in Gothenburg; and also, if possible, to take into account the quality criteria agreed on within the EERQI framework (Figure 2).
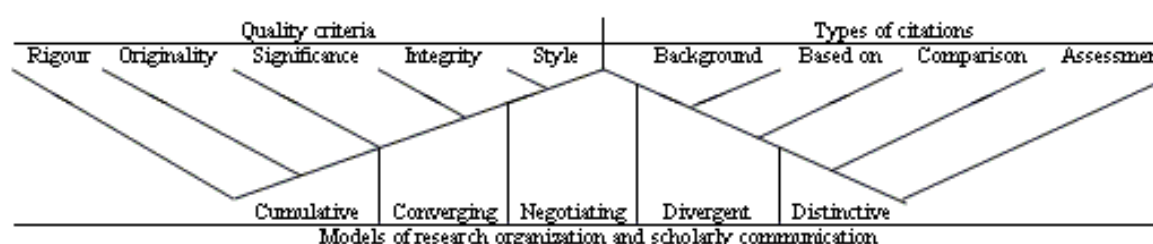

Figure 2. Model of dimensions for analysing citations.

The idea is to see what, e.g. background type citations have to say about the originality of the cited work; and are the relations between citing and cited document of a cumulative or negotiating nature? This will be important in telling us about the function of citations and how knowledge is produced within the field, but also, to what extent citations can be seen as an indicator on quality; not only by distinguishing between positive and negative citations, but also acknowledging that in e.g. a negotiating research organization, research being cited negatively can still be of high quality and with a significant impact for the field. However, to be able to use this to say anything about the quality of the literature in our content base; there must be citations to the texts we have in the content base and also, citations that are originating from other documents in the content base. In my tentative analyses on the ISI data, this 'intra-document set citation traffic' was scarce and primarily present in one particular area of educational research, but it remains to be seen if the case is the same in the EERQI content base.

The model of research organization still needs developing; and before having made the analyses, we are not sure to what extent we will be able to get results that are interpretable in relation to the suggested model. Hopefully, we will be able submit a paper based on this experiment to the International Society for Scientometrics and Informetrics conference in 2009.

References

Astrom, F. (2006). *The Social and intellectual development of library and information science*. Diss. Umeå: Doctoral theses at the Department of Sociology.

Etzkowitz, H. & Leydesdorff, L. (2000). The dynamics of innovation: From national systems and "mode 2" to a triple helix of university-industry-government relations. *Research Policy*, 29(1), 109-123.

Gibbons, M. et al (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: Sage.

Hellqvist, B. (2008). Bibliometri och humaniora: En relationsanalys [Bibliometrics and the humanities: A relational analysis]. *InfoTrend*, 63(3).

Landstrom, H. & Astrom, F. (fortcoming). A history of new venture creation research. In: K. Hindel & K. Klyver (eds.). *Handbook of research into new venture creation*.

Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford: University Press