

KEYNOTE ADDRESS

Citation Analysis in Research Evaluation¹

Henk F. Moed

moed@cwts.leidenuniv.nl

Centre for Science and Technology Studies (CWTS)
Leiden University, the Netherlands

This book is about scientific, or more generally, scholarly research. It focuses on a type of research that is characterised as ‘basic’, ‘fundamental’ or ‘strategic’. It recognises its crucial importance for global economic progress and social welfare, but at the same time it acknowledges that a firm political or societal basis for this type of research can be maintained only by further developing a system of internal quality control and performance enhancement. This book aims at showing that citation analysis is a useful tool in such a system.

It primarily concerns the assessment of the contributions scholars make in their research publications to the advancement of valid scholarly knowledge. It deals with the assessment of research performance of individual scholars, research groups, departments and institutions, scholarly journals and national scholarly systems, and with the analysis of general characteristics of global science and scholarship.

It explores the uses and limits of citation analysis, involving the construction and application of a wide range of ‘bibliometric’ indicators of the ‘impact’, ‘influence’ or ‘quality’ of scholarly work, and derived from citation data, i.e. data on references cited in footnotes or bibliographies of scholarly research papers. It focuses on the Citation Indexes produced by the Institute for Scientific Information (ISI, currently Thomson Scientific), but many findings are also relevant in the use of other citation indexes.

This book aims to provide useful information for members of the scholarly community and research policy officials about basic technical aspects of citation analysis, what it measures, and how it can be properly applied in research evaluation and policy processes, by systematically discussing numerous statements about its value made by scholars and policy makers, correcting misunderstandings and illustrating its strengths and limits, particularly in relation to peer review.

It is argued that the use of citation analysis in the evaluation of individuals, groups and institutions is more appropriate the more it is:

- Formal – i.e., previously known to evaluators or decision makers and to scholars or institutions subjected to evaluation that indicators are used as one of the sources of information.
- Open – those subjected to the bibliometric analysis have the opportunity to examine the accuracy of underlying data, and to provide background information that in their view is relevant for a proper interpretation of the quantitative outcomes.
- Scholarly founded – that bibliometric investigators present their outcomes within a scholarly framework, discuss issues of validity, explicitly state theoretical assumptions, and underline their potentialities and limits.

¹ Executive Summary from: Henk F. Moed. *Citation Analysis in Research Evaluation*. Springer, 2005.

- Supplemented with expert and background knowledge about the substantive contents of the work under evaluation, the conditions under which evaluated scholars operated, and their research objectives.
- Carried out in a clear policy context – i.e., applied in the framework of an evaluation procedure of which both the evaluative perspective and the objectives are clear to all participants.
- Stimulating users to explicitly state basic notions of scholarly quality, its dimensions and how they were operationalised and weighted.
- Enlightening rather than formulaic – the indicators are used to obtain insight in a particular aspect addressed in the process, rather than as inputs in formulas designed to algorithmically generate the process' outcomes.

Application of citation analysis in the assessment of past research performance in basic science and of scientific journals has reached a *high level of sophistication*. This book discusses numerous issues raised by scientists subjected to citation analysis, by journal editors and policy makers, and shows how such issues can in principle be accounted for or solved technically.

The extent to which citation analysis based on the ISI Citation Indexes can be validly applied in all domains of scholarship, including the applied and technical sciences, social sciences and humanities, is often debated. This book thoroughly examines differences in the *structure of the written communication systems* among the various domains of scholarship, and the extent to which these systems are *covered by the ISI Citation Indexes*.

The ISI Indexes do not claim to have complete journal coverage, but rather to include the most important. Their founder, Eugene Garfield, developed a powerful and unique criterion for expanding the database beyond the core of journals whose importance to a given field is obvious: the frequency at which journals are cited in those sources that are already included in the index.

Applying a 'database internal' criterion, this book shows that ISI coverage tends to be *excellent* in physics, chemistry, molecular biology and biochemistry, biological sciences related to humans and clinical medicine; *good*, yet not excellent, in applied and engineering sciences, biological sciences related to animals and plants, geosciences, mathematics, psychology and other social sciences related to medicine and health; and *moderate* in other social sciences including sociology, political science, anthropology and educational sciences, and particularly in humanities.

A principal cause of non-excellent coverage is the importance of sources other than international journals, such as books and conference proceedings. In fields with a moderate ISI coverage, language or national barriers play a much greater role than they do in other domains of science and scholarship. In addition, research activities may be fragmented into distinct schools of thought, each with their own 'paradigms'.

This book distinguishes and illustrates *four types of bibliometric studies* in which the ISI database plays different roles. The decision as to which type of study is appropriate in a discipline depends upon the extent to which it is covered by the ISI Indexes. Compared to a 'standard' analysis in fields with excellent coverage, this database may be expanded in several ways in fields with good but not excellent coverage, or it may play a limited role or no role at all when field coverage is moderate.

If the extent to which research findings reach beyond a purely national or local viewpoint and are exposed to criticisms from a wide international scholarly audience is considered as a relevant criterion of research quality in social sciences and humanities, a major task would be to develop for the various subfields valid indicators of this aspect of research performance. This book argues that it cannot be taken for granted that the ISI Citation Indexes provide such indicators in all subfields of these domains of scholarship. A challenge would be to systematically explore alternative data sources and methodologies. The expertise and perceptions of scholars active in the various subfields should play an important role in such an exploration.

As regards *journal impact factors*, this book provides a technical and historical explanation of how ISI impact factors are calculated, and highlights a number of problems affecting their accuracy and applicability. It illustrates how alternative journal impact measures solve many of these problems, but at the same time underlines that there is no single 'perfect' indicator of journal performance. Although the status of the journals in which a research group publishes is an aspect of research performance in its own right, journal impact factors should not be used as surrogates of citation impact of a group's publications.

Data accuracy is a next crucial issue. It is illustrated how uninformed data collection and analysis may substantially distort the outcomes of citation analysis. Use of inaccurate data may not only distort results for particular groups, but also affect the credibility and hence the usefulness of a bibliometric study as a whole. But accuracy problems can be overcome in advanced data handling and in data verification procedures involving evaluated scholars and their institutions.

The next key issue concerns *what citations measure*. Outcomes of citation analysis of basic science research groups tend to statistically correlate in a positive way with peer ratings of the groups' past performance. This book presents more empirical case studies revealing such a positive correlation. Findings provide a further theoretical justification for applying citation analysis in research evaluation, but correlations are not perfect.

It is argued that citation counts can be conceived as manifestations of intellectual influence, but the concepts of citation impact and intellectual influence do not coincide. Distinct notions of the concept of intellectual influence may exist, and evaluators assessing scholarly work may have different views upon which are the most crucial aspects to be taken into account. Outcomes of citation analysis must be *valued* in terms of a qualitative, evaluative framework that takes into account the substantive contents of the works under evaluation.

The interpretation of citation impact involves a quest for possible biases. It is therefore crucial at which level of aggregation citation analysis is carried out. Evaluating aggregates of entities can be carried out in such a way that the effects of special characteristics and circumstances of individual entities to some extent cancel out. It must be underlined that systematic biases as regards the aggregate as a whole may still occur and should be taken into account.

The conditions for proper use of bibliometric indicators at the level of individual scholars, research groups or departments tend to be more readily satisfied in a *peer review* context than in a policy context. It can therefore be argued that bibliometric analyses at such lower aggregation levels normally best find their way to the policy arena through peer assessments. But it does *not* follow that citation analysis is *merely* a tool to be used by peers.

This book illustrates the use of citation analysis as a tool to *assess peer review procedures* and to keep the peer review process honest. From the latter perspective, it is a tool for policy decision makers as well. It shows that citation analysis has its strengths and limits, and that the same is true for peer review. The challenge is to combine the two methodologies in a proper, productive way.

A study of *research assessment exercises*, in which a small peer committee evaluated research departments in an entire national discipline, raised the question whether such exercises are capable of identifying truly excellent or 'top' research departments. This finding underlines the need for research policy makers to thoroughly reflect upon the objectives of such exercises, taking into account their cost effectiveness.

This study also provided evidence that a peer rating system (e.g., in terms of 'excellent', 'good', 'less good', 'poor') tends to generate a distribution of ratings among departments that depends upon the rating system itself, and that is to some extent independent of the overall performance level of the departments under evaluation.

A study of funding procedures of a *national research council* provided evidence that proximity relationships between applicants and expert committees responsible for the evaluation of grant

proposals made their outcomes inequitable. It illustrates how quantitative, bibliometric methods can fruitfully contribute to an *internal* debate within a funding agency about funding procedures and evaluation criteria, and to a *public* debate between a funding agency and the national science policy sphere.

Citation analysis is a most valuable tool in policy studies addressing general issues regarding the academic system, with a complexity that reaches beyond the capabilities of expert panels. Studies of the global academic system and ‘macro’ studies of national academic systems are excellent examples. This book presents four studies that deal with ‘classical’ issues in the field of quantitative science studies and that have a high policy relevance:

- Did scientists’ global publication productivity increase during the 1980s and 1990s?
- How to measure trends in national publication output?
- Does international scientific collaboration pay?
- Do US scientists overcite papers from their own country?

A first macro study presented in this book examined trends during the 1980s and 1990s in *global publication productivity*, defined as the total number of articles published in a year per scientist active in that year. It was found that, although an ‘average’ individual scientist can justly claim to have published in recent years more research articles than in the past, from a global perspective scientific publication productivity did not increase during the past two decades. One interpretation is that raising both the internal productivity of the science system, its economic relevance and collaboration, are to some extent conflicting policy objectives for basic science.

Nowadays many countries publish National Science Indicators Reports and analyse what bibliometric macro indicators express about the state of *a nation’s research system*, and about the level of its research performance. Not infrequently, the various indicators and methodologies seem to lead to different conclusions. This makes bibliometric indicators vulnerable to selective use and manipulation. A second macro study presented in this book provides technical information as regards the construction and interpretation of publication based macro indicators.

Assessing the trend in a single country’s publication output, it explores a categorisation of publishing authors into *domestic* (i.e., working in institutions located in the country itself) and *foreign* (active in other countries). Indicators are considered that give an answer to the following questions: did the country’s scientific workforce expand or shrink, and did the number of papers in which it participated per domestic author increase or decline? It concludes that it is essential to calculate a *series* of indicators and to provide them with a consistent interpretation. Isolating a single measure from the others may distort the results and lead to biased conclusions.

A third macro study addressed the ‘classical’ issue ‘*Does international scientific collaboration pay?*’ It concludes that when scientifically advanced countries collaborate with one another, they profit in around 7 out of 10 cases from such bi-lateral collaboration, in the sense that both raise their citation impact compared to that of their ‘purely domestic’ papers. But when advanced countries contribute in bi-lateral international collaboration to the development of scientifically less advanced countries – and thus to the advancement of science in the longer term than the perspective normally adopted in research evaluation – this activity tends to negatively affect their short-term citation impact, particularly when their role is secondary.

It has been claimed that US authors excessively cite other US colleagues. This would lead to a *US bias* in the selection of journals for the ISI Citation Indexes and would distort the outcomes of citation analysis. This book argues that the crucial issue at stake is the adequacy of the norm against which referencing practices of US scientists is evaluated. A fourth macro study found *no* conclusive evidence that US scientists in science fields excessively cite papers originating from their own country.

Finally, this book discusses *recent trends* in the development of indicators and in scholarly publication. The need is emphasised to carry out systematic studies of the *conditions* under which citation analysis is actually applied in research evaluation, and of the *effects* of its use upon the scholarly community, its evaluators and the policy arena. Such insights may contribute to the further development of the 'critical' potential of citation analysis as a research evaluation tool.

Analyses of changes in publication and citation practices are illuminating, but the principal question is not whether or not scholars' practices change under the influence of the use of bibliometric indicators, but rather whether or not the application of such measures as a research evaluation tool enhances research performance and scholarly progress in general.

As more and more scholarly documents become available in electronic form through the World Wide Web, their use as sources in citation analysis is expected to increase in the near future. From the perspective of research evaluation, including more sources does not necessarily lead to more valid assessments of the contributions scholars make to the advancement of scholarly knowledge. The extent to which the sources' documents contain new knowledge and meet professional quality standards is a critical issue.

Outcomes of citation analysis are often presented to the 'outside world' in the form of *rankings* of entities such as individual scholars, research departments or institutions. This also occurs with outcomes of peer reviews. It is argued that the need for policy makers and the wider public to obtain insight into the scholarly quality of the various groups is legitimate, but that scholarly quality is not as straightforwardly measured and ranked as performance is in many other societal domains. Moreover, rankings disregard how the performance of one entity depends upon that of others. Bibliometric investigators should look for means to express these notions in the outcomes they produce.