Jointly published by Elsevier Science Ltd, Oxford
and Akadémiai Kiadó, Budapest

Scientometrics,
Vol. 36, No. 3 (1996) 397–420

ADVANCED BIBLIOMETRIC METHODS AS QUANTITATIVE CORE OF PEER REVIEW BASED EVALUATION AND FORESIGHT EXERCISES

A.F.J. VAN RAAN

Centre for Science and Technology Studies,
University of Leiden, Wassenaarseweg 52,
P.O. Box 9555, 2300 RB Leiden (The Netherlands)

(Received May 24, 1996)

This paper gives an overview of the potentials and limitations of bibliometric methods for the assessment of strengths and weaknesses in research performance, and for monitoring scientific developments. We distinguish two different methods. In the first application, research performance assessment, the bibliometric method is based on advanced analysis of publication and citation data. We show that the resulting indicators are very useful, and in fact an indispensable element next to peer review in research evaluation procedures. Indicators based on advanced bibliometric methods offer much more than 'only numbers'. They provide insight into the position of actors at the research front in terms of influence and specializations, as well as into patterns of scientific communication and processes of knowledge dissemination. After a discussion of technical and methodological problems, we present practical examples of the use of research performance indicators. In the second application, monitoring scientific developments, bibliometric methods based on advanced mapping techniques are essential. We discuss these techniques briefly and indicate their most important potentials, particularly their role in foresight exercises. Finally, we give a first outline of how both bibliometric approaches can be combined to a broader and powerful methodology to observe scientific advancement and the role of actors.

1. Qualitative and quantitative performance measures

1.1 Scope of this paper

The fundamental purpose of evaluation is to promote research quality. Therefore evaluation is without any doubt a necessity. In the first place, scientists themselves are responsible for quality control of their intellectual territory. Thus, review by colleague-scientists, 'peers', is applied to judge research proposals, appointments of research staff, evaluation of research groups or programmes, and so on (*Cole* et al,[1] ABRC,[2] *Daniel,*[3] an excellent overview is given by *Nederhof*[4]).

0138-9130/96/US $ 15.00

In this paper we discuss the application of advanced bibliometric methods for research performance evaluation and, more generally, monitoring scientific developments and the role of actors in these developments. Bibliometric methods should never be used in 'isolation'. Therefore, we put the application of bibliometric research performance indicators in the framework of peer-based evaluation procedures. Thus, we first discuss the main aspects of peer review, some examples in practice, advantages and disadvantages. Our main focus, however, is on the use of bibliometric performance indicators, in particular citation-analysis, as an indispensable support tool of peer review. Furthermore, we show how specific bibliometric methods -mapping or 'cartography'- can be applied in review procedures and, particularly, in 'foresight' exercises.

With this paper we intend to give an overview of the state-of-the-art in a comprehensible way. For a thorough discussion of the bibliometric research performance assessment methodology we refer to a recent publication of our group. All concepts and types of indicators (and their symbols) discussed in this paper are in accordance with the above mentioned publication (*Moed* et al[5]).

1.2 Peer review and research quality care

Scientific research, often closely interacting with technology, is undoubtedly a major driving force of our modern society. The demand for money is outstripping the supply of money. Choices have to be made. We would like to support the best work, to promote the best groups, to stimulate the best people. An evident but by no means easy task. As discussed above, the expertise for these choices is provided, in the first place, by knowledgeable colleague-scientists.

This peer review is one of the mechanisms that keep science in a healthy condition. All scientists agree that peer review, like democracy, may not be the perfect system, but it is the best we have (*Moxham* and *Anderson* [6]). And therefore we must do everything to keep it sound. Peer review, however, also has its typical shortcomings. Thus, it is necessary to explore the possibilities to improve peer review based evaluation procedures.

The central concept in all evaluation procedures and performance assessments is 'scientific quality'. Quality is a measure of the extent to which a group or an individual scientist contributes to the progress of our knowledge. In other words, the capacity to solve problems, to provide new insight into 'reality', or to make new technology possible. Ultimately, it is always the scientific community ('the peers', but now as a much broader group of colleague-scientists than only the peers in a review committee) who will have to decide in an 'inter-subjective' way about quality. Indeed, in this respect we may compare the peer system with democracy: the scientific community acts as a 'republic of science' (*Polanyi* [7]) in which a majority opinion about quality will exist and smaller or larger minorities will have (very) different opinions.

Peer review is typically a *qualitative* assessment of research performance. The bibliometric indicators discussed in this paper represent the *quantitative* side. Needless to say that quantitative elements are clearly present in peer review, e.g., number of publications in high prestige scientific journals. Conversely, citations given to research work can be seen as judgements, 'votes' of colleague-scientists in favour of the work cited. Therefore, peer review and bibliometric analysis will never be completely independent measures, 'orthogonal vectors in evaluation space', as they will always be related to some extent.

1.3 The practice of peer review

Let us describe the practice of peer review in a recent evaluation procedure in the Netherlands (for a review of different practices, see *Van de Kaa* [8]). This procedure is of particular importance, as it has now become an example for similar reviews within the European Union, especially Germany (*Müller-Böling* [9]).

A few years ago, the universities in the Netherlands jointly set up a new assessment system for research quality care (Society of the Netherlands Universities, see VSNU[10]). The Minister for Education and Science stays on armlength, with the role of 'superintendent'. The aim of the VSNU procedure is ambitious: evaluation within the next five years of all main disciplines (e.g., physics, chemistry, biology, psychology, sociology, linguistics, in total about 25 major disciplines) in all thirteen Dutch universities. Also, a certain 'foresight' element is included: an assessment of each group in terms of its 'long term viability'. Thus, a country-wide assessment of the whole spectrum of academic scientific work. Peer review is the heart of this provocative operation. But advanced bibliometric indicators form an essential part of the work. Their application becomes more and more sophisticated and useful for research management as well as directly for the working floor.

This Dutch joint-university research quality assessment system serves several purposes. First of all, a thorough analysis of scientific quality, in particular observation of strengths & weaknesses. Simultaneously this will induce a feed-back to university management in order to improve research quality. Furthermore it will be a contribution to accountability: are the resources used well? And finally, it provides the government with important information for science policy.

For each discipline one peer review committee of 5-7 members is set up. The majority must come from outside the Netherlands. The task of the committee is to give an independent and thorough assessment of each research group or programme within the discipline from an international perspective.

The four main aspects of this assessment are:

* Scientific quality in general;
* Scientific productivity;
* Scientific, and where appropriate, societal and/or technological relevance;
* Long-term viability.

These aspects are judged on a five-point scale: excellent/good/satisfactory-average/unsatisfactory/poor, together with a short explanatory report per group or programme. The final report of the peer review committee is available to the public. How do these VSNU peer review committees work? We mention the main characteristics.

* Assessment of the international standing of the publications of a group in the last five years, in particular the role of the best ten (as perceived by the group);
* Assessment of the progress report written by the leading scientist of a group (mostly a professor/head of the department), with (1) a comprehensive description of the position of the group in its field, particularly how the group contributes to knowledge progress (our central quality definition); and (2) a comprehensive description of future developments in relation to the capacities of the group and to the possibilities given its academic infrastructure and 'environment';
* Interviews with the leading scientist about the above mentioned elements in the progress report (international standing and future developments);
* Obtaining further information from experts outside the committee.

The whole procedure per discipline takes place in about six months. It is expected that with this procedure all Dutch universities will be evaluated within five years. The evaluation of several major disciplines –mechanical engineering (including marine technology), archaeology & history, psychology, educational psychology, and biology– is just finished. Currently, sociology, political science, chemistry, astronomy, and physics are being evaluated. The costs of such a VSNU peer review procedure is in the order of about $ 100,000 per major discipline.

Assessment of scientific performance by peers is, as discussed earlier, regarded as vital for scientific development. Important advantages are the reinforcement of self-regulating capacities of the scientific system, the high intellectual level of the whole review process, and the general consensus among scientists that we have to proceed in this way (*Moxham* and *Anderson*[6]). It is clear that the value of bibliometric indicators is determined by the extent to which they are able to improve substantially the peer review process. This means that they should 'compensate' for shortcomings of the peer system in general, and, particularly, in relation to the four above mentioned main aspects of quality assessment.

1.4 Why bibliometric indicators are necessary

Like any other human enterprise, peer review also has its shortcomings and disadvantages. I here mention the most heavily discussed problems (*Horrobin,*[11] *Moxham* and *Anderson*[6]).

* The quality of the peer review process and, particularly, the results of the peers' decisions are highly dependent on the selection of members in the committee. For the evaluation of established excellent or, on the other side, well-known poor groups, this 'member-dependancy' is mostly not a problem. But it might be a severe problem in the case of groups in emerging or interdisciplinary fields;

* In times of decreasing budgets there will be conflicts of interests. Peers from outside the country solve this problem only partly. Severe conflicts of interest may provoke unfair judgements or even fraudulent acts;

* Peers may not always be aware of the quality of younger people or new-comers to the field. Such a negative bias against young or new researchers is the caricatural but probably not always unrealistic feature of the so-called old-boys-system.

In this time of strong growth of emerging new fields -often at the cost of other fields- together with an increasing interdisciplinarity, it is indeed not easy for peers to form a valid opinion on the track record -the past performance- of those to be evaluated, and, equally important, to consider this past performance in terms of foresight with respect to the field concerned.

We simply cannot take the risk of deteriorating the peer review system by a lack of relevant information. It is therefore crucially important for peers to have access to consistent and objective as possible information on the past performance of research groups, as well as information about the position of a group or institute in the worldwide 'landscape' of the field and its recent developments. Here quantitative indicators come into the picture. And, of course, this is not a big surprise. We already discussed that an important aspect of scientific quality to be judged by the peers, is scientific productivity based on number and type of publications. So clearly quantitative indicators are already very common to peer review.

2. Application of bibliometric indicators

2.1 Approach of methodological and technical problems

What are useful quantitative indicators of scientific research performance? We here focus particularly on indicators based on data from the scientific literature. Therefore we call them bibliometric. We already mentioned the first type, commonly used by peers: indicators of scientific output, based on numbers of publications in international journals. But we can add an additional dimension: the international scientific influence or impact of a group as a function of time.

Furthermore, the work of a research group concerns a specific domain of scientific activities. Often this domain is characterized by a spectrum of different research fields. For instance, the scientific work of an astronomy department will undoubtedly be classified for the major part as 'astronomy' or 'astrophysics', but smaller parts of the department's output might be very relevant for, and also classified as 'atomic and molecular physics', 'nuclear physics', 'atmospheric chemistry', or 'fluid dynamics'. The larger the group to be analyzed (e.g., a whole institute) and the more interdisciplinary (e.g., environmental research), the broader such 'research profiles' of scientific output are (see for recent examples in environmental research: *Van Raan* and *Van Leeuwen* [12]). Thus, it is important to construct bibliometric indicators which allow a breakdown of impact assessment by the different (sub)fields in the institute's research profile. Such an impact assessment in relation to research profile can be operationalized by analysing different aspects of the world-wide citations to scientific work. The *Science Citation Index* (SCI) of the Institute of Scientific Information in Philadelphia is a central data source for this information. It is, however, not unique anymore. The high-energy physics community at Stanford Linear Accelerator Centre (SLAC) for instance operates their own, publicly and freely accessible citation database (SPIRES) on Internet.

Next to the bibliometric assessment of research productivity and impact, there is an entirely different application of bibliometric methods. It is the construction of maps or landscapes of scientific fields. These maps are of particular interest for the foresight part of research assessment. We briefly describe this type of bibliometric analysis in Section 2.5.

Before we discuss concrete examples of bibliometric indicators, and in particular how they are used, or can be used, we emphasize that design, construction, and application of bibliometric indicators is a research field in its own. Large and comprehensive databases such as the SCI have been built for information retrieval and not for evaluation purposes. Extensive methodological and technical work is absolutely necessary to construct reliable and useful indicators. Only then they can be used for evaluation. We mention a few important problems which have to be solved before bibliometric indicators can be applied in practice (*Van Raan* [13]). First, we have to solve many *technical* problems. To mention the most crucial ones:

* We must know database characteristics such as coverage, and, most importantly, changes over time in coverage;
* In citation analysis all authors must be involved instead of only the first one;
* Cleaning and unification of authors names, and in particular addresses of author

affiliations;

* Corrections for self-citations and/or 'in-house' citations;
* Assignment of publications to groups/institutes/organizations and other problems related to the choice of aggregation level;
* Allocation of the bibliometric data to the proper 'input' data;
* Verification with other sources (e.g., annual research reports, publication lists of departments) are necessary to check the 'completeness' of the information in the database;
* Analysis of publications in journals not covered as a source by the citation index but cited in journals that are covered (i.e., assessment of the 'non-SCI' impact);
* Last but not least reliable data collection algorithms are necessary to handle large amount of data in an efficient way, thereby taking into account as much as possible the solution of the above problems.

Second, there are many *methodological problems*. Major ones are:

* The (very!) different publication and citation characteristics in the different fields of science. These differences must be taken into account. An important consequence is that scientific fields can never be compared on the basis of absolute numbers of citations; field-dependent normalization is necessary;
* These field-dependent characteristics may change over time during the period of analysis;
* Even after field-dependent normalization of citation numbers it is not clear whether a certain normalized score is high or low for that specific field: a comparison with other, similar groups or with a world-wide reference value for that specific research field is necessary to get meaningful results;
* The above mentioned problems concern field-dependent characteristics. But in fact the definition of a scientific field is a problem in itself. Depending on the type of analysis we have to solve this problem of 'field delineation';
* In addition to the above problems we need measures of significance in order to decide whether impact is indeed higher or lower than a world-wide average;
* The 'size of the object to be evaluated', or aggregation level must be sufficiently high. Application of bibliometric indicators at a level too low, e.g., individual scientists, will be statistically problematic. For research groups the situation is much better;
* Citations are given after publication. How long must we wait, in other words: what citation window should we use? We found empirically (see for instance *Moed* et al,[5] *Van Raan* and *Van Leeuwen* [14]) that three years is a good choice;

* Related with the technical problem of coverage: in many disciplines other media than journals covered by one of the citation indexes play an important role in the dissemination of scientific results. Examples are books, reports, and electronic versions of preprints, etc.;
* A central criticism to citation analysis is the problem of 'time lag'. But also peers generally need time to see whether research results will 'take root'. It is a real challenge to make bibliometric indicators as 'topical' as possible;
* The skew distribution of publications as a function of citations urges to develop better statistics than those based only on 'mean values';
* Last but certainly not least we have the crucial question of validity: is the thing we are measuring, the same thing as we want to know? Or: do citations indeed measure at least an important aspect of scientific quality?

The empirical and theoretical work in our Leiden group is devoted to cope with the above problems. It is part of our long-standing and extensive experience in the application of bibliometric indicators. On the basis of this experience (see for instance *Van Raan* [13]) we can say, yes, bibliometric indicators based on methodologically thorough analysis generally gives you a good to even very good quantitative impression of at least an important aspect of quality, namely international impact.

Bibliometric assessment of research performance is based on one important assumption: the work to be evaluated must be published in the open, international journal literature. This means that bibliometric indicators are well applicable in the natural and life sciences. In the applied and engineering sciences as well as in the social and behavioural sciences and in the humanities, international journals are often not the primary communication channel. Then, bibliometric assessment becomes problematic. However, in fields like chemical engineering, psychology and even linguistics (*Nederhof* and *Noyons* [15]) bibliometric analysis can be applied successfully. For instance, we observe in recent years a striking increase of publications in international journals for the social and behavioural sciences as well as for the humanities in the Netherlands.

Nevertheless, it is important to stress again the 'journal-literature' based character of bibliometric indicators. Although publication in international journals is a major driving force of scientific development, it never encompasses the entire spectrum of the presentation and dissemination of research activities. Other measures of quality and esteem will be necessary in the evaluation of scientific performance.

2.2 *Advanced indicators based on an added-value data-system*

The application of bibliometric indicators in research performance evaluation will only be successful if an advanced bibliometric data-system is available with the following features:

* The above mentioned and all other relevant *technical problems* with respect to the basic publication and citation data must be solved. We again mention as a crucial example unification and 'hierarchization' (i.e., position in organizational structure) of affiliation addresses. Addresses form a heavily underestimated problem in bibliometric analysis. The size -and with that the importance- of this problem is clearly illustrated by the fact that in our group about 1 full-time equivalent staff position is dedicated to solve address problems and to add important additional address-related data.
  We regard the solution of these problems -or at least an acceptable first approach- as 'technical added values' of the bibliometric data system. They require a substantial amount of new, 'non-bibliometric' information from many other sources (e.g., addresses from university guide-books and research reports, 'Who is Who in Science', etc.). Moreover, they have to be included in the data system and 'matched' with existing data in an automated way in order to handle large amount of data effectively;
* The above mentioned and all other relevant *methodological problems* with respect to design, construction and calculation of appropriate indicators must be solved by advanced software routines enabling the choice of many relevant options. We again mention crucial examples: choice of aggregation level; choice of citation window; choice of field normalization; definition and 'delineation' of (sub)fields; choice of reference values; choice of particular statistics.
  We regard the solution of these problems as 'methodological added values' of the bibliometric data system. They also require large experience in bibliometric analysis and the 'encoding' of this experience in terms of automated algorithms and software;
* Based on experiences, our bibliometric data system 'grows' continuously by adding important *practical knowledge* in an 'encoded' way. For instance, there are many researchers with the same name and even the same (first) initials. Conversely, many researchers appear in databases under differently spelled names and initials. Especially for (married) women this can be a major problem. For each institute, university, R&D division, research organisation for which a bibliometric analysis

has been performed by us, we constructed a 'verified output-database' with a reliable link between names of researchers, their (corrected) address(es), and their (source) publications. Data from annual research reports as well as information given by individual scientists play an essential role. Routines for measuring the impact of 'non-SCI' publications appear to be more and more important for an increasing number of research fields, for instance the social and behavioral sciences, the applied sciences, mathematics, biology;

* Also based on experiences are the methodological improvements. For instance, new indicators designed on the basis of discussions with and suggestions from 'users'. A striking example is an appropriate reference value for international comparison and approaches to make bibliometric analysis as topical as possible (see our methodology-paper, *Moed* et al[5]);
* An advanced bibliometric data-system must be organized in such a way that developments in electronic publishing and in relevant Internet facilities -such as the earlier mentioned high-energy physics SPIRES database- can be included as soon as these developments prove to play a non-negligible role in bibliometric studies. Current research in our group is devoted to meet the 'bibliometric challenges' of the 'expanding data universe' (see, for instance, *Roosendaal* [16]).

The above shows that research performance evaluation by bibliometric methods requires an advanced data system with much more than only the basic publication and citation data alone. It took our Leiden research institute CWTS more than a decade to develop such a system, and we are just beginning to improve its applicability in terms of user-oriented facilities, as well as in terms of economy and prompt analytical results.

2.3 Practical use of advanced bibliometric indicators

With help of the following figures we illustrate several main elements of our bibliometric performance assessment system. The essential feature is an analysis of research impact in terms of time trends, international comparison, and research-profile. In order not to disturb the main line of this paper, we refer for a more thorough discussion of the underlying methodology to *Moed* et al.[5] In Figures 1, 2, and 3 we present the results of a real-life example: the Department of Astronomy at Leiden University (for more details, see *Van Raan* and *Van Leeuwen* [14]).

In Figure 1 we show (a) the scientific productivity in terms of the number of publications in international journals, and (b) also the total number of 'external' citations (within a three-year citation window), as a measure of the absolute size of the department's world-wide scientific impact. Two observations are important. First, one immediately sees that performance measurement must cover a wider range of years. Bibliometric 'snapshots' are useless, even periods of five years are too short. We state that this finding will also hold for peer review. So an important lesson is learned from empirical, bibliometric analysis: research groups need time to establish their position; it is incorrect to judge research performance on the basis of just a few years.

Second, as discussed earlier in the methodological problems, we clearly need a reference value to know whether the department's performance is high or low in its own research field. Therefore we look at Fig. 2. Here we make a comparison of the actual average impact of the department with the same indicator for the whole world.

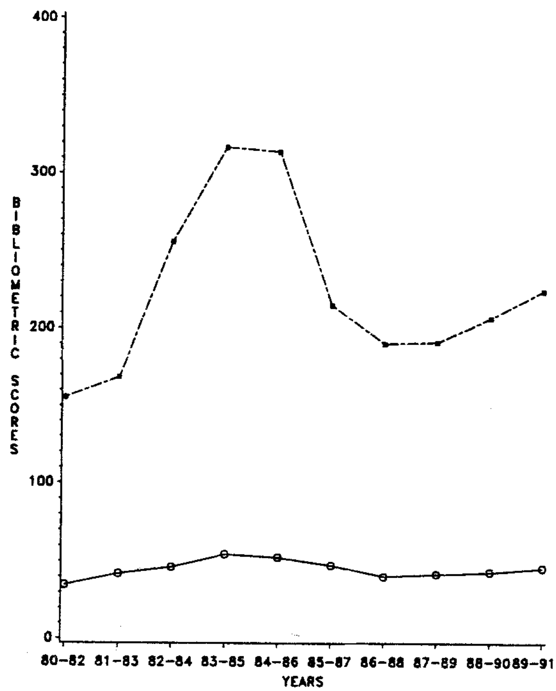

Fig. 1. Trends in numbers of publications and in ('short-term', i.e., three year window) citations for the Department of Astronomy, Leiden University, 1980–1991. Solid line: publications (three year average); dashed line: citations (three year average, for each publication year t, citations are counted for the years t, $t+1$, $t+2$, self-citations not included)

This latter indicator is based on all journals of that specific field (see *Moed* et al[5]). The result is clear: this department performs far above world level (which is in fact: Western-world level). It is an excellent department, and this finding corresponds completely with the opinion of world-wide astronomy.

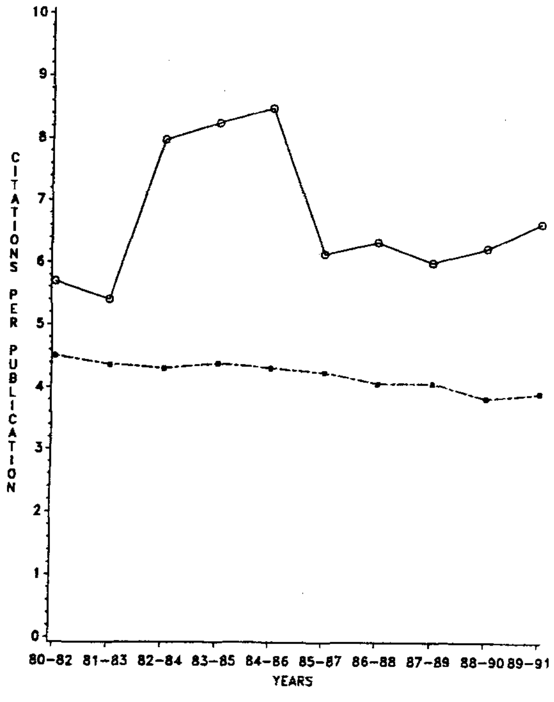

Fig. 2. Trends in short-term impact as compared to world average of the field, Department of Astronomy, Leiden University, 1980–1991. Solid line: citations per publications (oeuvre of the department); dashed line: citations per publication, world average of the field; self-citations included

Figure 3 shows the relation between the department's research profile and impact: a breakdown of the impact by the different (sub)fields covered by the astronomy department. In this case of a typical disciplinary research field such as astronomy, the research profile is sharply focused to just that research field. In the case of inter- or multidisciplinary work, an institute's research profile will be much broader. Thus, the breakdown of impact to the different (sub)fields yields interesting information about the strengths and weaknesses of groups and institutes with respect to their scientific basis. As an example we show in Figure 4 a similar breakdown of impact as in Fig. 3, but now for an environmental research institute (*Van Raan* and *Van Leeuwen* [12]).

Figure 5 is similar to Fig. 2, but for another department at a university in the Netherlands. We clearly see the difference with the Leiden astronomy department. The performance of this second department is certainly not bad, but not real top-performance.

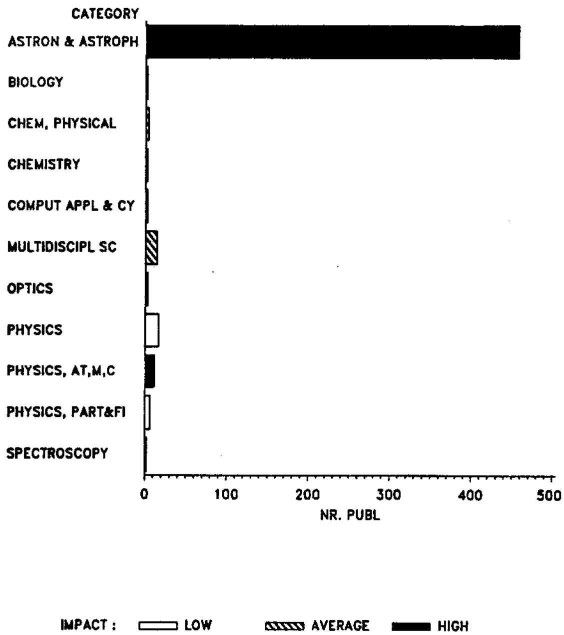

Fig. 3. Research profile in relation to impact, Department of Astronomy, Leiden University, 1980–1993. The department's impact is normalized to the world-average of the field(s) concerned. Fields are defined in terms of the SCI 'journal categories'

Important methodological elements concern citation windows and 'topicality' of the analysis. A possibility is to measure group's output and impact 'cumulatively' during a fixed time period (*Braun* et al[17]), based on *all* publications and citations related to this period. Thus, we count the total number of papers published by a group during a specific period, for instance 1985–1993, and the total number of citations received by all these papers during the *same* period. Consequently, for papers published in 1985, citations are counted during the period 1985–1993. For papers published in 1993, only citations received in 1993 are counted. As an example, we present in Table A1 (*Appendix*) the results for the Leiden Department of Astronomy. In this 'total block' indicator, publications up till 1991 are covered, but we enlarged the citation period up till 1993. We present the results for the entire period (1980–1993) and focus particularly on the more recent period 1986–1993. The table provides eleven main indicators which are our CWTS 'standard' indicators. Again, for a more detailed discussion of the indicators concerned we refer to *Moed* et al.[5]

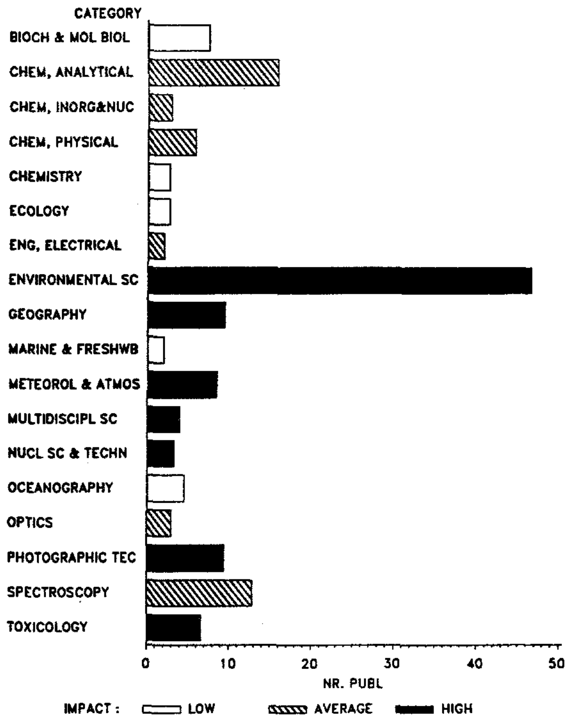

Fig. 4. Research profile in relation to impact, a European institute for environmental research, 1987–1993. The institute's impact is normalized to the world-average of the field(s) concerned. Fields are defined in terms of the SCI 'journal categories'

In a country-wide assessment of chemistry, we provided the above indicators for all groups, about 150. An overview of the international impact level of these chemistry research groups together, is given in Fig. 6.

Here we plot for each group the ratio of the actual impact versus each group's own world-wide level (ratio averaged over a five-year time period) against number of publications (*Moed* and *Van der Velde;*[18] *Moed* and *Hesselink* [19]). We see two striking features, which are directly important for national science policy. First, the majority of chemistry groups in the Netherlands is good or very good, with performance above (western-)world level. Second, those groups that are not able to reach a reasonable scientific productivity, also do not perform very well in terms of impact. The above shows that advanced bibliometric indicators can be applied successfully to monitor strengths and weaknesses in national research performance.

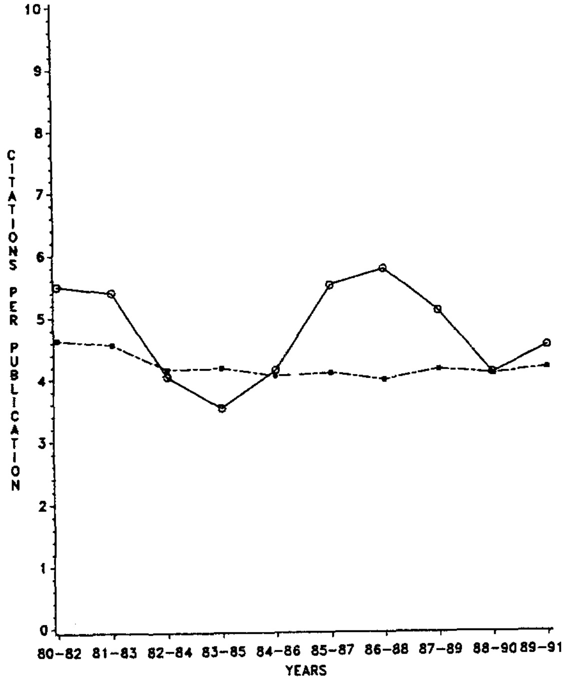

Fig. 5. Trends in short-term impact as compared to world average of the field, similar to Fig. 2, but now other department, 1980–1991. Solid line: citations per publications (oeuvre of the department); dashed line: citations per publication, world average of the field; self-citations included.

What are the costs of a performance analysis with advanced bibliometric indicators as described above (and in the appendix)? To give a reasonable indication: our work for one of Europe's oustanding academic institutions, the University of Louvain, covering all the natural and life sciences (Faculty of Mathematics and Natural Sciences, and the Faculty of Medicine), took about hundred working days, including extensive input analysis (e.g., categories of scientific personnel, financial resources), presentation of results and 'validation discussions'. Currently, a similar effort is necessary for two other universities with a high international reputation, our own University of Leiden, and the University of Utrecht. We stress that close cooperation with the institution is necessary, particularly with respect to the creation of the input data, the verification of first-round results, and the presentation, discussion and validation of the final results. Especially this latter 'advisory part', including opinions about 'implementation' of the results, comes next to the analytical part and is extremely important in the application of bibliometric indicators. Here we fully make use of our ample experiences. The permanent interaction of our application-work with academic basic research and developmental work in a university setting is highly valued by users. Improvements of data automation and of major parts of the analytical process will certainly reduce the costs of performance analysis.

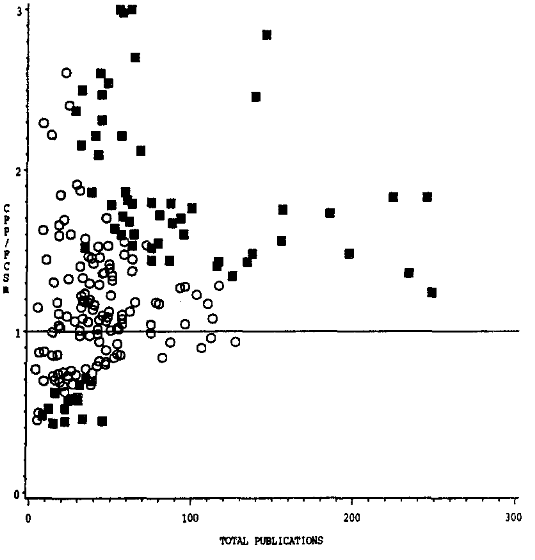

Fig. 6. Impact compared to world average for all academic chemistry research groups at universities in The Netherlands, 1986–1991 (from: *Moed* and *Hesselink*, Ref.19). Black coloured squares above (below) the horizontal reference line represent groups for which the actual impact (CPP) is significantly above (below) the world average for that group (FCSm). Ratio CPP/FCSm = 1 defines the world average. The CPP, FCSm, and CPP/FCSm indicators are discussed in detail by *Moed* et al, Ref.5

2.4 Peer review opinions versus bibliometric findings

How do our bibliometric results compare with peer judgements? From our experience with the peer review procedure discussed earlier and other comparisons we can say that there is generally a significant correlation between the opinion of peers and bibliometric indicators. This is clearly visible in the results of the recent VSNU peer review procedures for biology (VSNU[10]) and even for a social science research field as educational psychology (VSNU;[20] *Kroonenberg* and *Van der Veer*[21]). This confirms earlier observations (*Nederhof*[4]) that, generally, a strong correlation exists between qualitative or non-quantitative measures of quality or esteem and bibliometric findings.

It is, however, important to investigate the above mentioned correlation between quantitative and qualitative measures more thoroughly. Currently we are conducting an extensive bibliometric analysis project at our own university for the natural, life, and social sciences. Next to bibliometric indicators other, qualitative indicators such as editorships of outstanding journals, awards, memberships of prominent scientific societies, honorary degrees, guest professorships at renowned universities, and advisory committee memberships are used. These non-quantitative indicators are typical ele-ments of peer review. They are important for all fields of science as they do not directly depend on publication and citation 'cultures' and thus provide another view of scientific performance, for example a more 'scholarly' contribution to science. They are particularly important in those disciplines where bibliometric methods are difficult to use, for instance in the social sciences, the applied sciences, but especially in the humanities. We expect to present in a forthcoming study the results of an extensive analysis of correlations between these qualitative measures of esteem and bibliometric indicators.

Those groups or departments where a considerable difference is found between peer judgement and bibliometric results are interesting cases. What is going on? Are the peers wrong? Do the bibliometric indicators give an impression too optimistic or pessimistic? Our opinion is the following. If bibliometric indicators show a poor performance, but the peers' judgement is positive, then possibly, as discussed earlier, communication practices of the group concerned are such that bibliometric assessment may not work well. However, this has to be verified.

If bibliometric indicators show a good performance and the peers' judgement is negative, there is a good chance that the peers are wrong. The argument mentioned in the Boden Report (ABRC[2]) that in such cases "very poor 'peers' were being employed" is too cheap. It is virtually impossible to cover all issues relevant for a broad as well as deep review of research work by a (very) limited group of people. For example, peers may not sufficiently be aware of the role of the many different actors in recent developments in a specific field, and the position of the group to be evaluated in relation to all these actors.

We summarize the important aspects of bibliometric indicators as quantitative core of peer review. Most of these aspects confirm and reinforce earlier observations on the relation between peer review and bibliometric indicators (*Moxham* and *Anderson* [6]).

* Advanced bibliometric indicators as developed in our Leiden group provide up-to-date, detailed, 'objective' (though of course never 'complete'), and purposively structured information on the performance (particularly: 'impact') of a research group;
* Bibliometric indicators stimulate posing hard questions (for instance: what is going on in this group, we see a dramatic decrease of impact) and prevent the peer review process from becoming too soft or too 'uninterested';

* Thus, bibliometric indicators support peer review recommendations with 'hard' evidences, especially in the case where decisions will probably be unpleasant (for instance it is difficult to defend the positions of research groups or programmes with no results of any reasonable impact for many years);
* The learning effects for improvement of evaluation processes are strong when bibliometric indicators are applied, e.g., the role of international publications in a specific field;
* Bibliometric indicators allow comparison of performance within research fields but also between fields;
* They can be applied effectively in discussions between, for example, research councils and the government. This is especially the case when specific strengths or weaknesses on the national level are found;
* Finally, advanced bibliometric indicators provide a substantial insight into scientific communication practices. This is important for research management, for the work of individual scientists, and last but not least it contributes to a better understanding of the dissemination and use of knowledge.

We conclude that bibliometric performance indicators allow substantial improvement of the peer review based evaluation by supplying new elements in terms of objective analysis of research output and impact. The costs of advanced bibliometric analysis is comparable or even lower than that of peer review and further developments in automated procedures as well as the cumulation of 'cleaned' and 'value-added' data will certainly reduce these costs. Moreover, once a bibliometric performance analysis has been conducted for a specific institution, a permanent 'strengths and weaknesses monitoring system' can be relaized at relatively low costs.

2.5 Bibliometric cartography as a monitoring device

Finally, we discuss in this paper the second major line in our quantitative methods: bibliometric mapping, cartography of research fields. We limit ourselves to the main lines and refer for a more detailed discussion to *Tijssen* and *Van Raan.*[22]

The basic idea is the following. Each year about a million scientific articles are published. For just one research field, such as materials science, the amount of papers is already about 30,000 per year. This gives you an impression of the enormous size of nowadays scientific output. How to keep track of all these developments? Are there cognitive structures 'hidden' in this mass of publihed knowledge, at a 'meta-level'?

Suppose each research field can be characterised by a list of most important, say 100, keywords. For materials science such a list will cover words like ceramics, polymers, semiconductors, high-temperature superconductivity, alloys, and so on. Each publication can be characterized by a subset from the total list of keywords. For all 30,000 publications we compare their keyword-lists pairwise. In other words, these 30,000 publications constitute a gigantic network in which all publications are linked together by one of more common keywords. The more keywords two publications have in common, the more these publications are related (keyword-similarity) and thus belong to the same research area or research specialty. In mathematical terms, publications are represented as vectors in a high-dimensional word-space. In this space they group together, or take very distant positions when they are not related to each other. We developed mathematical techniques to unravel these publication networks and to map the underlying structures. The fascinating point is that these structures can be regarded as the cognitive, or intellectual structure of science. As discussed above, it is entirely based on the total of relations between all publications. Thus, the structures that are discovered are not the result of any pre-arranged classification system or whatsoever. Nobody prescribed these structures. The structures emerge solely from the internal relations of the whole universe of publications together. In other words, what we make visible by our mathematical methods, is the self-organised structure of science.

In Figure 7, we show the result of 'freezing-out' the underlying patterns in about 40,000 publications (years 1990/91). The map (*Tijssen* and *Van Raan* [22]) clearly shows the major subfields of material science in their mutual relationships. The closer word-clusters are, the more related the subfields, represented by those words. We observe the major subfields: instrumentation (central cluster), semiconductor research, superconductivity, ceramics, alloys, and properties of materials.

This bibliometric cartography has large potentials. First, as shown in the figure, it visualises the cognitive landscape of a scientific field. Second, by making these maps for a series of years, we are able to observe trends and changes in structure. So we gain insight into the dynamics of scientific development. Third, we are also able to put the position of major actors on the map. Thus we are creating a strategic map: who is where in science? For recent examples of bibliometric mapping we refer to *Noyons* and *Van Raan* [23] and to our institute's Internet home page at http://sahara.fsw.leidenuniv.nl/cwts/cwtshome.html.

The combination of both bibliometric methods, performance assessment and mapping, appears to be a very powerful tool in the evaluation of research activities. We recently applied this 'combined methodology' as a support for a governmental audit of the activities of an internationally renowned institute in the field of micro-electronics. The focus of this audit was on the position of the institute in its academic setting; on the research strategy of the institute with a special emphasis on interaction with industry; and on the potentials of the institute in terms of economy, particularly employment perspectives in the micro-electronics sector. Our analytical work, which included the analysis of 'competing' or 'benchmark' institutes, was highly valued by the governmental authority that commissioned the work as well as by the researchers in the institute concerned (*Noyons* et al[24]).

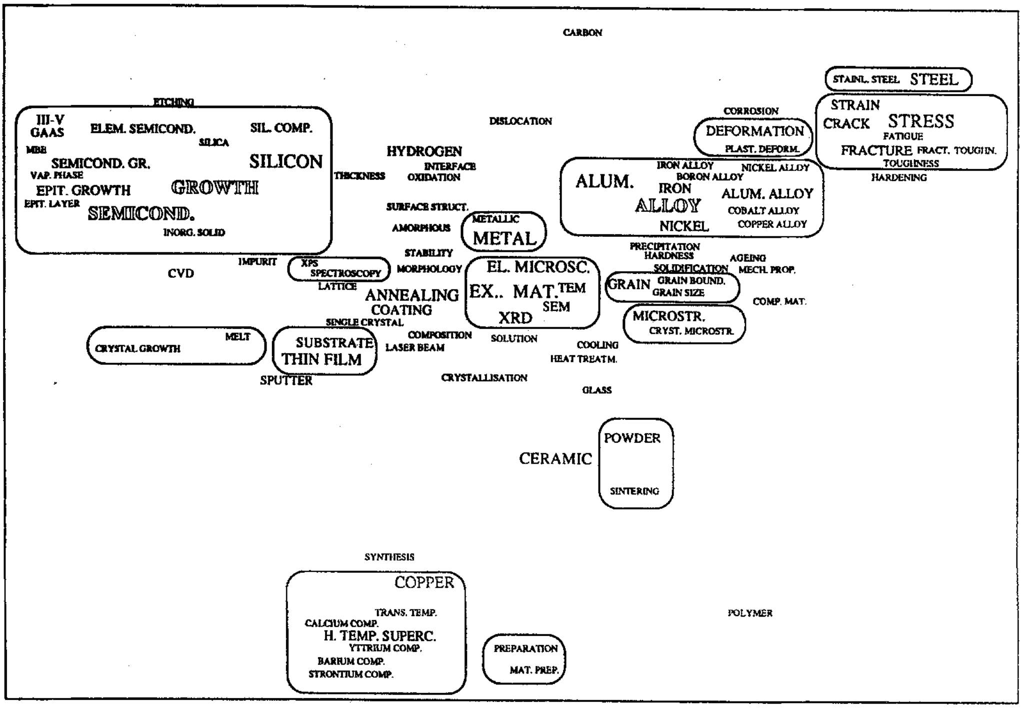

Fig. 7. Bibliometric map based on co-word analysis of materials science, resulting from 39,044 publications in 1990/91. Details are amply discussed in *Tijssen* and *Van Raan*, Ref.22

As we all know already, bibliometric performance assessment typically concerns the 'past'. We must realize, however, that also peers have nothing more than the 'past' to judge performance. Some 'time delay' is therefore not a specific characteristic of bibliometric methods. It is inherent of any performance assessment. Of course, peers may 'predict' or 'forecast' possible scientific developments. 'Delphi-procedures' are an example of such foresight activities (BMFT[25]). Another example is the 'long-term viability' aspect in the VSNU peer review as discussed in this paper (VSNU[10,20]). But all speculations whatsoever about future developments will always be, by nature, on yesterday's and today's impressions. It is precisely this phenomenon which makes 'bibliometrically-based' foresight exercises possible: a time-series of maps for recent years allow, to a certain extent, an 'extrapolation' of trends. For instance, the map shown in Fig. 5 can also be made for the years 1992 to 1995. Or even include the now available data for 1996. Such a time-series is in fact a film of the research field material science. If, for instance, two particular subfields are positioned more and more close to each other as a function of time, it is not unrealistic to 'predict' that a 'synthesis' of both subfields will occur in the coming years. We are now experimenting with this type of 'bibliometric cinematography' as a tool for foresight exercises (*Noyons* and *Van Raan* [23]).

3. Concluding remarks

We presented an overview of the potentials and limitations of bibliometric methods for the assessment of strengths and weaknesses in research performance, and for monitoring scientific developments. Two different approaches are distinguished. Research performance assessment is based on advanced analysis of publication and citation data. We showed that the resulting indicators are useful: they address significant concepts in the framework of evaluation and can therefore be considered as an indispensable element next to peer review in research performance assessment procedures.

Indicators based on advanced bibliometric methods, and particularly their trends as a function of time, offer much more than 'only numbers'. They provide insight into the international position of actors at the research front in terms of influence and specializations, as well as into patterns of scientific communication and processes of knowledge dissemination.

In the second application, monitoring scientific developments, bibliometric mapping techniques are essential. This 'bibliometric cartography' provides an instrument to create a 'landscape', i.e., a cognitive structure of scientific fields. As a function of time, this 'monitoring device' may even have foresight potentials in terms of extrapolation of significant patterns.

Both bibliometric approaches can be combined to a broader and powerful methodology to observe scientific advancement and the role of actors.

*

This work was supported in part by grants of the Netherlands Organisation for Scientific Research (NWO), Economic and Social Research Council (ESR).

Appendix

Table A1
Bibliometric 'Total block' Indicators. Example: Department of Astronomy, Leiden University (for more details see: *Van Raan* and *Van Leeuwen*[14])

Indicator	Symbol	Time Period	
		[1980–1993]	[1986–1993]
Number of publications	***P***	524	257
Number of citations received	***C***	10324	3788
Citations per publication (average)	***CPP***	19.7	14.7
Citations per publication, self-citations not included	***CPPex***	15.9	11.7
Percentage of papers not cited during the time period considered	***%Pnc***	7.6	9.0
Av. citation rate of journal set	***JCSm***	14.7	10.1
Av. citation rate of (sub)field(s)	***FCSm***	12.1	8.2
Citations per publ., compared to journal set world average	***CPP/JCSm***	1.34	1.46
Citations per publ., compared to (sub)field world average	***CPP/FCSm***	1.63	1.80
Ratio of journal set world av. and (sub)field world average	***JCSm/FCSm***	1.21	1.24
Percentage self-citations	***% SELFCIT***	19.5	21.0

References

1. S. COLE, L. RUBIN, J. R. COLE, *Peer Review in the National Science Foundation*. Washington DC: National Academy of Sciences, 1978 (ISBN: 0-309-02788-8).
2. ABRC. *Peer Review: A Report to the Advisory Board for the Research Councils from the Working Group on Peer Review* ('Boden Report'). London: Advisory Board for the Research Councils (ABRC), 1990.
3. H.-D. DANIEL, An evaluation of the peer review process at Angewandte Chemie, *Angew. Chem. Int.(Ed. Engl.)* 32 (1993) 234–238.
4. A. J. NEDERHOF, The validity and reliability of evaluation of scholarly performance. In: A. F. J. VAN RAAN (Ed.), *Handbook of Quantitative Studies of Science and Technology*, Amsterdam: Elsevier Science/North-Holland, 1988, p.193–228 (ISBN: 0-444-70537-6).
5. H. F. MOED, R. E. DE BRUIN, TH. N. VAN LEEUWEN, New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications, *Scientometrics*, 33 (1995) 381–422.
6. H. MOXHAM, J. ANDERSON, Peer review: a view from the inside, *Science and Technology Policy*, (1992) 7–15.
7. M. POLANYI, The republic of Science, its political and economic theory, *Minerva*, 1 (1962) 54–73.
8. D. J. VAN DE KAA, Picking the winners by consensus: grant-giving practice in the Netherlands, *Higher Education*, 28 (1994) 59–83.
9. D. MÜLLER-BÖLING (Ed.), *Qualitätssicherung in Hochschulen: Forschung, Lehre, Management* (Quality Care in Universities: Research, Teaching, Management). Gütersloh: Verlag Bertelsmann Stiftung , 1995 (ISBN: 3-89204-184-9).
10. VSNU, *Quality Assessment of Research. Netherlands Biology in the Nineties*. Utrecht: VSNU, 1994 (ISBN: 90-801015-7-5)..
11. D. F. HORROBIN, The philosophical basis of peer review and the suppression of innovation, *Journal of the American Medical Association (JAMA)*, 263 (1990) 1438–1441.
12. A. F. J. VAN RAAN, TH. N. VAN LEEUWEN, Evaluation of the performance of three European environmental research institutes. Policy and management relevance of advanced bibliometric methods, To be published, 1996.
13. A. F. J. VAN RAAN, Advanced bibliometric methods to assess research performance and scientific development. Basic principles and recent practical applications, *Research Evaluation*, 3 (1993) 151–166.
14. A. F. J. VAN RAAN, TH. N. VAN LEEUWEN, *A Decade of Astronomy Research in the Netherlands*. Research Report to the Netherlands Organisation for Scientific Research (Astronomy Division, NWO/ASTRON). Leiden: Centre for Science and Technology Studies, report CWTS-95-01, 1995. A shorter version of this report will be published as a journal paper.
15. A. J. NEDERHOF, E. C. M. NOYONS, International comparison of departments' research performance in the humanities, *Journal of the American Society for Information Science (JASIS)*, 43 (1992) 249–256.
16. H. E. ROOSENDAAL, Roles of bibliometrics in scientific communication, *Research Evaluation*, to be published, 1996.
17. T. BRAUN, W. GLÄNZEL, A. SCHUBERT, Scientometric indicator datafiles. A multidimensional machine readable database for evaluative purposes, *Scientometrics*, 28 (1993) 137–150.
18. H. F. MOED, J. G. M. VAN DER VELDE (1993), *Bibliometric profiles of academic chemistry research in the Netherlands*. Research report to the Netherlands Organisation for Scientific Research (Chemistry Division, NWO/SON). Leiden: Centre for Science and Technology Studies, report CWTS-93-08, 1993.
19. H. F. MOED, F. TH. HESSELINK, The publication output and impact of academic chemistry research in the Netherlands during the 1980's. Bibliometric analysis and policy implications, *Research Policy*, to be published, 1996

20. VSNU. *Quality Assessment of Research. Pedagogical and Educational Sciences. Past Performance and Future Aspects*. Utrecht: VSNU, 1995 (ISBN: 90-5588-020-5).
21. P. M. KROONENBERG, R. VAN DER VEER, Internationale publikaties en kwaliteit. Een onderzoek naar het publikatiegedrag van Nederlandse pedagogen en onderwijskundigen (International publications and quality. A study of publication patterns of educational psychologists in the Netherlands), *Pedagogische Studiën*, to be published, 1996.
22. R. J. W. TIJSSEN, A. F. J. VAN RAAN, Mapping changes in science and technology: bibliometric co-occurrence analysis of the R&D literature, *Evaluation Review*, 18 (1994) 98–115.
23. E. C. M. NOYONS, A. F. J. VAN RAAN, *Mapping the development of neural network research. Structuring the dynamics of neural network research and an estimation of German activity*. Research report to the German Federal Ministry of Education, Science and Technology (BMBF). Leiden: Centre for Science and Technology Studies, report CWTS-95-06, 1995. A shorter version of this report will be published as a journal paper.
24. E. C. M. NOYONS, M. LUWEL, H. F. MOED, *The Position of IMEC in the Field of Micro-Electronics*. Research report to the Ministry of the Flemish Community, Brussels. Leiden/Brussels: Centre for Science and Technology Studies/Ministry of the Flemish Community, Report D/1996/3241/002, 1995 (ISBN: 90-403-0052-6).
25. BMFT. *Deutscher Delphi-Bericht zur Entwicklung von Wissenschaft und Technik*. Bonn: Bundesministerium für Forschung und Technologie (BMFT, now BMBF), 1993 (ISBN: 3-88135-267-8).