Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 57, No. 2 (2003) 257–280

The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence

THED N. VAN LEEUWEN, MARTIJN S. VISSER, HENK F. MOED, TON J. NEDERHOF, ANTHONY F. J. VAN RAAN

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

Evaluation studies of scientific performance conducted during the past years more and more focus on the identification of research of the 'highest quality', 'top' research, or 'scientific excellence'. This shift in focus has lead to the development of new bibliometric methodologies and indicators. Technically, it meant a shift from bibliometric impact scores based on *average values* such as the average impact of all papers published by some unit to be evaluated towards indicators reflecting the *top* of the citation distribution, such as the number of 'highly cited' or 'top' articles.

In this study we present a comparative analysis of a number of standard and new indicators of research performance or 'scientific excellence', using techniques applied in studies conducted by CWTS in recent years. It will be shown that each type of indicator reflects a particular dimension of the general concept of research performance. Consequently, the application of one single indicator only may provide an incomplete picture of a unit's performance. It is argued that one needs to combine the various types of indicators in order to offer policy makers and evaluators valid and useful assessment tools.

Introduction

Increasingly, evaluation studies of scientific performance conducted during the past years focus on the identification of research of the 'highest quality', 'top research', or 'scientific excellence'. This shift in focus has lead to the development of new bibliometric methodologies and indicators. Technically, it means a shift from bibliometric impact scores based on average values – such as the average impact of all papers published by some unit – towards indicators reflecting the presence in the top of the citation distribution – such as the number of 'highly cited' or 'top' articles.

In this study we present a comparative analysis of a number of indicators of research performance, and particularly of 'scientific excellence', using techniques applied in studies conducted by CWTS in recent years (MOED et al, 1995; SCHUBERT et al, 1989; and TIJSSEN et al, 2002). It will be shown that each type of indicator reflects a particular

Received March 27, 2003

Address for correspondence:
THED N. VAN LEEUWEN
Centre for Science and Technology Studies (CWTS)
Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands
E-mail: leeuwen@cwts.leidenuniv.nl

0138–9130/2003/US $ 20.00

dimension of the general concept of research performance. Consequently, the application of a single indicator may provide an incomplete picture of a unit's performance. It is argued that one needs to combine the various types of indicators in order to provide policy makers and evaluators with valid and useful assessment tools.

This study is based on a national research performance assessment in the field of chemistry and related fields, initiated by the collaborating Dutch universities, organized in the Association of Universities in the Netherlands (VSNU).

Data and methods

In this paper outcomes are presented from a study of publication output and international impact of academic chemistry researchers in the Netherlands. The study was performed on behalf of the International Review Committee on Chemistry in the Netherlands. This Committee was established in 2001 by the Association of Universities in the Netherlands (VSNU) for a quality assessment of academic chemistry research. Using bibliometric techniques, the present study assesses the publication output and citation impact of senior scientists affiliated with chemistry research programmes at ten universities. The outcomes of the study add supplement to the qualitative instruments available to the Committee (VAN RAAN, 1996). The impact, as measured by citations, is compared with worldwide reference values.

The ten universities involved in the research assessment procedure are: Catholic University Nijmegen (KUN), Leiden University (LEI), State University Groningen (RUG), Delft University of Technology (TUD), Eindhoven University of Technology (TUE), Twente University (UT), Utrecht University (UU), University of Amsterdam (UVA), Vrije Universiteit Amsterdam (VU), and Wageningen University Research Center (WUR).

The covered period is 1991–2000 for both publications and their citation impact. In our experience, a period of about ten years is needed to assess research performance. This period allows most units to produce a number of publications sufficient for statistical analysis. The core of the study consists of scientific publications in chemistry and related fields, published in journals that were processed for the CD-ROM versions of the *Science Citation Index* and eight associated indices (CI): the *Science Citation Index* (SCI), the *Social Science Citation Index* (SSCI), and the *Arts & Humanities Citation Index* (A&HCI), recently extended with six specialty Citation Indices (Chemistry, Compumath, Materials Science, Biotechnology, Biochemistry & Biophysics, and Neuroscience).

The study is based on 18,160 papers, published by 600 senior researchers, who were associated with chemistry research programmes on December 31, 2000. The names of the senior scientists were provided by VSNU. The researchers were aggregated into about 160 research groups at 10 Dutch universities.

In a first step, for each senior scientist all relevant publications from 1991 - 2000 (the most recent available ten-year period at the start of the data collection round) were extracted from our Citation Index-based publication data system. This includes all publications listing the researcher either as first author or as co-author. In a verification-round, researchers were asked to verify that publication lists were correct and complete. About 60% of the scientists responded to our request to check their publication lists. Additions and corrections indicated by responding scientists were entered into our database. For scientists who did not respond, we performed a test ourselves, aimed at identifying and deleting publications authored by other scientists having similar names. As a result, we are confident that we obtained valid publication data for nearly all chemistry researchers in this study.

Standard indicators

The first indicator gives the total number of papers published by the research unit during the entire period (*P*). We considered only papers classified as *normal articles*, *letters*, *notes*, and *reviews* (from 1996 onwards, notes are no longer used as a separate document-type, and in general notes are treated as normal articles). The second indicator is the number of citations received, *C* (excluding author self-citations). The third indicator is the average number of citations per publication calculated while self-citations are not included (*CPP*). A fourth indicator is the percentage of articles not cited during the time period considered (*%Pnc*), excluding self-citations.

Next, two international reference values are computed. A first value represents the mean citation rate of the journals in which the research unit has published (*JCSm*, the mean Journal Citation Score). The second value relates to the fields in which the research unit has published (*FCSm,* the mean Field Citation Score). Our definition of subfields is based on a classification of scientific journals into *categories* developed by ISI. Although this classification is certainly not perfect, it is at present the only classification available to us, fitting the multidisciplinary nature of the ISI citation indexes. Both the *JCSm* and *FCSm* take into account the type of paper (e.g., normal article, review, and so on), as well as the specific years in which the research unit's papers were published. For example, with respect to the calculation of *FCSm*, the number of citations received during the period 1991 - 2000 by a *letter* published by a research unit in 1991 in field X is compared to the average number of citations received during the same period (1991–2000) by all *letters* published in the same field (X) in the same year (1991). Generally, a research unit publishes its papers in several fields rather than one. Therefore, we calculated a weighted average *FCS* indicated as *FCSm*, with the weights determined by the number of papers published in each field. Self-citations are excluded from the computation of *FCSm*. When a journal is classified in multiple subfields, citation scores are computed according to their number of field assignments. Basically, a paper in a journal classified in N subfields is counted as 1/N paper in each subfield, and so are its citations and *FCSm* scores.

On the basis of the above international reference values, two 'normalized' impact indicators are calculated. First is the indicator *CPP/JCSm*. This indicator compares the average number of citations to the oeuvre of a university (*CPP*) to the journal mean citation scores *JCSm*, by calculating the ratio for both. Self-citations are excluded in the calculation of the ratio *CPP/JCSm* to prevent that ratios are affected by divergent self-citation behavior. Next, in calculating *CPP/FCSm*, the average number of citations to the oeuvre of a university (*CPP*) is compared to the field mean citation scores *FCSm*, by calculating the ratio for both. Again, self-citations are excluded in the calculation of the ratio *CPP/FCSm.*

A limitation of a focus on the journals in which a unit is publishing is that low impact publications published in low impact journals may get a similar score as high impact publications published in high impact journals. However, the *CPP/FCSm* indicator is free from this limitation, because it takes the impact level of a units' journal set normalized to fields into account. Therefore, it seems the most suitable indicator of the international position of a research unit. If the ratio *CPP/FCSm* is above (below) 1.0, this means that the oeuvre of the research unit is cited more (less) frequently than an 'average' publication in the subfield(s) in which the research unit is active. Thus *FCSm* constitutes a *world subfield average* in a specific (combination of) subfield(s). In this way, one may obtain an indication of the international position of a research unit, in terms of its impact compared to a 'world' average. This 'world' average is calculated for the total population of articles published in Citation Indexed journals assigned to a particular subfield or journal category. As a rule, about 80 percent of these papers are authored by scientists from the United States, Canada, Western Europe, Australia and Japan. Therefore, this 'world' average is dominated by the Western world.

Another important international reference value is *JCSm/FCSm*, an indicator for scientific status, international visibility and 'quality'**.** If this indicator is above (below) 1.0, the mean citation score of the journal set in which the research unit has published exceeds the mean citation score of all papers published in the subfield(s) to which the journals belong. In this case, one can conclude that the research unit publishes in journals with a relatively high (low) impact.

Frequently cited publications

An additional set of impact indicators reflects the contribution to the most frequently cited papers worldwide (TIJSSEN et al., 2002). We illustrate the importance of these top-impact indicators with an example. Two research units may have equal impact scores on the *CPP/FCSm* indicator, but one produces a steady stream of publications that are cited well but fails to produce really high impact publications, while the other contributes considerably to the high impact publications (and also has a larger number of less well cited publications). To examine the distribution of frequently cited papers, we have ranked each publication by the number of citations received up to four years after publication, and identified those belonging to the 5% most frequently cited papers in a given year within their discipline/subfield. The use of such a 'fixed length' citation window implies that the analysis only involves papers published during 1991–1997. Moreover, letters were excluded, due to the fact that letters display a deviant citation pattern compared to articles and reviews.

Thus, the *P91-97* figure gives the number of review articles and normal articles published during 1991–1997. As described above, we determined for this set of publications the absolute number of papers that are represented among the top 5% most frequently cited of all papers published in a particular year, and subject category. On the basis of the total output in the period 1991-1997, we calculated an expected top 5% number of publications, taking into account also the deviations from the 95th percentile if tied values occur due to the discrete nature of the citation distribution. Finally, the *A/E(Ptop)* indicator marks the relative contribution to the 5% most frequently cited papers, and is calculated as the ratio of the *actual* and *expected* presence in the top of the citation distribution. Here, a value above (below) 1.0 indicates a relatively high (low) contribution to the 5% most frequently cited papers.

Distribution of publications over impact classes

In this study we also focus on the distribution of individual publications over five impact classes. Citation distributions are skewed, and are a topic of bibliometric research in itself (SEGLEN, 1992; AKSNES et al., 2001). This skewness of citation distributions creates a need for insight in the distribution of impact scores by the users of bibliometric data. As impact indicators, we used a contraction of CWTS standard indicators *C* and *FCSm* (resulting in *C/FCSm*), and *JCSm/FCSm*, and applied these to each individual publication. Normally, the *CPP/FCSm* score indicates the relative impact of a whole unit in all the fields in which the unit published, here *C/FCSm* is related to each individual paper. *JCSm/FCSm* is indicative for the status of the journals in which a research unit has published its research output, in this study we applied it as indicative for the journal status of each individual paper. The five impact classes are defined as follows: Class 0 consists of publications with no impact at all (*C/FCSm*=0), Class 0-0.8 consists of papers with *C/FCSm* scores ranging from 0 to 0.8, Class 0.8-1.2 consists of papers with *C/FCSm* scores ranging from 0.8 to 1.2, Class 1.2-2.0 consists of publications with *C/FCSm* scores ranging from 1.2 to 2.0, and the upper Class >2.0 consists of papers with *C/FCSm* scores above 2. A likewise system was applied to the journal-to-field impact indicator *JCSm/FCSm*. Main purpose of the analysis is to gain insight in the distribution of impact scores over papers, and more in particular, in the top class(es). Therefore, in the presentation we have focused on both the upper class (*C/FCSm* > 2.0), and on the two upper classes combined (*C/FCSm* > 1.2).

Determining top-journals in science

Searching for scientific excellence brings the bibliometric researcher to the main vehicle of scientific communication in many fields of science: the scientific journal. Bibliometric indicators have long been used as measures to indicate the level of esteem from the scientific community for specific scientific journals, based on the assumption that a high number of citations stands for scientific quality. As such, searching for scientific quality, one expects scientists to publish in the top-journals available in their field of science. Journals that are often mentioned in this respect include the well-known multidisciplinary journals like *Nature*, *Science*, the *Proceedings of the National Academy of Sciences of the US*, and more specialized journals like *Cell* (in the field of biochemistry and molecular cell biology), and multidisciplinary medical journals like the *New England Journal of Medicine*, *The Lancet*, and the *Physical Review*. However, the perception of these journals being 'top journals' is partly created by their position in rankings based on, for example, Journal Impact Factors. In these rankings, these journals have high values, thereby further contributing to the perception of being top-journals, especially when compared with journals in fields in which such high impact factor values do not occur. However, if one applies a journal impact indicator that is based on field normalization, each field will have its 'top-journals' simply because the journals are first compared with their own environment, before comparing them across fields. Such a field-normalized journal impact indicator, previously developed at CWTS, is the Journal to Field Impact Score (*JFIS*) (VAN LEEUWEN et al., 2002). The Journal to Field Impact Score (*JFIS*) is based on four types of documents (articles, letters, notes and reviews (from 1996 onwards ISI categorised notes as articles)). Next, *JFIS* is field-normalized. This means that a journal's impact is compared the world citation average in the fields it covers. This field average is currently based on the classification of journals in so-called Journal Subject Categories as defined by ISI. Finally, the *JFIS* value of a journal is based on flexible and variable citation and publication windows.

A remark needs to be made about the conceptual difference between *JFIS* and *JCSm/FCSm*. While these two indicators are methodologically related, their major difference can be found in the fact that *JFIS* is based on only one journal, while *JCSm/FCSm* generally represents an oeuvre of a research unit, and is based on various journals.

By applying *JFIS* to the field of Netherlands academic chemistry research, we were able to indicate the top-journals in chemistry and related fields. However, this indicator alone is not sufficient to make a distinction among journals. Therefore, we had to define some additional criteria. Due to the fact that *JFIS* is based on field normalization, the value 1 is indicative of the 'mean level' in each field. Therefore, a certain deviation from 1 is indicative of a top journal in its field(s), and a *JFIS* value of 3 times this field-specific standard was chosen to select top-journals. This results in 2160 journal-document type combinations, about 1% of all available journal-document type combinations, representing all fields of sciences, ranging from molecular biology to mathematics and the engineering sciences. However, this alone offers the opportunity for journals with a small volume of papers and a relative short coverage in the database to be characterized as high impact or top-journals. Therefore, we formulated two additional criteria, stating that a journal should at least receive *10* citations per paper on average (excluding self-citations), and contain more than *20* publications (articles, letters, or reviews) per year. This left us with 225 journal-document type combinations, about 2% of all available journal-document type combinations.

Weighted impact analysis

One of the criticisms often heard from scientists in research evaluation assessments is that it makes a difference whether one's work is cited by well-known high impact journals as opposed to mainstream disciplinary journals in a field. Until recently, this rather complicated (both technically and methodologically) aspect of citation analysis could not be integrated in the research assessment procedure of CWTS.

In the current study, the first results of weighted impact analysis are presented, by applying the *JFIS* of the *citing* journal as the parameter determining the weight of each citation. *JFIS* takes into account the length of the citation window, and type of document published in a journal.

For each paper in the set of publications per university, we identified the citing publications. Next, we determined for each citing publication the year of publication, the journal in which the citing publication appeared, and the document type. We then assigned *JFIS* values to each citing publication on the basis of corresponding publication year, journal and document-type. Thus a weight was attributed to each citation, corresponding to the characteristics of each citing paper.

Results

In this section we focus on results from the various approaches. First we will present the results from the study on academic chemistry research in the Netherlands. Next we will present the outcome of the application of the three new impact measures.

Impact analysis based on average values: results

In this section we present the indicators that are part of the standard CWTS analysis. Table 1 shows these indicators for the ten universities, based on the analysis of the publication output in the period 1991–2000.

Focusing on the impact compared to the world sub-field average (our ‘crown-indicator’ *CPP/FCSm*) in 1991–2000, we see in Table 1 that the impact of each of the ten universities is significantly above world average. The impact scores vary between 1.20 and 1.92. A comparison with the average impact of their journal set (*CPP/JCSm*) shows that nine out of ten universities as a whole have an impact significantly above their journal packet. Furthermore, on average, all universities publish in relatively high impact journals (see *JCSm/FCSm*). Three universities even surpass the average international journal impact level with about 50%.

Table 1. Bibliometric statistics of Dutch academic chemistry research, by university, 1991–2000

University	*P*	*Cx*	*CPP*	% *Pnc*	*CPP*/ *JCSm*	*CPP*/ *FCSm*	*JCSm*/ *FCSm*	% Self-citations
KUN	1,328	17,708	13.33	18%	1.16 +	1.64 +	1.41	23%
LEI	1,789	16,188	9.05	21%	0.99	1.20 +	1.22	30%
RUG	1,836	24,604	13.40	19%	1.25 +	1.92 +	1.54	23%
TUD	2,475	17,003	6.87	31%	1.20 +	1.41 +	1.18	27%
TUE	1,762	13,404	7.61	30%	1.24 +	1.90 +	1.53	28%
UT	1,429	11,971	8.38	24%	1.41 +	1.88 +	1.33	28%
UU	2,949	29,328	9.95	22%	1.12 +	1.43 +	1.27	27%
UvA	2,474	25,279	10.22	20%	1.13 +	1.51 +	1.33	28%
VU	1,045	12,037	11.52	16%	1.18 +	1.76 +	1.49	27%
WUR	2,658	23,702	8.92	23%	1.15 +	1.29 +	1.13	28%

P	Number of papers
Cx	Number of external citations
CPP	Average citation rate per paper
% Pnc,	Percentage of papers not cited within the time-frame 1991-2000
CPP/JCSm	Average impact of papers compared to the journal citation average
CPP/FCSm	Average impact of papers compared to the subfield citation average
JCSm/FCSm	Average impact of journals compared to the subfield citation average
% Self-citations	Share of author self-citations among all received citations.

Frequently cited publications: results

Table 2 presents the results of an analysis of the most frequently cited papers (top 5%).

The column *A/E(Ptop)* shows that all universities equal or surpass the expected number of top 5% papers. Four universities have at least twice the number of expected top 5% publications, three others exceed the expected score with at least 67%. The University of Amsterdam (UvA) and State University Groningen (RUG) have the largest number of top 5% papers in absolute numbers, while the State University Groningen, the University of Technology Eindhoven (TUE), and the University of Twente (UT), have relatively the largest number of top 5% papers. This shows that Dutch chemistry papers are not only cited well above average (as *CPP/FCSm* indicated in Table 1) but also occur much more often than expected among the 5% most frequently cited papers in their subfields.

Table 2. Overview of frequently cited publications (top-5%) by university, 1991–1997

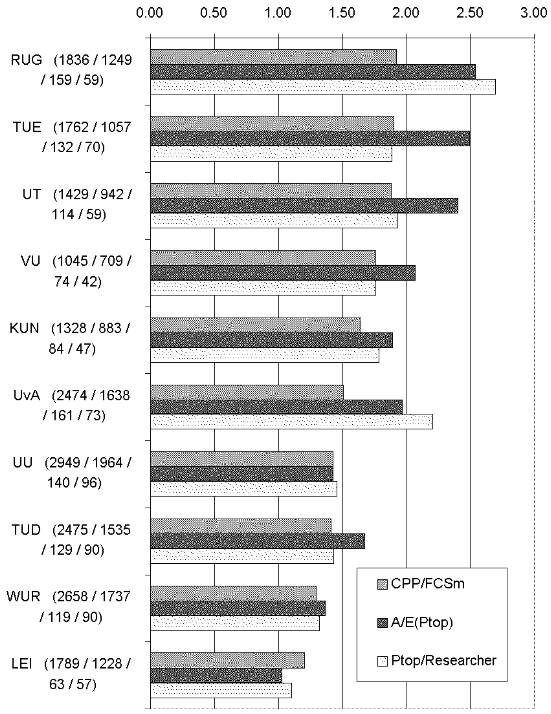
University	*P 91-97*	*Ptop*	*E(Ptop)*	*A/E(Ptop)*
KUN	883	84	44.4	1.89
LEI	1,228	63	61.3	1.03
RUG	1,249	159	62.6	2.54
TUD	1,535	129	77.0	1.67
TUE	1,057	132	53.0	2.49
UT	942	114	47.5	2.40
UU	1,964	140	98.1	1.43
UvA	1,638	161	81.8	1.97
VU	709	74	35.8	2.07
WUR	1,737	119	87.0	1.37
NL-CHEM	11,921	1,062	597.5	1.78

P91-97	Number of review articles and normal articles in 1991–1997
Ptop	Absolute number of papers among the top 5% most frequently cited articles (the top 5%) published in a particular year and subfield
E(Ptop)	Expected number of papers among the top 5%
A/E(Ptop)	Ratio of the actual and expected number of papers among the top 5%

The numbers in brackets in Figure 1 have the following meaning: the first number is the number of articles published during the period 1991-2000, and the second the number of articles published during the period 1991-1997. The third number is the number of top-publications, and the fourth is the number of senior scientists appointed at a university at 31-12-2000 and included in our analysis. The ranking of the universities in Figure 1 is based on a descending value of *CPP/FCSm.*

In Figure 1, the mean impact scores (*CPP/FCSm*) of the universities are compared to their ratio of actual and expected number of papers among the 5 percent most frequently cited articles published worldwide in the various subfields (*A/E Top*). A third parameter in this figure consists of the mean number of highly-cited publications per senior scientist appointed at 31-12-2000 at a university and included in our analysis (*Ptop/Researcher*).

In Table 1 we observed that the mean impact scores of all universities involved were well above worldwide citation average. Figure 1 shows that a strong correlation exists between mean impact scores and the ratio of actual and expected number of highly cited publications (*A/E(top))*: the universities ranking one to four on *CPP/FCSm* occupy the same positions in the ranking based on *A/E Top* (RUG, TUE and the UT, and the VU Amsterdam, respectively). Only at the positions five and six (the KUN and UvA Amsterdam), and seven and eight (UU and TUD), we observe changes in the ranking based on *A/E top*, compared to that based on *CPP/FCSm*. However, if we take into account the mean number of highly cited papers per senior scientist, we find a striking change in the ranking, namely the shift of the UvA from position 6 to 2 (having the second largest number of highly cited publications per senior researcher). No major changes in position are observed for the two universities in the lower end of the figure (WUR and LEI).

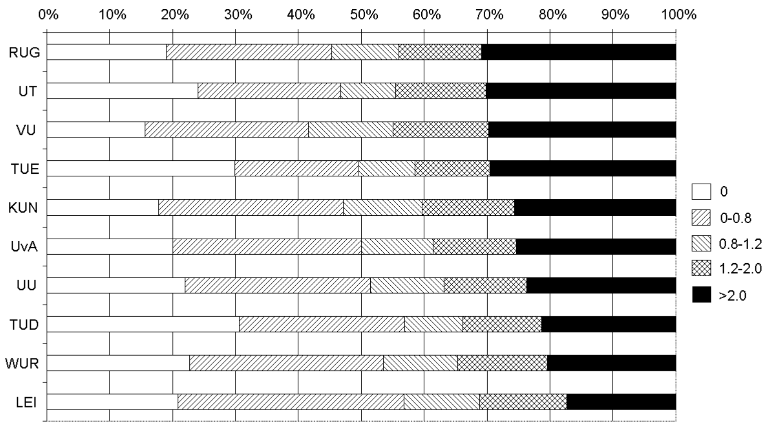


CPP/FCSm	Average impact of papers compared to the subfield citation average
A/E(Ptop)	Ratio of the actual and expected number of papers among the top 5%
Ptop/researcher	Number of papers among top 5% per researcher

Figure 1. Mean impact and top impact compared, 1991–1997/2000. Universities

Distribution of publications over impact classes: results

Figure 2 displays the distribution of research papers of academic chemistry research in the Netherlands over classes of impact scores. In this analysis, the applied impact score is *C/FCSm*, the comparison of the actual impact of each university in the study with the mean field average in which these universities have published their results.



C/FCSm	The relative impact of a journal paper compared to all the fields to which the journal belongs.
Classes of impact scores:	
Class 0	Output share with no impact at all (*C/FCSm*=0)
Class 0-0.8	Output share with *C/FCSm* scores ranging from 0 to 0.8
Class 0.8-1.2	Output share with *C/FCSm* scores ranging from 0.8 to 1.2
Class 1.2-2.0	Output share with *C/FCSm* scores ranging from 1.2 to 2.0
Class >2.0	Output share with *C/FCSm* scores above 2.

Figure 2. Distribution of publications of Dutch academic chemistry research over classes of impact scores (*C/FCSm*), ranked by highest class

While Figure 2 is ranked on the basis of the relative share in the highest class of impact scores (*C/FCSm* > 2.0), Table 3 displays the ranking based on the *two* highest classes of impact scores (*C/FCSm* > 1.2). This offers a more thorough insight in the distribution of output over impact classes, thereby contributing to the robustness of the chosen approach.

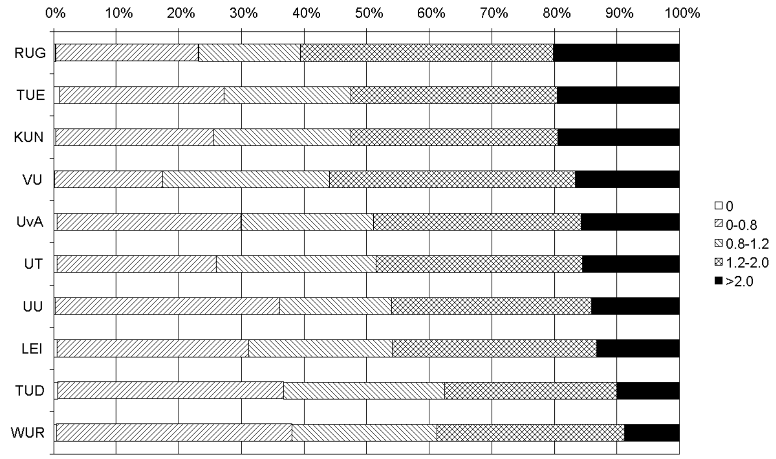
Table 3. Comparison of rankings, based on output shares and related impact scores, 1991–2000

	Ranked by Upper Class (*C/FCSm*> 2.0)	Ranked by upper 2 Classes (*C/FCSm* > 1.2)
RUG	1	3
UT	2	2
VU	3	1
TUE	4	4
KUN	5	5
UvA	6	6
UU	7	7
TUD	8	9
WUR	9	8
LEI	10	10

While in Figure 2 and Table 3 the RUG is ranked as number 1 university, with 30.9% of their output in the highest impact class, followed by the UT (30.2%) and the VU Amsterdam (29.8%), the comparison of rankings displayed in Table 3 shows that the VU Amsterdam has a relatively larger share of its output in the impact classes ranging from 1.2 upwards (45%), followed by UT (44.5%) and the RUG (44.1%).

Of the other seven universities, only the TUD and the WUR change positions, when comparing both rankings. Overall, when comparing both rankings, the changes in rank positions are relatively small, and based on relatively small differences in shares of output per class of impact scores. Another remarkable aspect of the comparison between both rankings is that the highest share of papers with no impact at all (*C/FCSm*=0, or the percentage of papers uncited) is found at the universities of technology in the Netherlands, even more remarkable is the fact that two of these universities are amongst the highest ranking universities (UT and TUE), while the TUD has both a large share of uncited publications and a low share of highly cited publications.

Figure 3 displays for each university the distribution of publications over classes of journal impact scores compared to the field(s) the journals belong. This *JCSm/FCSm* indicates the status of the journals in which the universities have published, compared to the field in which they published. In this analysis, the number of classes is in fact reduced to four classes due to the fact that Dutch chemists hardly publish in journals with no impact at all (*JCSm/FCSm* = 0).



JCSm/FCSm	The status of the journal in which the papers of a university were published compared to their fields. In this analysis we applied it as indicative for the journal status for each paper.

Classes of impact scores:

Class 0	Output share with no impact at all (*JCSm/FCSm*=0)
Class 0-0.8	Output share with *JCSm/FCSm* scores ranging from 0 to 0.8
Class 0.8-1.2	Output share with *JCSm/FCSm* scores ranging from 0.8 to 1.2
Class 1.2-2.0	Output share with *JCSm/FCSm* scores ranging from 1.2 to 2.0
Class >2.0	Output share with *JCSm/FCSm* scores above 2.

Figure 3. Distribution of publications of Dutch academic chemistry research over classes of journal impact scores (*JCSm/FCSm*), ranked by highest class

The university with the largest share of its output in both the highest journal impact class as well as in the two highest classes RUG (with respectively 20.2% and 60.6% of its output in these classes). When focusing on the highest class of journal to field impact scores, we find the TUE, KUN, and VU Amsterdam following Groningen in the top of the ranking. When the two upper impact classes are observed, the VU Amsterdam is ranking second, with TUE and the KUN as third and fourth.

Among the other universities we do not observe large differences in ranking, only the TUD and WUR change in position when we extend the analysis from the upper journal impact class to the two upper classes.

Both analyses indicate that the changes in rank position occur in the upper part of the ranking. However, the differences in shares of highly cited papers or papers in high impact journals between the universities in the top of the rankings are relatively small.

Determining top-journals in science: results

Of a total of 1,394 journals used by Dutch academics chemistry researchers, 47 qualify as top-journals. These 47 journals contained 697 papers by Dutch chemistry researchers in the period 1991-2000 (3.8% of the total volume of publications published in that period).

The ten top journals most frequently used by the chemistry researchers are displayed in Table 4. The *Journal of the American Chemical Society*, a multidisciplinary chemistry journal covers one third of all the papers in top-journals. The second ranking top-journal is *Physical Review Letters*, which is not surprising given the fact that physical chemistry is the largest discipline in Dutch academic chemistry research (12% of the output is published in journals belonging to Chemistry, Physical).

Table 4. Top journals used by chemistry researchers in the Netherlands

	P	% P
J AM CHEM SOC	236	33.9
PHYS REV LETT	83	11.9
ANGEW CHEM INT ED	57	8.2
NATURE	54	7.7
SCIENCE	36	5.2
ANAL CHEM	30	4.3
ADV MATER	29	4.2
EMBO J	29	4.2
PLANT J	16	2.3
CELL	14	2.0

In Table 5, the presence of Dutch academic chemistry research in the (ten most used) top-journals is displayed.

All universities in this study except WUR have their largest volume of papers in top-journals in the *Journal of the American Chemical Society*, while WUR is strongest in *Plant Journal*. RUG is strongly present in *Journal of the American Chemical Society*, *Physical Review Letters*, *Nature* and *Science*.

Table 5. Presence of Dutch academic chemistry research in top journals, 1991–2000

	KUN	LEI	RUG	TUD	TUE	UT	UU	UvA	VU	WUR
J AM CHEM SOC	23	25	26	13	31	25	32	39	17	5
PHYS REV LETT	5	4	23	2	7	1	17	11	12	1
ANGEW CHEM INT ED	6	3	9	4	11	7	8	6	3	
NATURE	3	2	10	10	9	2	5	9	1	3
SCIENCE	4	2	12	1	3	1	3	8		2
ANAL CHEM	6			3	2	1	1	9	7	1
ADV MATER			10	3	8	3	4	1		
EMBO J	3	5	3			1	8	6	2	1
PLANT J								2	1	13
CELL	7		1				1	3	2	
Sum of the 10 most-used top-journals	57	41	94	36	71	41	79	94	45	26
Sum all top-Journals	74	43	102	51	79	47	102	110	53	36

Table 6. Output of Dutch academic chemistry research in top-journals, 1991–2000

University	Output in top-journals	Percentage	Rank
KUN	74	5.6	1
LEI	43	2.4	8
RUG	102	5.6	2
TUD	51	2.1	9
TUE	79	4.5	4
UT	47	3.3	7
UU	102	3.5	6
UvA	110	4.4	5
VU	53	5.1	3
WUR	36	1.4	10

A comparison of the presence of universities in the top-journals and their total number of papers is displayed in Table 6. This further underlines the capability of a university to have its research published in the most prestigious high-impact journals in the field of chemistry and related areas. The last column contains the rank position, calculated on the basis of the share of output in the top-journals, compared to their total output. Remarkably, KUN is the top-ranking university in this analysis, closely followed by RUG. Universities with a higher share in the top-journals than the 3.8% national average are: KUN, RUG, TUE and both the universities from Amsterdam (UvA and VU). Much lower shares are observed for LEI, TUD and WUR.

Weighted impact analysis: results

As described above, the weighting of impact scores requires the attachment of a journal impact indicator to each individual citation. Two problems arose while attaching *JFIS* values to each citing paper. The first problem was that some citations to Dutch academic chemistry research originate from document types for which no *JFIS* values are calculated (such as editorials or meeting abstracts). This is due to the fact that these document types are too small in number to calculate reliable *JFIS* values, or they are published on a very irregular basis. The second problem relates to citations received by Dutch academic chemistry research that originates from journals for which no ISI Journal Subject Category is available. This results in a missing *JFIS* value, as *JFIS* is based on field-normalization. Together, these two problems lowered the total number of citations from 175,446 to 173,240, a difference of 1.2%, while for universities this difference ranged from 0.7% to 1.3%.

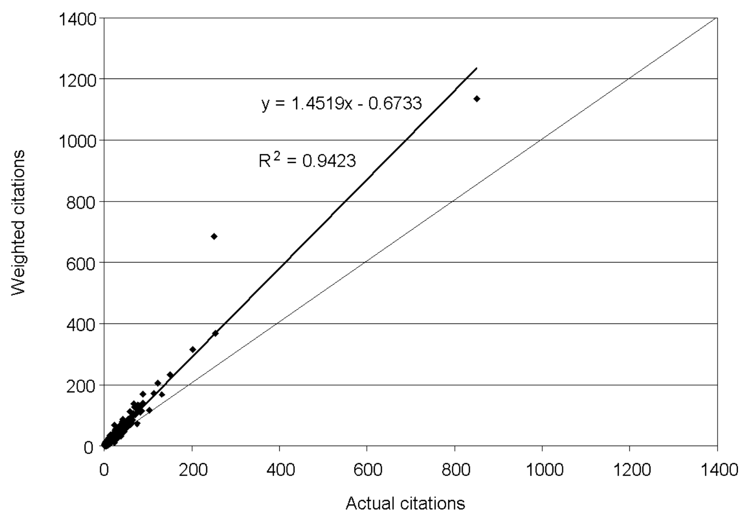


Figure 4. Distribution of the weighted versus actual citation scores for Leiden University chemistry research, 1991-2000

Figure 4 displays the distribution of *JFIS* values and actual citations scores per publication, plotted for all publications of Leiden University chemistry groups within the national research assessment. The diagonal indicates the score where the sum of *JFIS* values is equal to the sum of unweighted citations per publication. If the *JFIS* values of journals citing Leiden University research output in chemistry would be low, the scatter of papers would be positioned below the diagonal, while a high sum of *JFIS* values of papers citing Leiden University chemistry research leads to a positioning of the scatter of papers above the diagonal. As Figure 4 indicates, the sum of *JFIS* of papers citing Leiden University chemistry research was high, as the papers are positioned above the diagonal.

The comparison of the not-weighted and *JFIS* weighted impact of papers citing Leiden University chemistry research output answers the question: is Leiden University chemistry research cited by the high(er) impact journals ?

Table 7. Weighted impact analysis per university, 1991–2000

University	Number of citations	Citations * *JFIS*	Weighted Impact	Rank	Missing (%)
KUN	17,543	25,831	1.47	2	0.8
LEI	16,028	22,315	1.39	6	0.8
RUG	24,288	34,190	1.41	4	1.4
TUD	16,807	21,244	1.26	9	0.8
TUE	13,275	20,235	1.52	1	0.7
UT	11,834	16,240	1.37	7	1.0
UU	29,046	40,892	1.41	5	1.1
UvA	24,964	35,728	1.43	3	1.2
VU	11,954	16,316	1.36	8	0.3
WUR	23,493	27,211	1.16	10	1.0

In Table 7 the results of the weighted impact analysis is displayed. In the second column we find the number of citations received by papers from each of the universities. The third column contains the number of citations multiplied with the *JFIS* value of each paper citing the papers from a particular university. The fourth column contains the ratio of the two previous columns, indicative for the scientific status of the journals in which Dutch academic research per university is cited. The fifth column contains the rank of each university on this indicator, and finally the last column indicates the percentage of missing citations described above, which is caused by the discrepancy in received citations and the availability of *JFIS* values for each citing paper.

On this indicator, Eindhoven has the number one ranking position, followed by Nijmegen, Amsterdam (UvA), and Groningen. A remarkable aspect of this analysis is the small range of values obtained by the various universities: while we observed larger differences in bibliometric scores on previous – unweighted – indicators, this weighted indicator shows a more 'dense' distribution of the ten universitiesand also, on average, a lower impact level in this weighted impact analysis.

Synthesis

The previous sections have shown applications of bibliometric indicators already used and tested in earlier studies, as well as novel and still experimental ones. For purposes of clarity, the rankings on the various indicators are combined in Table 8. This synthesizes all previous sections, thereby showing the rank position of each university on the various indicators. A disadvantage of this combination, based on rank positions, is the loss of insight in the (sometimes small) differences between universities on some of the indicators, on which the rankings are based.

Table 8. Overview of rank positions per university on bibliometric scores

	KUN	LEI	RUG	TUD	TUE	UT	UU	UvA	VU	WUR
P 91-2000	9	6	5	3	7	8	1	4	10	2
Number Staff	9	8	6	2	5	7	1	4	10	3
C	5	7	3	6	8	10	1	2	9	4
CPP	2	6	1	10	9	8	5	4	3	7
% Cited	2	5	3	10	9	8	6	4	1	7
CPP/FCSm	5	10	1	8	2	3	7	6	4	9
JCSm/FCSm	4	8	1	9	2	6	7	5	3	10
Ptop 91-97	8	10	2	5	4	7	3	1	9	6
A/E Ptop 91-97	6	10	1	7	2	3	8	5	4	9
% P (C/FCSm>2)	5	10	1	8	4	2	7	6	3	9
% P (C/FCSm>1.2)	5	10	3	9	4	2	7	6	1	8
% P (JCSm/FCSm>2)	3	8	1	9	2	6	7	5	4	10
% P (JCSm/FCSm>1.2)	4	8	1	10	3	6	7	5	2	9
P in top-Journals	1	8	2	9	4	7	6	5	3	10
Weighted Impact	2	6	4	9	1	7	5	3	8	10

Rankings are arranged by descending order for each variable (highest value ===>>> rank=1)
See the Annex for a short explanation of the indicators

In Table 9, rankings of universities on the various indicators are compared on the basis of Spearman rank correlations for eleven indicators. By comparing the rank positions on various indicators, the reliability of combinations of indicators in research performance assessment studies becomes visible. So far, the study has shown that various indicators provide different rankings of universities. This poses the question whether application in research performance assessments of a single indicator provides sufficiently valid results. The answer to this question might be twofold, namely in the first place by using multiple indicators instead of only one, and secondly by investigating which combination of indicators provides the strongest correlations, thereby indicating the best combinations of indicators in research performance procedures.

Table 9. Spearman rank correlations: rank correlations calculated for scores on bibliometric indicators for academic chemistry research in the Netherlands

		1	2	3	4	5	6	7	8	9	10	11
P 1991-2000	(1)	*										
C	(2)	0.73	*									
CPP	(3)	-0.26	0.39	*								
CPP/JCSm	(4)	-0.39	-0.49	-0.15	*							
CPP/FCSm	(5)	-0.50	-0.24	-0.32	0.78	*						
JCSm/FCSm	(6)	-0.58	-0.12	-0.48	0.48	0.89	*					
Ptop/Researcher	(7)	-0.24	0.07	0.35	0.59	0.85	0.75	*				
A/E(Ptop 91-97)	(8)	-0.47	-0.26	-0.27	0.82	0.98	0.85	0.89	*			
% P (C/FCSm>2)	(9)	-0.55	-0.27	0.38	0.78	0.96	0.83	0.84	0.94	*		
P in Top-Journals	(10)	-0.59	0.04	0.73	0.26	0.72	0.90	0.59	0.64	0.93	*	
Weighted Impact	(11)	-0.26	0.20	0.33	0.03	0.50	0.68	0.53	0.44	0.35	0.72	*

See the Annex for a short explanation of the indicators

From the analyses, and the rankings based on the indicators displayed in Table 8, one might easily draw the conclusion that Groningen (RUG) and Eindhoven (TUE) are the best performing universities in the field of chemistry research. However, the situation becomes more complex if one tries to determine which university is their runner-up. For example, Utrecht (UU) has two rank 1 positions and a couple of lower rank positions, but their number 1 ranking position is on the indicators expressing absolute output and citation numbers and number of staff involved, while for example Amsterdam (VU) and the Twente (UT) have a couple of rank 3 positions on indicators expressing relative and / or normalized scores. So how can one determine which indicators are the most relevant with regard to the scope of a research performance assessment, in which aspects like quality, viability, relevance, and continuity play a central role? In order to prevent one indicator to be chosen as the only indicative means to express (some of) these aspects, we need to stress the variety and expressiveness of a combination of indicators. Therefore, we need to test the (combinations of) indicators, in order to determine which ones do fit the goals of the bibliometric part of a research performance assessment.

In Table 9, rank correlations for the rank positions of the ten universities in our study of Dutch academic chemistry research, are calculated for eleven indicators. Some of the indicators are mutually inter-dependent, therefore a somewhat high(er) correlation can be expected (e.g., between *CPP/JCSm*, *CPP/FCSm*, and *JCSm/FCSm*). However, between other indicators we observe a conceptual distance, e.g., the number of papers published (*P*) and the ratio of actual versus expected publications in the top 5% of the field (*A/E (Ptop 91-97)*).

Conclusions and discussion

In this study we have explored some new bibliometric indicators that could be helpful in distinguishing research of 'high quality', or 'strong international visibility'. These indicators should be complementary to the instruments used in research performance assessment procedures that already exist (for a broader discussion on this topic, see TIJSSEN, to be published).

In this study we have presented various types of indicators, such as absolute and relative indicators, input, output and productivity indicators, output versus impact indicators, based on citation averages as well as based on citation distributions.

The first new indicator is the analysis of the distribution of papers over impact classes (either *C/FCSm* or *JCSm/FCSm*). Based on aggregates of citation scores per individual paper, expressing the distribution of these impact scores of classes of impact compared to the total output, this approach combines the best of both worlds. Both impact indicators (*C/FCSm* and *JCSm/FCSm*) may provide an additional insight when analyzed in this way. In this study it had the advantage that it directs the user to other aspects of the citation distribution, namely the diversity of the character of universities in a research performance assessment study (e.g., the position of the three Dutch universities of technology (TUE, TUD, and UT) and the agricultural university (WUR). The expansion of the analysis from the highest impact class, to the two highest impact classes hardly provides additional information.

The second new indicator focuses on the capability to enter the most prestigious and prolific journals. As a result of this analysis, we observe some advantages. First, a method to determine the impact level of journal in a field, which creates a more objective definition on the top-journals in science as compared to the more subjective labels, in a way journals like *Nature* and *Science* are generally considered. Secondly, by focusing on the presence of researchers or research groups in these top-journals, without analyzing their actual impact in the papers published in these journals, scientific status and visibility become focal points in the analysis: being present and visible in these journals generally means passing the often strict rules of the referee process. A disadvantage of this approach might be that it is less beneficial for young researchers of research groups, while it is highly likely that scientists with a certain tenure track have less problems entering the journal and the referee process, based on the fact that due to their previous work, these more experienced scientists are well-known to both the editorial board's policy, and the referee process involved. A disadvantage on a more technical level deals with the fact that not all fields (as defined by ISI Journal Subject Categories) contain to the same extent journals with high impact scores (*JFIS* > 3). This problem might be solved by selecting the journals not on the basis of an absolute score (like *JFIS* > 3), but on the distribution of *JFIS* values within a field, thereby allowing also journals with a relative low (absolute) impact score, but a high relative impact score within their field, being qualified as top-journals.

The third new indicator focuses on the aspect of the origin of the received citations. This aspect of citation analysis has been recently introduced in the research program of CWTS, and resulted in citing research profiles, in which both the cognitive as well as the geographical origin of received citations were analyzed (VAN RAAN& VAN LEEUWEN, 2002). In this particular case, the impact value of the citing journals forms the basis of the analysis. This new approach provides information on the aspect of knowledge transfer in science, and more in particular whether or not the impact is collected from high impact journals. The underlying is based on the hypothesis that the high impact journals are the most prolific and influential journals, containing articles that deal with topics of the research front of the field to which a journal belongs. In this respect, one might state that the analysis in which the impact score of the citing journal is integrated in citation analysis, contributes to the more qualitative aspects of research performance assessment. In the analysis, we observed a narrowing down of the range across universities of weighted impact scores, a phenomenon that needs further research, because until now we do not know exactly whether this is characteristic for the method itself, or that it is caused by the fact that Dutch academic chemistry has a rather high impact in general, is published in high impact journals, and is cited by high impact journals (as can be seen by the correlation of 0.68 on both indicators).

An aspect that needs to be mentioned here and was referred to in the previous lines, is the fact whether or not the overall (high impact) position of Dutch academic chemistry influences the outcomes of the study, and in particular the comparison of the rankings between universities. This needs to be explored in more detail in future studies.

The diversity of the explored and tested indicators suggests that one or two indicators are probably not sufficient to provide reliable information that fits the multiple goals of research performance assessment procedures, aimed at gaining insight in various aspects of the research performance of researchers, research groups, institutions, or even countries.

An important aspect that needs mentioning here as well is the fact that comparisons have been made on rank-positions. This approach results in a clear distance between universities ranging from one to ten. However, in a number of cases, the distances are very small, and this should be taken into consideration in the analysis as well. By expressing the differences between universities in a subtle, relative way (different from a relative scale of one to ten), the rank positions might become equal. This can cause the Spearman rank correlations to become invalid, due to the fact that is supposes a certain number of rank positions (where ten observations in a rank correlation analysis is becoming somewhat critical).

References

AKSNES, D. W., G. SIVERTSEN (2001), The effect of highly cited papers on national citation indicators, *Proceedings of the 8th International Conference on Scientometrics & Informetrics*, Sydney, Australia, pp. 23–30.

MOED, H. F., R. E. DE BRUIN, TH. N. VAN LEEUWEN (1995), New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications, *Scientometrics*, 33 : 381–422.

SCHUBERT, A, W. GLÄNZEL, T. BRAUN (1989), Scientometric datafiles – A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfield, 1981-1985, *Scientometrics*, 16 :3–478.

SEGLEN, P. O. (1992), The skewness of science, *Journal of the American Society of Information Sciences*, 43 : 628–638.

TIJSSEN, R. J. W., M. S. VISSER, TH. N. VAN LEEUWEN (2002), Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics*, 54 : 381–397.

TIJSSEN, R .J. W. (to be published) Organizational scoreboards of research excellence.

VAN LEEUWEN, TH. N., H. F. MOED (2002), Development and application of journal impact measures in the Dutch science system, *Scientometrics*, 53 :249–266.

VAN RAAN, A. F. J. (1996), Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises, *Scientometrics*, 36 : 397-420

VAN RAAN, A. F. J., TH. N. VAN LEEUWEN (2002), Assessment of the scientific basis of interdisciplinary, applied research: Application of bibliometric methods in nutrition and food research, *Research Policy*, 31 : 611–632.

Annex: An overview of the indicators used in this paper

P 91-2000	Numbers of papers in the period 1991-2000.
Number staff	Numbers of senior staff researchers involved (per university).
Cx	Number of external citations.
CPP	Average citation rate per paper.
% Pnc,	Percentage papers not cited within the time-frame 1991-2000.
CPP/JCSm	Average impact of papers compared to the journal citation average.
CPP/FCSm	Average impact of papers compared to the subfield citation average.
JCSm/FCSm	Average impact of journals compared to the subfield citation average.
% Self-citations	Share of author self-citations among all received citations.
P 91-97	Numbers of papers in the period 1991-1997.
A/E(Ptop 91-97)	Ratio actual versus expected number of top-papers.
Ptop/Researcher	Number of top-papers per researcher.
C/FCSm	Relative impact of a journal paper compared to the field(s) to which the journal belongs.
JFIS	Journal to field impact score, for each journal/publication year/document type.
P in top journals	Output per university in top-journals.
Weighted impact	Impact multiplied by the normalized field impact score of the citing journal.