Handbook of Quantitative Science and Technology Research

The Use of Publication and Patent Statistics in Studies of S&T Systems

edited by

Henk F. Moed

Centre for Science and Technology Studies,
University of Leiden, The Netherlands

Wolfgang Glänzel

Steunpunt O&O Statistieken,
K.U. Leuven, Belgium

and

Ulrich Schmoch

Fraunhofer Institute for Systems and Innovation Research,
Karlsruhe, Germany

KLUWER ACADEMIC PUBLISHERS
DORDRECHT / BOSTON / LONDON

Library of Congress Cataloging-in-Publication Data

1. Van Leeuwen
2. Butler

ISBN 1-4020-2702-8 (HB)
ISBN 1-4020-2755-9 (e-book)

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Sold and distributed in North, Central and South America
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

Printed on acid-free paper

Printed in the Netherlands.

Chapter 16

DESCRIPTIVE VERSUS EVALUATIVE BIBLIOMETRICS

Monitoring and Assessing of National R&D Systems

Thed van Leeuwen

Centre for Science and Technology Studies (CWTS), Leiden University, the Netherlands
E-mail: leeuwen@cwts.leidenuniv.nl

Abstract: This paper covers the differences between two separate bibliometric approaches, labelled 'descriptive' versus 'evaluative', or top down versus bottom up. The most important difference between these two approaches is found in the level of validity of the underlying research output. Whilst the publications in a top down approach, having a descriptive character, are collected by following general characteristics of these publications (such as country names, or fields), the consequence is that findings from such studies have a 'meaning' that is limited with respect to actual research assessment. On the other hand, in a bottom up approach the publications are collected from individual oeuvres of scientists, including a process of verification by the researchers involved. This procedure contributes significantly to the validity of the publication material, and consequently research assessment procedures can be based on the results of this type of bibliometric analyses. A strong focus in the paper will be on the actual application of bibliometric analysis within research assessment procedures, in particular within the UK and the Netherlands.

1. INTRODUCTION

The last decades have shown a steady increase in the growth of assessments of science, at different levels in the science system, and by different actors within the science system. Whilst peer review was the long existing practice for assessing science and its actors, the last decade has shown a strong increase of the application of quantitative techniques in the 'assessment' of science. As a result of the development of the field of quantitative studies of science and technology, and the growing awareness in the western world of the necessity of the science system periodically to provide some sort of accountability of science and scientists to society, we observed a development in which modern day science is both monitored as well as assessed on a regular basis. In a number of countries (such as France and the Netherlands) national facilities were created to monitor the national science system, in an international perspective, whilst in a number of countries actual assessments at a national level of the science system was initiated (e.g., the UK, and the Netherlands). In this paper, we will discuss the various approaches in the monitoring and assessing on both the international and national level, as well the differences between actual assessments within the UK and the Netherlands (an excellent overview in a more general sense has recently been given by Geuna and Martin, 2003). The difference between both methods evolves around the distinction between descriptive and evaluative bibliometrics.

2. DESCRIPTIVE AND EVALUATIVE BIBLIOMETRICS

Bibliometrics is the field of science that deals with the development and application of quantitative measures and indicators for sciences and technology, based on bibliographic information. This bibliographic information is the representation of codified knowledge as can be found in a diversity of scientific output types, such as serial literature, books, and book chapters, conference proceedings, patents, etc. And although the use of serial literature is not evenly distributed over fields of science (for instance, the dominant use of conference literature within some fields of the technical sciences), bibliometric studies start from the assumption that the most important findings of scientific research finally end up in the international serial literature. This, however, means that, in general, bibliometrics is less applicable in those fields of science in which the internationally oriented scientific journal is not the main medium for communicating research findings to the (international) community in those fields. Next to this application in the analysis and assessment of the development of science, bibliometrics focuses on the development of technology as well. Here patents are the main source of analysis.

Evaluative bibliometrics, as introduced by Narin (1976), is the application of bibliometrics which focuses particularly on the evaluation of scientific activity, and more, in particular, on quality aspects of scientific performance. In general, evaluation in itself is focused on the control of quality, so that, more specifically, research evaluation is focusing on the safeguarding of scientific quality. Scientific quality is a rather diverse concept, and a synonym for several different meanings. From a bibliometricians point of view, scientific quality is the synonym for scientific merit (Moed, 1989), representing scientific influence, particularly (international) scientific visibility. Quantitatively, it is based on both scientific output and impact measurement. The circumstance that distinguishes evaluative bibliometrics from descriptive bibliometrics is the degree of validity and reliability of the publication data, underlying a bibliometric analysis. Whilst a focus on a whole country or even a whole university normally uses a so called top down approach, in which all output is collected using the address information of the publications in the ISI citation indexes, the sight on smaller, lower level organisational structures remains clouded. Therefore this approach only offers insight on rather high levels of aggregation, and does not allow for any detailed conclusions within a country or organisation. On the other hand, in a so called bottom up approach in which specific target groups are assessed, these groups should be asked to co-operate in collecting publication data, thereby seriously and significantly contributing to the validity and reliability of the resulting output and impact scores.

An important issue with respect to the distinction between evaluative and descriptive relates to the interpretation of bibliometric results from both approaches. Whilst it is clear to both the bibliometrician, as well as to the direct user of the results of a top down or descriptive study, to what extent conclusions can be drawn from the results of such a study, other users who are not very well versed in the ins and outs of bibliometrics in general or that specific study might overlook these limitations. Comparisons made at higher levels of aggregation still do not allow for conclusions at lower levels of aggregation. Therefore comparisons made in top down studies do not have the same meaning as the results from a bottom up study.

3. MONITORING SCIENCE THOUGH MACRO-BIBLIOMETRICS

In recent years bibliometrics in its broadest sense has become more and more important at the national or supra-national level, as can be concluded from the studies initiated by government agencies in a number of countries and the European Commission. Next to the long time existence of the National Science Foundation (NSF) in the USA, with its periodical reports containing analyses of the American science system, the early 1990s in Europe was a period in which national facilities were created to monitor the national science system. In France the national Observatory of Science and Technology (OST, Observatoire de Science et de Technologie) was founded in Paris in 1990, while in the Netherlands a similar, but more virtual, observatory of science and technology developments was created in 1992 (NOWT, which stands for the Dutch Observatory of Science and Technology). Next to bibliometric indicators of scientific and technological performance, both organisations also provided in their periodic reports other indicators of the science system they were describing, e.g., data on the financing of the national science system, knowledge workers, graduation figures, etc. The reports of both observatories have a strong focus on benchmarking their national system with other countries, and present analyses of regions within a country and across Europe, of societal sectors, and of actors (at the level of organisations) within the national science system. The existence of institutionalised (France) or 'virtual' (the Netherlands) observatories of the development of science and technology clearly indicates that bibliometrics can and does contribute to science policy. This development in France and the Netherlands was followed by the appearance of the publication by the European Commission of a series of European Science and Technology Indicators Reports, of which the third edition appeared in 2003. While the OST report was written in French and the NOWT report in Dutch, the European Report, with its much broader scope (in terms of the geographic area covered), is an English language report, and as such it is more accessible than its French or Dutch counterparts.

Owing to the interest of these agencies, bibliometric studies often have a comparative character, and compare countries, geographical regions, and, in the last couple of years, even universities. As these studies start from a macro level of aggregation, based on the names of countries, cities, postal codes, and universities from the address information attached to publications in the publications retrieved from the ISI databases, their outcomes often lack accuracy. Therefore these studies can only remain comparative and descriptive.

The type of bibliometric description supplied by this kind of study provides insight on a geographical level as well as a cognitive level, that is, at the level of major fields or disciplines of science. Whilst the geographical level is 'distilled' from the address information available in scientific publications, the level of research fields or disciplines is made available by linking the journals in which the publications occurred to research fields and disciplines. The data available for this enrichment is found in the Journal Subject Categories supplied by ISI for journals covered in their databases. Although this journal classification system is far from perfect, and subject to debate within the bibliometric community, it is currently the only system available for bibliometricians, which fits the multidisciplinary character of the ISI citation indexes best.

From the above we see that by linking journal publications to either (aggregated) address information or fields of science or disciplines the macro-level becomes two-dimensional: countries, regions or organisations versus fields or disciplines. However, one still needs to take care with respect to the 'reach' of these analyses. Whilst it is previously argued that some areas of science are far less well represented by the publications covered in the ISI databases (see for example Hicks (1999)), the problem of sufficient representation stretches to the bibliometric analysis at a macro-level as well. In general we find less international journal publication output for scholars in the arts and humanities, and to a lesser extent for those working in some fields of the social and behavioural sciences. Next, we find a somewhat stronger presence in these international ISI–covered journals by scientists from the Anglo Saxon world. For quite a large group of countries, publishing in English in the arts and humanities is less obvious. For instance, results of the research into Spanish culture and history and its influence on Latin America, research focusing on Italian history and archaeology, German literature and French philosophy, to mention only a few areas, appear in journals containing articles written in these respective languages. Only some of these non-English language journals are actually covered in the ISI citations indexes. This does not mean that these research fields are not internationally oriented, but rather reflect the situation that in some areas of scholarship, English is not necessarily the *lingua franca* of current day developments. Contrary to, for example, the Dutch and Scandinavian language areas, the Spanish, French, German end Italian language areas are large enough to allow scientific publishers to operate in a market that is economically sound enough to publish journals in these languages. Remarkably enough, the issue of the language of publication is not limited to the scholars working in the arts and humanities. A study for the German government showed the effect of language of publication in the fields of the medical sciences (Tijssen, van Leeuwen, and van Raan, 2002; van Leeuwen et al., 2000; and van Leeuwen et al., 2001). The discussion of publication language affected heavily the discussion on the value of journal impact measures within the German language scientific arena (Herfarth and Schurmann, 1996; Haller, Hepp and Reinhold, 1997; Rempen, 1998; Beller, 1999; and Kindermann, 1999).

As mentioned above, the macro level approach stretches out to the level of organisations. However, this is not an easy task. Whilst it often seems pretty obvious from the addresses attached to the publications in the ISI databases, which institutes or organisations are meant, many organisational links underlying the published address, indicating the actual structure of an university (e.g., the affiliated academic hospitals, related research institutes, etc) are not clearly visible, and much efforts should be put into cleaning and unifying address information on scientific publications. On the level of macro bibliometrics, simply not enough noise is filtered out to allow for far reaching conclusions at the level of the organisational level, let alone the level **below** the main organisational level.

4. RESEARCH EVALUATION EXERCISES IN THE UK

In the UK scientific research has been monitored on a large scale from 1986 onwards. With the Research Assessment Exercises (RAE), the British Higher Education Funding Councils assess periodically the research performance of British universities. The main purpose of the RAE is 'not just to enable funding to be allocated selectively but also to promote high quality: research in higher education institutions conducting the best research receive the largest portion of the grant' (RAE 2001). As a secondary effect, the results of the RAE inform other funding bodies in the UK and abroad, as the results of these assessments constitute a basis for science policy and strategy decisions. These assessments take place across so called 'units of assessment' (in 2001, 68 of these units were applied), which embrace research activities in broad scientific disciplines such as chemistry, sociology, etc. The grouping of research fields is dynamic, and is established in consultation with the higher education sector itself. The information playing a role in the assessment has developed over time. Whilst the first exercises consisted of mainly publication output assessment, in later exercises additional information was provide for the peer panel. This additional information is provided by the research groups and departments themselves, and includes the number of staff involved, a description of the research conducted and future plans, the funding received, and a short survey of various types of research output. In the assessments peer committees determine scores for university groups and departments, based on the input provided by the submitting groups and departments. As a consequence of the time consuming process, and the high costs involved with the RAE, the discussion on applying bibliometric analyses in the RAE has started. Probably a more important role in future RAEs is played by citation analysis, which has caused a vivid debate on the applicability and validity of citation analysis on the level of the individual researcher (e.g., Warner, 2000a, 2000b), and the undesirable effects citation analysis might have on the behaviour of scientists (Warner, 2003). Whilst the proponents of the application of citation analysis in the research assessment indicate that there exists a strong correlation between RAE scores and impact scores (Oppenheim, 2000; and Norris and Oppenheim, 2003), the (small scale)

citation analyses so far applied have certain weaknesses. These will be discussed below. A very thorough, but more general, critique of the RAE is given in a report by PREST, of the University of Manchester (PREST, 2000). Their criticism focuses on specific aspects of the UK situation, e.g., the funding structure or the composition of research assessment, no longer following disciplinary borders, but rather problem and application oriented assessment of research.

The most recent plans of the Higher Education Funding Council for England (HEFCE) intend to strongly reward 'world class' research. This might lead to a further undermining of the research system in the UK, simply because not enough money becomes available to keep research groups and departments at a viable level. The idea is to focus on excellent research, and support, if possible, for the other research. This might lead to concentration of research funding, which in its turn leads to a cyclic process in which the stronger ('funded') groups and institutes become stronger, and the weaker might eventually disappear.

5. PAST EXPERIENCE IN THE NETHERLANDS

In the Netherlands large scale research assessment procedures were initiated by the VSNU (Association of Universities in the Netherlands) in almost all disciplines of the sciences, social sciences and arts and humanities, according to clearly described procedures from a protocol (VSNU, 'Assessment of Research Quality', 1998). This protocol was in principle applicable to all fields of science and scholarship. However, the main characteristic of all the research assessment procedures in which evaluative bibliometrics was applied, was the choice of the research field underlying the assessments. For instance, in chemistry several bibliometric analyses were applied, and in cultural anthropology none. Another common feature of the research evaluations by VSNU was a focus on the research group level. In all these VSNU research evaluation assessments bibliometrics was used as an instrumental supporting tool for the peer review process. An international peer review committee was provided with both qualitative and quantitative information. The qualitative part included a report written by the groups themselves in which an overview was given of the scope of the research involved, publication output, ability to raise funds, whereas the quantitative part comprised a bibliometric report, focusing on the research group level. The committee judges each group on four aspects of their scientific performance (*quality*, *productivity*, *relevance*, and *viability*) on a five-point scale.

Ranging from astronomy and astrophysics, physics, chemistry, biology, the life sciences, electrical engineering, and psychology, bibliometrics was applied in addition to peer review. In the latter field, particularly within the theoretical psychology community, a discussion started about the validity and reliability of bibliometric indicators within the field.

Within these disciplinary research evaluation procedures initiated by the VSNU, bibliometric procedures have been applied in an open, transparent way. Scientists are asked to control and verify journal publications identified in the citation index based data system, and to comment on the results of the bibliometric data collection. The results of this process are discussed with the peer review committee (or its chairman) and are finally reported to the committee. The peer committee has meetings with the leaders of all university research groups involved in a research performance procedure, giving the research group leaders the opportunity to comment on the findings in the bibliometric procedure.

6. FUTURE PROSPECTS FOR EVALUATION IN THE NETHERLANDS

The research performance assessments initiated by VSNU had a nation–wide disciplinary and cyclic (each five or six years) character, focusing on research groups within all universities. Each research group had to provide its own self–evaluation study, containing next to a survey of input data, or personnel, a qualitative description of the research conducted and of the future plans and prospects. This created a considerable 'evaluation bureaucracy' because research groups had to provide much paperwork, similar to the information they had to provide in the paperwork requested by granting organisations (such as national research organisations) but, of course, always in another format, for another time period, etc.

This situation stimulated the argument to revise the evaluation procedure, in such a way that many of the efforts by scientists and science managers within universities would serve both purposes.

While the research performance procedures initiated by VSNU were based on a protocol from 1998, the assessment of publicly funded research is now subject to a new protocol, designed by VSNU, NWO (the national research council, funding research in various disciplines on a proposal basis) and the KNAW (the Royal Dutch Academy of Sciences) (Standard Evaluation protocol (2003 – 2009) for Public Research organisations, VSNU, NWO, and KNAW, January 2003). The purpose of this extension is to evaluate, next to academic research, the research done on behalf of NWO and KNAW (both their funded research, and the research conducted in their

research institutes). The new protocol is strongly influenced by a report from the Commission chaired by professor van Bemmel ('Kwaliteit verplicht', 2000).

As mentioned above, the newly applied protocol covers the full range of publicly financed scientific research. Whilst the main goal of the previous system was to monitor the quality of national academic scientific research in comparative perspective, the new system adds another goal to the research evaluation procedures. This goal relates to the management of scientific research itself, within the organisational structure in which it is carried out in. Another important difference between both procedures is the level of focus: the new protocol is more flexible in such a way that it also allows evaluation at levels other than the research group level. As a consequence the initiative and responsibility for research assessment procedures has shifted from the different Chambers (national disciplinary 'boards') within the VSNU organisational structure, to the top of the organisations (universities) subject to research assessment.

A critique which regularly arose in the previous evaluation procedure dealt with the composition of the review committee. It was either too small (in number) or it lacked knowledge about a specific (sub)field. The new procedure allows a more flexible composition of the peer review committee, taking into consideration the specific type of research conducted in a specific organisation. How this will work out in practice is not yet clear, owing to the recent nature of these developments.

The new evaluation procedure should (try to) focus more on the future perspectives of a specific research unit, thereby replacing the previous main characteristic, namely 'accountability', which leads by nature to a more 'historical' result. A perspective on future developments should be included in the self–evaluation study that is obligatory in each evaluation procedure. Interestingly enough, within bibliometric analyses such a distinction between 'looking farther back in the past', and a 'future perspective' has been long included, and is part of the CWTS methodology applied in national assessment procedures under the previous protocol.

In the report by the van Bemmel commission quantitative studies supporting the evaluation procedure have hardly any priority. The term 'citation analysis' is mentioned only once ('Kwaliteit verplicht', pag. 25), thereby indicating the very modest position of quantitative measures in the new evaluation procedure. However, some of the new goals (for instance, a future perspective of the research in a research group) are hard to grasp by means of only qualitative information. Bibliometric analysis has at least the advantage that it can, based on recent past performance, give some objective insight into future perspectives, starting from the hypothesis that results in the recent past are the best predictors for the near future. This raises the question of the competence of peer committees, and the level of the qualitative information provided in the evaluation process, in the sight of the defined goals in future research assessment procedures

Ironically enough, the 2003 Science Budget of the Netherlands indicates that the current research assessment protocol finds little enthusiasm within cycles of policy makers, especially because the national perspective is abandoned within the current approach, an aspect that was already satisfactorily covered in the 1998 Research Assessment Protocol. So little can be said about the direction in which research assessments will be heading in the very near future in the Netherlands.

7. SYNTHESIS

As stated above, the approaches in the UK and the Netherlands with respect to research assessment were different, and especially in the application of bibliometric indicators, some remarkable differences are observed. Whilst the RAE is bibliometrically based on publication assessment, as can be concluded from the statement that 'All forms of research output (books, papers, journals, recordings, performances) are treated equally' (HERO, 2002, Guide to the 2001 Research Assessment Exercise), the research assessments in the Netherlands were, if applicable in the 'field under assessment', full scale analyses of both the output and the impact of research groups, but bibliometrically restricted to journal publications because they can be found in the citation indexes from ISI.

A very important difference between the actual research assessments in both the UK and the Netherlands is found in the relation between the assessment and research funding. Whilst one could argue that this is beneficial for 'the best scientists', the position and application of quantitative techniques in the process of assessment are of major importance. Given the statement above, the question arises of how peer panels are capable of assessing the four items of research output, in an international context, and of how they treat different research output types equally. Here citation analysis might be helpful, but then other problems occur. In a citation analysis of research output that is supposed to treat different output types equally, citation analysis might be helpful to analyse internationally refereed journal publications as covered by the citation indexes of the ISI adequately, but immediately raises the issue of the limitations of citations analysis of the other types of research output mentioned above. As the ISI citation indexes give, in general, an excellent survey of the most important serial literature, the degree with which citations to non-serial literature are representative for the actual impact of these types

of sources, remains to be seen. As one might expect, these types of sources have a citation cycle of its own, with a stronger focus on the same type.

Within the current constraints of bibliometric analysis, which are predominantly indicated by the boundaries of the ISI citation indexes, with their focus on journal publications, a number of guidelines can be determined for research evaluation assessment procedures. If the results of the research assessments, expressed in peer panel judgements, are going to form the basis for the funding, and, based on bibliometric data, the applicability of quantitative techniques on the assessed field should be assessed first. In other words, if the types of research output cannot be measured adequately with bibliometric techniques, owing to the nature of the field under assessment, current existing bibliometric techniques hardly contribute to the final assessment of the research conducted in this field. Only in the last couple of years have systematic attempts been made to develop quantitative indicators that can be used within the social sciences and the humanities, in close consultation with scientists in those fields (Luwel et al., 1999). However, this is not yet as developed as the quantitative techniques applied in the natural, life and technical sciences.

Once you have established the applicability of bibliometric techniques in a certain field, a next guideline would be the level of bibliometric analysis. Whilst some proponents of bibliometric techniques in the British RAE propose citation analysis of individual output, which can be aggregated to group scores (Norris and Oppenheim, 2003), a direct group approach involving the complete journal publication of a research group or research department would be more fruitful. Instead of a selection of only four research outputs, a systematic analysis of the complete oeuvre of a group would provide a more thorough insight into the volume and development of a research group, in terms of its research output in journal publications. And although many scientists in the Western world nowadays have access to the ISI citation indexes (mainly through the *Web of Science*), and therefore are able to estimate which of their papers belong to their most highly cited publications, the number of only four publications is a very small number to base a research assessment on, which determines research funding for the upcoming period. The usage of a somewhat larger number of publications, which follows nearly automatically from an approach in which the research group is the focal point, is less sensitive to statistical problems which occur when the bibliometrician focuses on the (small) output of any individual researcher. Another problem is the overlap that can occur between the research outputs of staff members of one group. This might lead to unrealistically high citation scores.

Once the decision is made to apply the bibliometric analysis to only those fields in which bibliometrics can be applied adequately, and to focus the research assessment on the research output at any level above the individual researcher's level, one needs to take care with respect to the data collection procedure. In the data collection phase the distinction between descriptive and evaluative bibliometrics can be expressed most clearly: with the implementation of a verification process, in which the researchers under assessment check and control the collected publication output, the necessary reliability and validity are added to the process. This step makes the whole process more transparent, and provides a certain space in which scientists can argue about the process, add and/or delete publications to their oeuvres, and, in general provide valuable background information on the publications characteristics of a field. Furthermore, this step allows the users of the outcomes of such bibliometric analyses to draw conclusions from the material, that were not possible if the publications were collected 'from a much larger distance' between assessed research field and bibliometricians.

Finally, if the publication data are collected in such a way that assessments can be carried out, perhaps the most important part of the whole process starts, namely, the calculation of bibliometric indicators. Whilst the RAE impact analyses are based on crude citation counts, citation per publication counts, and the number of papers amongst the most highly cited publications in a field, more sophisticated indicators should be calculated. For example, normalised indicators expressing the ratio between the actual impact and expected values are strong indicators (Moed, de Bruin, and van Leeuwen, 1995), but also indicators describing aspects of scientific publishing like the percentage of self-citations, or the percentage of publications not cited can contribute to a more balanced judgement of the research performance of a group.

8. CONCLUSIONS AND DISCUSSION

This paper focused on the distinction between descriptive bibliometrics, as resulting from top down analyses, and applied in national facilities monitoring national R&D systems from a somewhat larger distance, the results of which are normally publicly available, and evaluative bibliometrics, resulting from bottom-up approaches, applied in (national) research assessment procedures of disciplines and research organisations such as universities or research institutes. Here the results of such exercises are mostly kept confidential. For the former type of bibliometric analyses peer information is hardly used, consequently indicating the larger distance of these analyses to the scientific research itself. In the latter type, peer judgements are a necessary ingredient of research assessment. There is a long-standing, ongoing discussion on advantages and disadvantages of peer

review in general (a very good survey is given by Nederhof (1988), and by van Raan (1996). The most important advantages relate to the self–organising principles of science as a quality — and particularly reputation — oriented community, in combination with a reasonable degree of consensus about the direction in which scientific developments should/could proceed. On the other hand, important drawbacks of peer review relate to the composition of the review committee: aspects of subjectivity, conflicts of interest, high costs involved, and insufficient recognition of young promising scientists and/or recent promising scientific developments by the members of a peer review committee. And whilst the application of bibliometric techniques in assessment procedures will not solve all these problems or disadvantages, there most certainly can be a future for the application of bibliometrics in research assessment procedures. This requires a stronger consensus amongst the people in the bibliometric community, a better presentation of the indicators of use in those processes, and also a testing of the robustness of indicators applied in research performance assessment exercises (van Leeuwen et al., 2003).

An example of the difficulties for bibliometricians, let alone laymen, to fully understand the handling of underlying data in the construction of indicators can be found in the calculation of the Impact Factor. Although defined quite clearly, the actual reconstruction of the composing parts for interested users is quite difficult. In an article in *Nature* in 2002 by Moed, the limitations of journal impact factors for the bibliometric practice were shown. However, some of the arguments which were addressed against the application of journal impact factors in the bibliometric practice have a farther reaching meaning, which goes beyond the application of only Impact Factors. The application of bibliometrics, in general, within a research performance measurement procedure requires answers to major questions. For the bibliometrician the first question relates to the issue of visibility of its subject within the international scientific literature, in other words, to what extent is the application of bibliometrics (based on bibliographic data in the ISI databases) focusing on a substantial part of the research output of the subject of study? Normally a bibliometric analysis contains both *output* and *impact* indicators, and the latter can be calculated in a reliable manner only on the basis of the citation index data. These indexes of the ISI are, up until now, the only multi-disciplinary databases containing the complete reference lists to previous scientific journal literature. This allows the bibliometrician to calculate impact indicators, in which the visibility or influence of these previous scientific publications can be expressed. So a first issue that needs to be addressed before starting a bibliometric study relates to the focus on the (English language) international journal literature, and its coverage within the databases of ISI.

If this first issue is answered in such a way that bibliometric data can be collected, the bibliometrician should ask him/herself the question: 'What percentage of the publications of a research group is indeed covered in the international journal literature, and to what extent is this influenced by the specific characteristics of either the (sub)field concerned, or of the research group?' Within a large discipline, such as e.g., chemistry, differences may exist within that discipline. For instance, in an evaluation of chemistry as a whole, the typical 'classical' fields of chemistry such as analytical chemistry are analysed in combination with chemical engineering, biotechnology, and biochemistry. We are then confronted with considerable diversity in publication practices amongst these fields within a discipline, such as a stronger focus on conference proceedings, higher publication rates, a more 'dense' citation traffic, etc. Thus, the intra-disciplinary variation can be as broad as the interdisciplinary or intra-science variation as a whole.

Once these topics are adequately dealt with, a bibliometric analysis should comprise more than only crude publication counts, citation counts, or mean citations per publication ratios, exactly because these indicators do not indicate the performance of a research groups in such way that it can be compared to other units within the assessment, e.g., how does a group in biology then compare with a group in the social sciences? Therefore the bibliometric community should stress the necessity of a certain standardisation in which normalised indicators play a central role. This is a general improvement; other, more conceptual, improvements could be found in a broadening of the scope of bibliometric analyses, in which other aspects of the performance of a research group or institute become visible. One of such techniques can be found in the so called research profiles, in which a spectral analysis is given of the distribution of the output of a group over research fields, in combination with impact scores. In the current bibliometric practice of CWTS these profiles are generated on the basis of the ISI Journal Subject Categories, but this information can be replaced by other information indicating the research topics of a certain field. These profiles then indicate the multi-disciplinary character of research groups from one group to another, and as such indicate a certain resemblance or difference between groups in a research assessment process.

Another useful instrument is formed by scientific cooperation profiles. These profiles, based on the addresses attached to a publication in the citation indexes, indicate the different types of orientation research groups can have in the world outside their own 'environment', and especially indicate the international scope of both the group and the research field as a whole, by comparing scientific cooperation profiles among groups. These analyses could be extended by network analyses, indicating the partners of research groups, and the success of this cooperation.

Finally, the assessment of research groups performance can be extended with analyses of the 'reception' of a groups output. Here the focus can be on who is citing your recent research output, in which journals, and to what extent do we see knowledge transfer between fields.

REFERENCES

Beller, F.K. (1999). Der Zusammenhang zwischen Index Medicus, dem Impact Factor und der deutschen Sprache. *Geburtshilfe und Frauenheilkunde*, 59, 53–56.

Geuna, A., Martin, B.R. (2003). University Research Evaluation and funding: an international comparison. *Minerva*, 41, 277–304.

Haller, U., Hepp, H., Reinhold, E. (1997). Tötet der 'Impact Factor' die deutsche Sprache ? *Gynakologisch–geburtshilfliche Rundschau*, 37, 117–118.

Herfarth, C., Schurmann, G., (1996). Deutsche klinische Zeitschriften und der Impact Factor. *Chirurg* 67, 297–299.

Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, 44, 193–215.

Jimenez–Contreras, E., Lopez–Cozar, E.D., Ruiz–Perez, R., Fernandez, V.M., (2002). Impact–factor rewards affect Spanish research. *Nature*, 417, 898–898.

Kindermann G., (1999). Hat Deutsch noch Zukunft als Wissenschaftssprache? *Geburtshilfe und Frauenheilkunde*, 59, 188–190.

KNAW Commission, chaired by Professor van Bemmel *"Kwaliteit verplicht"*, report. Amsterdam, 2000.

van Leeuwen T.N., Moed, H.F., Tijssen, R.J.W., Visser, M.S., van Raan, A.F.J., (2000). First evidence of serious language–bias in the use of citation analysis for the evaluation of National Science Systems. *Research Evaluation*, 9, 121–122.

van Leeuwen T.N., Moed, H.F., Tijssen, R.J.W., Visser, M.S., van Raan, A.F.J., (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics*, 51, 335–246.

van Leeuwen, T.N., Visser, M.S., Moed, H.F., Nederhof, A.J., van Raan, A.F.J., (2003). The Holy Grail of Science Policy: exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics*, 57, 257–280.

Luwel, M., Moed, H.F., Nederhof, A.J., De Samblanx, V., Verbrugghen, K., van der Wurff, L.J., (1999). *Towards indicators of research performance in the social sciences and humanities.* An exploratory study in the fields of Law and Linguistics at Flemish Universities. Report of the Flemish Inter-University Council (V.L.I.R.), Brussels, Belgium / Centre for Science and Technology Studies (CWTS), Leiden University, the Netherlands / Ministry of the Flemish Community, Brussels, Belgium. V.L.I.R. Brussels, Belgium.

Moed H.F., *Thesis*, Leiden University, 1989.

Moed, H.F., De Bruin, R.E., van Leeuwen, T.N., (1995). New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications. *Scientometrics*, 33, 381–422.

Moed H.F., (2002). The impact factors debate: the ISI's uses and limits. *Nature*, 415, 731–732.

Narin, F., (1976). Evaluative bibliometrics. *The use of publications and citation analysis in the evaluation of scientific activity.* Computer Horizons, Inc., Cherry Hill, New Jersey, USA.

Nederhof, A.J. (1988) *The validity and reliability of evaluation of scholarly performance*. In A.F.J. van Raan (Ed.), Handbook of Quantitative Studies (pp. 193–228). Amsterdam: North-Holland.
NOWT, *Science and Technology Indicators report 1994* (in Dutch).
NOWT, *Science and Technology Indicators report 1996* (in Dutch).
NOWT, *Science and Technology Indicators report 1998* (in Dutch).
NOWT, *Science and Technology Indicators report 2000* (in Dutch).
NOWT, *Science and Technology Indicators report 2003* (in Dutch).
OST, *Science & Technologie Indicateurs – Edition 1992*, Economica, Paris, 286 pp.
OST, *Science & Technologie Indicateurs – Edition 1994*, Economica, Paris, 425 pp.
OST, *Science & Technologie Indicateurs – Edition 1996*, Economica, Paris, 473 pp.
OST, *Science & Technologie Indicateurs – Edition 1998*, Economica, Paris, 551 pp.
OST, *Science & Technologie Indicateurs – Edition 2000*, Economica, Paris, 512 pp.
OST, *Science & Technologie Indicateurs – Edition 2002* ,Economica, Paris, 467 pp.
PREST (2000). *Impact of the Research Assessment Exercise and the Future Quality Assurance in the light of changes in the research landscape.* Report to the Higher Education Funding Council for England (HEFCE), by PREST, University of Manchester, United Kingdom.
RAE 2001. *Guide to the 2001 Research Assessment Exercise.*
Van Raan, A.F.J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight studies. *Scientometrics*, 36, 397–420.
Rempen, A., (1998). Leserbrief zur Haller, U., H. Hepp and E. Reinhold Tötet der 'Impact factor' die deutsche Sprache? *Gynakologisch–geburtshilfliche Rundschau*, 38, 54–54.
Saiz–Salinas, J.I., (1996). Failed professor. *Nature*, 381, 186.
Tijssen, R.J.W., van Leeuwen T.N., van Raan, A.F.J. (2002). *Mapping the scientific performance of German medical research. An international comparative bibliometric study.* Report to the German Federal Ministry of Education and Research (BMBF).
VSNU, *Standard Evaluation Protocol (1998) for Public Research organisations*, Utrecht, 1998.
VSNU, NWO, and KNAW, *Standard Evaluation Protocol (2003 – 2009) for Public Research organisations*, January 2003.
Warner, J. (2000a). A critical review of the application of citation studies to the Research Assessment Exercises, *Journal of Information Science*, 26, 453–460.
Warner, J. (2000b). Research assessment and Citation Analysis, *The Scientist*, 14, October 30.

Chapter 17

WHAT HAPPENS WHEN FUNDING IS LINKED TO PUBLICATION COUNTS?

Linda Butler

Research Evaluation and Policy Project, Research School of Social Sciences, The Australian National University, Australia
E-mail: linda.butler@anu.edu.au

Abstract: Many countries are placing a greater emphasis on public accountability for government research funding and are starting to use quantitative performance indicators for the distribution of funds. In Australian universities the use of quantitative formulas to allocate the research component of university block grants to institutions has been in place for a decade, and thus the system provides fertile ground for using bibliometrics to examine the effects of such policies on academic output. An analysis of Australian data from the Institute for Scientific Information's major citation indexes clearly demonstrates the academic response to the linking of funds, at least in part, to productivity measures undifferentiated by any measure of quality — publication numbers jumped dramatically, with the highest percentage increase in the lower impact journals. The trends were apparent across all fields of research in the university sector, but were not present in other sectors active in research (such as hospitals or government research agencies). The trends were not, however, uniform across all institutions.

1. INTRODUCTION

In most OECD countries increasing emphasis is being placed on greater public accountability, with a need to demonstrate the effectiveness and efficiency of government supported research. A workshop held by the OECD in 1997 characterised the recent evaluation of basic research as "a rapid growth industry"(OECD, 1997).

This new demand for research evaluation cannot be fully serviced by the finite capacity of traditional peer review. Researchers, particularly the more senior ones, have many calls on their expertise, such as reviewing journal articles, assessing grant applications, sitting on selection and promotion committees, being co-opted to national or institutional review bodies. They can only devote a limited proportion of their time to such activities before their own research begins to suffer. Partly as a consequence of the pressures on peer review, there has been an increased use of quantitative performance indicators as an alternative method for evaluating research performance, which has the added advantage of being more cost efficient. There is also an increasing trend to link such measures directly to the distribution of research funds.

For Australian universities the allocation of funds earmarked for research is based on a formula encapsulating a number of performance measures (graduate student numbers or completion rates, research income, and publications). Spanish scientists are directly rewarded with a salary supplement for increasing their output in the major English language international journals (Jiménez–Contreras, Anegón and López–Cózar, 2003). In Finland part of the funding for university hospitals rests on publication points, weighted according to the impact factor of the journals carrying the work (Adam, 2002). While in the British Research Assessment Exercise the link between research rankings and performance measures, and hence funding, is less direct, they nevertheless play an important role in the deliberations of the review panels.

The link between research funding and quantitative performance measures has now been in place in Australian universities for a decade, and thus provides fertile ground for using bibliometric data to examine the effects of this policy on academic output. Since performance measures relating to publications are limited to aggregate productivity counts, the expectation would be that Australian university publication output would increase significantly in response to the signals embodied in the funding formula. As there is no attempt to weight for the quality of either the output itself, or the publication in which it appears, there would also be an expectation that any increased journal output is likely to be concentrated in lower ranked journals where it may be easier to place additional articles. Both these anticipated outcomes are clearly visible in the data for Australian universities in a number of major journal citation indexes.

2. POLICY BACKGROUND

The Australian government has a dual system for funding research in universities. A significant amount of money is distributed by the two research councils, the National Health and Medical Research Council and the Australian Research Council, via a peer reviewed assessment system. Both agencies distribute the bulk of their funding support in the form of project grants, which can vary in length from one to five years, with three years being the most common duration. Secondly, a proportion of the block operating grant to universities (of the order of 5%) is earmarked for research and research training, and since the beginning of the 1990s this has been distributed via a formula. The formula aimed at taking account of a broad range of measures of research performance when making allocations to universities. Initially this formula was based only on external earnings, but subsequently student and publication components were added.

Australian universities began supplying details of their research output to the Department of Education, Science, and Technology (DEST[1]) and its predecessors in 1993, initially through the Australian Vice Chancellors Committee (AVCC), and more recently directly to the department. The research funding formula was expanded in 1995 to include output measures — publication counts and higher degree loads and completions — and was also used in the allocation of postgraduate awards. The components of the formulas, the funding schemes they were applied to, and the weighting given to each element, are shown for a sample of years in Table 17.1.

From 2001, as a result of a review of higher education research, the amount of funds allocated on the basis of formulas has nearly trebled, and now accounts for more than half the funding specifically targeted to research and research training through the education portfolio (DEST, 2002a). The Small Grants scheme, not previously funded by this method, was rolled in with the Research Quantum (RQ) and became the Institutional Grants Scheme. Postgraduate awards continued to be funded under this arrangement and, in addition, a new Research Training Scheme was introduced which more than doubled the funds distributed via formulas. None of the more recent changes represented 'new' money, merely a change in the method by which some of the funds were distributed, and a greater reliance on formula driven schemes.

Australia's approach in this area of higher education policy is not common. A recent survey of 14 countries by Geuna and Martin only identified two that used ex-post quantitative evaluation for allocating core research funds, Finland and Australia (Geuna and Martin, 2003) Unlike Australia's mechanistic system of quantitative measures, Finland employs a

[1] The Australian Government department which encompasses the education portfolio has had several name changes in the period referred to in this paper — the Department of ... Employment, Education, Training, and Youth Affairs; Employment, Education, and Training; and Education, Science, and Training — but I will use the acronym for the department in its current form (DEST) throughout this chapter

series of agreed indicators focusing on the quality and impact of teaching and research. The Australian experience is not mirrored in other countries, and may well be part of the explanation for the publication trends seen in Table 17.1.

Table 17.1. Formulas that distribute research funds to Australian universities through block grants

Funding Scheme	Total funds ($mil)	Weight given to each element (percent): Publications	Higher degree load	Higher degree completions	Research income
1996					
Research Quantum	218.6	12.50		5	82.50
Postgraduate awards (2 schemes)	91.7	5.26	40	20	34.74
2000					
Research Quantum	223.0	10.00		10	80.00
Postgraduate awards (2 schemes)	96.2	4.44	40	20	35.56
2002					
Institutional Grants Scheme	271.3	10	30		60
Postgraduate awards (2 schemes)	102.0	10		50	40
Research Training Scheme	515.6	10		50	40

Source: Australian Vice Chancellors' Committee (AVCC), 2002.

3. THE REWARDS FOR PUBLISHING

Determining the 'value' of a publication unit to a university is a simple calculation and it was not long before figures became commonly referred to in the sector. Taking the data given in Table 17.1, together with the publication counts on which the distribution of funds was based, Table 17.2 details the calculations for the three sample years. The distribution of funding for the publications element was based on data for the most recently available two years.

Table 17.2 demonstrates the effect that adjustments to the coverage of publications in the collections, and/or the amount of funding distributed in this way, can have on calculations of the unit value. For example, the 1996 distribution was based on 1993 and 1994 publications. The 1993 data covered 8 publication types; the 1994 data covered 22. After a sample audit of the universities' lists of 1994 publications, the number of categories covered was reduced to just four for subsequent collections: books, book chapters, refereed journal articles, and refereed conference papers[2]. As a result the number of publication units in subsequent collections dropped significantly, with a consequential increase in the value of each unit. This occurred despite a reduction in the weight given to publications in the formula from 12.5% to 10%, and a reduction in the amount of funds distributed on this basis.

Table 17. 2. Value of a publication unit: 1996, 2000 and 2002

Funding year	*Funds tied to publication counts (AUD$million)*	*Publication counts**	*Value per publication unit*
1996	32.1	42,259	$761
2000	26.6	24,390	$1,089
2002	88.9	26,877	$3,307

Source: Department of Employment Education and Training, 1996; Department of Education, Training and Youth Affairs (DETYA), 2000; Department of Education, Science and Training (DEST), 2002b.
* Weighted by type of publication

From 2001 the funds distributed via the formulas were increased significantly, leading to a three-fold increase in the value of a publication unit. Every refereed journal article is now 'worth' over AUD$3,000 to a university, and a book is now 'worth' AUD$15,000.

4. IDENTIFYING THE EFFECTS OF INTRODUCING FUNDING FORMULAS

As the categories covered by the Australian collection have been refined and reduced in number, the importance of journal publications indexed by the *Institute for Scientific Information* (ISI) has increased. The collection is externally audited, and universities must prove, among other things, that the journals carrying the articles they are claiming are peer reviewed. A journal that is indexed by ISI is accepted as peer reviewed without question, but universities must prove that any other journal meets the definition. Publishing in ISI-indexed journals is obviously the easiest course of action to take. The data contained in ISI's three main indices, the *Science Citation*

[2] In recent collections books receive a weighting of five in the calculations, while the other three categories are all given the base weighting of one

Index (SCI), the Social Sciences Citation Index (SSCI), and the Arts and Humanities Citation Index (A&HCI), therefore provide fertile ground for examining the impact that introducing the funding formulas had on Australian university output.

The Research Evaluation and Policy Project (REPP) maintains a database which contains all Australian publications in these ISI indices. Considerable effort is expended in standardising the addresses listed for each publication, thus enabling accurate analysis to be undertaken at the sectoral (university, hospital, government, etc), institutional, and even lower levels of aggregation, such as faculties and departments.

4.1 The University Sector in Aggregate

An analysis of Australia's presence in the SCI was the first step taken to investigate whether it was possible to demonstrate the apparent effect of the introduction of the funding formulas in the 1990s. In the analysis SCI journals were allocated to quartiles based on the average citation per publication rates of the publications they carried. Mean journal citation rates were calculated for each five year window from 1981–85 through until 1996–2000. For both publication counts and citation totals, the calculation was limited to publications classified by ISI as articles, notes, reviews and proceedings papers, and to the specified five year period. As a separate calculation was made for each period, journals were free to move between quartiles over time.

Australian universities' presence in these four quartiles was then tracked over the full twenty year period. Their share of total publications in each of the four quartiles is shown in Figure 17.1.

The response of the academic community appears very clear, and in line with expectations. Until the period 1989–93 there had been virtually no movement in the institutions' presence in the SCI journal set, with the exception of an increase in the third quartile. Since that period university output has jumped dramatically, particularly in journals allocated to the bottom two quartiles. The sector's share of publications in journals allocated to the top two quartiles increased by 28% and 15% respectively; their share of publications in the third quartile rose at double those rates, i.e., by 55%; and in journals allocated to the bottom quartile their share doubled.

With no attempt made to differentiate between the quality, visibility or impact of the different journals when funding is allocated, there is little incentive to strive for publication in a prestigious journal. Whether a publication reports ground breaking research or is a more pedestrian piece; whether it appears in a highly visible journal such as *Nature* or a lower impact outlet, the rewards are identical.

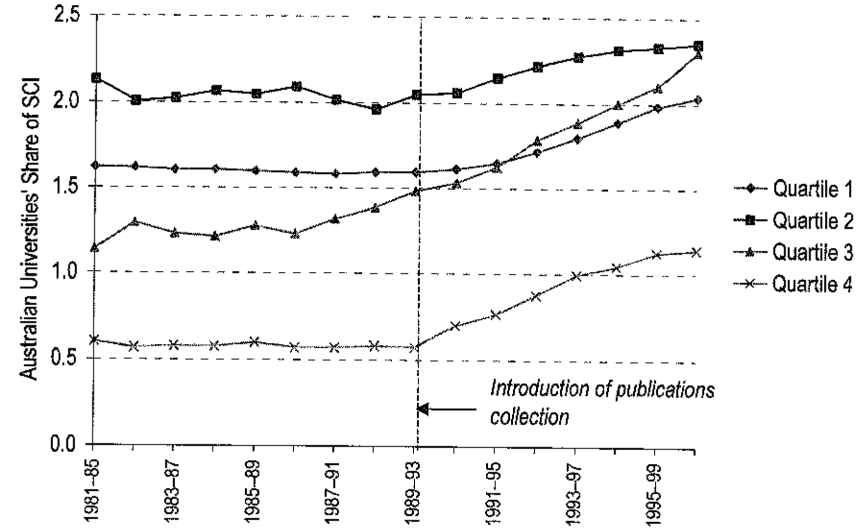

Figure 17 1. Australian universities' share of publication in the SCI, by journal impact quartile: five year windows, 1981–1985 to 1996–2000.

The trends shown in Figure 17.1 are not proof in themselves of a direct link between funding formulas and increased productivity. However, they did occur at a time when funds to the sector are extremely tight. A detailed analysis was undertaken when these trends first became apparent to determine whether the increased output could be explained by either the entry of new institutions into the sector, or an increased number of researchers (Butler, 2001a) Results showed that while the new institutions had increased the sector's research capacity, they accounted for less than one third of the expanded output — the bulk came from the older, established universities. Nor were increased staff numbers the explanation. They had risen in the period after the introduction of the publications collection, but the increase was no greater than it had been prior to this time.

To be more confident that the trends are a result of the introduction of funding formulas it is necessary to examine the data in more detail in order to determine whether the following three scenarios also exist:

1. The trends are specific to the university sector. No other Australian research sector is faced with the same funding drivers, so the trends for other sectors should not mirror that for universities.

2. The trends are present in all fields of research. The formulas are applied across the university sector, so all fields of research, including those less traditionally reliant on journal outlets for their research, should exhibit similar trends.
3. Another university system faced with similar incentives, exhibits similar trends. The Spanish research system is subject to funding drivers based on journal output, and the effect of this should also be apparent in ISI data.

The results of these analyses are given in the following sections.

4.1.1 Comparison of sectors

The three largest sites of research in Australia outside the universities are the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the hospitals, and government research agencies. None of the institutions comprising these sectors are subject to funding formulas of the type present in the university sector, although all have strong collaborative links with it. Figure 17.2 shows the trend in publication output for these three sectors using an identical analysis to that applied to the university sector.

It is quite clear that the 1989–93 period does not mark a turning point in trends for any of these sectors. CSIRO, with an increasing emphasis on seeking external funds for a significant share of its operating costs, has seen its overall share decline (although actual publication numbers have remained steady). The hospital sector's share of output in the top quartile has been increasing steadily across the whole period, while its presence in the journals allocated to the bottom quartile has increased but remains very low. There are considerable fluctuations in its share of the other two quartiles, and the mirror image in movement between these quartiles suggests some journal movement between the two sets. The government sector's share of output in the top and bottom quartiles has remained relatively steady across the twenty year period covered by our data. As with the hospital sector, their presence in quartiles 2 and 3 is more volatile, and presents a mirror image in movement.

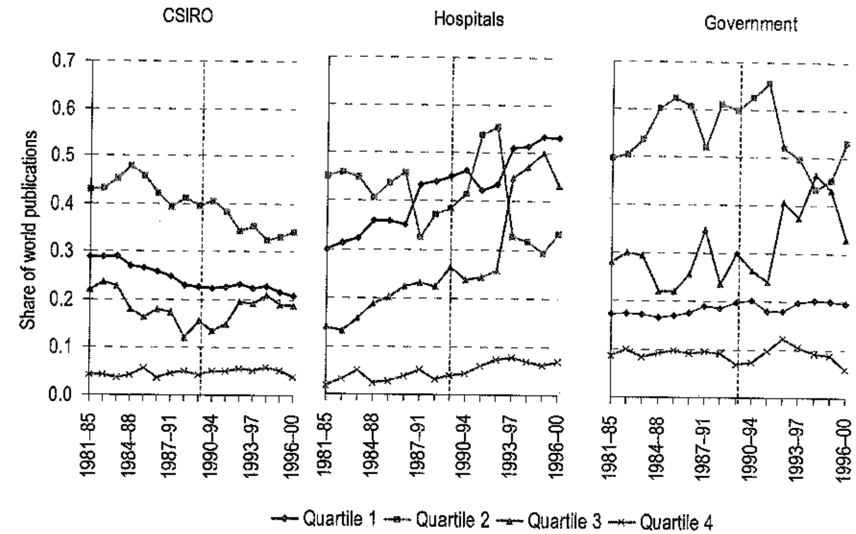

Figure 17. 2. Share of publications in the SCI by other Australian sectors, by journal impact quartile: 1981–1985 to 1996–2000.

4.1.2 Comparison of fields

To disaggregate the trends and examine what was occurring in different fields of research, the methodology used for the SCI as a whole was applied to subsets of journals. For this analysis ISI subject category journal sets were used, and translated into the Australian Research fields, Courses and Disciplines classification scheme. Within each field journals were allocated to quartiles on the basis of the five year average citation impact of the publications they carried. As expected, the average citation per publication (cpp) threshold varied considerably between fields. For example, to be in the top quartile in chemistry in the period 1996–2000, a journal needed a cpp rate of 3.61, while a mathematics journal required only 1.86.

Table 17.3 shows the increase in Australian universities share of world publications by field in two periods of equal length: the increase between 1981–85 and 1988–92; and the increase between 1989–93 and 1996–2000, the period after the introduction of the publications collection.

2. The trends are present in all fields of research. The formulas are applied across the university sector, so all fields of research, including those less traditionally reliant on journal outlets for their research, should exhibit similar trends.
3. Another university system faced with similar incentives, exhibits similar trends. The Spanish research system is subject to funding drivers based on journal output, and the effect of this should also be apparent in ISI data.

The results of these analyses are given in the following sections.

4.1.1 Comparison of sectors

The three largest sites of research in Australia outside the universities are the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the hospitals, and government research agencies. None of the institutions comprising these sectors are subject to funding formulas of the type present in the university sector, although all have strong collaborative links with it. Figure 17.2 shows the trend in publication output for these three sectors using an identical analysis to that applied to the university sector.

It is quite clear that the 1989–93 period does not mark a turning point in trends for any of these sectors. CSIRO, with an increasing emphasis on seeking external funds for a significant share of its operating costs, has seen its overall share decline (although actual publication numbers have remained steady). The hospital sector's share of output in the top quartile has been increasing steadily across the whole period, while its presence in the journals allocated to the bottom quartile has increased but remains very low. There are considerable fluctuations in its share of the other two quartiles, and the mirror image in movement between these quartiles suggests some journal movement between the two sets. The government sector's share of output in the top and bottom quartiles has remained relatively steady across the twenty year period covered by our data. As with the hospital sector, their presence in quartiles 2 and 3 is more volatile, and presents a mirror image in movement.

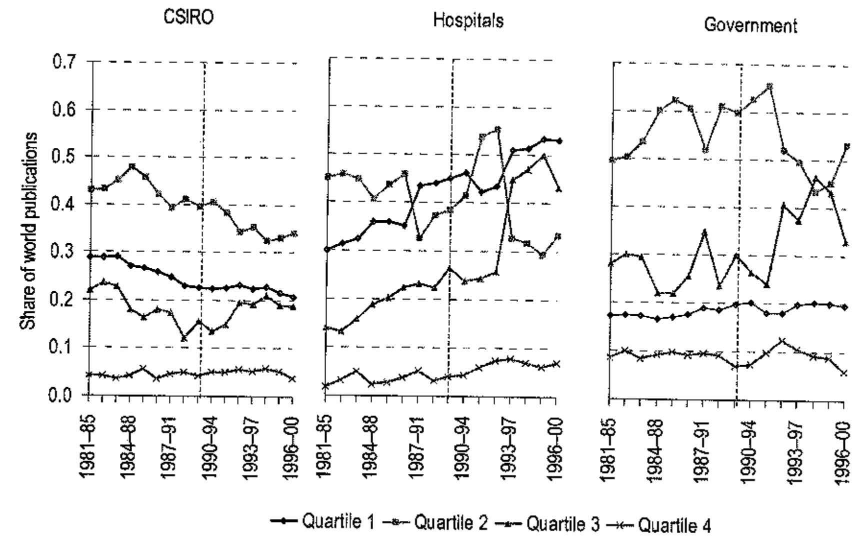

Figure 17. 2. Share of publications in the SCI by other Australian sectors, by journal impact quartile: 1981–1985 to 1996–2000.

4.1.2 Comparison of fields

To disaggregate the trends and examine what was occurring in different fields of research, the methodology used for the SCI as a whole was applied to subsets of journals. For this analysis ISI subject category journal sets were used, and translated into the Australian Research fields, Courses and Disciplines classification scheme. Within each field journals were allocated to quartiles on the basis of the five year average citation impact of the publications they carried. As expected, the average citation per publication (cpp) threshold varied considerably between fields. For example, to be in the top quartile in chemistry in the period 1996–2000, a journal needed a cpp rate of 3.61, while a mathematics journal required only 1.86.

Table 17.3 shows the increase in Australian universities share of world publications by field in two periods of equal length: the increase between 1981–85 and 1988–92; and the increase between 1989–93 and 1996–2000, the period after the introduction of the publications collection.

Table 17 3. Percentage increase in publication output by field — two periods

	% Change: 81–85 to 88–92				% Change: 89–93 to 96–00			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
All sciences	–2	–8	22	–4	28	15	55	100
Mathematical sciences	–13	–16	–3	14	1	43	34	77
Physical sciences	–8	–25	137	–32	42	63	18	85
Chemical sciences	–20	–8	47	13	24	–17	124	137
Earth sciences	7	15	38	19	4	28	31	88
Biological sciences	–7	–5	41	–17	18	25	27	74
Engineering and technology	–10	0	16	–1	37	42	75	117
Agric, vet, environ	–16	14	78	–21	14	48	52	144
Medical and health sciences	0	9	0	29	22	18	84	82
Social sciences	4	–19	56	–25	13	63	28	65

Most fields of research demonstrate relatively stable publication shares between 1981–85 and 1988–92, with movements contained within 25%. The exception is increases in the third quartile — in line with overall trends.

These data show, with the one exception of a decrease in university publications in the second quartile in chemistry, that universities have significantly increased their output in all fields and in all quartiles in the second period studied (1989–93 to 1996–00). In the medical and health sciences the increase in share of the bottom two quartiles is at a similar level; in all other fields the largest increase is in the bottom quartile, usually by a significant margin. As with trends in the preceding period, in this later time frame the fields exhibit trends similar to the aggregate ones, although inevitably there is some variation. In most cases quartile 3 accounted for the second largest increase, with those in the top two quartiles much more modest.

The increase in output in the physical sciences is more evenly spread across the four quartiles. Notably, universities increased their share of the highest impact journals by 40%, a greater margin than for any other field. The two possibilities which immediately suggested themselves as an explanation for this trend — the influence of astronomy in which Australia is particularly strong, and the movement of major Australian journals in the field between quartiles — were found to have no impact on the trends.

4.1.3 The Spanish experience

Since 1989 a research incentive system has existed in Spain, administered by the National Commission for the Evaluation of Research Activity (CNEAI). Researchers were rewarded with salary bonuses for publishing in prestigious journals, principally articles appearing in a relatively high position (approximately the top one third) in ISI's Journal

Citation Report lists by subject category. Unlike the Australian system, the focus is clearly on the individual rather than the institution. But the message is clear — it is increased productivity that is important. A recent study has clearly demonstrated the effect of this policy on Spanish publication output in the ISI–indexed journals (Jiménez–Contreras, Anegón and López–Cózar, 2003).

Their work demonstrates clearly that Spanish researchers have also responded to funding stimuli by increasing their output well above the long-term trend line for Spanish publications in the ISI indices. However, in the Spanish case CNEAI achieved its stated aims, which were to increase productivity and the internationalisation of Spanish research. In contrast, the Australian funding formulas were designed to reward quality, but in fact reward quantity.

4.1.4 Interpretation of trends

The similar trends found in university output in all fields of research, the lack of similar trends in other research sectors, and the Spanish experience, all support the hypothesis that the increased university output in Australia, and the pattern of its distribution across impact quartiles, is a direct result of the introduction of the DEST funding formulas.

There are differing interpretations which can be placed on these trends. In discussions which followed the release of the data, there were those who argued it was 'good news' — that the large jump in output in low impact journals was of little concern because the Australian presence in high impact journals had also increased. While this may be true, there is an overriding objection to the use of undifferentiated publication counts in this instance, and that is one of intent. The formulas, and in particular the publications component, were conceived as a means of distributing research funds on the basis of the quality of research in Australian universities. Publication counts are not measures of quality.

4.2 Institutional Analysis

While the trends in Australian publication output were similar across the different fields of research, it is perhaps not surprising that trends in individual institutions are not as uniform. This is largely because of the disparate signals which individuals within these institutions are receiving from a variety of sources, their judgment on which carry the most weight, and their subsequent reaction to these signals. Researchers face one set of performance measures when applying for grants; another when seeking promotion; yet another when applying for a job at a new institution; a series of community standards set by the peers in their own discipline — all in addition to any sector–wide signals which their institution may be receiving and passing on down through faculties and departments. Some of the signals received will inevitably be contradictory.

Table 17.4 shows publication trends for individual institutions calculated in the same manner as for fields. To provide some indication of the nature of each institution the universities have been classified by type and by the size of their output for the two periods. Australian universities are often classified into four categories:

- 'Go8' (Group of Eight) universities are a self–selected group with a strong research focus and a wide coverage of disciplines. Most are among the oldest of the nation's universities, the first institutions to be established in the major State capital cities. The exceptions are New South Wales, Monash, and the Australian National University, although all three have been established over 50 years;
- 'pre-1988' universities are more recent, but were in existence prior to the major higher education reforms of 1988 which saw the abolition of Institutes of Technology or Colleges of Advanced Education as distinct types of tertiary institutions;
- 'ex-IT' universities are those which, prior to the 1988 reforms, were solely undergraduate institutes of technology. A few of the larger, older establishments had already been granted university status just prior to the major reorganisation of the sector; and
- 'ex-CAE' universities are those which, prior to the 1988 reforms, existed primarily as small, undergraduate institutions focusing on the professions, such as teaching and nursing, with little research capacity.

Table 17.4 has been limited to those institutions with at least a modest publication profile in the 1980s — those with less than 100 publications in the five year period 1988–1992 were excluded.

The institutions with the greatest overall increase in publication output are the 'ex-CAEs' and the 'ex-ITs'. For both groups this is to be expected, because their capacity to undertake research, and the number of staff qualified and experienced to do so, increased significantly after the change in status of their institutions.

Only four institutions showed a greater growth in publication output in the first period (1981–85 to 1988–92) than in the second period (1989–93 to 1996–00). Two were 'ex-CAEs' which started from a low publication base — University of Western Sydney and Northern Territory University. The other two institutions were 'pre-1988' universities — Deakin University and University of New England. Deakin University's publication trends are

unique among Australian universities with more growth in the earlier period, and the highest increase in the second period to be found in the top quartile.

All other institutions in the analysis showed a significantly greater growth in publication output in the second period. In fifteen instances the highest growth rate was in the bottom quartile, while in another four cases the highest growth was recorded in quartile three. In the remaining five cases four recorded their highest growth in quartile two and just one institution, James Cook University, recorded its strongest growth in the top quartile.

Table 17. 4: Publication output trends for Australian universities — two periods

	Type	*No. pubs*	*% change: 81–85 to 88–92*				*No. pubs*	*% change: 89–93 to 96–00*			
University		*88–92*	*Q1*	*Q2*	*Q3*	*Q4*	*96–00*	*Q1*	*Q2*	*Q3*	*Q4*
All Universities		37,721	–2	–8	22	–4	60,014	28	15	55	100
U Sydney	Go8	10,620	20	41	36	13	17,628	51	54	71	93
UAdelaide	Go8	6,048	1	83	12	3	8,350	23	10	64	206
Australian Natl U	Go8	5,595	–8	49	6	–3	7,536	35	24	42	71
U Queensland	Go8	3,987	33	42	20	–6	7,514	78	55	103	82
U Melbourne	Go8	5,170	22	37	28	–4	7,490	28	37	74	104
New S Wales	Go8	4,270	28	71	10	15	6,628	35	45	60	120
Monash	Go8	3,438	5	57	–6	3	5,386	49	33	67	103
U W Australia	Go8	3,054	1	70	74	26	5,052	47	60	48	112
Queensland U Tec	ex–IT	668	4	243	126	79	2,554	202	162	150	453
La Trobe U	pre1988	1,643	0	81	18	3	2,235	15	43	33	63
Flinders U	pre1988	1,930	13	17	36	–2	2,119	–3	6	77	76
U Tasmania	pre1988	1,143	16	71	14	23	2,021	60	66	69	128
U Newcastle	pre1988	1,222	12	65	7	0	1,891	17	47	89	158
Macquarie U	pre1988	1,138	35	87	18	–21	1,700	41	88	30	30
U Wollongong	pre1988	697	40	100	56	50	1,537	73	168	86	70
James Cook U	pre1988	680	1	95	7	144	1,451	115	51	95	94
Griffith U	pre1988	756	–2	116	65	100	1,250	35	66	122	76
U New England	pre1988	1,059	2	46	13	3	1,115	–5	–2	–10	17
Curtin U	ex-IT	409	36	52	50	120	1,113	56	156	154	241
Murdoch U	pre1988	805	–23	61	93	97	1,064	8	9	114	23
Deakin U	pre1988	491	44	115	142	97	880	80	75	56	68
RMIT	ex-IT	261	10	144	10	53	821	91	212	248	248
U Tec Sydney	ex-IT	250	25	116	4	40	772	97	174	168	172
U South Australia	ex-IT	170	53	15	59	–8	695	113	436	251	174
U Western Sydney	ex-CAE	155	617	180	750	375	680	275	289	197	305
Charles Sturt U	ex-CAE	113	20	157	88	29	364	226	248	173	400
Nth Territory U	ex-CAE	122	1150	2400	200	283	279	133	120	117	20
U Canberra	ex-CAE	128	21	293	58	–55	272	235	70	49	242

The universities with the most even growth across quartiles subsequent to the introduction of the funding formulas were University of Sydney, Australian National University, University of Queensland, La Trobe University and Deakin University, all with a standard deviation of less than 20.

5. DISCUSSION

Problems with the composite index, and in particular with the publications component, were raised soon after its introduction (Anderson, Johnson and Milligan, 1996). Most of the discussion concentrated on the Research Quantum (RQ) as it was the largest scheme. These concerns were taken on board in a ministerial discussion paper on higher education research and research training, issued in June 1999:

> "The publications component of the Composite Index has been subject to a range of criticisms since its implementation in 1995. These concern the reliability of the information provided by institutions, the costs of data collection and the incentives created by the inclusion of a publications component in the index. It seems likely that the publications component of the Composite Index has stimulated an increased volume of publication at the expense of quality … on these grounds, the Government proposes ... to drop the publications measure in any future indices used to allocate block research funds" (Kemp, 1999a).

Not all universities were keen to see the removal of the publications element. The notional proportion of the RQ to be distributed via the publications component was 10% in 1999. However, over half the universities, particularly smaller institutions, received more than 10% of their RQ allocation through publications. For one university the proportion was above 40%; for another five it was more than 20%. It was predominantly the research intensive older universities that were at, or even under, the 10% benchmark (DEST, 1999).

It was therefore hardly surprising that in its response to the discussion paper, the AVCC, representing all 36 institutions which received funds via the RQ, argued for the retention of the publications component:

> "… of the quality measures that might be utilised, 'publications' is the only measure able to fulfil all the requirements … for a driver of sector–wide funding" (AVCC, 1999).

The government was swayed by the submission of the AVCC and others, and in its final policy statement all talk of removing the publications component had disappeared (Kemp, 1999b).

Concerns also surfaced about the direction in which the publications component of funding formulas currently in place in the higher education sector was driving universities, when data produced by the *Institute for Scientific Information* (ISI) confirmed the marked increase in Australian output in the journal literature but pointed to a significant decline in citation impact relative to many OECD countries (Butler, 2001b).

The concerns raised back in 1999 about the use of an undifferentiated publication count are re-surfacing in the context of the latest review of the Australian higher education system. A number of submissions to the government review established to evaluate the Knowledge and Innovation reforms have suggested the removal or modification of the publications component (DEST, 2002c). Two questions stakeholders were specifically asked to address related to the publications collection:

> "Should the research publications element be removed from the formulae? Should the research publications element of the formulae include quality measures"(DEST, 2002c).

In their submissions the majority of institutions remain committed to the continuation of the collection. A number would like to see the introduction of quality measures, although generally this approach has been rejected because few have any knowledge of possible performance measures that could be used to approximate the notion of quality. Most appear to assume it means weighting publication counts by ISI's journal impact factor, or using aggregate citation counts, and have no knowledge of the more complex and sophisticated bibliometric methods that have been developed in recent years.

The University of Central Queensland highlights another problem with the collection in its existing form:

> "The resources used in collecting, submitting and verifying publications by institution exceed the income received for publications at Central Queensland University".

It is clearly apparent that before any alternative could be adopted there needs to be detailed assessment of possible measures. Several questions need to be examined. Do the proposed indicators come close to measuring the aspect of the research endeavour the government is targeting? Is the measure suitable for the level of aggregation being assessed? Is the measure applicable to all fields of research? Is the necessary data readily available and independently verifiable? Is it more effective to combine a suite of indicators, perhaps varying by field, rather than relying on a single measure? Only when these and other questions have been answered, and the effects of their introduction demonstrated, would there be any chance of gaining broad acceptance for the replacement of existing measures.

ACKNOWLEDGEMENTS

Certain data included herein are derived from the *Australian National Citation Report* prepared by the *Institute for Scientific Information®*, Inc. (ISI®), Philadelphia, Pennsylvania, USA: ©Copyright Institute for Scientific Information® 2000. All rights reserved. This chapter updates and extends research first reported in *Research Evaluation* in 2003 (Butler, 2003).

REFERENCES

OECD. (1997). *The evaluation of scientific research: selected experiences*. Paris: OCDE/GD(97) 194.
Jiménez–Contreras, E., Anegón, F.M., López–Cózar, E.D. (2003). The evolution of research activity in Spain: the impact of the National Commission for Evaluation of Research Activity (CNEIA). *Research Policy*, 32 (1), 123–142.
Adam, D. (2002). The counting house. Nature, 415, 726–729.
Australian Vice Chancellors' Committee (AVCC). (2002). Time series data relating to DEST higher education research data collection (http://www.avcc.edu.au/ australias_unis/ statistics/research/index.htm).
Department of Education, Science and Training (DEST) (2002a): (http://www.detya.gov.au/ highered/research/index.htm#funding.
Guena, A., Martin, B.R. (2003). University Research Evaluation and funding: an international comparison. Minerva, 41 (4), 277–304. http://www.sussex.ac.uk/spru/publications/ imprint/sewps/sewp71.html.
Department of Employment Education and Training. (1996). *Higher education report for the 1996 to 1998 triennium*. Canberra.
Department of Education, Training and Youth Affairs (DETYA). (2000). *Higher education report for the 2000 to 2002 triennium*. Canberra.
Department of Education, Science and Training (DEST).(2002b). *Higher education report for the 2002 to 2004 triennium*. Canberra.
Butler, L. (2001a). *Monitoring Australia's scientific research*. Canberra: Australian Academy of Science, 20.
Anderson, D., Johnson, R., Milligan, B. (1996). *Performance–based funding of universities*. Canberra: Commissioned Report No.51, National Board of Employment Education and Training.
Kemp, D. (1999a). *New knowledge, new opportunities*. Canberra: Department of Education, Training and Youth Affairs.

Department of Education, Science and Training (DEST) (1999). *Characteristics and performance indicators of higher education institutions 1999*, http://www.detya.gov.au/ archive/highered/statistics/characteristics/30_researchquantum.htm.
AVCC. (1999). *Discussion paper on higher education research and research training*. Canberra: AVCC.
Kemp, D. (1999b). *Knowledge and Innovation*. Canberra: DETYA.
Butler, L. (2001b). *Monitoring Australia's Scientific Research*. Canberra: Australian Academy of Science, p.11.
Department of Education, Science and Training (DEST) (2002c). *Setting firm foundations: financing Australian higher education*. Canberra: DEST.
DEST. *Evaluation of knowledge and innovation reforms: issues paper*. http://www.detya. gov.au/highered/ki_reforms/issues_paper.rtf.
Butler, L. (2003). Modifying publication practices in response to funding formulas. *Research Evaluation*, 12 (1), 39–46.