CHAPTER 5

JOCHEN GLÄSER AND GRIT LAUDEL

THE SOCIAL CONSTRUCTION OF BIBLIOMETRIC EVALUATIONS

A CONSTRUCTIVIST APPROACH TO BIBLIOMETRIC EVALUATION TOOLS

While national systems of research evaluation vary in many dimensions, they all need to rely on very few methods of evaluating research performance. These methods constitute a crucial interface between the science system and science policy through which information about research is translated into strategic knowledge for policy decisions. They therefore merit specific attention. This chapter is motivated by the observation that one type of evaluation methods – bibliometric evaluations – is rapidly spreading. Our aim is to identify the processes that drive this growth by creating an image of bibliometric evaluation methods as valid, reliable and therefore legitimate means of performance evaluation. We thus take up Woolgar's (1991) initiative for 'a sociology of measurement technologies and their use in science policies' and apply it to one important type of evaluation techniques, namely bibliometrics. However, we share neither Woolgar's interest in the reasons why bibliometricians have not accepted the fundamental criticisms levelled at citation analyses, nor the *radical* constructivist perspective he applies. Instead, we want to understand why bibliometric methods are being embraced by an ever-increasing number of users with an ever-decreasing regard for validity and reliability. Thus, we are not discussing professional – careful, methodologically sound, state-of-the-art – bibliometric evaluations, which in our opinion can be as valid or invalid as evaluations by peer review. We are interested in the rapid spread of these evaluations beyond the boundaries of this professionalism.

In order to understand this process, we will treat the application of bibliometric methods as a process of knowledge construction and apply insights of the constructivist sociology of scientific knowledge to their analysis. The premise that the interests and powers of actors co-shape construction processes (e.g. Barnes 1977; Pickering 1982) suggests identifying the actors who participate in the construction

R. Whitley and J. Gläser (eds.), The Changing Governance of the Sciences, 101–123.

of bibliometric evaluations, their interests concerning bibliometric evaluations, and their power to influence them. A second important insight concerns the construction of scientific facts. Ethnographic studies (Latour and Woolgar 1986) and citation context analyses (ibid.; Cozzens 1985) have demonstrated that the construction of scientific facts includes a process of decontextualisation. In this process, information surrounding the conditions under which scientific knowledge has originally been produced is gradually omitted. The decontextualisation has been described in the sociology of scientific knowledge as a process of omitting modalities. We apply this heuristic by treating statements about the applicability of bibliometric methods as modalities and by looking at the fate of these modalities in the hands of the different actors. These modalities are omitted (neglected, ignored, or denied) by interested and powerful actors who take part in the application of bibliometric methods for evaluative purposes.

We begin our discussion by clarifying what we regard as the important modalities of bibliometric methods by sketching why they can be regarded as valid under certain conditions, and what statements on limitations of bibliometric methods are embedded in the state of the art of bibliometric research. We then proceed by demonstrating how these modalities disappear because of the interests and powers of the four main actors involved, namely the users in science policy and science management, the commercial owner of the major citation databases, Thomson Scientific's Institute for Scientific Information (ISI), the community of professional bibliometricians, and amateur bibliometricians, i.e. academics, manager and politicians who apply bibliometrics in all sorts of ways without having the necessary professional background.

THE MODALITIES OF BIBLIOMETRIC EVALUATION METHODS

In the following analysis, we treat bibliometric methods as instructions for the collection and manipulation of empirical data. These instructions also contain information about the conditions under which a specific method is applicable, i.e. will yield valid and reliable results, and about the information that is produced by a specific method. This information can be thought about as 'when and how to use' – statements, i.e. as 'modalities' that are part of the instructions.

The concept of 'modalities' was introduced by Latour and Woolgar in their analysis of the construction of scientific facts and has been subsequently expanded by Latour (Latour and Woolgar 1986: 75-90; Latour 1987: 23-29). Modalities are 'statements about statements', i.e. statements that modify the validity and reliability of statements about scientific findings. For example, in the sentence "The structure of GH.RH was *reported to be* X" the modality "reported to be" modifies the statement about the structure of GH.RH. If the modality is dropped, we get "The structure of GH.RH is X", which is a different statement (Latour and Woolgar 1986: 78).

Latour later distinguished between positive modalities "that lead a statement away from its conditions of production, making it solid enough to render some other conditions necessary" and negative modalities "that lead a statement towards its conditions of production and that explain in detail why it is solid or weak" (Latour

1987: 23). A fact is seen here as a statement that is devoid of negative modalities and is linked to positive modalities. It is not accompanied by any information about the persons who produced the knowledge or about the conditions under which the knowledge was produced because this information is regarded as unnecessary. Latour and Woolgar describe the construction of facts as processes in which modalities are removed from statements, and the controversies about facts as processes in which interested parties add or remove modalities to statements in order to weaken them or to make them more solid (Latour and Woolgar 1986: 88; Latour 1987: 25-26).

This general description of fact construction can be applied to methods. Scientific methods contain statements about the conditions under which they can be applied and the way in which they must be applied in order to produce valid and reliable results. These statements are part of the 'prescription of operations' that represents the method. However, they can also be considered as modalities because they modify the methodical statements by limiting their applicability to specified conditions and ways of use. In scientific contexts these modalities are thought of as inseparable from the methods because any application of a method rests on assumptions about applicability and proper procedure. They are implicitly tested in every application of a method, but simply confirmed in most cases. Together with the methodical statements, they are the subject of methodological research, where they are challenged and changed in the attempts to apply a method to new problems. Laboratory studies observed that these modalities can be replaced by locally created ones, a practice which can lead to significant variations of what is thought of as a standard procedure. Such variations, when observed, are subject to serious criticisms by scientists (Knorr-Cetina 1981: 37-40; Latour and Woolgar 1986: 158).

In the remainder of this section, we will identify some of these inherent methodological modalities of bibliometric methods. As mentioned in the introduction, we start from the presupposition that bibliometric methods actually can measure aspects of research quality. Establishing this ability is difficult for the fundamental reason that it is not at all clear what research quality 'really' is. The discussions that circle around this topic usually settle on some statement like 'you know it when you see it' or, more scientifically, that the quality of research is a complex, field-specific property that ultimately can only be judged by members of the scientific community.

This would deny the possibility of any other analysis of research performance than peer review. However, bibliometrics has a strong point in stating that scientists cite their colleagues' work when it is useful for their argument, which means that the cited work has had a certain impact on the citing author's work, and that the usefulness and impact are aspects of research quality. The observation that citations indicate use, and therefore usefulness as well as impact, is the basic argument for using them as an indicator of quality. The reconciliation between the complex and indeterminate concept of research quality and the measurable property of impact has been achieved by regarding impact as *an aspect of* research quality that can be measured by bibliometric indicators, and that bibliometric indicators can therefore serve as 'partial indicators' of quality (Dieks and Chang 1976; Martin and Irvine 1983; Moed et al. 1985a; Weingart et al. 1988; Phillimore 1989; Van Raan 1996).

This interpretation provides the justification of bibliometric evaluations as they are conducted by professional bibliometricians. It also contains the first and most fundamental modality of all citation-based performance measures, namely the principal limitation of this type of evaluative technique that it does not measure quality per se but rather "an important aspect of quality" (Van Raan 1996: 404; see also Moed et al. 1985a: 133-135). Although impact is meanwhile regarded as "[t]he most crucial parameter in the assessment of research performance" (Van Raan 2000: 303), i.e. the modality has been weakened, bibliometricians are aware of the fact that *they do not measure quality*, which is one of the reasons why they regard it as essential that their results need to be interpreted by experts of the field (Moed et al. 1985a: 147; Van Raan and Van Leeuwen 2002: 614-615).

As readers would probably expect from their own experiences with citing and citation, many objections have been raised against the premise that citations indicate use or impact. It is not surprising that one major argument in the debate on bibliometric methods has always been that citations cannot be used as an indicator because they are given and withheld for a great variety of reasons (e.g. Gilbert and Woolgar 1974; MacRoberts and MacRoberts 1989). We refrain from discussing these reasons because we think that there is a convincing rebuttal of the whole argument:

> That is statistically only the case if all researchers refer to earlier work completely arbitrarily. But nobody can seriously maintain that the references in, for instance, this paper are totally unreasonably and completely arbitrary ….valid patterns in citations will be detected if a sufficiently large number of papers is used for analysis. Furthermore, it is statistically very improbable that all researchers in a field share the same distinct reference-biases. (Van Raan 1998: 134-135)

This argument is consistent with all the empirical work about reasons for citations of which we are aware. It is also consistent with both the 'normative' and the 'constructivist' theories of citation. These theories state that the act of citing is governed by compliance with norms ('to give credit where credit is due') respectively by micropolitical interests (to convince – 'enrol' – readers, see Cozzens 1989 for a description of both positions).[1]

The argument that citations *statistically* represent impact contains a second modality. Bibliometric methods must be applied to a larger number of publications for the statistics to become reliable. If enough publications exist and citation data can be accessed, it is possible to measure the one aspect of research quality that is measurable by bibliometrics, namely international impact. This modality has been

[1] Bibliometricians strongly prefer the 'normative' theory because it emphasises the impact of the cited on the citing publication. Their recent attempts to prove the 'normative' theory and to disprove the 'constructivist' theory misconstrue the latter as stating that authors cite highly reputed colleagues (Baldi 1998: 833, 835; White 2004: 93, 96-97) respectively highly cited papers (Moed and Garfield 2004: 295-297) in order to persuade their readers. This is an oversimplification because according to the 'constructivist' theory authors would not restrain themselves to citing authoritative colleagues or papers but would opportunistically cite anything that supports their arguments. Since the most useful tools of 'persuasion' (Gilbert 1977) or 'allies' (Latour 1987) are those that are thematically relevant to the author's research, the 'constructivist' theory would predict the same publications to be cited as the 'normative', namely publications that either have been used in the research leading up to the publication or are used in the publication to support an argument.

explicitly discussed and acknowledged by bibliometricians (Van Raan 2000: 307-309; Butler 2001: 49).

A third important modality of citation studies is technical in nature. For valid conclusions about research performance to be drawn, the whole research output of the evaluated unit must be covered by the analysis. In particular, the citation databases used for bibliometric analyses must enable counting citations to all publications of evaluated units. This is particularly important when bibliometric methods are applied at lower levels of aggregation, i.e. close to the minimal number of publications required to perform reliable statistical analyses. At the level of research groups, 99% completeness of publication data is necessary (Moed et al. 1985a: 139-140), and one missing well-cited publication can create a significant error (Smith 1981: 93; Nederhof 1988: 204).

A fourth modality of the application of bibliometric measures concerns time. Citations that indicate the use of publications and thus their quality occur some time later, and the validity and reliability of citation-based evaluation measures therefore depends on the time-frame chosen for analysis. Since publications reach their highest citation score after three years in many fields of the natural sciences, this time span is considered as the minimal 'citation window' in these fields, and bigger citation windows may be necessary in others (Moed et al. 1985a: 136; Van Raan 1996: 403). This means in turn that the most recent publications for which reliable citation data can be obtained are three years old at the time of the bibliometric evaluation. This trade-off between reliability and timeliness of bibliometric evaluation methods is an inevitable consequence of the modality concerning time.

The fifth basic modality is caused by the specificity of knowledge production in different fields. Since the practices of knowledge production vary between fields, so do publication and citation practices (Moed et al. 1985b). This implies that the results of bibliometric measurements are field specific and can neither be compared nor aggregated without normalising the results with field-specific reference values (Van Raan 1996: 403). Furthermore, the delineation of fields becomes a crucial task on which the validity of bibliometric evaluations depends.

These are the most basic and most important modalities of bibliometric methods. In the following section we will add several others after a closer look at the actual source of citation data. With these modalities taken into account, bibliometric methods can offer information about aspects of research quality. Sociologists of science were the first to use bibliometric methods for evaluative purposes in their studies of the reward system of science and of processes of stratification (Cole and Cole 1967, 1972, 1973). From these early studies, a line of research emerged that develops and uses bibliometric indicators to assess the quality of publications, scientists, organisations, and countries.

MARKETING BY DROPPING MODALITIES

Citation studies require data on the publications of scientists, research organisations, or academic fields, and on the frequencies of their citation. To date, the only databases that offer this sort of data are the Science Citation Index, the Social Science

Citation Index, and the Arts and Humanities Citation Index.[2] These databases are the property of the firm Thomson Scientific that sells online access to the databases, online access to specific indicators, and data derived from these databases to anyone who is interested.

This is not only an absolute monopoly, which is very rare in the economy, it also creates the unusual situation whereby a whole scientific community (the bibliometricians) depends on data that are not a public good but need to be bought. While private property of data occurs in other fields as well (in particular in the life sciences, see e.g. Marshall 1997; Brickley 2002), the problem faced by bibliometrics is much more severe because there is no competing public production of data. The major source of data for the community is privately owned.

This situation suppresses the central mechanism of quality control in scientific communities, namely the use of knowledge in subsequent knowledge production processes. Any such use is also a test of the knowledge that results in explicit or implicit confirmation or in the proposal of changes to that knowledge. Since the bibliometric data is shielded from the community by the property rights of Thomson Scientific, the usual feedback loops of using data, finding and reporting errors, and subsequent improvement of the data do not occur. Any flaws in the data remain there, and there is little improvement over time. This limits the research opportunities for bibliometricians, which was commented on by Glänzel and Schoepflin who wrote that "the databases fall short of the expectations of bibliometricians" (Glänzel and Schoepflin 1994: 380) Similarly, Barre mentions "a fragile and unstabilized situation regarding the supply and quality of data" (Barre 1994: 423; see also Bookstein 1994: 459). Furthermore, peer review of the data generation is impossible. Moed has summed up the results of an extensive analysis of Thomson Scientific's citation indices as follows:

> In our institute's huge analysis of more than 20 million cited references matched to 8 million target articles extracted from the *Science Citation Index* (*SCI*) and related ISI citation indexes, we found that when data are derived from 'simple' or 'standard' citation-matching procedures, citation statistics at the level of individuals, research groups, journals and countries are strongly affected by sloppy referencing, editorial characteristics of scientific journals, referencing conventions in scholarly subfields, language problems, author-identification problems, unfamiliarity with foreign author names and ISI data-capturing conventions. (Moed 2002: 731)

The many criticisms of products sold by Thomson Scientific do not lead to changes in the data. Meanwhile, many bibliometricians are convinced that criticising the owner of the database is a risky business as long as one still needs to buy data from them. While there is no evidence that Thomson Scientific exerts any pressure on the bibliometrics community, the belief that it is better not to contradict them is powerful enough, and certainly shapes actions according to the famous Thomas theorem: "If man believe a situation to be real, it is real in its consequences" (Thomas and Thomas 1928: 572). Since only one source of data is available, database-related

[2] A recently created competitive product (Elsevier's *Scopus*) has the potential to support bibliometric analyses as well.

modalities enter the description of bibliometric methods. The most crucial of these modalities are:[3]

- The databases' coverage of the literature and the changes over time in coverage must be taken into account. The literature of the social science, arts and humanities, but also of some of the engineering sciences and for example mathematics is not well covered by the databases. This undermines the validity of results that are obtained by the use of these databases for bibliometric evaluations in those fields.
- Publications must be unambiguously assigned to authors, which can be difficult because of spelling errors, or homonyms, i.e. different authors having the same last name and initials.
- Publications must be unambiguously assigned to organisations, which is difficult because of incomplete and erroneous information.

Thomson Scientific drops these and many other modalities when offering its own products for research evaluation such as *Essential Science Indicators*, *Highly-Cited.com*, *National Science Indicators* that contain evaluative information about specific countries; *Citation Laureates* (the most highly cited academics) who have been named in award ceremonies in several countries (e.g. Japan, Australia, and Denmark); and evaluations of all sorts published in the bimonthly newsletter *Science Watch*. Examples of dropped modalities include:[4]

- The journal impact factor which is calculated as the average number of times articles from the journal published in the past two years have been cited in the current year. The serious flaws of this factor have been criticised for twenty years now without success (for a recent review, see Glänzel and Moed 2002). Here it suffices to notice that the modality concerning time is dropped and an extremely short citation window (less than two years for one half of the publications and less than one year for the other half) is used. Thomson Scientific has implicitly acknowledged this modality by adding a guide for the calculation of a five year impact factor to its website.
- *Essential Science Indicators'* citation ranking of scientists ignores homonyms. The helpfile indicates that each name in the list refers to one scientist. However, it also adds the interesting comment that "[s]cientists having the same last name and initials may represent multiple individuals". Thus, the modality of citation analysis – that unambiguously assigning publications to scientists is a prerequisite for valid analyses – is acknowledged, but the data and ranking lists remain unchanged.
- In 1995, bibliometricians published the following critique:

> In 1992, several publications appeared in the journal *Science Watch* – published by the Institute for Scientific Information – in which impact indicators have been calculated for universities or even research departments in the field of chemistry. […] We have strong indications that ISI has made several severe errors in their assignment of papers to universities. ISI has not taken into account all variations under which the name of a

[3] An extensive list of modalities that includes the modalities discussed by us and many more has been provided by van Raan (1996: 402-404).

[4] Examples are taken from the websites of Thomson Scientific's "Web of Knowledge" (Homepage: http://www.isiknowledge.com), which was accessed on April 28th, 2006.

> university appears in the addresses. For instance, ISI seems to have missed important variations in the name of the University of Leiden. As a consequence, the number of papers assigned to this university […] is much too low. (Moed et al. 1995: 411-412).

Four years later, the same criticism – that error in addresses produce invalid rankings of universities – was published with regard to the then-marketed product *University Indicators on Diskette*, which proved to be inaccurate in the case of Australia (Butler 1999). Today, *Essential Science Indicators'* citation ranking of institutions still largely ignores variations in addresses. The helpfile states under "Name variations" that "[i]nstitutions may appear with different name abbreviations, in which case more than one entry may need to be consulted". However, some articles are not counted at all due to name variations. Using the 'advanced search' feature of the *Web of Science*, we found 212 Papers that contained only 'ANU' (the common acronym of 'Australian National University', that is not listed in the *Essential Science Indicators*) and checked for the first ten papers and for the most highly cited paper (196 citations) that the addresses in question indeed referred to the Australian National University. Again, the modality of unambiguous assignment is acknowledged, but not acted upon.

The marketing efforts of Thomson Scientific and their impact on science policy have been recently dealt with by Weingart (2005). For our purposes it is important to notice that the products that are marketed are largely devoid of modalities, which means that the validity of the products is questionable. Naturally, the impact of dropped modalities is not mentioned by the seller. Even the information about modalities mentioned above is well hidden. It can be found on the 'Help' sites, i.e. on sites one usually turns to when encountering problems with the intended search. For unsuspicious users, the bibliometric information marketed by Thomson Scientific looks simple and straightforward because most modalities have been removed.

THE DEMAND FOR MODALITY-FREE BIBLIOMETRIC EVALUATIONS

As early as 1975, science policy and management recognised the potential of bibliometric evaluations. In what is probably the first report on this use of bibliometric methods outside the bibliometrics community, Wade (1975: 429) mentioned among other examples the use of bibliometric information by US universities in their decisions on promotion and tenure and by the US National Science Foundation in its assessment of the funding of chemistry departments.

Since then, bibliometric evaluations have diffused into many policy and management processes. Over the last two decades, this diffusion has been driven by the adoption of the paradigm by science policy. This paradigm has at its core the belief that market competition and market exchange are the best way of conducting *any* public task regardless of its content, and that there is no better way of solving allocation problems or producing efficiently. Market competition is supposed to improve efficiency because only the most efficient producers survive, and the market pressure on efficiency forces producers to adopt the most efficient internal structures. Thus, in order to achieve the most efficient conduct of public tasks, these tasks need to be assigned to autonomous units that compete for the resources needed to produce the required outcome. The application of the market paradigm to public

administration has led to the conviction that its organisations should adopt corporate structures and practices, and that competition between them needs to be introduced in order to make sure that public money is only given to the best performers, and is used efficiently by them (James 2001: 233).

The application of 'new public management' to science has resulted in a change of the major governance instruments (e.g. Slaughter and Leslie 1997; Henkel 2000; Marginson and Considine 2000). Governments are consciously or subconsciously introducing market structures into their science systems (see Engwall and Nybom, this volume). The autonomy of research organisations, albeit always limited by their dependence on public funding, is increasing. Competition between research organisations and between researchers is increasing because more and more money is distributed in the form of competitive funding. In many countries, performance-based funding of universities has been introduced over the last three decades or is currently being introduced (see Whitley, this volume). Universities employ their newly gained autonomy by managing research performance and research conditions, a practice which creates a strong trend towards hierarchical corporate structures that weaken the traditional academic self-governance (Morris 2002; Schimank 2005; Gläser and Laudel, this volume).

Both the market constellations and the corporate structures that are evolving in the science system require comparative assessments. The functioning of markets rests on the comparability of products and prices because without comparability a competition for exchange opportunities would be impossible. The management of organisations requires comparisons of the performance of a subunit with that of comparable subunits of competitors, other subunits within the organisation, and the past performance of the same subunit. These needs for comparable performance assessments are the reason why evaluations have been spreading throughout many science systems. Evaluations are the only way to render the idiosyncratic processes of knowledge production comparable, and to make them manageable to outsiders who lack the scientific knowledge to 'manage by content'.

The two basic methods of research evaluation – peer review and quantitative indicators – can both be used in comparative evaluations. In order to achieve comparability with peer review processes, reviewers are usually asked to rank the subjects of their evaluation, thus providing a relationship between the applications that abstracts from their idiosyncratic content and can be used by non-scientists. The now classical example of this use of peer review is the UK's Research Assessment Exercise, in which more than 60 panels of assessors evaluate the outcomes of university research.

The other, more recent evaluation strategy that is being applied more and more widely uses quantitative indicators of research performance. There are three main reasons for the recent rapid growth of the demand for 'metrics', i.e. for quantitative and among them bibliometric methods (see Weingart 2005: 122 for a similar argument). Firstly, the increasing demand for evaluations cannot be met by peer review. The ability and willingness of academics to serve as reviewers of their colleagues is limited, and is already strained by the increasing demand for everyday review activities due to the rise of competitive funding and publications. It is unlikely that peer review alone could meet the demand for evaluation. A related reason for preferring

bibliometric evaluations over peer review is that the latter appear to be cheaper. While it is not entirely clear that carefully conducted professional bibliometric evaluations are necessarily cheaper, the costs of a congregation of assessors are obviously significant.

Secondly, bibliometric evaluations appear to be legitimated by scientific practices as well as being objective and therefore more trustworthy than peer review. Their legitimacy stems from the fact that they rest on cumulated peer judgements. Most publications must pass peer review, and citations indicate peer judgements on usability and impact. Being based on a multitude of those judgements made by a large number of researchers, bibliometric methods appear to overcome the idiosyncrasies and biases of small, all-powerful groups of ad-hoc assessors. Their objective character makes them an ideal tool for routine intraorganisational evaluations, where any kind of peer review would either be very costly (if external assessors are employed) or would put an enormous strain on intraorganisational personal relations if scientists from the same organisation acted as assessors. After all, the peer review of colleagues violates norms of the scientific profession (Schimank 2004).

Thirdly, bibliometric evaluations give the impression of being accessible by politicians and managers without the involvement of scientists. In peer review assessments, recommendations are inextricably linked to idiosyncratic judgements of quality and potential. The numbers produced by bibliometric evaluations appear to be decontextualised and thus can be more easily processed than qualitative judgements by assessors who are outsiders to the policy and management processes.

These perceptions, which make bibliometric evaluations appealing to science policy and management, already miss some of the previously described modalities. They are convincing only as long as it is ignored that the results of bibliometric measurements need to be interpreted by scientists from the field, that they only provide information about past performance, and that they are not equally applicable at all levels of aggregation or to all fields. These and other modalities are readily dropped by science policy and management when requesting or conducting bibliometric evaluations. A leading bibliometrician described this demand:

> Quite often I am confronted with the situation that responsible science administrators in national governments and in institutions request the application of bibliometric indicators that are not advanced enough. They are aware of this insufficient quality level, but the want to have it 'fast', in 'main lines', and not 'too expensive'. (Van Raan 2005a: 140)

While bibliometricians can stand firm in these situations and decline providing 'cheap and dirty' evaluations, science policy and management can often use their discretion to initiate modality-free bibliometrics. We provide examples from three different countries, which cover the policy, intraorganisational collective and individual levels (for more examples, see Adam 2002; Cameron 2005: 112-115).

(1) The following quote from a report on the use of performance indicators in the evaluation of the humanities and social sciences in the Netherlands describes a radical change of mind of Dutch science policy concerning the modalities of bibliometric methods. After observing that in the 2001 Observatory, bibliometric methods

were not applied to the humanities because of their divergent publication culture, the report continues:

> In its 2003 Observatory, however, the Ministry presents tables of 'relative citation-impact scores of Dutch universities, by discipline', as related to the world average by discipline. In these calculations, literary studies in the Netherlands, for example, score much higher than the world average, whereas this indicator is for example much lower for law. Apparently, the Ministry's officials are already applying their own research evaluation methods, while the researchers are still reflecting on the adequacy of such methods, thereby also seemingly neglecting the cautionary notes that are contained in the methodological appendix to the 2000 Observatory. (KNAW 2005: 9)

The initial refusal to apply bibliometric methods to the humanities was based on an acknowledgement of modalities concerning the coverage of the Dutch humanities by Thomson Scientific's databases, which is generally low in the humanities, even lower for journals in languages other than English, and does not include books. Two years later, the modalities were dropped.

(2) Germany is currently introducing performance-based funding in universities and research organisations. A Committee of the Deutsche Forschungsgemeinschaft, Germany's major Research Council, issued recommendations on the performance-based resource distribution in medical university departments (DFG 2004). The recommendations acknowledge the modalities of quantitative performance measurement in general and Thomson Scientific's journal impact factor in particular, and propose the development of procedures and criteria for the assessment of the content of publications. Observing that the development and trial of such procedures is a research project in itself and requires time, the report proposes to use the impact factor as a substitute in the meantime because it is an "albeit imprecise, but relatively inexpensive fallback solution" (ibid.: 15, our translation). The conclusion to this section reads as follows:

> For the calculation of the performance-based funding … the evaluation of original publications can be conducted in terms of a stepwise introduction of quality criteria by using the unweighted impact factors of the respective journals. (ibid., our translation)

Thus, the most important modality of all – that bibliometric methods do not measure quality – and the numerous modalities of the impact factor are dropped for administrative convenience.

(3) In an ongoing project on the evaluation-based funding of Australian university research (see Gläser and Laudel, this volume), an historian described aspects of the promotion procedure (note the shocked responses of the interviewer, who knew about the modalities):

> Historian: … In fact, I have now - for promotion, I've had to look at my citations.
>
> Interviewer: What?
>
> Historian: What you have to do - - -
>
> Interviewer: You go into these citation indices?
>
> Historian: More than that. More than that, because that citation index is not very useful. It's limited. I mean, so I found about 330 citations of my work, but only probably 50 off that. I've found them in other places.

Since the literature in history is not at all well covered by Thomson Scientific's *Arts and Humanities Citation Index*, the historian collected the publications of his field in which citations to his work were likely to occur, and counted these citations manually. The interviewee's promotion was at stake, and he had no choice but to comply with the bureaucratic requirement that ignored all modalities about the validity of citation counts at the individual level, the coverage of the literature by existing databases, the field-specificity of publication and citation practices, etc.

WORKING ON MODALITIES

To conduct a valid bibliometric analysis requires extracting data from the databases of Thomson Scientific, cleaning them, and analysing them carefully by applying state-of-the-art methods. This is what professional bibliometricians do. Many bibliometricians conduct this kind of service as consultancies for political actors or research organisations. However, this is not their major interest. Bibliometrics is an academic field that is aimed at the advancement of knowledge about quantitative analyses of the published output of science (in the widest sense). In the context of our analysis this means that the bibliometrics community works on the modalities of bibliometric methods. Bibliometricians change modalities of existing methods by refining the methods, or they overcome modalities by inventing new methods. Their designing and testing of methods and data establishes, reshapes and erases modalities. The bibliometrics community is the gatekeeper of bibliometric methods and their modalities.

Unfortunately, the bibliometrics community suffers from structural problems that severely limit its opportunities to act as a watchdog for the correct applications of bibliometric methods. The following account is mainly based on the published contributions in a discussion about a 'crisis' of bibliometrics, which occurred in 1993 and 1994.[5] While many discussants disagreed with the diagnosis of a crisis, the structural problems were more or less confirmed in this discussion and in later contributions. The major problems that have an impact on the work with modalities of bibliometric evaluations are:

Epistemic fragmentation and isolation. Glänzel and Schoepflin noticed "failing communication" leading to "parallel studies on the same issue but with completely divergent conclusions and a Babylonian chaos in terminology". They also observed "a lack of consensus in some fundamental questions" and the "drifting apart" of the sub disciplines of scientometrics (Glänzel and Schoepflin 1994: 377). The authors later confirmed the centrifugal trend by a bibliometric study of the journal *Scientometrics* (Schoepflin and Glänzel 2001).[6] The fragmentation is partly driven

[5] Glänzel and Schoepflin presented a paper on the crisis theme in the closing session of a bibliometrics conference in Berlin 1993. The paper and responses from many bibliometricians were subsequently published in a dedicated issue (No. 2-3 of Volume 30) of the journal *Scientometrics* in 1994.

[6] The 2001 article lists six "categories", namely "(1) Bibliometric theory, mathematical models and formalisation of bibliometric laws; (2) Case studies and empirical papers; (3) Methodological papers including applications; (4) Indicator engineering and data presentation; (5) Sociological approach to bibliometrics, sociology of science; and (6) Science policy, science management and general or technical discussions" (Schoepflin and Glänzel 2001: 305). These categories are later referred to as

by the unequal access to data. We already commented on the problematic situation created by the private ownership of the community's major data source (problematic for both the data and the community). Glänzel and Schoepflin hinted at a resulting split of the community:

> On the economic side, there are in fact two classes of bibliometric research groups by now: the ones who can afford to buy expensive data sets, process complex data analyses and plan long-term bibliometric research programs, and the others who cannot do all this. (Glänzel and Schoepflin 1994: 379)

This situation contributes to epistemic fragmentation because some groups develop advanced bibliometric methods that cannot be applied by anyone else, and analyses conducted with these methods cannot be replicated, as the following quote indicates:

> Our study shows that, even for an extensive analysis covering many institutions and thousands of publications, it is possible to extract citation data for non-source items. One major caveat applies to this. For any such analysis to be undertaken efficiently and effectively, direct access to the raw data behind the ISI (or similar) indexes is an essential pre-requisite. (Butler and Visser 2006: 340)

This access only exists in a very few places in the world. We know of no bibliometric research group that has this access and uses it to advance bibliometric methods – except for the bibliometrics group at Leiden University, to which one of the authors of the article (Visser) belongs. The epistemic fragmentation has reached a level where the most advanced findings cannot be replicated by members of the bibliometrics community.

The epistemic fragmentation is accompanied by an epistemic isolation of bibliometrics from other fields of science studies. This isolation has been implicitly acknowledged by Glänzel and Schoepflin (1994: 377, 381), and has later been the subject of several bibliometric analyses of science studies journals (Leydesdorff and Van den Besselaar 1997; Van den Besselaar 2000, 2001). Two attempts to renew the links between quantitative and qualitative science studies (see Leydesdorff 1989; Leydesdorff and Wouters 1996) have not led to changes in the situation. While this isolation of bibliometrics from more qualitatively and theoretically oriented science studies is lamentable, it can by no means be attributed to epistemic changes of bibliometrics alone. The microsociological turn of the sociology of science plays a major role in the separation because the current sociology of science largely neglects aggregate units of analysis such as scientific specialties, to which the statistical approach of bibliometrics applies (Gläser 2001; see also Van Raan 1998).

Weak institutionalisation. Glänzel and Schoepflin wrote that there are only "very few … educational programs for informetrics/ bibliometrics/ scientometrics/ technometrics at universities and colleges" (Glänzel and Schoepflin 1994: 378) and mentioned a split between research groups with and without access to bibliometric raw data (see above). This assessment was confirmed by two respondents to their paper for the situation in Latin America and in the USA.

> Because of the multi- and interdisciplinary nature of our field, it does not easily fit into the hierarchical structure of traditional knowledge fields, and even less so, when

"sub-disciplines" (ibid.: 311). The authors observe that by 1997, the field has become dominated by the sub-disciplines (2) and (3).

> administratively and functionally universities, such a those in Latin America, are bifurcated into science on the one side and humanities on the other. (Russell 1994: 408)

> [In the United States there] are, at the bottom line, few people and virtually no resources for scientometrics; it is a small field with little perceived relevance to major societal goals. (Griffith 1994: 490)

This situation has not changed, and has even become worse in some countries because of the financial pressures on universities.

Lack of standards. The epistemic fragmentation described above already suggests that there might be a problem with missing standards. Glänzel and Schoepflin reinforce this issue by stating that the lack of consensus among the different research groups renders them unable to actively defend their scientific standards (Glänzel and Schoepflin 1994: 381). The bibliometrics community has indeed been troubled by a lack of standards for some time, which becomes clear from the responses to the Glänzel and Schoepflin paper by Vinkler (1994: 499), van Raan (1994: 531), and Luukkonen (1994), and from a workshop dedicated to the topic (see Glänzel et al. 1996 for an overview ; Glänzel 1996; Vinkler 1996). This lack of standards becomes apparent in the current discussion of citation-based performance measures for the social sciences. The most important modality concerns the coverage of the published output in the social sciences by Thomson Scientific's databases, which is regarded as insufficient for valid evaluations by most bibliometricians. However, the position of the bibliometrics community is by no means unambiguous, as the following statements indicate:

- In a report for the UK's Economic and Social Research Council, Katz did not conduct citation impact analysis for most of the social science fields. He stated that "[s]ocial science research is published in a wider variety of publication types and addresses more national issues than natural science research. This makes the construction of internationally comparable bibliometric indicators somewhat problematic." (Katz 1999: i)
- The contrary position is taken by Godin: "one of our main conclusions is that *bibliometrics is as good a tool for measuring the social sciences as it is for the natural, biomedical and engineering sciences*" (Godin 2002: 4, emphasis in original). Godin acknowledges the problem of coverage but counters that "the internationalization of the social fields is changing researchers' publication practices" (ibid.: 11). He backs his argument by referring to Hicks (1999). Hicks indeed observes a trend towards internationalisation of social science (ibid.: 206-208) but states in the summary of her review that "[n]one of the authors discussed here believe SSCI-based bibliometric indicators alone can form a basis for evaluation" (ibid.: 212), which is exactly what Godin has done (Godin 2002: 4).
- Van Leeuwen uses the report by Katz and the paper by Godin as references to back the following statement: "The last couple of years, studies for example in the UK and Canada have indicated the possibilities and advantages of applying bibliometrics in the evaluation of social sciences." (Van Leeuwen 2006: 133) Shortly thereafter, he states that "[t]he main argument against applying bibliometric techniques in the evaluation of social sciences research has always been

(and still is) the (poor) coverage of the social sciences by ISI's SSCI …." (ibid.: 133-134).

- When challenged by one of us (Gläser 2006) for letting a bibliometric evaluation of sociology departments pass regardless of its ignorance of numerous modalities (Sternberg and Litzenberger 2005), the reviewer for the journal scientometrics responded by stating *inter alia* "Without a doubt, the research contributions of economists and social scientists can be gauged from the attention their work receives in journals indexed by the SSCI and EconLit." (Daniel 2006: 332).
- A review of the literature on bibliometric evaluations of the social sciences reaches the conclusion that in the social sciences, "the same bibliometric methods can be applied as in science, but with several extensions. In particular, a broader range of both publications (including non-ISI journals and monographs) and indicators is needed in many social sciences and humanities." (Nederhof 2006: 96)

These contradicting statements indicate that there is no consensus on bibliometric evaluations of the social sciences and humanities. Given the epistemic fragmentation and weak institutionalisation of the community, it is difficult to see how a consensus could emerge. Until it exists we can expect some bibliometricians to conduct bibliometric evaluations which are considered invalid by others.

Commercialisation. The perception of bibliometrics being dominated by science policy and business interests was a major concern of Glänzel and Schoepflin, who considered it to be one of the causes of fragmentation and diminishing quality of some bibliometric studies. The increasing commercialisation of bibliometrics has been confirmed by several respondents, not all of whom, however, interpreted commercialisation as causing a crisis.

> Scientometric research has become indirectly dominated by the interest group "science policy and business". Its interest is clearly focused on "prompt" and "comprehensible" indicators, while the state of knowledge would allow the application of more sophisticated methods. Moreover, such research-reports tend to be only partially published and without the necessary methodological enhancements, which reduces its value for the bibliometrics community. As a consequence a clear shift away from basic and methodological research towards applied bibliometrics can be observed. (Glänzel and Schoepflin 1994: 380)

> When "commercialization" entered the field of scientometrics, it is certain that the commercial value helped develop the discipline. However it is also certain that the field which had originally been "supply oriented" (or "hobby like" affair) has been transformed to "demand-pull" affair. Analysis has to meet the clients' (decision makers) needs, in some cases must even limit observations to what the clients want to see, "science" thus becomes "a good" which must sell. (Miquel 1994: 444)

> Yes, too rapid commercialization is probably one of the biggest threats for our research field. Quick and dirty citation analyses, rainy-Sunday-afternoon publication analyses, stupid hit-lists, completely irresponsible applications of specific methods like co-citation analysis, already did a lot of harm to our research community. (Van Raan 1994: 531)

> Latin American bibliometric research has often been of the kind sponsored by government and scientific bodies interested in raw (and often unqualified) data on

> national scientific achievement for feeding science policy. The production of data without sound theoretical and methodological foundations which is then used to back-up policy decisions implicates the bibliometricians and scientometricians supplying the original data. (Russell 1994: 409)

Bibliometrics has some features of a research technology as described by Joerges and Shinn (Joerges and Shinn 2001), particularly a "generic quality" and the development of a "metrology". However, bibliometric methods are scarcely used as research methods in fields other than bibliometrics itself. The technology does not diffuse into fields of application, either. The bibliometrics community is not an 'interstitial' community but homogeneously institutionalised in the academic sector. Therefore it is better seen as a specialty whose common body of knowledge consists of specific data and methods. The problems of epistemic fragmentation and isolation, weak institutionalisation, lack of standards, and commercialisation, albeit not necessarily symptoms of a crisis, contribute to the inability of the bibliometrics community to guard bibliometric methods. They are partly due to the variety of audiences and funding sources (see Whitley, this volume). However, because of its weak institutionalisation the bibliometrics community strongly depends on the 'soft money' from a variety of sources rather than being able to use this money to shield itself from adverse effects, as has been predicted by Whitley. The diverse audiences and funding sources 'reach through' and fragment the community. Modalities of methods and data are *local* constructs by bibliometric research groups rather than a *global* state of the art of the community. For many bibliometricians, commercial clients are a more influential reference group than the bibliometrics research community. Interests are heterogeneous, too. While there is a widespread epistemic interest in the advancement of bibliometric methods, the dependence of many groups on income from clients creates a parallel commercial interest in satisfying these clients.

Without a consistent interest in defending the modalities or even an agreement on what they are, the bibliometrics community is in no position to fulfil the role usually ascribed to professional experts – at least not beyond the local spheres of influence of a few advanced research groups.

AMATEUR BIBLIOMETRICS

The image of bibliometric methods has undergone a significant change. Initially, they were met with 'hostility' and the threat of legal action (Garfield 1979: 360; Weingart 2005: 117-118). Ten years later Glänzel, Schoepflin and others worried about damage to an otherwise good image of bibliometric methods. Today, according to Weingart, we observe "a dramatic shift away from the well founded skepticism to an uncritical embrace of bibliometric numbers" (ibid.: 119). One of the reasons for this climate change that has not yet been sufficiently acknowledged is 'grass roots bibliometrics' or, more precisely, 'amateur bibliometrics'. We use the latter term to denote the practice of producing bibliometric analyses of an evaluative character by actors with little or no professional background in the field, and with little or no knowledge or regard for the modalities involved.

This practice is more widespread than even the most pessimistic outsider would assume. The use of individual-level bibliometrics in organisational decisions about tenure and promotion has already been mentioned. Today, many organisations routinely use bibliometric indicators for the performance evaluation of their academics (for the US American Library and Information Science, see Meho and Spurgin 2005). Funding agencies in the biomedical sciences often request the submission of journal impact factors or citation counts for the publications contained in the investigators' CVs, thus forcing applicants to conduct amateur bibliometrics. At higher levels of aggregation, ranking exercises that are partly or entirely based on bibliometric indicators proliferate at all levels of aggregation.

Being amateur bibliometrics, all these practices have in common the ignorance of the modalities of bibliometric analyses. They represent extreme cases in which the perceived need for evaluations has overridden any concerns about validity. A characteristic case in point is the 'Shanghai ranking', whose gross misfit of validity and international political impact had moved one of the leading bibliometricians to expose the serious methodological flaws of the exercise (Van Raan 2005a). The authors of the ranking answered that any technical flaws in the database (ISI's citation databases) are the sole responsibility of the authors of the articles contained in the database, and refused to acknowledge any responsibility of the owner of the database or the producer of the evaluation (Liu et al. 2005). Van Raan responded by emphasising the responsibility of the evaluator:

> Of course, the use of a certain description of an affiliation is and will always remain the sole responsibility of authors. But this author-related responsibility is for articles only. As soon as somebody wants to make 'constructions' (i.e., indicators) on a higher aggregation level that constitute an entirely new added value in order to transform 'the world of individual authors and publications' into 'the world of evaluation', it is the responsibility of this person or institute to define an affiliation as good as possible on the basis of any available information that must go beyond the 'local' responsibility of an author for his or her article. (Van Raan 2005b: 111)

Amateur bibliometrics reflects the rapidly increasing demand for evaluations, which it promotes at the same time. It is supported by two interwoven trends. Firstly, the growing competition for funding drives actors to use any resource that might improve their position. Thus, any evaluation of questionable validity will be actively marketed by those organisations that find themselves on top or at least above their major competitors (see Weingart and Maasen, this volume). Similarly, scientists who perform well become increasingly interested in opportunities to 'objectively' demonstrate that they are better than their colleagues. Suppose the historian who collected citations to his publications as part of his application for promotion is successful, and on a later occasion perceives himself as competing with colleagues who he thinks are less well cited than him. There is at least a chance that he will advocate citation analysis (possibly by conducting it for his own work) in order to win the competition for funding or to advance his career. Secondly, the increasing specialisation of research forces scientists to use second-order criteria when judging research performance. In many fields, specialisation has reached a stage where scientists must rely on proxy criteria for performance such as the reputation of colleagues, organisations, and journals. Apart from the informal communication in

which reputation is discussed, numbers of citations, ranks of journals and external funding are adopted as surrogates for the judgement of content. When asked what he would regard as an extraordinary scientific contribution in his field, a scientist answered:

> Our measure comes from impact factors of journals. You know, if you are looking for a global measure, that's how you do it: is your work published in *Science* or *Nature*. They are the gold standard. And they are the gold standard because they accept even for review a very small number of the submitted materials. I think it is much more difficult to apply another standard because standards vary within certain branches of disciplines. So, I am an experimental plant biologist, essentially a plant physiologist and you know there are twenty or thirty branches.

As the quote indicates, scientists are left with the second-order criteria because there is no other way of assessing the quality of contributions from areas which are too remote to understand them but are nevertheless necessary to know about. This occurs in peer review evaluations when assessors cannot judge the content of research (as they should) because it lies outside their area of competence. A common response is resorting to properties of the publication such as the rank of the journal, citations, etc. (Gläser and Laudel 2005).

The need to use crude bibliometric indicators to judge research performance as part of their everyday work and the need to constantly market performance jointly contribute to the acceptance and even active introduction of some person counting another person's or an organisation's citations, that is amateur bibliometrics. An unwanted side effect of amateur bibliometrics is that it lowers the methodological standards of bibliometrics in general, because for amateur bibliometricians, none are needed. The widespread acceptance of bibliometric evaluations resulting from amateur bibliometrics is due to the personal experience of the numerous amateur bibliometricians (anybody can count citations) and of those who do not believe in the validity of these exercises but feel overwhelmed by the ubiquity of amateur bibliometrics. The major danger of this development is that peer review might loose its character as an independent counterpoint to bibliometric evaluations when the assessors – being amateur bibliometricians themselves – simply trust the numbers.

CONCLUSIONS

The discussion of the actors who jointly promote bibliometric evaluations has demonstrated that their interests align well. Science policy and management want quick and easy-to-handle evaluations, and get them because (a) they distribute resources in the science system in a way that requires this sort of evaluation and (b) they are powerful patrons of bibliometrics, which in turn is not coherent and independent enough to create and defend standards of bibliometric conduct. Since neither the clients, nor the owner of the data source, nor the background chorus of amateur bibliometricians care for the modalities of bibliometric methods, and since there is no strong counterweight, modalities are easily dropped. This is not to say that there are no considerate, well constructed bibliometric evaluations. There are many of

those, but they are not as cheap as many clients need them, and the few groups who can produce them cannot meet the demand.

The strong demand for bibliometric evaluations that contribute to the disappearance of modalities is at least partly caused by bibliometrics itself. In a dynamics that has been described by Rip as the "promise-requirement cycle" (Rip 1997: 628-632), bibliometrics has promised methods that can be used to evaluate research performance (e.g. Garfield 1979), a promise that has been turned into a performance expectation by science policy. Bibliometrics thus confronts the ghosts it called.

Our discussion of the conditions needed for evaluations explains why bibliometric evaluations in general are growing so popular.[7] They are valid in certain areas and under certain conditions; they give the impression of being easy to understand because they come as numbers; they are actively marketed by a commercial enterprise; professional bibliometricians depend on conducting them; and amateur bibliometricians are spreading the message without caring about risks and side-effects. We have also made it clear why in this process modalities of the methods are dropped. All the actors involved implicitly contribute to the construction of bibliometric evaluations as a universal tool that measures research quality, even though none of them would describe them as such. The whole – the use of bibliometric techniques as a universal, modality-free evaluation tool – is certainly more than the sum of its parts, i.e. of the numerous individual applications of bibliometric methods and discussions about them. The 'subtracted value' in form of missing modalities is the result of a complex construction process whose elements we have traced in this article.

What can be done? It is difficult to imagine a way in which the urge to evaluate could be contained. Therefore, it becomes important to make clear to everyone that appropriate bibliometric evaluations are not cheap, and to discuss the use that is made of them. Apart from that, two major steps suggest themselves. First, science policy should support the *professionalisation of bibliometric evaluations* by securing the independence of the profession from both its source of data and its customers, and by strengthening 'disinterested' academic work on the qualification of bibliometric methods. Systematic training of bibliometricians as well as some dedicated 'watchdog activities' could be expected in return. The latter should include the introduction of a code of ethics and quality control guidelines and mechanisms. Secondly, the creation of a public citation database would enable the quality control of data and thus help to overcome many of the described problems. Both steps require major policy efforts, international coordination, and investments. They would therefore only be undertaken if detrimental effects of 'quick and dirty' bibliometric evaluations make themselves felt, e.g. when powerful actors in academia or the public lose their trust in science policy because of invalid evaluations. Since the validity of all research evaluation exercises is inherently limited, and since excellent research is 'robust' enough to come out top in most

[7] In this article, we have focused on bibliometric methods. The situation concerning the second major quantitative evaluation method – calculating external grant income – is even worse. No systematic research on this indicator has been conducted, and we are just learning about its modalities (Laudel 2005). For very much the same reasons as described in section 4, the indicator is nevertheless widely applied by science policy and management.

evaluations, it is unlikely that the problems of misconstrued bibliometric evaluations will ever become pressing enough to warrant major investments.

This means that we will continue to live with the status quo for a longer time. But we should not despair! In all of its history, the science system as a whole has proven to be very adaptable to all sorts of external changes. It will be interesting to see how it adapts to bibliometric evaluations.

REFERENCES

Adam, David (2002), 'The counting house', *Nature*, 415, 726-729.

Baldi, Stephane (1998), 'Normative Versus Social Constructivist Processes in the Allocation of Citations: A Network-Analytic Model', *American Sociological Review*, 63, 829-846.

Barnes, Barry (1977), *Interests and the Growth of Knowledge*, London: Routledge & Kegan Paul.

Barre, R. (1994), 'Do Not Look for Scapegoats - Link Bibliometrics to Social Sciences and Address Societal Needs', *Scientometrics*, 30, 419-424.

Bookstein, A. (1994), 'Scientometrics: new opportunities', *Scientometrics*, 30, 455-460.

Brickley, Peg (2002), 'A Scrap over Sequences, Take Two', *The Scientist*, 16, 13 May 2002.

Butler, Linda (1999), 'Who 'Owns' this Publication? Problems with assigning research publications on the basis of addresses', *Proceedings of the Seventh Conference of the International Society for Scientometrics and Informetrics, Colima, México, July 5-8, 1999,* pp. 87-96.

Butler, Linda (2001), *Monitoring Australia's Scientific Research: Partial indicators of Australia's research performance,* Canberra: Australian Academy of Science.

Butler, Linda and Martijn S. Visser (2006), 'Extending citation analysis to non-source items', *Scientometrics*, 66, 327-343.

Cameron, Brian D. (2005), 'Trends in the Usage of ISI Bibliometric Data: Uses, Abuses, and Implications', *portal: Libraries and the Academy*, 5, 105-125.

Cole, Jonathan R. and Stephen Cole (1972), 'The Ortega Hypothesis', *Science*, 178, 368-375.

Cole, Jonathan R. and Stephen Cole (1973), *Social Stratification in Science,* Chicago: The University of Chicago Press.

Cole, Stephen and Jonathan R. Cole (1967), 'Scientific Output and Recognition, a Study in the Operation of the Reward System in Science', *American Sociological Review*, 32, 377-390.

Cozzens, Susan E. (1985), 'Comparing the Sciences: Citation Context Analysis of Papers from Neuropharmacology and the Sociology of Science', *Social Studies of Science*, 15, 127-153.

Cozzens, Susan E. (1989), 'What Do Citations Count? The Rhetoric-First Model', *Scientometrics*, 15, 437-447.

Daniel, Hans-Dieter (2006), 'The referee's comments on Gläser's criticism', *Scientometrics*, 67, 331-333.

DFG [Deutsche Forschungsgemeinschaft] (2004), *Empfehlungen zu einer 'Leistungsorientierten Mittelvergabe' (LOM) an den Medizinischen Fakultäten - Stellungnahme der Senatskommission für Klinische Forschung der Deutschen Forschungsgemeinschaft*, Bonn: DFG.

Dieks, Dennis and Hans Chang (1976), 'Differences in impact of scientific publications: Some indices derived from a citation analysis', *Social Studies of Science*, 6, 247-267.

Garfield, Eugene (1979), 'Is Citation Analysis a Legitimate Evaluation Tool?', *Scientometrics*, 1, 359-375.

Gilbert, G. Nigel and Steve Woolgar (1974), 'Essay Review The Quantitative Study of Science : An Examination of the Literature', *Science Studies*, 4, 279-294.

Gilbert, G. Nigel (1977), 'Referencing as Persuasion', *Social Studies of Science*, 7, 113-122.

Glänzel, Wolfgang (1996), 'The need for standards in bibliometric research and technology', *Scientometrics*, 35, 167-176.

Glänzel, Wolfgang, Sylvan Katz, Henk Moed and Urs Schoepflin (1996), 'Preface', *Scientometrics*, 35, 165-166.

Glänzel, Wolfgang and Henk F. Moed (2002), 'Journal impact measures in bibliometric research', *Scientometrics*, 53, 171-193.

Glänzel, Wolfgang and Urs Schoepflin (1994), 'Little Scientometrics, Big Scientometrics ... And Beyond', *Scientometrics*, 30, 375-384.

Gläser, Jochen (2001), 'Scientific specialties as the (currently missing) link between scientometrics and the sociology of science', in Mari Davis and Concepción S. Wilson (eds.), *Proceedings of the 8th International Conference on Scientometrics & Informetrics,* Sydney, Australia, July 16-20th, BIRG, UNSW, pp. 191-210.

Gläser, Jochen (2006), 'Letter to the Editor', *Scientometrics*, 67, 327-329.

Gläser, Jochen and Grit Laudel (2005), 'Advantages and dangers of 'remote' peer evaluation', *Research Evaluation*, 14, 186-198.

Godin, Benoit (2002), *The Social Sciences in Canada: What Can We Learn From Bibliometrics?' Project on the Measurement of the Social Sciences*, Working Paper no. 1, Quebec: INRS.

Griffith, Belver C. (1994), 'Little Scientometrics, Little Scientometrics, Little Scientometrics, Little Scientometrics ... And So on and So On', *Scientometrics*, 30, 487-493.

Henkel, Mary (2000), *Academic Identities and Policy Change in Higher Education*, London: Jessica Kingsley.

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

James, Oliver (2001), 'Business Models and the Transfer of Businesslike Central Government Agencies', *Governance: An International Journal of Governance and Administration*, 14, 233-252.

Joerges, Bernward and Terry Shinn (2001), 'A Fresh Look at Instrumentation: An Introduction', in Bernward Joerges and Terry Shinn (eds.), *Instrumentation Between Science, State and Industry*, Dordrecht: Kluwer, pp. 1-13.

Katz, J. Sylvan (1999), *Bibliometric Indicators and the Social Sciences*, Sussex: SPRU.

KNAW [Koninklijke Nederlandse Akademie van Wetenschappen] (2005), *Judging research on its merits: An advisory report by the council for the Humanities and the Social Sciences Council,* Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Knorr-Cetina, Karin (1981), *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science,* Oxford: Pergamon Press.

Latour, Bruno (1987), *Science in Action*. Cambridge, Mass: Harvard University Press.

Latour, Bruno and Steve Woolgar (1986 [1979]), *Laboratory Life: The Construction of Scientific Facts*, Princeton: Princeton University Press.

Laudel, Grit (2005), 'Is external funding a valid indicator for research performance?' *Research Evaluation*, 14, 27-34.

Leydesdorff, Loet (1989) 'The Relations between Qualitative Theory and Scientometric Methods in Science and Technology Studies: Introduction to the Topical Issue', *Scientometrics*, 15, 333-347.

Leydesdorff, Loet and Peter van den Besselaar (1997), 'Scientometrics and communication theory: Towards theoretically informed indicators', *Scientometrics*, 38, 155-174.

Leydesdorff, Loet and Paul Wouters (1996), 'Quantitative Measuring or Qualitative Understanding: Is it possible to bridge the divide in STS?' *EASST Review*, 15, 20-24.

Liu, Nian Cai, Ying Cheng and Li Liu (2005), 'Academic ranking of world universities using scientometrics - A comment to the "Fatal Attraction"', *Scientometrics*, 64, 101-109.

Luukkonen, Terttu (1994), 'Are We Longing for the Golden Era Lost or for the One to Come', *Scientometrics*, 30, 481-485.

MacRoberts, Michael H. and Barbara R. MacRoberts (1989), 'Problems of Citation Analysis: A Critical Review', *Journal of the American Society for Information Science*, 40, 342-349.

Marginson, Simon and Mark Considine (2000), *The Enterprise University: Power, Governance and Reinvention in Australia,* Cambridge, UK: Cambridge University Press.

Marshall, Eliot (1997), 'Snipping Away at Genome Patenting', *Science*, 277, 1752-1753.

Martin, Ben R. and John Irvine (1983), 'Assessing Basic Research. Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, 12, 61-90.

Meho, Lokman I. and Kristina M. Spurgin (2005), 'Ranking the Productivity of LIS Faculty and Schools: An Evaluation of Data Sources and Research Methods', *Journal of the American Society for Information Science and Technology*, 56, 1314-1331.

Miquel, J.F. (1994), 'Little scientometrics, big scientometrics ... and beyond', *Scientometrics*, 30, 443-445.

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

Moed, Henk F., W.J.M. Burger, J.G. Frankfort and A.F.J. Van Raan (1985a), 'The Use of Bibliometric Data for the Measurement of University Research Performance', *Research Policy*, 14, 131-149.

Moed, Henk F., J.M. Burger, J.G. Frankfort and Anthony F. J. Van Raan (1985b), 'The application of bibliometric indicators: important field- and time-dependent factors to be considered', *Scientometrics*, 8, 177-203.

Moed, Henk F., R.E. De Bruin and Thed 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.

Moed, Henk F. and Eugene Garfield (2004), 'In basic science the percentage of 'authoritative' references decreases as bibliographies become shorter', *Scientometrics*, 60, 295-303.

Morris, Norma (2002), 'The developing role of departments', *Research Policy*, 31, 817-833.

Nederhof, A. J. (1988), 'The validity and reliability of evaluation of scholarly performance', in Anthony J. F. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology,* North-Holland: Elsevier, pp.193-228.

Nederhof, Anton J. (2006), 'Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review', *Scientometrics*, 66, 81-100.

Phillimore, A.J. (1989), 'University Research Performance Indicators in Practice: The University Grants Committee's Evaluation of British Universitites, 1985-86', *Research Policy*, 18, 255-271.

Pickering, Andrew (1982), 'Interests and analogies', in Barry Barnes and David Edge (eds.), *Science in Context. Readings in the Sociology of Science,* Milton Keynes: The Open University Press, pp. 125-146.

Rip, Arie (1997), 'A Cognitive Approach to Relevance of Science', *Social Science Information*, 36, 615-640.

Russell, Jane M. (1994), 'Back to the Future for Informetrics', *Scientometrics*, 30, 407-410.

Schimank, Uwe (2004), 'Leistungsbeurteilung von Kollegen als Politikberatung', in Thomas Brüsemeister (ed.), *Die beratene Gesellschaft: Zur gesellschaftlichen Bedeutung von Beratung*, Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 39-55.

Schimank, Uwe (2005), ''New Public Management' and the academic profession: Reflections on the German situation', *Minerva*, 43, 361-376.

Schoepflin, Urs and Wolfgang Glänzel (2001), 'Two decades of "Scientometrics": An interdisciplinary field represented by its leading journal', *Scientometrics*, 50, 301-312.

Slaughter, Sheila and Larry L. Leslie (1997), *Academic Capitalism: Politics, Policies, and the Entrepeneurial University,* Baltimore: The Johns Hopkins University Press.

Smith, Linda C. (1981), 'Citation Analysis', *Library Trends*, 30, 83-106.

Sternberg, R. and T. Litzenberger (2005), 'The publication and citation output of German faculties and disciplines based upon SSCI data', *Scientometrics*, 65, 29-53.

Thomas, William I. and Dorothy Swaine Thomas (1928), *The Child in America: Behavior Problems and Programs*, New York: Knopf.

Van den Besselaar, Peter (2000), 'Communication between science and technology studies journals: A case study in differentiation and integration in scientific fields', *Scientometrics*, 47, 169-193.

Van den Besselaar, Peter (2001), 'The cognitive and social structure of STS', *Scientometrics*, 51, 441-460.

Van Leeuwen, Thed N. (2006), 'The application of bibliometric analyses in the evaluation of social science research. Who benefits from it, and why it is still feasible', *Scientometrics*, 66, 133-154.

Van Raan, Anthony F. J. (1994), 'Little scientometrics, big scientometrics ... and beyond', *Scientometrics*, 30, 529-531.

Van Raan, Anthony F. J. (1996), 'Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises', *Scientometrics*, 36, 397-420.

Van Raan, Anthony F. J. (1998), 'In matters of quantitative studies of science the fault of theorists is offering too little and asking too much - Comments on: Theories of citation?', *Scientometrics*, 43, 129-139.

Van Raan, Anthony F. J.(2000) 'The Pandora's Box of Citation Analysis: Measuring Scientific Excellence - The Last Evil?' in Blaise Cronin and Helen Barsky Atkins (eds.) *The Web of Knowledge,* Medford, New Jersey: Information Today Inc., pp. 301-319.

Van Raan, Anthony F. J. (2005a), 'Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods', *Scientometrics*, 62, 133-143.

Van Raan, Anthony F. J. (2005b), 'Reply to the comments of Liu et al', *Scientometrics*, 64, 111-112.

Van Raan, Anthony F. J. and Thed 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.

Vinkler, P. (1994), 'Words and indicators. As scientometrics stands', *Scientometrics*, 30.

Vinkler, P. (1996), 'Some practical aspects of the standardization of scientometric indicators', *Scientometrics*, 35, 237-245.

Wade, Nicholas (1975), 'Citation Analysis: A New Tool for Science Administrators', *Science*, 188, 429-432.

Weingart, Peter (2005), 'Impact of bibliometrics upon the science system: Inadvertent consequences?', *Scientometrics*, 62, 117-131.

Weingart, Peter, R. Sehringer and M. Winterhager (1988), 'Bibliometric indicators for assessing strength and weaknesses of West German science', in Anthony J. F. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology,* North-Holland: Elsevier, pp. 391-430.

White, Howard D. (2004), 'Reward, persuasion, and the Sokal Hoax: A study in citation identities', *Scientometrics*, 60, 93-120.

Woolgar, Steve (1991), 'Beyond the Citation Debate: Towards a Sociology of Measurement Technologies and Their Use in Science Policy', *Science and Public Policy*, 18, 319-326.