

Judging research on its merits

**An advisory report by the Council for the Humanities
and the Social Sciences Council**

Royal Netherlands Academy of Arts and Sciences
Amsterdam, May 2005

© 2005. Royal Netherlands Academy of Arts and Sciences

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photo-copying, recording or otherwise, without the prior written permission of the publisher.

P.O. Box 19121, 1000 GC Amsterdam, the Netherlands

T +31 20 551 07 00

F +31 20 620 49 41

E knaw@bureau.knaw.nl

www.knaw.nl

ISBN 90-6984-449-4

The paper in this publication meets the requirements of  iso-norm 9706 (1994) for permanence.

Preface

Scientific research is creative work performed by qualified researchers in an international setting. Researchers meet their peers at scientific conferences, have their work reviewed by the editorial boards of scientific journals, and submit proposals for research grants or tenured appointments to peer committees for assessment. Consequently, it is fair to say that research is subject to *ex ante* peer evaluation on an almost continuous basis. It is hardly a matter of dispute that the work of researchers in all disciplines should also be subject to systematically peer evaluation. However, the evaluation systems and the use of particular performance indicators are topics of heated debate.

The Council for the Humanities (RGW) and the Social Sciences Council (SWR) has observed a number of recent developments that have implications for the evaluation of research in the Netherlands. They include the following.

It is expensive to evaluate research, especially if we count the amount of time researchers spend on the research evaluation process. They may even be involved in several research evaluation procedures at once. One can only hope that the benefits of the evaluations outweigh the costs.

The Dutch Minister of Science is studying an assessment system based on performance indicators, to be used throughout the sciences. The system may have consequences for the way government funds university research.

The use of performance indicators has become a topic of vigorous debate in the international scientific community. One can scarcely expect researchers in the humanities and the social sciences to accept the performance indicators used in the natural sciences as valid in their own field. As a result, alternative methods are required. The success of any evaluation procedures and instruments depends on their being accepted by the relevant researchers.

In the light of the above, the RGW and the SWR have articulated the need to assess the procedure by which research in the humanities and the social sciences is evaluated. A joint *ad hoc* committee was asked to analyze the specific characteristics of this procedure and to consider the role of bibliometric indicators. The basic principle is that research in the humanities and the social sciences should be judged on its own characteristics and features.

Although the committee had a difficult task, it has drafted a report with a number of interesting and practical recommendations. We trust that the report will be useful to researchers in the humanities and the social sciences, as well as to the research leaders and directors of research institutes in these fields. It will provide them with sound instruments for performing independent but fit-to-purpose research evaluations.

On behalf of the Council for
the Humanities
Wim Blockmans
Chair

On behalf of the Social Sciences
Council
Jacques Thomassen
Chair

Contents

1	Introduction	9
2	Some complexities of research evaluations	13
2.1	Introduction	13
2.2	Evaluating research in the humanities and the social sciences	14
2.3	The role of bibliometric indicators	17
2.4	Conclusion	20
3	Contours of an evaluation method for research in the humanities and the social sciences	21
3.1	Introduction	21
3.2	A tentative list of target groups	22
3.3	A tentative list of indicators	22
3.4	Conclusion	24
4	Recommendations	25
	Aanbevelingen	27
	References	29
	Appendices	31
	1. Abbreviations	33
	2. Bibliometric indicators and peer review	34

There appears to be a growing demand for evaluations of research activities in the public domain. The competition for scarce resources for research stimulates the search for generally accepted criteria enabling granting agencies and authorities to make rational and objectively arguable choices in the allocation of these resources. Apart from such extrinsic factors, an intrinsic reason for researchers to welcome evaluation is that it may help to raise the quality of future research, depending on the suitability of the evaluation method.

For example, in the medical and natural sciences, bibliometric indicators and especially the measurement of the impact of articles written by researchers, either directly through the number of citations, or indirectly through the impact factors of the media in which these articles are published, have become standard practice, applied in all kinds of research evaluations, from the individual to the institutional level. Although some fields in the humanities and the social sciences do conform to this practice to a greater or lesser extent, there is no generally accepted standard for evaluating research activities in the public domain. Reasons for this situation will be reviewed in the second section of this report. The third section contains an outline of an approach to research evaluation that should be adequate with respect to all fields of research - the medical and natural sciences, and the humanities and the social sciences alike -, albeit with differences in emphasis according to the research field, the institute, or even the individual researcher to whom it is applied.

In its biannual Observatory of Science and Technology, the Dutch Ministry of Education, Culture and Science (ocw) noted for the year 2000 that:

‘... the humanities are not taken into consideration, not for reasons of any lack of international orientation or lack of quality, but because of their divergent publication culture, putting a much greater emphasis on the use of Dutch and books as means of scholarly communication. Because of these fundamental differences, the humanities (...) are not taken into account in the following analyses, with the exception of linguistics and literary studies where publication in international specialist reviews is relatively frequent.’¹

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.² 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.

¹ Nederlands Observatorium van Wetenschap en Technologie. *Wetenschaps- en technologie-indicatoren 2000/2001*. OC&W, Zoetermeer 2001, p. 44.

² Nederlands Observatorium van Wetenschap en Technologie. *Wetenschaps- en technologie-indicatoren 2003*. OC&W, Zoetermeer 2003, table 5.14.

³ More fundamentally, the ministerial procedure is at odds with the principle that peer reviews, not indicators, constitute the core of research evaluations.

The need for recognized research evaluation methods is broadly accepted in the scientific community. It is also recognized that unsuitable methods can only lead to debatable conclusions. The research council in the Netherlands (NWO) observed that law, if assessed by the same international criteria as other disciplines, would miss opportunities, in favor of, for example, psychology and economics.⁴ NWO chairman Nijkamp therefore called for a European 'design of benchmark procedures for research assessment'. A reflection on the function of the humanities and the social sciences, and on the methods used to evaluate performance in this regard, is most urgently required. The ongoing research evaluation processes would acquire greater authority if they were based upon generally accepted methods.

In actual practice in The Netherlands, ex post research evaluations are performed within the context of the universities and research institutes, following a protocol. In 2003, the Association of Universities in the Netherlands (Vereniging van Samenwerkende Nederlandse Universiteiten, VSNU), the Netherlands Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, NWO) and the Royal Netherlands Academy of Arts and Sciences (Koninklijke Nederlandse Akademie van Wetenschappen, KNAW) published a new protocol for evaluating research activities, the SEP (Standard Evaluation Protocol 2003-2009 For Public Research Organizations). However, the SEP does not present a rigid research evaluation method. It indicates that evaluating research activities is a far more complex matter than for instance counting articles in high-ranking journals:

'It has to be noted (...) that the elaboration of these (evaluation) criteria may differ for different fields. Because publication traditions and contextual relations vary among different fields, articles in high-ranking journals, for example, are much more telling and accepted as indicators in some fields than in others. This goes for the distinction at large between scientific areas (natural sciences, social sciences, humanities, medical sciences, agricultural sciences, technical sciences) as for sub fields in these areas. Having said that, the main criteria are elaborated as a guideline for the evaluators.'⁵

The SEP mandates Dutch researchers and research directors to elaborate the evaluation criteria. The present report may be viewed as a step in that direction.

This report is aimed primarily at researchers in the humanities and the social

³ See page 1: 'De publicatiedekking is beperkt in het geval van de technische wetenschappen, de sociale wetenschappen en zeker bij de alfawetenschappen. In deze gebieden wordt de kennisverspreiding, internationale (h)erkenning en wetenschappelijke invloed in mindere mate bepaald door internationale tijdschriftartikelen'.

⁴ Dr. J.K. Koppen, Director of the NWO Social and Behavioral Sciences Board, as quoted in *Mare*, 4 November 2004, p. 5.

⁵ VSNU, NWO, KNAW 2003, p. 9.

sciences, but also at the scientific directors who are research leaders. The report is written in English, thereby enabling the foreign members of (future) research evaluation committees to examine its deliberations. Furthermore, this report aims to be instrumental in the context of the internationalization of research evaluations. Research is increasingly being evaluated within an international context. The development of a keen interest in the EU for research in the humanities and social sciences makes a reflection on the function of - and research evaluation methods for - these fields of science even more urgent.

This report has been drafted by a joint committee from the Council for the Humanities (RGW) and the Social Sciences Council (SWR) of the KNAW. The task of the committee was to analyze the specific characteristics of the evaluation procedure of research activities in the humanities and the social sciences, including the role of bibliometric indicators.⁶

The RGW-SWR committee consisted of:

- Dr. K.A. Algra, professor of ancient and medieval philosophy, University of Utrecht.
- Dr. J.M. Bensing, chair; professor of psychology of health, University of Utrecht; director, Netherlands Institute for Research on Health Care (NIVEL).
- Dr. W.P. Blockmans, professor of medieval history; rector, Netherlands Institute for Advanced Study in Humanities and Social Sciences (NIAS).
- Dr. J.C.H. Blom, professor of history; director, Netherlands Institute for War Documentation (NIOD).
- Dr. E.E.C. van Damme, professor of economics, University of Tilburg.
- Dr. W.K.B. Hofstee, em. professor of psychology, University of Groningen.
- Dr. A. Vollering, secretary of the SWR, served as the secretary of the committee.

The final edition of this report was prepared by Dr. W.P. Blockmans and Dr. W.K.B. Hofstee.

⁶ 'De commissie heeft als taak te analyseren op welke wijze geesteswetenschappelijk en sociaal-wetenschappelijk onderzoek dient te worden beoordeeld en welke rol sciëntometrisch onderzoek hierbij dient te hebben.'

2.1

Introduction

As a matter of principle, research evaluations are a form of peer review: a committee of scholars assesses the merits of the research program. In contemporary practice, however, such judgments are supported by systems of indicators, spelled out in part by evaluation protocols.

When a research evaluation system is established – whether it is based on bibliographic indicators, on peer review, a combination of both methods, or another method – sooner or later the actors in the system will be affected by the characteristics of that system. This will be especially true when their performance with respect to these characteristics is pivotal to the future allocation of research funding.

An example of this norm-oriented behavior among researchers can be observed in Dutch official statistics. As the probable consequence of the premium awarded to the number of dissertations and ‘scientific publications’, these categories in particular underwent an increase in a period of fifteen years. The number of dissertations rose by 120% (1986: 1153, 2001: 2534), that of ‘scientific publications’ by 60% (1986: 31988, 2001: 51192), and the ‘specialist publications’ (‘vakpublicaties’) by a mere 33% (1986: 12078, 2001: 16065).⁷ The research policy did not only lead to a higher overall ‘productivity’, but also to shifts in emphasis on types of publications.

Another example has come to be known as the ‘knowledge paradox’. In some fields of the social sciences, such as economics, researchers are encouraged to compete with their colleagues on a global scale. This competition forces them to focus on numbers of dissertations, ‘scientific publications’ in leading journals, and the like. However, the resulting pattern of specialization does not necessarily match the economic knowledge needs of the local societies in which these researchers are working. It is not therefore surprising that the concept of the knowledge paradox (on the one hand an abundant supply of fundamental knowledge, and on the other, huge demand for applied research) has emerged as a concern, not only in the Netherlands but also at European level.

In this respect, the comments in Holcombe (2004) on the consequences of the rankings published by the National Research Council (NRC) in the US are also revealing.⁸ Holcombe writes:

The NRC rankings affect research activity within some departments because the administrations at some universities have set an explicit goal of improving their NRC rankings (p. 499).

The NRC rankings create an incentive to hire a particular kind of scholar. Doing interesting work, being intelligent and working hard are all good things, but if they lead to publications, citations and grants, as measured by the NRC, that is

⁷ VSNU DOC Onderzoek, Table 24 ff. Although the 2002 numbers of dissertations, scientific and specialist publications are available at VSNU, these numbers suffer from a disturbing factor and are therefore not comparable with the numbers of 2001 and earlier.

⁸ Randall Holcombe: ‘The National Research Council Ranking of Research Universities: Its Impact on Research in Economics’, *Economic Journal Watch* 1 (3), December 2004, 498–514.

even better (p. 512).

This emphasis on research and graduate education has also robbed academic economics of some of its relevance to real-world economic phenomena.

Research departments give virtually no credit to faculty who write magazine articles, newspaper editorials, policy reports or other material designed to enlighten the general public (p. 512).

In general, the effectiveness of research evaluation methods depends on the effect of allowing incentive-related indicators into research funding schemes. One should be aware of the direct and indirect effects of incentives before introducing them into the research evaluation process, especially when these incentives have consequences for the future allocation of research funding.

2.2

Evaluating research in the humanities and the social sciences

This section emphasizes some of the specific features of research in the humanities and the social sciences that should be taken into account in research evaluations. Above all, one should consider the function these disciplines have in society. A reflection of this societal function is to be found in the types of publications produced in various disciplines.

An interesting study in this regard was performed by Kyvik (2003) with respect to faculty members at Norway's four universities. Table 1 presents some results from his study. It shows the percentage of faculty members who published popular science articles and contributions to public debate, and average number of these articles in the period 1998-2000, by field of learning.

Table 1

	Popular science		Contributions to public debate	
	%	No.	%	No.
Humanities	64	2.9	45	2.2
Social sciences	60	2.4	52	2.2
Natural sciences	44	1.4	23	0.8
Medicine	43	1.8	32	1.0
Technology	38	1.1	25	0.8
Total	51	2.0	36	1.4

Source: Kyvik, 2003

In the 1998-2000 period, half of the faculty members published at least one popular science article and more than one third published at least one contribution to the public debate. There are large variations between research fields, however. Researchers in the humanities and the social sciences are more engaged in such activities than their colleagues in the natural sciences, medicine, and technology.

The same pattern of involvement in popular science and contributions to public debate could probably be observed in faculty members at universities in the

Netherlands, but no research has been conducted on this aspect. An obvious source is to be found in the universities' and research institutes' annual reports, listing publications by categories. The vsNU and the Ministry statistics discern three types of relevant publications: dissertations, 'scientific' publications, and 'specialist' publications ('vakpublicaties'). A fourth category, 'other publications', which might include contributions to the public debate, reviews, and the like, is excluded from the official statistics. The emphasis on these three types of publications shows that the public authorities have a keen interest in the quantifiable output of the research investment, and that some kind of categorization is applied already, even if it remains unclear for what purpose all these figures might be used. They make it equally clear that the authorities have not recognized the value of contributions to the public debate, as counted in the Norwegian research.

Table 2.
Numbers of publications in The Netherlands in 2002, by field of learning.

Fields	Number of 'specialist' publications	Numbers of 'scientific' publications	'Disciplinary' publications in relation to 'scientific' publications
Humanities	2152	4648	46.3
Social Sciences	6696	13416	49.9
- Economy	1162	3430	33.9
- Law	2820	3766	74.9
- Behavioral / Social	2714	6220	43.6
Natural Sciences	744	7318	10.2
Medicine	2243	14405	15.6
Technology	2236	8405	26.6
Agriculture	381	2403	15.9
Miscellaneous	150	280	53.6
Total	14602	50875	28.7

Source: vsNU

In Table 2, a clear differentiation in publication culture is evident between law, the humanities and the social/behavioral sciences whose publishing is much more slanted toward 'specialist' channels (> 40%), economics, and notably, also technology, which have intermediary positions, and the natural sciences, medicine and the agricultural sciences, which are distinctly much less orientated toward this type of media (<18%).

Debackere, Glänzel and Schoepflin⁹ provided convincing quantitative material to expand these observations. They demonstrated among other things the disproportionate representation and coverage of fields and sub-fields in the humanities and social sciences and the problematic selection criteria in the existing indexes. This is

⁹ Debackere and Glänzel, 2004; Debackere and Glänzel, conference publication; Glänzel and Debackere, 2003; Glänzel and Schoepflin, 1999.

particularly relevant since the publication types in these disciplines are much more varied than in, for example, the natural sciences. They also illustrated the fact that journals represent more than 50% of the publications in mathematics, chemistry, physics and biology, but only around 30% in the humanities and social sciences.¹⁰

The mission of the humanities and the social sciences is different from, for example, the mission of the natural sciences and medicine. Researchers in the humanities and the social sciences study cultural and societal issues that may have a direct impact on policy makers, managers, the judiciary, and public opinion. They respond to a demand that requires specific forms of communication, such as a public report, a legal advisory or policy paper, or a publication for the general public. Those publications targeted at professionals, politicians, or the general public, are frequently in the national language. For some domains, such as national law, national politics, social and economic policy, addressees are also compatriots for whom the use of an international medium does not make much sense in general. This does not preclude national issues from also being debated within an international scientific context.

As examples, one might quote the scientific reports such as the research on the Srebrenica genocide.¹¹ Exhibitions often attract hundreds of thousands of visitors, and catalogues containing new scientific research outcomes are sold by tens of thousands. The recently published authoritative volume on the history of Amsterdam, authored by eminent researchers, had an equal success.¹² The same applies to some well-written books in psychology, history or literary studies. The conclusion must be that these disciplines have a much broader and more stratified audience than merely international scientific reviewers. The communication of knowledge to this audience has to be evaluated in an appropriate manner. Even though researchers in the humanities and the social sciences may communicate primarily with their peers, they also have a role in the communication of knowledge to a non-peer audience, and they tend to be more involved in the public debate than their colleagues in other research fields.

Particular target groups for scientific results have to be taken into account in the research evaluation process, in order to weigh the impact of research in these domains.¹³ Target groups also correspond to specific content and forms of publication. For example, research on Spanish literature, or on Islamic law, does not necessarily reach its most expert and concerned target groups if it is published in an English-language journal. It has to be recognized in the research evaluation process that such journals also have their particular focus, priorities, and preferences.

Target groups may partly be found outside the scientific community, as is usual for legal studies or policy-oriented research. Target groups are not simply to be considered as passive consumers of research, since they are in many cases the best

¹⁰ Debackere and Glänzel 2004, 6.

¹¹ J.C.H. Blom a.o., eds., *Srebrenica, een 'veilig' gebied: reconstructie, achtergronden, gevolgen en analyses van de val van een 'Safe Area'*. 3 vols., Amsterdam: Boom 2002, 3393 pp. An English translation is available on www.srebrenica.nl, a summary in Serbo-Croatian is ready for publication.

¹² M. Carasso-Kok, ed., *Geschiedenis van Amsterdam*, vol. 1, Amsterdam: SUN 2004, 540 pp.

¹³ Council for Medical Sciences, 2002; Spaapen and Dijkstra, 2005.

informed and concerned users of the information. Another example of this is demonstrated in the case of health services research:

The main groups of users (of health services research) are the political community, the civil service, and organized groups within the health care field.¹⁴

In conclusion, research fields differ with respect to research issues, methodologies, target groups and communication patterns. Research evaluations therefore have to be tailor-made. Evaluation standards are required which are based upon the current or desirable forms of communication within disciplines. The great variety of the target groups and, consequently, of the most appropriate communication means, needs to be recognized, following the mission of the humanities and the social sciences.

2.3 The role of bibliometric indicators

In the year 2000, the Standing Committee for the Humanities of the European Science Foundation (ESF-SCH) launched an inquiry into the value of the existing Arts and Humanities Citation Index (AHCI), and was forced to conclude that it does not qualify as an adequate research evaluation instrument. The selection of reviews is heavily biased towards the English-speaking world, and moreover, other media such as monographs and collective volumes are not taken into account.

The AHCI of the ISI (Institute for Scientific Information, Philadelphia) is obviously deficient and should not be used by scientific policy makers in Europe. (...) such indexes rarely include the best journals published outside the USA, especially those in languages other than English (...) There is an urgent need for a European Citation Index in the Humanities, even if it is only an additional tool for research evaluation, and would not be the only way of evaluating research quality.¹⁵

While similar strong statements ('should not be used') have not been made with respect to ISI's Social Science Citation Index, Glänzel and Schoepflin (1999) have demonstrated the heavy overrepresentation of US and UK publications in the existing indexes, and thus of the English language in these indexes. They also pointed to the different role of citations in the various research fields: in medicine, chemistry and physics 84 to 94% of the total references refer to serials, whereas this figure is only 40% in sociology. Furthermore, the mean reference age is considerably higher in the latter than in the former research fields. They concluded that:

'... the model of information transfer from scientific literature to scientific (journal) literature assumed by standard bibliometrics requires substantial revision before valid results can be expected through its application to social science areas.'

¹⁴ J.M. Bensing, W.M.C.M. Caris-Verhallen, J. Dekker, D.M.J. Delnoij and P.P. Groenewegen, 'Doing the right thing and doing it right; toward a framework for assessing the policy relevance of health services research', *International Journal of Technology Assessment in Health Care*, 19:4, 611-2.

¹⁵ A. Peyraube, Project for Building a European Citation Index for the Humanities. In *Reflections*, ESF, Strasbourg, December 2002, 14-5.

More generally, one has to observe that bibliometric indicators are derived from communication patterns in sciences other than the humanities and the social sciences, and that the reflection of these indicators is not synonymous with quality. In particular, the impact factor has been overestimated as an evaluation instrument, since the applied citation window (one to two years) is too short in almost all research fields, especially in the humanities and the social sciences. Next, impact factors reflect the citation culture in a certain research field. Because in practice citation cultures between research fields are very different, comparing impact factors across research fields is like comparing apples and oranges. In all research fields review articles are much more frequently cited than research articles. Impact factors therefore tend to overemphasize journals that give more attention to review journals (journals containing more review articles) than to research journals (journals containing more research articles).¹⁶

With the input from its member organizations, the ESF-SCH initiated the creation of a European Reference Index in the Humanities, which is in the first place intended as a categorization of reviews, but should in a second phase also include books, volumes and conference papers. Responding to the ESF-SCH's initiative, national research councils have undertaken the classification of scientific reviews, as identified by the researchers in each research field in the humanities. Linguistics, for example, may be considered to be one of the most formalized and universal among the humanities. *The Linguistic Bibliography/Bibliographie Linguistique* currently lists about 300 peer-reviewed high-quality journals. It includes publications in a great number of languages, but does not yet cover the field of neurolinguistics.

The Modern Language Association produces an MLA Directory containing information about the number of subscriptions, the review process and the current success rate for submissions of scores of American and British journals. The European continent is less well represented.

These endeavors have not led to a clear ranking, as some fields in the humanities and the social sciences are bound to the social, economic and political system in which they operate or to the particular culture that is the object of study. As social and cultural phenomena are diverse, publication of research results in a few leading journals on a global scale cannot be the most suitable method of dissemination. The variety of means of communication is a fundamental characteristic of many fields in the humanities and the social sciences, which makes the application of bibliometric indicators unrepresentative and not at all cost-effective.

The value of bibliometric indicators for evaluating research in some fields in the humanities and the social sciences is, in general, considered to be rather limited. The main arguments are as follows:

- By their very nature and mission, these fields in the humanities and the social sciences address themselves largely to their particular society and culture. In this

¹⁶ Van Leeuwen 2004, 145-6.

context, a report issued by the KNAW has emphasized the value and functions of publications in Dutch, as well as in other languages, as they are related to their subject matter.¹⁷

- In some research fields of the humanities and the social sciences researchers do not communicate with their peers in English but in another language. Their peers might well be their compatriots or colleagues belonging to the linguistic or cultural sphere of the subject matter. The languages in which the international community of researchers communicate effectively, have been called the *forum-languages*,¹⁸ which might very well be Italian, for example, in some fields of classical and renaissance studies, as well as in musicology, not to mention research on Italian language, literature, history and culture.
- English-speaking researchers are not always well-informed about the relevant research tradition in other languages such as French, German, or Italian, which might be of primordial importance in some disciplinary specializations such as oriental studies or classical archaeology. Scholarship in English is thus not in all cases the highest level of excellence.
- Sometimes, communication with peers in the humanities and the social sciences is not primarily by means of articles. Book chapters, collective volumes and monographs are generally highly regarded means of scientific communication, to which suitable selection and reviewing processes are applied.¹⁹
- Book reviews in the humanities and the social sciences fulfill an evaluating function that reaches a far greater depth than sheer quantitative indicators.
- In the field of law, sentences and judgments of courts form an authoritative interpretation referring to the relevant literature, as available in the forum-language which is that of the national legal system.²⁰
- The life span of research in the humanities and the social sciences is sometimes relatively long, while citation indicators are based on citations of an article within a limited number of years after publication of this article.
- Interdisciplinary research, both in the humanities and in the social sciences, needs to be covered by a very varied set of bibliographical media.²¹
- Quantifying bibliometric indicators is relatively expensive and less significant for smaller research fields, and thus less reliable than indicators for larger research fields.

¹⁷ Commissie Nederlands als wetenschapstaal. *Nederlands, tenzij... Tweetaligheid in de geestes- en de gedrags- en maatschappijwetenschappen*. Amsterdam: KNAW, 2003.

¹⁸ Billiet, J. e.a. *Bibliometrie in de Humane Wetenschappen*. Brussel: Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten, Standpunten 3, 2004, 9.

¹⁹ Dr. Henry Small of the ISI at Philadelphia, calculated that 61.3% of the references in the field of the history and philosophy of sciences in a selected data-set, were to non-journal publications (paper presented at the Royal Flemish Academy for Sciences and Arts in Brussels on 26 January 2005).

²⁰ Billiet 2004, 23.

²¹ See the appendix to this report and Weingart, 2003.

See, for example: Council for Medical Sciences, 2002; Spaapen and Dijkstra, 2005; vvcw, 1988.

Other facilities are becoming available through search engines on the internet. For example, www.Amazon.com offers the option of listing citations in footnotes in books for an individual author. Similarly, the Social Science Research Network offers researchers the opportunity to list their discussion papers and it records the number of each paper that is listed (www.ssrn.com). It remains unclear, however, how much of the relevant literature is being covered in this way.

2.4

Conclusion

From the above, one can conclude that the evaluation of research activities in most fields in the humanities and the social sciences cannot be based on simple and uniform bibliometric indicators, and certainly not only on journals in English. Methods applicable in one discipline are not automatically valid for other disciplines, given the variety of communicative patterns. For a more detailed reflection, see the appendix to this report. Indicators always have to be interpreted by experts within the framework of a peer review. This leads us, in the next section, to the proposal for a process which should be more suitable for evaluating both the profile and the performance of institutes (i.e. research group or research groups) in the humanities and the social sciences.

3.1

Introduction

A research evaluation that does justice to research activities starts out with the writing of a self-evaluation report. At its heart are the mission of the institute (i.e. research group or research groups) and its audience, potentially ranging from top experts in the research field to the broader public. For an institute, basic research is a *sine qua non*, and its primary audience should consist of the international scientific forum. Other target groups – such as students, policy makers, the business community, professionals, and the broader public, see below – have a complementary position: addressing them is not necessarily essential, but any neglect of such target groups should lead to reflection and specific argument. The mission statement in these terms is subject to discussion by the research evaluation committee.

The institute's management sets up the self-evaluation report. The report presents, among other things:

- A full description of the mission statement of the institute.
- Academic reputation, as indicated by, for example, bibliometric indicators, citations of important scientific results, previous peer reviews, and awards.
- Where non-peer target groups belong to the audience of the institution: an evaluation of the effects of collaboration and dissemination of research results outside the scientific community. The effects can be discovered by means of, for example, an analysis of the institute's environment and its appreciation of the conduct and results of the institute.

The latter evaluation is an important tool for researchers and research directors in the humanities and the social sciences. It emphasizes their collaborative and disseminative efforts with respect to the target groups for their research. Researchers communicate with the target groups via specific communication channels, for example, a book, an exposition catalogue, or a policy report. These channels are very diverse and sometimes small-scale. It is of supreme importance that the research results are received and appreciated by the specified target groups.

Furthermore, the management of the institute describes in the self-evaluation report what the benchmarks are for the institute's academic reputation and the reputation of the institute outside the scientific community. The self-evaluation report should make clear whether or not the institute has fulfilled the expectations of the institute's management with respect to both the institute's academic reputation and the reputation of the institute outside the scientific community.

In an earlier section of this report it was concluded that one should be aware of direct and indirect effects of incentives before introducing them into the research evaluation process, especially when these incentives have consequences for the allocation of research funding. From this perspective, the institute's board and research management have the responsibility for deciding on the target groups and the indicators (see below), to be quantified in the self-evaluation report. The board and

management have to keep in mind that researchers may interpret these indicators as incentives and, as a consequence, will alter their behavior.

3.2 A tentative list of target groups

Researchers in the humanities and the social sciences can be orientated towards a selection or all of the following target groups:²²

- *Peers*: researchers are by definition orientated toward this target group. Selection of this target group is therefore a condition *sine qua non* for the research evaluation.
- *Students*: students represent the future of the discipline; the humanities and the social sciences have at times even been called ‘teaching professions’. Any separation of research and teaching is often artificial. Integrative texts (books), for example, may be valuable research products. Mere reproduction, of course, is not.
- *Policy makers*: the independent designing, analysis, and evaluation of policies represent an important research contribution in the social sciences and related research fields.
- *Business community*: the same holds true with respect to private organizations. The often-cited objection that the integrity of the research would suffer, should be met by scrutinizing the research in this respect when performing the substantive evaluation.
- *Professionals* (the ‘*learned profession*’). Instruments and procedures, for example, psychological tests and questionnaires, and other contributions to the work of practitioners, may be at least as valuable as theoretical innovations.
- *Broader public*: in many fields in the humanities and the social sciences, popular (not: vulgar) scientific texts, lectures, and the like, constitute a perfectly respectable output of research.

The list of target groups is meant to be indicative rather than complete. Sometimes target groups overlap, for example, students and the broader public, or policy makers and the business community.

3.3 A tentative list of indicators

According to customary practice, peer evaluations are supported by systems of indicators. The system of indicators representing the peer group is by far dominant in these evaluations. The institute's management and the evaluation committee will have to deliberate in an early stage of the peer evaluation process about the appropriate systems of indicators representing audiences other than the peer group.

The following list may be found inspirational. It is not meant to be exhaustive. It is also not meant to be mandatory in any way. In all indicator-supported evaluations, both the suitability of indicators and the costs involved in collecting the data have to be weighed against their contribution to the evaluation process.

²² See, for example: Council for Medical Sciences, 2002; Spaapen and Dijkstra, 2005, vvcw, 1988.

Indicators with respect to the target group of *peers*:

- Publications in scientific journals.
- Citations of publications by peers in scientific journals.
- Reviews of publications by peers on the internet.
- Cooperation with peers, for example, contributions to courses.
- Scientific awards.
- Keynote speeches and invited lectures.
- Editorship of scientific journals.
- Invitations by journals to review scientific publications.
- Invitations to contribute to special issues or collections.
- Received grants from research councils.

Indicators with respect to the target group of *students*:

- Text books and lecture materials sold.
- Reviews of publications by students on the internet.
- Courses for students abroad.
- Graduate PhD students.
- Graduate masters students and their first jobs.

Indicators with respect to the target group of *policy makers*:

- Publications via dissemination channels of policy makers.
- Citations of publications by policy makers in their dissemination channels, for example, reports.
- Reviews of publications by policy makers on the internet.
- Cooperation with policy makers, for example, contributions to courses for policy makers or setting up research programs with policy makers.
- Lectures for audiences of policy makers.
- Memberships of bodies advising policy makers.
- Grants received from policy makers.

Indicators with respect to the target group *business community*:

- Publications via dissemination channels of the business community.
- Citations of publications by the business community in their dissemination channels.
- Reviews of publications by the business community on the internet.
- Cooperation with the business community, for example contributions to courses for the business community or setting up research programs with the business community.
- Lectures for audiences of the business community.
- Memberships of bodies advising the business community.
- Grants received from the business community.

Indicators with respect to the target group of *professionals*:

- Publications for professionals.
- Citations of publications by professionals in journals for professionals.
- Reviews of publications by professionals on the internet.
- Cooperation with professionals, for example contributions to courses for professionals or setting up research programs with professionals.
- Awards by professionals.
- Lectures for audiences of professionals.
- Memberships of prestigious organizations of professionals.
- Grants received from professionals.

Indicators with respect to the target group *broader public*:

- Publications via dissemination channels of the broader public.
- Citations of publications by the broader public in the media
- Reviews of publications by the broader public on the internet.
- Cooperation with the broader public, for example contributions to public meetings and exhibitions.
- Awards by the broader public.
- Lectures for audiences of the broader public.
- Grants received from the broader public.

3.4

Conclusion

The contours of a method for evaluating research in the humanities and the social sciences have been outlined in this section. However, many choices still have to be made by the institute's board, management and researchers. The above target groups and indicators for appreciating by the specified target group(s) should contribute to this decision-making process.

- The mission statement, target groups for research and means of communication of research groups vary enormously within and between research fields. Research is primarily addressed to the scientific community, but it can also be addressed to _ and be useful for _ other target groups: students, policy makers, the business community, professionals, and the broader public. All relevant target groups should be taken into account when answering the following questions in the research evaluation: did the institute do the right thing and has the institute been doing this in the right way?
- The Standard Evaluation Protocol (VNSU, NWO, KNAW, 2003) offers researchers the opportunity and the task of formulating their mission statement, the target groups for their research and the means of communication with these target groups. Researchers in the humanities and the social sciences should actively take this opportunity and formulate, in the context of their institute: the mission statement, the target groups and means of communication with these target groups. The institute's research director should translate these into the terms of reference for evaluating the research activities.
- It is evident that these terms of reference are the point of departure for the ex post research evaluation process. In particular, the board, the management, and the researchers of institutes in the humanities and the social sciences should put a lot of effort into the development of the terms of reference for the evaluation of their research activities. The relation between the mission statement, the target groups and the means of communication should be properly translated into performance indicators. These indicators should be tested and approved of by the research evaluation committee. At this point, the usefulness of bibliometric indicators will also be discussed. The use of indicators should contribute to the objectivity of the peer review.
- Bibliometric methods cannot be made fully operational for some research fields. Attention should therefore also be paid to the methodology of other performance indicators.
- It will always be difficult to judge publicly funded research activities, because an institute is not able to enter into a performance contract: research is concerned with creative processes that cannot be quantified in advance. The research evaluation committee will not be able to assess such a performance contract. For this reason, the institute should set out its benchmarks in the self-evaluation report. The mission statement, target groups and means of communication are driving forces for formulating these benchmarks. On this basis, the research evaluation committee should be able to formulate its recommendations with respect to the mission statement of the institute.
- The institute needs to make careful use of performance indicators with respect to its research activities. Not all performance indicators can be applied in all research evaluations. A reasoned choice should be made.
- The research evaluation committee should concentrate on the boundary

between sufficiency and insufficiency of the research activities, and on other boundaries (for example, between very good and excellent) that are consequential. The main contribution of the research evaluation committee is not in a comparative ranking of institutes, but in judging them on their merits.

- The evaluation efforts should remain proportionate to the goal and expected revenues of the research evaluation.

Aanbevelingen

- De missie, doelgroepen en wijzen van communicatie van onderzoeksgroepen kunnen binnen en tussen onderzoeksterreinen enorm variëren. Onderzoek richt zich in eerste instantie op wetenschappelijke onderzoekers, maar kan zich ook richten op – en zinvol zijn voor – andere doelgroepen: studenten, beleidsmakers, het bedrijfsleven, vakspecialisten en het bredere publiek. In de onderzoeksevaluatie moeten alle relevante doelgroepen in aanmerking worden genomen bij het beantwoorden van de volgende vragen: heeft het instituut het juiste gedaan en heeft het instituut dit op de juiste manier gedaan?
- Het Standard Evaluation Protocol (VNSU, NWO, KNAW, 2003) geeft onderzoekers de gelegenheid maar ook de taak om hun missie, hun doelgroepen en de wijzen van communicatie met deze doelgroepen af te bakenen. Onderzoekers in de alfa- en gammawetenschappen moeten deze kans actief benutten en moeten, vanuit de context van hun instituut, de missie, doelgroepen en wijzen van communicatie met deze doelgroepen formuleren. De *onderzoeksdirecteur* van het instituut moet dit vertalen in de *terms of reference* voor de evaluatie van de onderzoeksactiviteiten.
- Het is duidelijk dat deze *terms of reference* het uitgangspunt vormen voor de onderzoeksevaluatie achteraf. Met name het bestuur, de directie en de onderzoekers van instituten in de alfa- en gammawetenschappen moeten veel aandacht besteden aan het uitwerken van de *terms of reference* voor de evaluatie van hun onderzoeksactiviteiten. De relatie tussen de missie, de doelgroepen en de wijzen van communicatie moet duidelijk worden vertaald naar prestatie-indicatoren. Deze indicatoren moeten door de onderzoeksevaluatiecommissie worden getoetst en goedgekeurd. Hierbij zal ook het nut van bibliometrische indicatoren aan de orde komen. Het gebruik van indicatoren moet bijdragen tot de objectiviteit van de *peer review*.
- Bibliometrische methoden kunnen voor sommige onderzoeksterreinen niet volledig worden geoperationaliseerd. Daarom moet er ook aandacht zijn voor de methodiek van andere prestatie-indicatoren.
- Het beoordelen van door de overheid gefinancierde onderzoeksactiviteiten blijft altijd moeilijk, omdat een instituut geen prestatiecontract kan afsluiten. Bij onderzoek gaat het immers om creatieve processen die niet vooraf te kwantificeren zijn. De onderzoeksevaluatiecommissie zal dus geen prestatiecontract kunnen beoordelen. Daarom moet het instituut zijn benchmarks aangeven in de zelfevaluatie. De missie, doelgroepen en wijzen van communicatie vormen de basis voor het formuleren van deze benchmarks. Aan de hand hiervan moet de onderzoeksevaluatiecommissie in staat zijn haar aanbevelingen over de missie van het instituut te formuleren.
- Het instituut moet bij zijn onderzoeksactiviteiten zorgvuldig gebruikmaken van prestatie-indicatoren. Niet alle prestatie-indicatoren kunnen in alle onderzoeksevaluaties worden toegepast. Er moet een gefundeerde keuze worden gemaakt.
- De onderzoeksevaluatiecommissie moet zich concentreren op de afbakening

tussen het voldoende en onvoldoende zijn van de onderzoeksactiviteiten, en op andere grenzen (bijvoorbeeld tussen zeer goed en uitstekend) die relevant zijn. De voornaamste taak van de onderzoeksevaluatiecommissie is niet om een vergelijkende ranglijst van instituten op te stellen, maar om ze op hun eigen merites te beoordelen.

- De evaluatie-inspanningen moeten in verhouding staan tot het doel en de verwachte opbrengsten van de onderzoeksevaluatie.

References

- Bensing, J.M, W.M.C.M. Caris-Verhallen, J. Dekker, D.M.J. Delnoij and P.P. Groenewegen. 'Doing the right thing and doing it right: Toward a framework for assessing the policy relevance of health services research'. In: *International Journal of Technology Assessment in Health Care*, (19:4) 2003, 604-612.
- Billiet, J. e.a. *Bibliometrie in de Humane Wetenschappen*. Brussel: Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten, Standpunten 3, 2004.
- Blom, J.C.H. et al., eds., *Srebrenica, een 'veilig' gebied: reconstructie, achtergronden, gevolgen en analyses van de val van een 'Safe Area'*. 3 vols., Amsterdam: Boom 2002, 3393 pp.
- Carasso-Kok, M., ed., *Geschiedenis van Amsterdam*, vol. 1, Amsterdam: SUN 2004, 540 pp.
- Commissie Herziening Outputmeting Economie. *Towards a new system of output measurement in Dutch economics*, 1999.
- Commissie Nederlands als wetenschapstaal. *Nederlands, tenzij... Tweetaligheid in de geestes- en de gedrags- en maatschappijwetenschappen*. Amsterdam: KNAW, 2003.
- Council for Medical Sciences. *The societal impact of applied health research. Towards a quality assessment system*. Amsterdam: KNAW, 2002.
- Debackere, K., W. Glänzel. Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key. In: *Scientometrics* (59:2) 2004, 253-276.
- Debackere, K., W. Glänzel. 'Expanding the use of WoS-based indicators towards funding decisions for humanities and social sciences: An analysis and a reflection' (conference presentation).
- Glänzel, W. and K. Debackere. 'On the opportunities and limitations in using bibliometric indicators in a policy relevant context,' In: R. Ball (ed.), *Bibliometric analysis in Science and Research: Applications, Benefits and Limitations*, Publications Forschungszentrum Jülich: 225-236, 2003.
- Glänzel, W., U. Schoepflin, A bibliometric study of reference literature in the sciences and social sciences. In: *Information Processing and Management* (35) 1999, 31-44.
- Hackmann, H., P.J.D. Drenth, J.J.F. Schroots, *Evaluating for Science: Processes & Protocols*. Amsterdam: ALLEA, 2004.
- Hofstee, W. *Beoordelen, wetenschap of kunst?* Amsterdam: KNAW, 1995.
- Holcombe, R. 'The National Research Council Ranking of Research Universities: Its Impact on Research in Economics', *Economic Journal Watch* 1 (3), December 2004, 498-514.

- Kortmann, C.A.J.M. In: *Nederlands Juristenblad*, 43, 26 November 2004, 2243.
- Kyvik, S. 'Changing trends in publishing behavior among university faculty', 1980-2000'. In: *Scientometrics*, (58) 2003-1:35-48.
- Leeuwen, Th. N. van. *Second generation bibliometric indicators. The improvement of existing and development of new bibliometric indicators for research and journal performance assessment procedures*. Leiden, PhD thesis 2004.
- Mare*, 4 November 2004, p. 5.
- Nederlands Observatorium van Wetenschap en Technologie (NOWT). *Wetenschaps- en technologie-indicatoren 2001*. Den Haag: Deltahage, 2001.
- Nederlands Observatorium van Wetenschap en Technologie (NOWT). *Wetenschaps- en technologie-indicatoren 2003*. Den Haag: Deltahage, 2003.
- Peyraube, A. Project for Building a European Citation Index for the Humanities. In: *Reflections*, ESF, Strasbourg, December 2002, 13-15.
- Spaapen, J., H. Dijstelbloem. *Evaluating Research in Context. A method for comprehensive assessment*. The Hague: Commissie van Overleg Sectorraden 2005.
- VSNU, NWO, KNAW. *Standard Evaluation Protocol 2003-2009 For Public Research Organizations*. 2003.
- VVCW. *Naar een Meet- en Monitoringsysteem van Productiviteit / Kwaliteit van het onderzoek in de Economische Faculteiten*. 1988.
- Weingart, P. Keynote address given at the second conference of the Central Library Forschungszentrum Jülich, 5-7 November 2003, Conference Proceedings *Bibliometric analysis in Science and Research*, Schriften des Forschungszentrums Jülich vo. 11, 2003, 7-19.
- Wissenschaftsrat. *Empfehlungen zu Rankings im Wissenschaftssystem. Teil 1: Forschung*. Hamburg 2004.
- Wood, F.Q. *The peer review process*. Canberra: Australian Government Publishing Service, 1997.

Appendices

Appendix 1 Abbreviations

AHCI	Arts and Humanities Citation Index
ESF	European Science Foundation
ESF-SCH	Standing Committee for the Humanities of the European Science Foundation
ISI	Institute for Scientific Information
KNAW	Royal Netherlands Academy of Arts and Sciences
MLA	Modern Language Association
NIAS	Netherlands Institute for Advanced Study in Humanities and Social Sciences
NIOD	Netherlands Institute for War Documentation
NIVEL	Netherlands Institute for Research on Health Care
NOWT	Netherlands Observatory of Science and Technology
NRC	National Research Council
NWO	Netherlands Organization for Scientific Research
OCW	Dutch Ministry of Education, Culture and Science
RGW	Council for the Humanities
SEP	Standard Evaluation Protocol 2003-2009 For Public Research Organizations
SWR	Social Sciences Council
VSNU	Association of Universities in the Netherlands

Appendix 2 Bibliometric indicators and peer review

This appendix goes into more detail regarding the advantages and disadvantages of bibliometric indicators and peer review as research evaluation methods (see, for example, Commissie Herziening Outputmeting Economie, 1999).

Bibliometric indicators as a research evaluation method

Examples of bibliometric indicators refer to publication counts, citation analysis, and impact factors as developed by Computer Horizons Inc. (Wood, 1997). The best known are publication and citation indicators.

Bibliometric indicators are considered to be reflections of researchers communicating in journals. For practical reasons, bibliographic analyses are performed on only a selected number of journals. Bibliometric indicators, resulting from these analyses, have clear advantages in evaluating research activities:

- They contribute to the objectivity and transparency of the research evaluation process (Wood, 1997).
- Bibliometric measures provide the ‘bigger picture’ (Weingart, 2003), revealing macro-patterns in the communication process that cannot be seen from the highly limited and selective perspective of the individual researcher. From bibliometric analysis one for example can discover that research fields are connected, or that a research field has declined or grown (Wood, 1997).

However, the indicators also have limitations with respect to their use in evaluating research activities:

- When selecting a set of relevant journals to define a research field, for example an interdisciplinary research field, some of these journals are not included in the database. It will therefore not be possible to define a research field properly through bibliometric indicators (Weingart, 2003).
- Other reflections of communication between researchers are not present in bibliometric indicators. Researchers not only communicate with their peers in journals but also via other channels, for example monographs. Furthermore, in some research fields researchers do not communicate in English – the dominant language in journals - because the lingua franca in that field is other than English (Commissie Nederlands als wetenschapstaal, 2003). Researchers communicating with their peers via non-English journals hardly ever find the reflections of their communication in the journals database. Bibliometric indicators therefore reflect only a part of the communication of researchers.
- Bibliometric analyses are costly and time-consuming.

Citation analysis has even more limitations with respect to research evaluation.

Some of the limitations are:

- It is not clear how researchers actually use citations in their articles. Is it a positive citation? Or a negative citation? In principle, it is possible for an article in a journal to be cited many times in a negative manner, whereas the citation indicator ignores the negative character of these citations. Only the number of citations is reflected in this indicator but not the qualification of these citations.
- There are distortions caused by self-citations (Wood, 1997).
- For practical reasons, citation analysis usually sets a time window for publications (for example, only citations within three years after publication of the article are counted). However, sometimes the time window is too small. This will be the case when an article needs time to be cited.
- The number of citations of, for example, individual researchers or small institutes can be too limited. In this case, comparison with a previous period or with, for example, sister institutes will not give reliable judgments.

Different research fields have developed very different customs of citing. Moreover, the number of specialists in a particular research field matters. Articles in basic biomedical research are being cited six times more often than articles in mathematics (Weingart, 2003). Comparison of bibliometric indicators across disciplinary boundaries is therefore very tricky.

Peer review as research evaluation method

There is sufficient empirical evidence in testing and measuring the reliability of peer judgments, leading to the conclusion that peers do not agree with each other, and do not remain consistent over time. The evidence, however, applies to peer judgments in decisions about research proposals submitted to funding agencies and articles to be published, rather than ex post research evaluations (Weingart, 2003).

Clear disadvantages of peer review as a research evaluation method are:²³

- The judgment of peers is not transparent to the outside world. For this reason, peer review as a research evaluation method lacks complete openness.
- Peers tend to use a limited range of evaluation criteria, in particular, only criteria from a scientific point of view.
- There may be biases due to limitations of personal knowledge of peers.

An advantage of peer review is:

- Peers can interpret the macro-patterns in the communication process that are provided by bibliographic analysis (Weingart, 2003).

²³ See also the cynical comment by C.A.J.M. Kortmann, Professor of Public and Administrative Law at Radboud University, Nijmegen, in *Nederlands Juristenblad*, 43, 26 November 2004, 2243: ‘... ik noch binnen het topdepartement van OCW, noch in de KNAW, NWO, de universiteit, de faculteit of mijn eigen, eenvoudige sectie ooit een au sérieux te nemen persoon heb ontmoet, die in staat is een ook maar enigszins ‘objectief’ oordeel te vellen over de ‘kwaliteit’ van een zusterfaculteit, een zusterinstelling of van daar werkzame individuen.’

Conclusion

The existing or proposed citation indexes do not reflect publication practice in most of the humanities and the social sciences. Transfer of evaluation methods from one discipline to another does not take into account the specific communication patterns within disciplines, nor their role in society. It seems advisable to combine bibliometric and other indicators and the peer review method in such a way that the clear advantages of both methods are optimally embedded in an evaluation system and the disadvantages are neutralized.