Chapter 9

The underestimated humanities and social sciences

Wim Blockmans[1]
Rector, NIAS (Netherlands Institute for Advanced Study in Humanities and Social Sciences), Meijboomlaan 1, 2242 PR, Wassenaar, The Netherlands

The spread and intensification of evaluation procedures throughout the research and educational institutes of most countries in Western Europe is placing an ever-increasing burden on senior researchers. They are required to contribute to the functioning of the evaluation system at several levels: they are responsible for designing and running research projects, while, at the same time, the task of assessing such projects also falls to them, although usually on a broader scale. Moreover, they are actively involved in managing projects and institutes and in reporting to evaluation commissions; often the same commissions that they are asked to be members of. The proliferation of evaluation processes has led to a research landscape, both national and international, that is beginning to resemble a fairground where the actors frequently change roundabouts and keep bumping into each other along the way.

Understandably, the introduction of evaluations often arouses vehement criticism. In the U.K. especially, where RAE (Research Analysis and Evaluation) has been in practice for the longest and has real direct consequences for the allocation of resources to universities, the system is losing credibility because of the continual alterations and adjustments that are made. Internationally, there is no uniformity with respect to approach or methods either. On the one hand, there are countries that carry out intense, even extremely intense, evaluations, while, on the other, there are also those that have hardly begun to carry out such an exercise. Unlike in the U.K., in most cases, evaluations have no discernable effect, partly because of the civil servant status of established researchers. The evaluation system is thus fighting a battle of legitimacy that relates to questions of methodology and intensity as well as effect. This makes the system vulnerable to criticism from those who consider that they are at a disadvantage and also discourages researchers and managers from seriously investing the time and resources necessary to carry out laborious evaluation procedures[2]. Nevertheless, academic research is probably one of the most rigorously and consistently evaluated sectors in modern society. For example, take a look at all of the assessment intervals in a researcher's career: it starts with having to compete for the first research project, followed by annual appraisals and progress reports, examination of the doctoral dissertation, review of articles by editorial boards and peers, submission of postdoctoral proposals, job applications, competition for project funding and the evaluation of institutes, and so on and so forth.

It is clear that this whole evaluation system was set up to address two very valid arguments: to guarantee quality and to control spiralling costs. Especially in the medical and natural sciences, which implement expensive technologies and equipment involving large research teams, it is perfectly justified to keep a sharp eye on the allocation of resources and resulting outcomes. The universal application of these disciplines means that they can be implemented on a global scale. Therefore the applicability of their

[1]email: blockmans@nias.knaw.nl

[2]In 2005, the Dutch and Flemish governments set up an organization for the evaluation of higher education (NVOA), for which 21 positions were made available and a yearly budget of €2 million was allocated. Universities and research facilities also have their own budgets to carry out evaluations.

findings has to be assessed globally and, in fact, researchers in these disciplines across the world mainly focus on the same issues and problems. This explains why these disciplines have built up such an extensive worldwide research forum with its own publication culture and associated standardized format and language, used in specific channels of academic communication.

The standard has become the short highly specialized article in an English language journal written by a number of authors who are listed in a specific order. Quality is ensured by a highly qualified group of staff editors who make a preliminary selection. These editors decide whether the paper meets the general standards and goals of the journal, and, if so, they send it on to a selection of expert *peers* for review. In principle, these peers are researchers of the highest repute who consider it an honour to be asked to be part of the Editorial Board of a prestigious journal or to be asked for advice on a particular article. They are expected to provide an objective, well-considered, knowledgeable and usually anonymous review. The evaluation process leads to feedback about the article, acceptance with or without adjustments required or a verdict of rejection. In the latter case, authors usually resubmit their article to a lower-ranking journal. This form of knowledge exchange is fast, reaching a wide audience, and the results are generally incremental, being superseded by new work within a few years.

The number of times an article from a journal is cited, either critically or as a source of knowledge, in another published scholarly article affects the *impact factor* of the journal, as well as the authors. The journal itself derives its prestige from the sum of its articles' citations. The impact factor of a journal in a particular discipline is calculated by relating the current year's citations to the number of source articles published in a set earlier period. However, the number of scholars in the various disciplines and therefore the number of publications that they can produce varies greatly. And, although impact factor calculations are designed to reduce bias, it is a fact that the larger the total number of source articles, the greater the number of citations. The more a field or discipline has standardized parameters with a clear hierarchy of journals and a broad academic forum, the more effective the bibliometric method of evaluation will be.

Assessment institutions need to know how results, research groups and individual researchers in a particular field are valued and esteemed by peers. In this sense, assessments by the editorial boards of journals and the positive or negative reception by the academic community contribute to the evaluation process. A commercial institute, the Institute for Scientific Information (ISI) in Philadelphia, saw a market in this area and began to screen citations systematically in selected journals per discipline. The assumption is that, by counting all the citations in relevant journals objectively, the large numbers will counteract distortion caused by dubious citations. Clearly, citations are not always objective: an article may cite another in order to reject their results, authors may cite themselves or do deals with others to cite each other and the selection of journals may be biased or too narrow. The impact factor system is not an assessment instrument itself; it reflects the assessment of the peers. It is widely used in the medical and natural sciences as an indicator in allotting resources, assessing groups and selecting personnel. In the absence of simple indicators for other kinds of achievement and in other fields of research, the bibliometric instrument, the impact factor in particular, has gradually become the standard for academic excellence. However, this is a highly questionable overestimation of this instrument. It is, after all, a reflection of a pattern of communication, not necessarily a measure of the quality of performance. Richard Smith ([1], reviewed in [2]), the editor-in-chief of the *British Medical Journal* recently published a candid self-criticism questioning the flaws of peer review, the amateurism of editors, the very concept of authorship and ethical issues such as not publishing negative findings.

Researchers from a variety of disciplines have objected to the application of the impact factor instrument as it does not accord with their work in practice. It has been remarked that interdisciplinary research, which is often the most innovative, cannot be adequately measured by a discipline-based evaluation system [3–5]. Mathematics and the technological sciences have different publication systems or performance indicators such as patents

[6]. But other scholars, especially in the humanities and social sciences, are feeling increasingly pressured. In the procedures for granting research funding and awarding prestigious grants or prizes, where the relative merits of proposals from various disciplines have to be compared and evaluated, the humanities and social sciences are finding it increasingly difficult to compete with the medical and natural sciences. This is causing demotivation and mounting frustration at all levels of humanities and social sciences research. The fundamental reason for this discrepancy is the current dependence on the bibliometric instrument as the standard for the objective and universal measurement of academic quality. The publication culture and the social relevance and application of humanities and social sciences research are so fundamentally different from those in the medical and natural sciences, that it is completely unjustified to claim that the evaluation instruments of the one sort of discipline automatically apply to the other. Currently, there is no question of a valid and broadly accepted application of impact factors or citation scores in the humanities and social sciences generally, nor is there yet an international classification of journals. In other words, the basis for what is currently seen as an objective evaluation system is lacking.

This conclusion would not be so alarming if not for the fact that various governing bodies such as the European Commission, national governments, university boards and research councils use impact factors, the number of publications of various categories and citation scores to support their decision-making. They publish these data in comparative tables and use these results to make policy. Even worse, rankings of universities are also constructed using such data, which means that an institute's reputation can be claimed to be good or bad on a very debatable basis[3]. As there is no consensus among researchers in the field and the research-evaluation experts with respect to the application of these instruments across the whole spectrum of disciplines, it seems that there is a very real danger that estimates of productivity will be systematically distorted, as will insights into the quality of research in certain disciplines.

The assumption that methods valid to evaluate a variety of medical and natural science disciplines can simply be applied to other fields fails to recognize the specific research traditions and goals of other disciplines, particularly the humanities and social sciences. They are being measured according to standards that are unsuitable for their methodology and working practices. The reasons for this have been investigated by the Standing Committee for the Humanities of the European Science Foundation since 2000. The Arts and Humanities Citation Index (AHCI) of the ISI was judged to be unsuitable as an evaluation instrument for these disciplines and therefore should not be used by European policy-makers. Moreover, the AHCI is biased towards English language journals, it includes only a few of the best journals published outside of the U.S.A., and in no way takes into account the humanities' distinctive publication culture which revolves more around books and volumes. The ISI is focused on important articles in the first few years of publication, whereas seminal publications in the humanities and social sciences generally remain relevant for decades [7]. For that matter, Nobel Prizes including those in medicine and natural sciences are not awarded on the basis of recent publications, but on the impact the research has after several decades in all its wider manifestations.

Experts on research evaluation acknowledge that there are a myriad of methods and techniques in circulation, but that "nothing approaching a dominant institution or methodology exists" [8]. Organizations such as the Wissenschaftsrat [9] in Germany, the Royal Flemish Academy of Sciences [10], the Royal Dutch Academy of Sciences [11] and the Deans of the Law Faculties in the Netherlands [12] have all come to the same devastating conclusion with respect to the humanities and social sciences. In economics, it has also been pointed out that, for opportunistic reasons, there has been a tendency to reduce the socially relevant tasks of researchers to aspects of the discipline that fall within the

[3]For example, the annual publication of the Netherlands Observatory of Science and Technology (NOWT), located in The Hague, which, in turn, relies on other publications from the European Commission.

criteria held by the National Research Council of the U.S.A. [13]. It is understandable that the ISI is not interested in developing a truly representative citation index for the humanities and social sciences: it is not a profitable sector for them as the research communities are much smaller because of the wide diversity of language groups and disciplines, because the study topics are less universal and because, on the whole, there is far less financial investment. These are also the reasons that a standardized publication culture has not developed. In mathematics, chemistry, physics and biology, more than half of what is published is produced in journals, but in the humanities and social sciences, this is less than one-third [14]. This greater diversity in knowledge distribution means that citation counts are less significant because average citation scores are lower and less representative.

Does this mean that the humanities and social sciences should conform to the rules that apply within the medical and natural sciences? Although it is certainly so that such rules can stimulate greater attention to the quality of publications and the internationalization of research debate [12], it must be remembered that the humanities and social sciences fulfil a different *mission* for which other communication channels are more appropriate. Whosoever does not acknowledge this is missing the core of the contribution that these disciplines make to the societies in which they are embedded. This, indeed, is the greatest difference from the medical and natural sciences. The humanities and social sciences do not study universal truths that overarch the individual thinking actor, but look at specific societies and *cultures* of which the researchers are often themselves a part. Their research observations are primarily intended for society with the aim of providing deeper insights into people's actions and policy-making and to help direct changes in behaviour. This means that the primary target is, on the whole, not the international, but the national *research community*; moreover, non-scientists can also be part of the *target groups*, such as policy-makers, civil servants, judges, lawyers, teachers, journalists and informed citizens. This links to the conclusion that, in these study areas, English language publications are not necessarily the best informed sources, nor is the international *forum* always the best qualified to judge the quality of research. It also means that the channels of communication for these target groups are more diverse and go beyond the specialized research paper: books, volumes, reports and non-specialist publications also contribute to the mission of the humanities and social sciences disciplines.

Because much humanities and social science research is produced in the vernacular and has not developed a highly specialized jargon, it is more readily accessible to non-specialists. However, this close contact also exists between technical, agricultural and pharmaceutical research and the industrial, economic and societal organizations to which they are of direct relevance [12]. There is a large audience ranging from, depending on the regional language and the topic, between ten and hundreds of thousands of non-specialists who are hungry for the latest scientific insights in the form of popular articles, books, exhibitions, documentaries and tourist information. Issues of consuming societal importance, such as health policy, ethical and legal matters, cultural identity, religious thought, cultural changes related to technologies and social mobility, the distribution of resources and wealth demand a mode of direct communication between the researchers and the various social stakeholders. Every society is entitled to a deep and thorough analysis of the way it functions and has the right to research-based information on this matter; and, in highly developed societies, decisions in all areas are based on academic research. Novotny et al. [15] called this "socially robust knowledge".

For those disciplines of the humanities and social sciences that specifically focus on problems within their own society, the *primary forum language* should be that of that particular society. This concept was introduced by Billiet et al. [10]. This principle also holds for cultural phenomena and other societal developments that are under study. For example, Italian and German are not only the designated *forum languages* to study Italian and German languages, societies and cultures of the past and present, but are also most appropriate to study the history of music. In fact, the paltry foreign-language skills of many English and North American academics means that they cannot draw on the huge *body of knowledge* in

various great European languages that has been built up over the ages especially with respect to, for example, oriental studies and classical archaeology. These disciplines, in particular, should not be disadvantaged by the dogma that a publication in the English language reflects a higher standard of knowledge than in another national language.

The legal system is traditionally closely linked to the state. Obviously, it therefore needs to employ the national language to embody the various definitions and institutions. Lawyers have to deal with citizens and are obliged to provide them with a clear explanation of current laws and regulations. Lawyers help to shape the judicial system of a country by reflecting on new laws, sentencing and the verdicts of judges. Politicians and judges draw on the notes and comments of lawyers [10,12]. To do justice to this clearly socially relevant discipline, one should not focus primarily on the number of citations in international journals, but must take the field's own professional *channels of communication* into account. The evaluation of each discipline should therefore take its *societal function* as a starting point and look at the specific communication channels that are necessary to fulfil that function. It remains to be seen how much room the highly esteemed English language journals are willing to relinquish to subjects related to so many different nations. American legal journals only occasionally publish articles by foreign authors and so, with few exceptions, the legal research community is still a divided world [16].

Alongside functions that are directly relevant to society, each discipline will also study areas that can benefit from international comparison or analysis at a supranational level. Staying with the legal system, a growing proportion of a country's legislation is now founded on European regulations, not national parliament. The internationalization of law naturally leads to expansion of the international forum. Within the humanities and social sciences, there are also a number of disciplines which focus on more universal domains, for example, economics, psychology and linguistics. For some fields such as sinology and Arab studies, publishing in English is currently the most practical mode. Nevertheless, for all of the above examples, the acknowledgement of the culturally specific differences and their relationship to the particular societal context should always be part of the process. In this sense, one could speak of a *secondary forum language*.

In the evaluation of research performance, the number and status of publications is currently a main parameter. Striving for objectivity by means of quantifiable *indicators* is not questioned as long as the diversity of contexts and forums is acknowledged. However, there are dangers in restricting an evaluation process to purely quantitative indicators and the simple relationship between input in resources (money as well as human resources measured in *full-time equivalents*) and output in terms of publications. The special value that research in the humanities and social sciences can contribute to society risks being underestimated. Every research group, every university department and every institute aims to fulfil a *mission* that meets the requirements of particular reference groups in the community. These consist of interested parties who can use this knowledge, including other researchers, consumers and colleagues. Ultimately, all aspects of such a mission statement should be manifest in explicit ways so that qualified expert *peers* can be asked to judge whether goals have been successfully attained and the ways in which they have been achieved. Quantitative indicators will not always be at hand for these particular evaluation dimensions. It will then depend on the wisdom of the peer experts to decide in practice how to weigh up and compare dissimilar assessments carefully [5,10,11].

The humanities and social sciences have responded positively to the challenge of citation scores by paying more attentions to international publication and evaluation standards [16]. Peer review is being employed on an increasing scale. Nevertheless, the humanities and social sciences, like all research disciplines, will not just have to justify their use of resources to their peers. They are also accountable to the social field in which they work: the other users and consumers of their knowledge must be involved in the evaluation of a research group or institute. Humanities and social science researchers in particular, more than the average natural scientist, seem to be able to make their results appeal to a wider audience via accessible books and a variety of mass media. They are less enclosed in isolated professional

circles and tend to communicate more directly to a widely ranging target audience in policy-making, social organizations and education, as well as an extensive group of interested people among the general public.

Evaluation instruments should therefore incorporate a far greater breadth of indicators than just the bibliometric parameters which tend to measure communication between peers over the short term. A so-called 'radar model' has been developed for this purpose, and was applied to the agricultural and pharmaceutical sciences [5]. It is clear that we must relinquish the notion that there will ever be a universally applicable system that can measure and reduce to simple numbers such a complex process as the evaluation of the quality and dissemination of research knowledge. Such a system can only disadvantage all disciplines alike and would gravely underestimate the achievements of researchers in all fields.

References

1 Smith, R. (2006) *The Trouble with Medical Journals*, RSM Press, London

2 Light, D.W. (2007) Struggling for the Soul of Academic Publishing. *The American Journal of Bioethics*, in the press

3 Council for Medical Sciences of the Royal Netherlands Academy of Sciences (2002) *The Societal Impact of Applied Health Research: Towards a Quality Assessment System*, Royal Netherlands Academy of Arts and Sciences, Amsterdam

4 Weingart, P. (2003) *Bibliometric Indicators in the Context of Science Policy: Numbers Game or Valuable Tool for Research Assessments?* Bibliometric Analysis in Science and Research: Applications, Benefits and Limitations Conference, Forschungszentrum Jülich, Jülich, Germany, 5–7 November 2003

5 Spaapen, J. and Dijstelbloem, H. (2005) *Evaluating Research in Context: a Method for Comprehensive Assessment*, Commissie van Overleg Sectorraden, The Hague

6 van Leeuwen, Th. N. (2004), *Second Generation Bibliometric Indicators: the Improvement of Existing and Development of New Bibliometric Indicators for Research and Journal Performance Assessment Procedures*. Ph.D. Thesis, Leiden, The Netherlands

7 Peyraube, A. (2002) Project for Building a European Citation Index for the Humanities. In *Reflections*, pp. 14–15, ESF, Strasbourg

8 Shapira, P. and Kuhlmann, S. (eds) (2003) *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*, Edward Elgar, Cheltenham

9 Wissenschaftsrat (2004) *Recommendations for Rankings in the System of Higher Education and Research. Part 1: Research*, Wissenschaftsrat, Hamburg

10 Billiet, J., Bocken, H., Bruynseraede, Y., Clarysse, W. De Dijn, H., De Mey, M., De Schutter, G., Schamp, N., Simons, L., Storme, M. et al. (2004) *Bibliometrie in de Humane Wetenschappen*, Royal Flemish Academy of Belgium for Sciences and Arts, Brussels

11 Council for Humanities and the Social Sciences Council (2005) *Judging Research on its Merits*, Royal Netherlands Academy of Arts and Sciences, Amsterdam

12 Commissie Voorbereiding Onderzoeksbeoordeling Rechtsgeleerdheid (2005) *Oordelen over Rechten*, VSNU, The Hague

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

14 Debackere, K. and Glänzel, W. (2004) Using a bibliometric approach to support research policy making: the case of the Flemish BOF-key. *Scientometrics* **59**(2), 253–276

15 Novotny, H., Scott, P. and Gibbons, M. (2001) *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*, Polity Press, Cambridge

16 Stolker, C.J.J.M. (2005) Legal Journals: in pursuit of a more scientific approach. *European Journal of Legal Education* **2**(2), 77–94