

TOWARDS A FRAMEWORK FOR THE QUALITY
ASSESSMENT OF SOCIAL SCIENCE RESEARCH



© 2013 Royal Netherlands Academy of Arts and Sciences
Some rights reserved.

Usage and distribution of this work is defined in the Creative Commons License, Attribution 3.0 Netherlands. To view a copy of this licence, visit: <http://www.creativecommons.org/licenses/by/3.0/nl/>


Royal Netherlands Academy of Arts and Sciences
PO Box 19121, NL-1000 GC Amsterdam
T +31 (0)20 551 0700
F +31 (0)20 620 4941
knaw@bureau.knaw.nl
www.knaw.nl

pdf available on www.knaw.nl

Basic design: edenspiekermann, Amsterdam
Typesetting: Ellen Bouma, Alkmaar
Translation: Balance Amsterdam/Maastricht

Illustration cover: Shutterstock

ISBN 978-90-6984-668-2

The paper for this publication complies with the  iso 9706 standard (1994) for permanent paper

TOWARDS A FRAMEWORK FOR THE
QUALITY ASSESSMENT
OF SOCIAL SCIENCE RESEARCH

Royal Netherlands Academy of Arts and Sciences
"Quality Indicators in the Social Sciences" Committee
March 2013

FOREWORD

Until the end of 2010, the Netherlands Observatory for Science and Technology (NOWT) drafted the biennial publication *Wetenschaps- en Technologie Indicatoren* [Science and Technology Indicators] on behalf of the Ministry of Education, Culture and Science. The indicators, which are based on bibliometric analysis, also play a role in assessing the quality of research.

Three different research communities – the design and engineering disciplines, the humanities, and the social sciences – believe that the indicators used in the Observatory’s reports do not provide a complete picture of developments in their fields and are therefore less appropriate for quality assessment. They have urged the Board of the Royal Academy to specify the assessment criteria for their fields.

The present document is the third panel of the resulting triptych. The first panel consists of the Academy’s advisory report on quality assessment in the design and engineering disciplines. The second is the Academy’s report on quality indicators in humanities research. In this third and final report, the Academy advises on the quality assessment of social science research.

If we unfold the triptych, we cannot help but be impressed by the magnificent diversity of research. However, the advisory reports also show that research should be subject to a quality assessment at regular intervals, not least because society needs the assurance that publicly funded research is of sound scientific quality.

If we look more closely at the three panels of the triptych, we see that, in essence, they all take the same approach to the process of quality assessment. In each case, peers play a vital role. Indeed, peers define the quality culture of a field. They therefore bear a heavy burden of responsibility: they must clarify the mission, target groups, and communication methods of research groups active in their discipline. And if they have any doubts, they must raise the alarm. By doing so, they can help engender trust in research across all segments of society.

Another similarity between the three panels is that in each one, quality assessment focuses on two quality domains – “scientific quality” and “societal relevance” – and on three assessable dimensions within each of these domains – “demonstrable output”, “demonstrable use”, and “demonstrable recognition”. Within this set of general assessment criteria, researchers and research groups have the right, but also the obligation, to define and account for quality indicators appropriate to their discipline and their weights. I trust that the social sciences will get down to the business of working with these six assessment criteria.

Hans Clevers

President of the Royal Netherlands Academy of Arts and Sciences

CONTENTS

FOREWORD 5

SUMMARY 8

1. INTRODUCTION 11

2. THE QUALITY OF SOCIAL SCIENCE RESEARCH 14

Informal assessment: peers play an essential role 14

Formal assessments: comparing groups 15

Quality comparisons are custom work 15

3. CRITERIA FOR ASSESSING THE QUALITY OF SOCIAL SCIENCE RESEARCH 18

Shared responsibility 18

Assessment system 19

Quality domains 19

Assessment dimensions 20

Criteria for assessing the quality of social science research 21

4. EXAMPLES OF QUALITY INDICATORS FOR THE SOCIAL SCIENCES 23

5. APPLYING THE FRAMEWORK: SOME EXAMPLES 26

6. RECOMMENDATIONS 30

Follow-up needed 30

General principles for quality assessments 31

SOURCES 34

APPENDICES

1. Resolution inaugurating the “Quality Indicators in the Social Sciences” Committee 35

2. “Societal Relevance” as a criterion in quality assessments: examples from abroad 40

SUMMARY

Background

Various shifts have taken place in research in recent years that are influencing the formal assessment of research quality, including research in the social sciences. The first is the growing awareness that quality assessment procedures need to allow for differences between and within fields. That is especially important in the social sciences because it encompasses a wide variety of different fields and because those fields differ enormously. Second, the growing interest in “knowledge valorisation” has led to a broader, deeper discussion of the societal relevance of research, and it has also raised awareness of the need to make the scientific quality and societal relevance of research more visible. The interest in “valorisation” is felt most strongly in discussions of assessment in the social sciences; that is because the societal relevance of research in this field cannot be expressed solely – or even primarily – in terms of “pecuniary advantages”. Third, researchers working “in the field” are growing increasingly dissatisfied with assessors’ tendency to express the quality of research in one-dimensional and overwhelmingly quantitative terms, and to adhere to the principle of “more is better”.

These shifts make it necessary for us to reconsider how research quality is assessed in the social sciences. Against this background, the Royal Academy installed a “Quality Indicators in the Social Sciences” Committee in October 2011 whose assignment was to prepare an advisory report offering organisations that assess research in the social sciences a set of guidelines for defining sound assessment instruments, in particular quality indicators in the social sciences. The intention is for the organisations charged with assessment to elaborate on this advice and make use of better assessment instruments.

In the present advisory report, the Committee describes a framework for the quality assessment of social science research that the various social science fields can fill in as needed. It has done this based on the available literature, a limited survey of examples abroad, and discussion meetings with deans, research directors, and representatives of professional associations in the social sciences, and with members of the Academy's Social Sciences Council. There were also lively discussions amongst the members of the Committee.

There is a good reason for calling this advisory report *Towards a framework for the quality assessment of social science research*. It does not provide a ready-made assessment system, but a frame within which administrators and researchers alike can shape the form and content of quality assessments. The background lies in the Committee's wish to do justice to the many different quality cultures in the social sciences, but also in its observation that researchers in the various fields must take responsibility for making quality in their field visible and verifiable. That does not mean that this report does not have any implications for them, however.

Basic principles generally applicable in quality assessments

The advisory report takes a clear position on the basic principles for assessing the quality of social science research. In its view, the following basic principles are generally applicable in quality assessments:

- The quality assessment is based on a set of *generally* applicable assessment criteria in which *specific* quality indicators and their weights can be identified and defined. This does justice to two understandable wishes that at first glance appear to be contradictory: the research group's wish to be assessed on its own mission and quality culture, and the administrators' wish to compare research groups in order to support strategic decision-making about staffing and funding.
- The quality assessment should focus on two quality domains: "scientific quality" and "societal relevance", and three assessment dimensions within each of these domains: "demonstrable output", "demonstrable utilisation", and "demonstrable recognition".
- Within this set of generally applicable assessment criteria, researchers and research groups have the right to identify quality indicators suitable for their own field, and the obligation to account for these indicators and how they are weighted.
- Performance on quality indicators should be transparent, traceable and verifiable.
- Assessors should judge the chosen quality indicators and assigned weights on their level of ambition, for example, and assess performance on these indicators. In that assessment, it is not the quantitative measure on a certain parameter that is important, but the balance between all six criteria. That applies less for individual researchers than it does for research groups.
- The assessment should allow for the researcher's career phase or the research group's stage of development.

- Performance on the quality indicators for societal relevance should only count when based on and traceable to qualitatively sound research output.

These principles provide the framework of the assessment system. Whether and how that system will be put into practice depends on how the social sciences work with it in the field. This advisory report is intended to provide guidelines for different types of quality assessment: annual appraisal interviews with research staff, appointments of professors, faculty investment policy, national research reviews, and so on. It is up to the social sciences to experiment with the methodology described here, so that they can draw on their experience when working with the new “Standard Evaluation Protocol” (SEP) that will enter into effect in 2015.

Follow-up needed

In order to streamline quality assessments at national level, the advisory report calls on deans and directors of research in the social sciences – but equally on external review committees and, last but not least, researchers themselves – to adhere to the basic principles for quality assessment in the social sciences and to get down to the business of working with the proposed quality assessment framework.

The bodies commissioning quality assessments, usually the (joint) deans in the social sciences or their research directors, are told: “The bodies that commission quality assessments are responsible for ensuring that a quality assessment system actually exists and that it allows for content-related differences between fields”. The research units to be evaluated – generally research groups at universities or research institutes active in the social sciences – are told: “The research units concerned have the right to choose their own quality indicators and the relevant weights, and the obligation to account for their choices”. And the review committees are told the following: “The review committees assess the quality of scientific research within the context of the research unit’s mission and determine the tools for doing so in advance. Quality is not the same as quantity”.

1. INTRODUCTION

Research, politics and society need objective standards for determining the quality of scientific research. Quite often, decisions must be taken in all these areas that draw on information concerning the quality of researchers' work, for example about staffing and funding, research priorities, new innovation programmes, project funding, or about appointments to professorships, the award of research prizes or other forms of recognition. Without sound indicators of research quality, there is a risk of such decisions being taken on casual, pragmatic or political grounds. In short, both the academic world and society as a whole have a huge, collective interest in making the quality of research transparent and verifiable.

In 2005, the Royal Academy published the advisory report *Judging research on its merits*, concerning formal research quality assessment. In that report, the Academy observed that research groups often varied enormously in their mission, target groups and communication methods, both within and between different fields. Researchers tend to focus on their peers, but their findings may also be relevant to other target groups, for example students, policymakers, the business community, professionals, and the general public. This calls for them to utilise a wide range of different publication channels. The advisory report noted that *all* relevant target groups should be considered when assessing the quality of research (KNAW, 2005, p. 27).

Since the advisory report was published, various shifts have occurred in research, and certainly in the social sciences,¹ which have influenced the formal assessment of research quality. The first is the growing awareness that quality assessment procedures need to allow for differences between and within fields. That is especially important in the social sciences because they embrace a wide variety of fields, and

1 In this document, we define the social sciences as a broad cluster of sciences ranging from economics to political science, from psychology to demographics, and from law to anthropology.

because those fields² differ enormously from one another. That means, for example, that university administrators require an assessment framework for social science research that takes this variety into account. In addition, the growing interest in “knowledge valorisation”³ has led to a broader, deeper discussion of the societal relevance of research, and it has also raised awareness of the need to make the scientific quality and societal relevance of research more visible. The interest in “valorisation” is felt most strongly in discussions of assessment in the social sciences; that is because the societal relevance of research in this field cannot be expressed solely – or even primarily – in terms of “pecuniary advantages”. Finally, researchers working “in the field” are growing increasingly dissatisfied with assessors’ tendency to express the quality of research in one-dimensional and overwhelmingly quantitative terms, and to adhere to the principle of “more is better”. That means that the quality of the research *content* is disregarded.

These shifts make it necessary for us to reconsider how research quality is assessed in the social sciences.

Following in the footsteps of the report *Begrip van evalueren* (De Groot, 1986), published under the auspices of the Academy’s Social Sciences Council, the Academy believes it would be useful to differentiate between:

- a. *Ex ante* quality assessment of research: estimating in advance whether a new research project or programme will make a significant contribution to the body of knowledge in the field or can help solve problems in society. The main purpose here is to recognise and acknowledge “promise”.
- b. *Ex post* quality assessment of research: retrospective assessment of the work of researchers or research groups, for example as part of an external review, an award procedure or some other form of recognition. The main aim here is to evaluate and appreciate research that has *already been carried out*.

In reality, the two are often combined, for example when assessing a researcher’s work as part of an appointment procedure or individual grant application such as NWO’s Innovational Research Incentives Scheme, or during a mid-term review, when an ongoing research project or its objectives are adjusted based on a quality assessment.

This advisory report was prepared by the “Quality Indicators in the Social Sciences” Committee. It provides guidelines for organisations involved in assessing research in the social sciences. The report can help them draw up sound assessment tools and quality indicators (see the *Resolution inaugurating the Committee* in Appendix 1).

As part of its assignment, the Committee has developed a framework that the various social science fields can fill in as needed. The framework allows for diversity within the social sciences and considers the scientific quality and societal relevance of social science research. Because quality assessment may involve assessing the work of

2 This can also refer to “fields of research”, “disciplines”, or “subdisciplines”, for example.

3 For example, one of the targets for research “valorisation” is to set aside 2.5% of the public funds earmarked for research in 2016 for “valorisation purposes” (Rijksbegroting Onderwijs, Cultuur en Wetenschappen, 2012, p. 17).

either individual researchers or research groups, the framework can be used for both categories.

The Committee has studied the literature (a selection of the sources consulted can be found in “Sources”) pertaining to quality assessment in the social sciences and other fields. Two key documents were the advisory reports *Kwaliteitsbeoordeling in de ontwerpende en construerende disciplines* [Quality assessment in the design and engineering disciplines] and *Kwaliteitsindicatoren voor onderzoek in de geesteswetenschappen* [Quality indicators for research in the humanities] (KNAW, 2010 and 2011). For the social sciences, the Committee surveyed some of the lessons learned abroad in assessing the societal relevance of research (see Appendix 2). It also discussed the draft version of this report with deans, research directors, and representatives of professional associations in the social sciences, and with members of the Academy’s Social Sciences Council. Finally, the Committee made grateful use of remarks and suggestions contributed by Prof. Keimpe Algra, Prof. Maarten Kleinhans and Prof. Patti M. Valkenburg, who were invited to review the draft report on behalf of the Academy Board.

In accordance with the *Resolution inaugurating the Committee*, the advisory report focuses on organisations involved in assessing research in the social sciences. It addresses deans and directors of research in the social sciences, as well as external review committees and researchers.

This report offers guidelines for the *ex post* quality assessment of social science research. With respect to *ex ante* assessment, the report indicates that the organisations charged with this task, such as NWO, should consider field-specific factors and ensure, as much as possible, that assessors – especially those unfamiliar with the field (for example members of multidisciplinary assessment committees) – understand these factors sufficiently. Field-specific factors concern the relevant research groups’ mission, target groups and communication methods. In so far as *ex ante* quality assessment includes an evaluation of past performance, the organisations concerned can naturally make use of the guidelines offered in this report for *ex post* quality assessment.

The proposed framework builds on the advisory report *Judging research on its merits* (KNAW, 2005) and later Academy advisory reports concerning the formal assessment of research quality (KNAW, 2010, 2011 and 2012a). It is structured such that it can be applied within the general framework for *ex post* research quality assessment laid down in the “Standard Evaluation Protocol 2009-2015” (SEP) developed by NWO, the Academy and the Association of Universities in the Netherlands (VSNU). The SEP gives the relevant research unit the opportunity to carry out a self-evaluation, using a predetermined template. The review committee then undertakes the actual assessment based on the self-evaluation report, in some cases supplemented by a site visit and inspection of the relevant documents. The proposed framework allows for specific quality indicators to be identified; these may differ from one field to the next.

2. THE QUALITY OF SOCIAL SCIENCE RESEARCH

The quality of research is not only assessed in formal situations such as external reviews or the appointment of a professor. Research is constantly being judged on its quality. Indeed, that is the strength of the research process: it is continuous dialogue with peers and external target groups that makes research smarter, more thorough, more innovative, and more relevant both to science and to society in general.

This section first discusses informal and then formal quality assessments. The latter are the focus of this report.

Informal assessment: peers play an essential role

Researchers are accustomed to a constant stream of critical questions and evaluations by peers. All researchers also serve as peers who assess the work of other researchers from time to time. For example, they present their research at peer-reviewed conferences and discuss their findings with peers on the Internet. They publish their results in peer-reviewed journals and books and sit on PhD committees. In other words, the quality of research is assessed largely by peers.

Reviewers and researchers regularly switch places in more or less informal national and international networks, many of which are based outside their own university or research organisation. These informal networks are vital to scientific culture and play a key role in the development of that culture's implicit and explicit quality standards.

Researchers must be aware of their role and responsibility in this respect; they must clarify the mission, target groups and communication methods of research groups in their field. There is another reason for clarifying the "standards" of a

particular field – including the standards related to quality indicators: it helps researchers active in other fields who are involved in the *ex ante* quality assessments carried out by NWO and other organisations understand how quality cultures vary from one field to the next.

Finally, peers are needed to assess the quality of the research content. They can judge the extent to which research is innovative, generates new knowledge or insights, or could even lead to the Eureka effect or scientific breakthroughs. It is much harder, if not impossible for relative outsiders – even if they are members of the academic community but operating in a different field – to identify research that is truly outstanding.

Peers also play a role in assessing the societal relevance of research, but they share that role with external target groups, who can and do have an important part to play.

The system of continuous quality assurance described above – in which it is mainly peers who, consciously or unconsciously, define the quality culture in their field – automatically gives them a role in more formal quality assessment. They do more than act as formal quality assessors (in external reviews and on juries and assessment committees); they are also the subjects of assessment. In other words, they constitute the forum responsible for defining field-specific quality indicators that can be verified externally.

Formal assessments: comparing groups

In addition to the process of continuous quality assessment and improvement described above, which drives research forward, the research community also needs “consolidation” at regular intervals. At such times, a formal review is carried out of a particular field, research group or researcher for the explicit purpose of making comparisons. This is what is understood by a formal quality assessment, for example an external review of research. Peers once more play a role in formal quality assessment procedures, just as they did in informal ones. One major difference between the two, however, is that formal assessments are not concerned (solely) with the intrinsic quality of the research work, but with a comparison between research groups. If necessary, they can serve to substantiate strategic decisions concerning staffing and funding or whether a research group should be supported or dismantled. It is precisely the comparative nature of these formal quality assessments that necessitates generally applicable, reliable and valid quality indicators, preferably ones that are also easy to use. The tendency to reduce the quality of a researcher’s or research group’s work to a single quantitative bibliometric indicator is entirely understandable in this respect.

Quality comparisons are custom work

It is, however, a far from simple matter – and a controversial one almost by definition – to establish the quality of research by drawing comparisons.

In the first place, fields differ in their nature and mission, in their publication

culture, and in the type and size of the target groups that the researcher wishes to address. The community of biomedical researchers, for example, is many times larger than that of social science researchers. For that reason alone, the work of biomedical researchers is cited much more often than the work of social science researchers. Such differences can be found both between and within fields. For example, the publication culture in some subdisciplines of psychology closely resembles that of medicine, whereas other subdisciplines of psychology feel more kinship with the publication culture in the humanities. These differences do not automatically imply a difference in quality. There are also many other factors that put the value of bibliometric indicators into perspective (KNAW, 2012a, pp. 38 and 39). Because the sense or nonsense of using bibliometric indicators in the social sciences differs from one field to the next, it is important that research in this field should always be assessed within its own context (KNAW, 2005 and other publications).

The second reason is that every quality indicator is only a proxy measure of quality. No single indicator can cover everything implied by the term “quality”. It is not a one-dimensional concept; analysing the various facets of research quality always requires multiple indicators. That is certainly the case in the social sciences, which are, after all, expected to elucidate human behaviour and developments in society and explain their backgrounds. “Valorisation” here does not necessarily mean economic gain, but instead involves making improvements in the public and semi-public domain, for example in health care, public safety, education, public administration, and so on. The “mission” of the social sciences must be reflected in its quality indicators. Allowing assessment procedures to place too much emphasis on quantifying a single, key quality indicator (for example the number of articles published in international peer-reviewed journals) may lead researchers to behave strategically; they would then fail to invest sufficiently in other facets of the multidimensional concept of quality. This is one reason for the common complaint that there is enough knowledge floating around, but not enough being put to practical use. Most clinicians, politicians, policymakers and business people simply do not read peer-reviewed journals, after all. They require other publication methods. The significance of this will differ from one field to the next.

The third reason is that much social science research is conducted in a multidisciplinary context. And ironically enough, even though it is precisely collaboration between different fields that can lead to astonishing breakthroughs, multidisciplinary research only gets moderate scores on traditional quality indicators. The relevant journals simply have smaller circulations and a lower impact factor.

The foregoing makes clear that there are no routine methods for assessing the quality of research. Every assessment procedure must take the nature of the field into account, something that should also be expressed in the choice of quality indicators. What typifies social science research is that it generally focuses not only on peers but also on external target groups. This means that the quality indicators used for social science research should also reveal its societal relevance. Quality indicators are any identifiable signs of scientific quality and societal relevance.

Finally, the foregoing shows that research also deserves to be assessed on the quality of its *content* – in fact, that is the most important assessment of all. Of course, that does not mean that peers assessing a research group should have read every single one of its members' publications. That is not feasible, and frankly unnecessary. The peers could, for example, decide to assess a small number of key publications suggested by the researchers themselves (for their scientific quality), as well as cases studies (for scientific quality and societal relevance) about which they would ask the researchers critical questions during site visits. Like more informal forms, formal quality assessments should always break down into four stages: content-related assessment, argumentation, consultation, and decision-making (De Groot, 1986). Appropriate quality indicators for social science research should therefore always refer, at least in part, to an assessment of research content.

Our conclusion must be that a set of generally applicable quality indicators will not make enough allowance for the diversity of social science research.

3. CRITERIA FOR ASSESSING THE QUALITY OF SOCIAL SCIENCE RESEARCH

In developing a framework for assessing the quality of social science research, we must search for an assessment system that has enough general components to facilitate a certain comparison between fields and research groups/researchers, but that leaves enough room for field-specific quality indicators that allow for differences between fields. We can do this by identifying a limited number of general assessment criteria that basically apply in each field while leaving their interpretation – in the shape of specific quality indicators – to the field that is being assessed. The same should apply, in particular, for the fields falling within the social sciences.

Shared responsibility

It is then the task of the assessors to determine whether research in the relevant field/ by the relevant group is inadequate, up to standard or outstanding, according to its/ their performance on the quality indicators identified for the field in question. That means that assessors must take these quality indicators (and not their own personal standards) as the starting point for assessment. Assessors must naturally also form an opinion as to whether the quality indicators are ambitious enough. Ideally they would do so before the actual assessment procedure and document this in bilaterally agreed “terms of reference”.

An assessment system of this kind also holds the researchers responsible; it requires them to identify their mission, target groups and communication methods with related quality indicators, and to account for their choices. No researcher can duck out of this process by claiming that the research group’s mission and the chosen

quality indicators do not apply to his or her research. The prevailing quality culture within the researcher's own field basically sets the standard.

Assessment system

This advisory report differentiates between two domains for assessing the quality of social science research: “scientific quality” and “societal relevance”. In a properly weighted quality assessment of social science research, both domains are essential. The proposed assessment system also encompasses three different quality dimensions: (1) knowledge output, (2) knowledge utilisation by others and (3) external forms of recognition and appreciation. These three dimensions cut across both quality domains, making it possible to identify six overall assessment criteria. That brings this advisory report into line with two similar reports, in the design and engineering disciplines (KNAW, 2010) and in the humanities (KNAW, 2011 and 2012a). Both the quality domains and the assessment dimensions are explained below. This section concludes with a diagram summarising the six assessment criteria.

Quality domains

The value of scientific research depends on a combination of scientific quality and societal relevance (De Groot, 1986). Both domains have their own target groups and thus their own communication methods, justifying separate assessment processes. Sometimes the communication methods in the two quality domains may conflict. That is why scientific quality and societal relevance should also be assessed on their mutual relationship.

In order to assess the **scientific quality** of research, the researcher focuses in the first instance on his or her peers. This target group encompasses all researchers (of varying seniority and experience), both fellow countrymen and those abroad, as well as PhD students and research Master's students in the researcher's own (broader) field. Assessor and assessed may switch places regularly. It is therefore important for the assessed to view the assessor as impartial and the quality assessment as transparent.

In terms of the **societal relevance** of the research, the researcher focuses on external target groups. These groups may have a vested interest in the research questions posed and in the results of the researcher's work. They include policymakers, the business community, professionals, patients (in medical research), NGOs, cultural institutions and other stakeholders. The researcher may focus on external target groups in the Netherlands and elsewhere. Journalists and other media channels may play an important role as mediators between science and society, and as such are another important target group within the “societal relevance” quality domain. Not all of the researcher's media performances should be marked down as “socially relevant”, however. In deciding whether certain forms of communication can be considered performances in the “societal relevance” quality domain, it is important for assessors to consider whether the researcher has been able to showcase the scientific aspects of his or her work.

Researchers should guard against being too guileless in their communication with potential external target groups. Because social science researchers generally tend to focus on external target groups as well as their peers, assessors should take the perspective of these target groups and the choice of communication strategies into account when determining the societal relevance of the research. The communication strategies can be extremely varied; they are also influenced by media dynamics. They can take many different forms, from docudrama to infotainment and scripted reality. The possibility of disseminating social science knowledge on the Internet and through the social media is also relevant when selecting and establishing quality indicators in the field. That does not mean, however, that every statement that a social science researcher makes on the Internet should be regarded as evidence of societal relevance.

The foregoing shows the pointlessness of performing a simple quantitative analysis for the “societal relevance” quality domain. We must be wary of a certain tendency to compare apples and oranges. It might be useful to involve external target groups in assessing the societal relevance of research. Because the results of social science research can have both positive and negative consequences for external groups, it is vital to select the right persons to assess societal relevance and preferably familiarise external assessors with the research culture.

This advisory report assumes that quality assessment in the social sciences is **primarily** a matter of assessing scientific quality. Research (in other words, a research *project* or *programme*) that is scientifically inferior cannot be socially relevant by definition, because it would be misleading for society. Society needs the assurance that publicly funded research is of sound scientific quality; if not, its trust in science will be undermined. This implies a certain hierarchy in quality assessment: if research is judged to be scientifically sound by peers, then it is worth assessing its societal relevance. However, in the researcher’s entire communication with external target groups, the peer-reviewed source of the knowledge he or she is conveying must be traceable, either in his or her own work or elsewhere in the research literature.

Assessment dimensions

In both quality domains (“scientific quality” and “societal relevance”), this report differentiates between three assessment dimensions:

1. (Demonstrable) output of scientific knowledge (including articles and books).
 2. (Demonstrable) knowledge utilisation (including citations and references).
 3. (Demonstrable) recognition (including awards and prestigious appointments).
- These dimensions are stratified. Scientific knowledge can only be utilised after it has been produced. Recognition is only possible after a researcher has done outstanding scientific work and others have utilised that work. Good quality therefore necessarily begins with the (demonstrable) output of scientific knowledge. But a high rate of output does not automatically mean a positive quality assessment. A researcher with ten publications to her name, each one cited more than fifty times on average, may be

rated more highly than a researcher with one hundred publications to her name, only twenty of which have been cited. And if a researcher is awarded NWO's Spinoza Prize, there will be very few who feel the need to count his publications. The point, then, is to look at the ratio between the three assessment dimensions. What that ratio and the precise quality indicators should be depends almost entirely on the field and its publication standards. Some fields in the social sciences (for example numerous psychological disciplines) have adopted the publication standards of science and medicine (i.e. a large number of short publications by multiple authors), whereas others prefer monographs and articles by a single author. The quality indicators are up to the researchers. It is they who must ensure the transparency of the "standards" in their field, including the transparency of the quality indicators.

Another important factor is the researcher's life phase. Demonstrable recognition is tied to the researcher's career path. For example, a *cum laude* or thesis award is a demonstrable form of recognition for PhD candidates, whereas an honorary professorship can be regarded as very prestigious demonstrable recognition for professors, and one that is difficult, if not impossible, to achieve early in one's career.

Criteria for assessing the quality of social science research

Diagram. Six assessment criteria

		QUALITY DOMAINS	
		SCIENTIFIC QUALITY: Conducting scientific research: carrying out research within the context of the research organisation and ensuring that peers consider this research of outstanding quality	SOCIETAL RELEVANCE: Communicating and/or collaborating with external target groups about one's own peer-reviewed research
ASSESSMENT DIMENSIONS	DEMONSTRABLE OUTPUT	(1) Demonstrable output that peers regard as of outstanding quality	(4) Demonstrable output for external target groups
	DEMONSTRABLE UTILISATION	(2) Demonstrable utilisation by peers of researcher's output	(5) Demonstrable utilisation by external target groups
	DEMONSTRABLE RECOGNITION	(3) Demonstrable recognition by peers for researcher's output	(6) Demonstrable recognition by external target groups

The diagram applies both to “a researcher” and to “a research group”. The words “researcher” and “research group” can be used interchangeably.

For “a researcher”, the ideal career course is “from talented student to professor”, moving from top left to bottom right in the diagram. At the end of his or her career, an outstanding researcher should have filled in each of the cells in the diagram. The road there can be an uncertain one, however.

The “research group” also follows a particular sequence in time: from top to bottom and from left to right in the diagram. Recognition follows knowledge utilisation and utilisation follows output. If it sticks to this “royal road”, the research group will only enter into dialogue with external target groups about the scientific knowledge it has developed after that knowledge has gone through a peer-reviewed output process. The first step is to generate outstanding scientific output, and only then communicate it to external target groups. It is also important to achieve an even balance within the assessment diagram when evaluating the scientific work of the research group: focusing exclusively on scientific quality is as undesirable as focusing only on societal relevance. It is precisely the combination of the two quality domains that makes them valuable. Nevertheless, in larger research groups some members may specialise in one area while others concentrate on another. The ideal scenario is for one pioneering researcher or a small group of pioneers to develop into a successful research group as more of the cells in the assessment diagram are transparently, traceably and verifiably completed.

4. EXAMPLES OF QUALITY INDICATORS FOR THE SOCIAL SCIENCES

This advisory report deliberately refrains from defining a set of fixed quality indicators for the social sciences. Quality indicators and their assigned weights differ from one field to the next and between differing orders of scale (project, programme, person, group, institute). The indicators also depend on the individuals' careers and the development path of the relevant groups and institutes.

This section offers examples of quality indicators for each of the six assessment criteria. The indicators related to education are shown in italics because they are often "forgotten", even though renowned researchers should be appreciated for teaching classes or tutoring students. PhD candidates and research Master's students are considered peers because they can play a key role in knowledge dissemination. Bachelor's and Master's students are counted as external target groups; they will apply the scientific knowledge in their future work and in society.

Examples for assessment criterion (1): "Demonstrable output that peers regard as of outstanding quality"

- Articles in science journals (including online ones)
- Monographs
- Books/sections of books
- Conference proceedings
- *Teaching/lecture hours for research Master's students*
- *Dissertations* (a form of "knowledge output")

Examples for assessment criterion (2): “Demonstrable utilisation by peers of researcher’s demonstrable output”

- Reviews
- Citations⁴
- Utilisation of models or methods developed
- Contribution to theory/modelling, to methodology, to developing a particular school of thought
- *Dissertations* (as “offspring” and for the development of a “school”)
- *Research Master’s graduates*
- *PhDs in first research job after taking their doctoral degree*

Examples for assessment criterion (3): “Demonstrable recognition by peers for researcher’s demonstrable output”

- Science prizes (including thesis awards)
- Prestigious national and international research grants, personal funding
- Membership of national/international juries, review committees, editorial boards of science journals
- Membership of scientific and scholarly organisations such as the Royal Netherlands Academy
- Membership of research programme committees
- Keynote lectures at science conferences
- Honorary doctorates
- *Visiting professorships*

Examples for assessment criterion (4): “Demonstrable output for external target groups”

- Books/sections of book aimed at a broad readership
- Monographs aimed at a broad readership
- Articles in specialist journals, in the science pages of a newspaper, or on digital knowledge sites
- Projects commissioned by or carried out in cooperation with target groups
- Lectures for professionals or a wide audience
- Contributions to public debates and appearances in the national or international media as an expert
- Fund-raising
- *Textbooks for Bachelor’s or regular Master’s degree courses*
- *Training of professionals*
- *Teaching/lecture hours for Bachelor’s and regular Masters courses*

4 It may someday be possible to gather statistics on downloads of online articles, similar to the citation indices. Such statistics would naturally say something about the amount of interest generated by a publication and they are becoming increasingly important thanks to the open access movement (De Jonge Akademie, 2012).

Examples for assessment criterion (5): “Demonstrable utilisation by external target groups”

- Demonstrable utilisation by professionals (citations in specialist journals, guidelines, textbooks, etc.)
- Demonstrable utilisation by politicians and policymakers (parliamentary Q&A, citations in parliamentary and policy documents, citations by European Union bodies and international organisations, evidence-based policy)
- Demonstrable contribution to social issues (citations, references, knowledge utilisation in public media, including TV/radio science programmes)
- *Bachelor’s and regular Master’s degree graduates*
- *PhDs in their first job outside research*

Examples for assessment criterion (6): “Demonstrable recognition by external target groups”

- Public prizes
- Membership of national organisations such as the Scientific Council for Government Policy (WRR), the Netherlands Health Council (*Gezondheidsraad*) and the Dutch Safety Board (*Onderzoeksraad voor de Veiligheid*), and of international organisations.
- Membership of the scientific advisory councils of civil society organisations (patient associations, NGOs, cultural institutions, and so on)
- *Membership of juries for thesis awards*
- *Prizes for Bachelor’s and regular Master’s theses*

5. APPLYING THE FRAMEWORK: SOME EXAMPLES

This section offers a few practical examples of how the proposed framework can be applied. The examples clarify the decision-making procedure. The first step is to look at whether the researcher's or research group's work is adequate, and the second is to look at whether that work is outstanding.

Four examples follow: a young researcher; a candidate for a professorship; a young research group; and a field.

EXAMPLE 1: YOUNG RESEARCHER WHO WILL EVENTUALLY MOVE FROM SENIOR LECTURER TO PROFESSOR?

A young economist was recently promoted to senior lecturer, and has thus been given a permanent appointment. He obtained his PhD in 2007 and has since published two articles in the prestigious *American Economic Review*. He was one of a number of co-authors, but they made clear that he was the driver behind the entire research project. The two articles have already been cited frequently. When he was promoted, he had five papers in the pipeline at various peer-reviewed journals, one of them another top publication. In addition to his research publications, the young economist has been teaching third-year Bachelor's students for several years and also tutors students writing their Bachelor's and Master's theses. Without exception, students rate his teaching as outstanding. He has given dozens of presentations, most of them abroad, both at conferences and more exclusive seminars. He reviews manuscripts for several journals, once again including top journals in his field. He also contributes to public debate (either when asked to do so or on his own initiative) by writing opinion pieces for the faculty newsletter/website

and for Dutch-language media outlets such as *Economisch Statistische Berichten*. His letters of reference refer to him as “an original thinker” and “a warm-hearted person”. The dean has let the young economist know that he can reckon on promotion to a full professorship in a few years’ time if he continues to build his portfolio both qualitatively and quantitatively in all the aforementioned areas, and if he presents himself as a contact person for talented young researchers, making it clear that this is the level of quality that the School wishes to employ.

EXAMPLE 2: CANDIDATE FOR PROFESSORSHIP?

The appointment committee has nothing to complain about when it comes to the number of Dutch and foreign candidates interested in a vacancy for a political science professor. The task now is to compile a short list of candidates and invite them for an interview with the committee and a presentation for the entire staff. In addition to scientific quality, administrative experience and management skills are important in this position. In assessing educational quality, the committee looks at the candidates’ portfolio of courses taught, the relevant students’ evaluations and experience in supervising PhD students. One of the candidates has already supervised three PhD candidates successfully, but he has no experience lecturing before large classes, and a committee member reports that his presentation at a recent conference was rather shaky. Fortunately, there are other candidates who appear to be better balanced. Then the committee compares the publication lists. In view of the publication and citation culture in the field of political science, there is little point in calculating the candidates’ H index, but at this level of the game, candidates are expected to have published a reasonable number of articles in international peer-reviewed journals. Increasingly, candidates add the journal’s average impact factor for the past five years to their publication lists. Even if they do not, however, the committee has a fairly unanimous opinion of the reputations of the leading journals. Candidates who have not published in leading journals will not go far, but chapters in edited anthologies are almost as important. In that case, the committee considers the editors of the anthologies and, in particular, the publisher. Major university presses and commercial publishers such as Palgrave or Routledge also subject submissions to a strict review process. Co-authorships certainly count, but the committee also expects aspiring professors in this field to have a number of single-authored publications to their name, and a book if possible. The committee has asked all the candidates to submit a few publications. One of the candidates is eliminated because she only submitted co-authored articles in response, and further investigation revealed that even her dissertation consisted of articles and papers that she had written together with her supervisor and co-supervisor. The two most promising candidates, one foreign and one Dutch, have been active for about the same length of time. Both have a nicely filled list of publications, and the publications they submitted to the committee have been enthusiastically received, including by committee members active in other subdisciplines. The foreign candidate has also co-authored an introductory textbook in his field, in addition

to publishing various articles and chapters. He is a member of the editorial board of various leading journals. The Dutch candidate has not only authored international publications but also a few articles in the only two refereed Dutch journals in her field, as well as a Dutch-language book intended for a wider readership than her peers alone. She plays an important role in the public debate. She just missed receiving a personal grant award, whereas her foreign rival did receive a prestigious grant in his own country. The committee is aware, however, that a smaller percentage of political scientists are awarded funding in the Netherlands. The committee intends interviewing a few more candidates, but it suspects that the decision will come down to these two.

EXAMPLE 3: CUTTING A YOUNG RESEARCH GROUP'S BUDGET?

The group emerged in connection with the activities of two professors of social and cultural anthropology. Both have been successful at promoting their graduate students and post-docs within the context of individual applications and larger-scale projects (NWO, European funding). They encourage lecturers and senior lecturers on permanent appointment to submit funding applications and to publish in prestigious international journals such as *American Ethnologist*, *Cultural Anthropology*, *Current Anthropology*, *Social Anthropology*, *Journal of the Royal Anthropological Institute* and leading journals in area studies and with a theme-based focus. The group publishes monographs with renowned publishers (university presses, Brill, Berg, Berghahn, Blackwell, Routledge, and Palgrave). Some PhD candidates manage to publish in anthologies and refereed journals even before they defend their dissertation; they then receive their degrees *cum laude*. Post-docs with a well-established list of publications are encouraged to submit an application for a "Veni" grant. The research climate is stimulating: the group organises at least two international workshops a year and monthly seminars for its members, maintains a popular Dutch-language weblog that presents research results to a wider audience, and publishes single-authored articles, books and anthologies. The group has a good reputation both in the Netherlands and abroad, as shown by the many invitations that its senior researchers and post-docs receive for keynotes, lectures, contributions to workshops, and citations. The professors work actively to promote the careers of their PhD students, who publish under their own names. In addition, members of the group participate in public debates in newspapers, websites and the social media, and also advise organisations and institutions on the subject of development cooperation and world culture. In other words: the research group demonstrates an ability to perform at a very high level. The dean has therefore decided to shield the group from the latest round of budget cuts.

EXAMPLE 4: NATIONAL OPINION OF A FIELD?

A national review committee for the study of law recently issued its report. After presentation, all the parties involved expressed their satisfaction with the assessment process. University administrators, research directors and the research groups that had been reviewed had some remarks about the outcomes, but the general feeling was that the committee had worked transparently and allowed for the various research groups' stages of development. There was also obvious delight that the committee had considered the thematic and methodological differences between the research groups while simultaneously applying a uniform system of assessment concerning how research is conducted in the field of law, how third parties utilise that research (peers and society), and the recognition that research groups received in return. Early on, the committee had made clear that allowing for variety (legal/dogmatic, interdisciplinary/multidisciplinary, national/international) by no means implied that it would not use the same framework to assess the groups' scientific quality and societal relevance. The committee thus challenged the discipline to clarify, prior to the review, what it wished to be assessed on. After the deans of the law faculties had accepted this challenge, they made clear, for example, that the quality indicators should consider not only multi-author articles on a research group's publication list but also books published by prestige publishing houses, with single-authored books serving as an extra indication of quality. They also argued that number of citations was, at best, a poor indication of peer knowledge utilisation (and in fact impossible to count given the thematic breadth of the discipline, the wide range of journals and books that it encompasses as a result, and the many purposes of citation), and that a better indicator was a demonstrable contribution to establishing a national or international school of thought. Stimulating and visible interplay between research and education was further evidence of both scientific quality and societal relevance, the deans told the committee. The quality of the research groups should therefore also be assessed on the contribution they make to Bachelor-level programmes, and most certainly to research Master's programmes. This also meant, conversely, that in its assessment of scientific output, the committee should consider the extent to which research groups were unable to devote all their time to research. The deans observed that quality could also be measured in terms of societal relevance (demonstrable by looking at the memberships of key organisations in jurisprudence or law and policy), but it should also be clear that the dividing line between scientific quality and societal relevance could not always be sharply drawn because certain outputs combined the two. They gave the example of annotations, which sometimes went beyond discussing the practical aspects of a judgement or case to share new scholarly insights. Once the research groups had the deans' comments on the indicators, they were able to account for the quality of their research in light of both their own research tradition and external scientific standards.

6. RECOMMENDATIONS

Follow-up needed

There is a good reason for calling this advisory report *Towards an outline for the quality assessment of social science research*. It does not provide a ready-made assessment system, but a frame within which administrators can shape the form and content of quality assessments. The background lies in the Committee's wish to do justice to the many different quality cultures in the social sciences, but also in its observation that researchers in the various fields must take responsibility for making quality in their field visible and verifiable. That does not mean that this report does not have any implications for them, however. It takes a clear position on the basic principles for assessing quality in social science research and calls on those to whom it is addressed – the deans and directors of research in the social sciences, as well as external review committees and the researchers themselves – to adhere to these principles and to work with the proposed quality assessment framework.

Identifying quality indicators in each of the six assessment criteria is something that the researchers must do themselves. Practitioners in the relevant fields are basically obligated to adhere to the quality indicators and their assigned weights as agreed within that field, unless they can argue effectively why their research should be allowed to deviate from those norms. Basically, the motto is “comply or explain”. Non-compliance should also be arranged prior to the assessment procedure in the “terms of reference” agreed between the assessors and those assessed.

The advisory report thus argues that the fields should, in the foreseeable future, establish quality indicators and assign weights for assessing the quality of social science research, for example in field-specific consultative bodies or at the research schools. A further advantage to doing so is that a field's quality culture – and along with it, the “standards” set for the mission, target groups, communication methods

and quality indicators – will be made more explicit and thus more visible. Researchers active in other fields who are involved in *ex ante* quality assessments by such organisations as NWO will then have a better understanding of how quality cultures differ from one field to the next, and take this into account in their assessments. In that way, they can allow for assessment standards that are specific to a field.

The advisory report further argues that those commissioning quality assessments should make field-specific quality assessment possible in the social sciences. According to the report, the relevant frame of reference should be established in a separate phase at the start of the assessment procedure, organised in close consultation with those being assessed, so that both they and the assessors have access to the same assessment tools and therefore know where they stand.

General principles for quality assessments

The following basic principles are generally applicable in quality assessments:

- The quality assessment is based on a set of *generally* applicable assessment criteria in which *specific* quality indicators and their weights can be identified and defined. This does justice to two understandable wishes that at first glance appear to be contradictory: the research group's wish to be assessed on its own mission and quality culture, and the administrators' wish to compare research groups in order to support strategic decision-making about staffing and funding.
- The quality assessment should focus on two quality domains: "scientific quality" and "societal relevance", and three assessment dimensions within each of these domains: "demonstrable output", "demonstrable utilisation", and "demonstrable recognition".
- Within this set of generally applicable assessment criteria, researchers and research groups have the right to identify quality indicators suitable for their own field, and the obligation to account for these indicators and how they are weighted.
- Performance on quality indicators should be transparent, traceable and verifiable.
- Assessors should judge the chosen quality indicators and assigned weights on their level of ambition, for example, and assess performance on these indicators. In that assessment, it is not the quantitative measure on a certain parameter that is important, but the balance between all six criteria. That applies less for individual researchers than it does for research groups.
- The assessment should allow for the researcher's career phase or the research group's stage of development.
- Performance on the quality indicators for societal relevance should only count when based on and traceable to qualitatively sound research output.

These principles provide the framework of the assessment system. Whether and how that system will be put into practice depends on how the social sciences work with it. This advisory report is intended to provide guidelines for different types of quality assessment: annual appraisal interviews with research staff, appointments of

professors, faculty investment policy, national research reviews, and so on. It is up to the social sciences to experiment with the methodology described here, so that they can draw on their experience when working with the new “Standard Evaluation Protocol” (SEP) that will enter into effect in 2015.

In an attempt to streamline the national quality assessments, this report offers commissioning bodies, research units and review committees the following recommendations.

RECOMMENDATIONS FOR BODIES COMMISSIONING QUALITY ASSESSMENTS

The bodies commissioning quality assessments are usually the (joint) deans in a particular field of the social sciences, or their research directors. They are told the following: “The bodies that commission quality assessments are responsible for ensuring that a quality assessment system actually **exists** and that it allows for content-related differences between fields”. The following recommendations apply in their case:

- Assemble the review committee such that it can issue an opinion on the six general assessment criteria. Make sure that the review committee is capable of forming an opinion of the quality culture in the field concerned. This is to prevent a “closed shop” from developing in the field. One idea would be to appoint relative outsiders to the review committee who are in any event not involved in the actual content of the field. External target groups may have an important role to play in assessing the societal relevance of social science research.
- Make sure that the review committee’s assignment is clear.
- Assemble a review committee that is knowledgeable about and has a feel for both quality domains (“scientific quality” and “societal relevance”), and provide it with enough relevant information about the assessment procedure and the assessment framework it is to apply.
- Make clear to the review committee that it can call on subject specialists in the relevant fields for the *content-related* assessment of the relevant research group’s work.
- Commence the assessment procedure well in advance so that there is time to negotiate the quality indicators and their weights, resulting in explicit agreement between the assessors and those assessed about the “terms of reference”.

RECOMMENDATIONS FOR THE RESEARCH UNITS

The research units to be evaluated are generally research groups at universities or research institutes active in the social sciences. They are told the following: “The research units concerned have the right to choose their own quality indicators and the relevant weights, and the obligation to account for their choices”. The following recommendations apply in their case:

- When drafting the self-evaluation report, be aware of the assignment that the review committee has been given.

- Identify your mission, target groups and related communication methods so that they are transparent, traceable and verifiable.
- For each of the six assessment criteria, define the quality indicators and their weights. The quality indicators and weighting system developed in the research group's own field should serve as the default option. The motto "comply or explain" should apply. If the field has yet to develop quality indicators and weighting systems for assessing the quality of social science research, it should do so in the foreseeable future, for example in discipline-specific consultative bodies or at the research schools.
- Contribute to developing the quality culture in the field by encouraging discussion of quality assessment and quality indicators as peers.
- Take responsibility by systematically updating relevant information on your own work and by accounting for (or being prepared to account for) the quality of your own work.

RECOMMENDATIONS FOR REVIEW COMMITTEES

The review committees are told the following: "The review committees assess the quality of scientific research within the context of the research unit's mission and determine the tools for doing so in advance. Quality is not the same as quantity". The following recommendations apply in their case:

- Prior to the assessment procedure, verify whether there may be differences in quality cultures and whether the quality indicators and their weights have been specified for each of the six assessment criteria.
- Establish whether/that these quality indicators and their weights are acceptable and provide timely feedback if problems can be anticipated.
- Relate the assessment to the research unit's mission and to the quality indicators and weights agreed.
- Begin by distinguishing between what is and is not adequate in the research unit's work, and then decide whether or not that work is outstanding.
- It is never enough to simply count. Researchers deserve an assessment based on the content of their work.

SOURCES

- AO-consult (2012), *Onderzoek publicatieculturen NWO-MaGW*.
- De Groot, A.D. (1986), *Begrip van evalueren*. (pp. 137-152).
- De Jonge Akademie (2012), *De wetenschap van het publiceren*.
- Klandermans, Bert (October 2009), *Gedrags- en maatschappijwetenschappen ondergewaardeerd*.
- Koninklijke Nederlandse Akademie van Wetenschappen (May 2005), *Judging research on its merits*.
- Koninklijke Nederlandse Akademie van Wetenschappen (November 2010), *Kwaliteitsbeoordeling in de ontwerpende en construerende disciplines (Quality assessment in the design and engineering disciplines)*.
- Koninklijke Nederlandse Akademie van Wetenschappen (May 2011), *Kwaliteitsindicatoren voor onderzoek in de geesteswetenschappen (Quality indicators for research in the humanities)*.
- Koninklijke Nederlandse Akademie van Wetenschappen (2012a), *Naar een adequaat systeem voor de beoordeling van kwaliteit en relevantie van geesteswetenschappelijk onderzoek*.
- Koninklijke Nederlandse Akademie van Wetenschappen (2012b), *Zorgvuldig en integer omgaan met onderzoeksgegevens*.
- Koops, Willem (June 2012), *Bekentenissen van een decaan*.
- Ministerie van Onderwijs, Cultuur en Wetenschappen (2012), *Rijksbegroting*.
- Vlaamse Raad voor Wetenschap en Innovatie (May 2011), *Naar waarde geschat. Valorisatie van onderzoek in de humane en sociale wetenschappen*.

APPENDIX 1.

RESOLUTION INAUGURATING THE “QUALITY INDICATORS IN THE SOCIAL SCIENCES” COMMITTEE

(Reference SWR/AV/3137)

Having regard to Article 8 of the Academy’s Regulations, and having considered that the Academy advisory reports *Maatwerk in de kwaliteitsbeoordeling van de ontwerpende en construerende wetenschappen* and *Kwaliteitsindicatoren voor onderzoek in de geesteswetenschappen* were issued in November 2010 and May 2011 respectively, and that the Division for the Social Sciences (MaGW) at the Netherlands Organisation for Scientific Research (NWO) and the deans in the social sciences recently indicated their desire for better assessment instruments for social science research, the Board of the Royal Netherlands Academy of Arts and Sciences has decided on the recommendation of the Social Sciences Council to install the “Quality Indicators in the Social Sciences” Committee (hereafter: “the Committee”).

(For clarification of the problem that prompted the advisory report, see the annex (“Proposal for Work programme for Academy advisory reports in 2011”, 7 January 2011), which is an integral part of this Resolution.)

Article 1. Assignment

The Committee’s assignment is to prepare an advisory report offering organisations that assess research in the social sciences a set of guidelines for defining sound assessment instruments, in particular quality indicators in the social sciences. The intention is for these organisations to elaborate on this advice by developing better assessment instruments.

The Committee will present the report to the Board by December 2012 at the latest.

Article 2. Composition and term

The following persons are appointed to the Committee in a private capacity:

- *Chairperson:* Prof. Jozien Bensing, (psychology, Utrecht University)
- *Other members:*
 - Prof. Rudy Andeweg (political science, Leiden University)
 - Prof. Philip Hans Franses (chairman of the Erasmus School of Economics, Erasmus University Rotterdam)

Prof. Birgit Meyer (anthropology, Utrecht University)

Prof. Corien Prins (law, University of Twente)

Prof. Kees Schuyt (sociology, University of Amsterdam)

The Committee's term will conclude on 1 January 2013. The Committee will be assisted by the Academy's Bureau⁵ in accordance with the Director-General's instructions.

Article 3. Quality management

Prior to their appointment, the members of the Committee familiarised themselves with the *Preamble op Belangenverklaring Adviescommissie* [Preamble to the Declaration of Interests for Advisory Committee] and completed and returned the *Formulier Belangenverklaring Adviescommissie KNAW* [Declaration of Interests for Academy Advisory Committee]. The members have stated that their participation in the Committee will not give rise to a conflict of interests.

The peer review policy is described in the document *Beleidskader Kwaliteitsborging Adviezen* [Policy Framework for Quality Assurance in Advisory Reports]. There will be no deviation from this policy.

Article 4. Follow-up and communication

The Committee will follow up and provide for communication concerning its findings.

Article 5. Costs and remuneration

The Committee members will be reimbursed for their travel expenses in accordance with Article 18(2) of the Academy Regulations.

Article 6. Confidentiality

The Committee will treat as confidential all information that may be regarded as such and that becomes known to its members within the context of this resolution.

Adopted in Amsterdam on 11 October 2011 by the Board of the Royal Netherlands Academy of Arts and Sciences.

On behalf of the Academy Board

Dr K.H. Chang

Director General

⁵ Dr Ans Vollering (senior policy officer, Academy) acted as the Committee's secretary.

To the Board of the Royal Academy
From the Social Sciences Council
Date 7 January 2011
Reference SWR/AV/323.not

Proposal for the work programme for Academy advisory reports in 2011

Title

"Quality assurance in the social sciences"

1. What question or problem has prompted the advisory report?

The disciplines that fall within the social sciences are much more heterogeneous than those in the natural sciences or humanities, for example. No one questions that a criminal lawyer's profession is entirely different to that of a neuropsychologist, even if our growing understanding of the human brain is posing new challenges in the field of criminal law.

So far, however, research assessment procedures have taken very little account of this heterogeneity, making it impossible to arrive at a fair evaluation of the quality of social science research. The issue that arises when assessing such research can be broken down into two interrelated questions:

1. How can we do justice to differences between publication cultures in the various disciplines when assessing research, and prevent a "one-size-fits-all" assessment from prevailing? The problem is not only how to allow for the many different forms of scientific output (annotations and other outputs in addition to articles and books) but also, and in particular, what the relevant quality criteria should be. Another factor is that some social science disciplines study phenomena within the national context. It should be noted that the Academy advisory report *Judging research on its merits* (2005) has already addressed the more general issue of differences in publication culture. That report requires further elaboration for the social sciences.
2. How can assessment take the societal relevance of social science research into account? The question here is how to measure and evaluate the quality of the impact that social science research has on society. This is an issue not only *between* disciplines (business and government utilise economic knowledge differently to psychological knowledge), but also *within* them (human rights lawyers address a different forum than specialists in administrative law). Scientific knowledge finds its way into society through many different channels, from professional journals and authoritative advisory councils to contributions to the op-ed pages of newspapers. Knowledge utilisation in the context of society raises important questions concerning scrupulousness and the potential for misuse. In many cases, those who

utilise the results are also the objects of the research.

Research funding bodies and university administrators have indicated that they require better assessment instruments for social science research. They recognise that different disciplines within the social sciences have differing publication cultures, but the indicators used in *ex ante* and *ex post* assessment of social science research are largely the same across the board. They therefore feel that there is a gap in the assessment of social science research.

The advisory report will consider criteria for developing adequate assessment instruments for social science research. On the one hand, the report will need to expand on the Academy's 2005 advisory report *Judging research on its merits* by examining the social sciences in greater depth. On the other, it will emphasise the social impact of social science research and answer such questions as "How can we measure the effects of social science research on society" and "What mechanisms contribute to the social impact of social science research?"

The report will build on the body of thought set out in: (a) the Academy advisory report *Kwaliteitsbeoordeling in de ontwerpende en construerende disciplines* (2010); (b) the work of the Council for the Humanities concerning quality assessment in the humanities (Academy advisory report, publication pending); (c) the Evaluating Research in Context platform, in which the Academy is participating; (d) the ad hoc committee of the European Science Foundation concerned with the problem of quality indicators in the social sciences.

There is a striking parallel between the social sciences and the design and engineering disciplines when it comes to output diversity, but the outputs in the social sciences are of an entirely different order. The parallels between some social science and humanities disciplines is that most are context-related, making their international orientation endogenous. With respect to social impact, there are far fewer parallels between the design and engineering disciplines and the humanities on the one hand and the social sciences on the other. This is because in the social sciences, human behaviour and social processes are themselves always the objects of study. That not only makes other demands on the scrupulousness of knowledge transfer, but is also more likely to lead – as indicated in the foregoing – to "contested knowledge" due to vested interests.

2. Aim and impact: what do we wish to achieve with this advisory report?

The aim of the advisory report is to offer organisations that assess research in the social sciences a set of guidelines for defining appropriate assessment instruments. The intention is for these organisations to elaborate on this advice by developing better assessment instruments.

3. Who is the report for?

The advisory report is intended for organisations involved in the assessment of social science research. The main target group consists of the deans in the social sciences in the broad sense. The secondary target group consists of organisations that assess research, such as the Netherlands Organisation for Scientific Research (NWO). Both target groups are currently addressing the question of diversity in publication culture and valorisation.

4. Nature of the advisory document

[report, memorandum, foresight study]

Advisory report.

5. What is the deadline for the advisory report?

[Month in which the Academy advisory report is to be published, and any other important dates for the target group]

The target deadline is late September / early October 2011⁶. The discipline-specific consultative bodies in the social sciences will hold meetings close to this date.

⁶ Date no longer applicable upon signature of the Resolution inaugurating the “Quality Indicators in the Social Sciences” Committee.

APPENDIX 2.

“SOCIAL RELEVANCE”

AS A CRITERION IN QUALITY ASSESSMENTS:

EXAMPLES FROM ABROAD

The main question addressed in this appendix is: how is societal relevance dealt with abroad? The question is especially pertinent for the following quality assessments:

- Appraisals by universities of individual researchers within the context of promotions and recruitment.
- Assessments and external reviews of research groups.
- Assessments of research proposals within the context of competitive funding.

Individual career appraisals

There has been very little systematic research on how different countries deal with the concept of societal relevance in individual career appraisals. Some information is available, however, on the approach taken by foreign universities. Practices abroad vary: some universities use a diagram such as the one given in this advisory report; others use more general guidelines that refer to research, teaching and management. Some examples:

- Harvard University (USA) differentiates between three types of professors, including those who make a major contribution to the development of “clinical expertise”. Harvard distinguishes between expertise, recognition and scholarship.
- In Sweden, the law dictates the criteria for appointments to professorships. The law refers only to research and education. Universities can add other criteria at their discretion. Gothenburg University does this, for example, but most of its additional criteria refer to management skills.
- University College Dublin (Ireland) has a “senior management pathway” that also involves outreach activities such as:
 - “Making a substantial contribution to intellectual, cultural, social, or economic life at institutional, regional, national or international level.
 - Having a substantial beneficial influence on the development or application of public or professional policy or practice at institutional, regional, national or international level.

7 This appendix is based on information provided by Dr Barend van der Meulen (Head of Science System Assessment at the Rathenau Institute).

- Significant engagement with and positive impact on areas of community activity at local and international level arising from scholarship activity.
- Developing and maintaining innovative and/or creative relationships, including secondments or other forms of service, with external bodies at a regional, national or international level, bringing a substantial benefit to the University as well as a reciprocal benefit to society.
- Establishing a successful company of benefit to the local or national economy”.
- Massey University (Wellington, New Zealand) has a “Pathways to Professor” booklet describing “excellence and leadership in service both to the university and the external community”.

Assessments and external reviews of research groups

National assessment and external review systems for research groups differ in the extent to which:

- assessment is based on quantitative performance or on external reviews by panels; and/or
- results are linked to research funding.

The Netherlands leads the way with its system of external research reviews and the approach developed in the early 1990s by NWO, the Academy and VSNU (with a protocol, self-evaluation reports, a review committee and, in some cases, bibliometric analysis). This system has many imitators abroad. This sometimes takes the form of a national exercise (such as in Norway, organised by the Research Council of Norway); at other times, it is organised by individual universities (such as recently at Finland’s Helsinki University). As in the system applied in the Netherlands, societal relevance is often included in the protocol.

There is no evidence that better methods of fleshing out this aspect have been developed abroad.

Other systems are performance-based, with the outcomes sometimes being linked to funding.

Examples include:

- Flanders, where money from a “special research fund” is distributed according to a particular output formula. The system does not include any consideration of societal relevance. There is, however, much discussion of how the “non-compliant” output of the humanities and social sciences should be included in the formula.
- The United Kingdom, where the Higher Education Funding Councils use the outcomes of the Research Excellence Framework (formerly the Research Assessment Exercise) to distribute funding to the universities (in the Netherlands this is known as the *eerste geldstroom voor onderzoek* or direct research funding mechanism). The British system pays close attention to impact and it is increasingly coming to resemble the Dutch system. Research units up for review are assessed by experts on their “overall quality”, “outputs”, “impact” and “research environment”. The

manual for the panel responsible for the social sciences (C2) provides a detailed explanation of possible outputs that may be included, a specification of the assessment scale, and a description of the criteria (source: www.ref.ac.uk/panels/assessmentcriteriaandleveldefinitions):

“The main panel acknowledges that impact within its remit may take many forms and occur in a wide range of spheres. These may include (but are not restricted to): creativity, culture and society; the economy, commerce or organisations; the environment; health and welfare; practitioners and professional services; public policy, law and services. The categories used to define spheres of impact, for the purpose of this document, inevitably overlap and should not be taken as restrictive. Case studies may describe impacts which have affected more than one sphere.

Impact of any type may be local, regional, national or international, in any part of the world. The beneficiaries of impact may include (but are not restricted to) community/ies, the environment, individuals and organisations. The panel will treat all forms and spheres of impact and any beneficiaries described on an equal basis, assessing them according to the generic REF criteria of reach and significance.

HEIs are reminded that impacts on research or the advancement of academic knowledge within the higher education sector (whether in the UK or internationally) **are excluded**. Other impacts within the HE sector that meet the definition of impact for the REF **are included** where they extend significantly beyond the submitting HEI. (See “Guidance on submissions”, Annex C.)”

A table gives examples of what can be counted as impacts in various subdisciplines.

Assessments of research proposals within the context of competitive funding

Although this advisory report does not discuss the assessment of research proposals within the context of competitive funding, it is possible to learn something from the way other countries deal with the concept of societal relevance in that regard.

Assessment based on societal relevance is important in the English-speaking and Scandinavian science communities. The NWO’s sister organisations in these countries do not have an easy time considering the societal relevance of research proposals, however. They usually do so by including a separate dimension for societal relevance in the protocol that is then specified in more detail. The importance of this dimension also depends on the funding system (e.g. programme funding versus open competition).

One example is the situation in the United Kingdom. The ESRC is going through a learning process concerning the evaluation of what used to be called “user relevance” and is now known as “impact”. This has resulted in a series of studies and attempts to measure and evaluate the impact of social science research. Important observations and conclusions that can be drawn from the ESRC’s learning process are:

- The ESRC does not expect all research to have a socio-economic impact. Outstanding science projects that do not have such an impact will not be “punished”.

- Factors conducive to impact are:
 - Institutionalised relationships and networks with user communities.
 - Well-organised user involvement in research.
 - A portfolio of research projects relevant to users.
 - Involvement of intermediary organisations, platforms, and so on.
- It is very difficult to quantify impact and doing so “remains challenging”.
- It is important for researchers to include a robust impact plan in their research proposal so as to increase the likelihood of socio-economic impact as much as possible. If no impact can be anticipated, the reasons should be explained in this plan.

Conclusions

Although “peer review” is often used as an overall term, suggesting a level of uniformity in assessment practices, in reality these practices vary hugely. This appendix offers a brief impression of different attempts made to flesh out the concept of societal relevance in three types of quality assessments.

The most important conclusions are:

- In countries where research councils and universities are under heavy pressure to be socially relevant, assessment criteria reflect this.
- That has led in some countries to a more systematic framework for quality assessment that includes societal relevance.
- Claiming societal relevance is not enough; the claims must be backed up by demonstrable efforts and results (utilisation, reputation).

