Assessing the qualities of science

On June 28th 2022 the Advisory council for science, technology and innovation (AWTI) received a request from the President of the House of Representatives of the Dutch Parliament, to assess and clarify how the quality of science can be determined. The AWTI responded to this request in the form of a letter that was offered to the President on December 19th 2022. An update was sent on January 11th 2023. The current document is the English translation of those letters.

The House of Representatives of the Netherlands
Attn. Drs. V.A. Bergkamp
President of the House
PO Box 20018
2500 EA The Hague
The Netherlands

date:	December 19th 2022	appendix:	--
about:	Assessing the qualities of science	e-mail:	t.vanviegen@awti.nl
feature:	038/22/sk	Your feature:	Your letter dated June 28th, 2022

Dear Mrs. Bergkamp,

Introduction

Science provides new knowledge, makes it possible to interpret and explain phenomena in the world around us, helps us find answers to complex social issues, contributes to the development of upcoming talent and strengthens the innovation and earning capacity of our economy. Scientists, administrators of scientific institutions, grant providers, policymakers and politicians pay a great deal of attention to the quality and value of science and scientists. With this letter, the Advisory Council for Science, Technology and Innovation (AWTI) responds to the request of the committee for Education, Culture and Science to clarify how the quality of science can be determined.[1]

The quality and value of science are important for the significance of science for society, for the quality of decision-making and for public trust. In addition, the interpretation of the quality of science is decisive for:

- the career of scientists;
- the distribution of funding for science;
- the attractiveness of Dutch science to international partners; and
- government policy aimed at science and innovation.[2]

The question arises how the quality of science can be determined in a way that does justice to the nature of science, the different types of scientific research, the different types of activities and roles that scientists play, the intended goals and the use of scientific activities. In this letter the AWTI outlines the various ways in which the quality of science can be determined, what the advantages and disadvantages are and what new developments are taking place.

Science is a global activity: a strong international position matters

Collaboration with international scientists, social organizations and companies is of great importance. In this way, researchers in international collaborations push the boundaries of science and enrich the knowledge base. In international cooperation, knowledge circulates faster, and knowledge has more

[1] Bergkamp, V. (2022). Brief aan de voorzitter van de Adviesraad voor wetenschap, technologie en innovatie (AWTI) over een adviesvraag over het objectief vaststellen van de kwaliteit van wetenschap, nr. 2022Z12914, 28 juni 2022.
[2] Adviesraad voor het Wetenschaps- en Technologiebeleid (2014). *Boven het Maaiveld – focus op wetenschappelijke zwaartepunten* [link]. English summary: *Standing out from the crowd* [link].

Visiting address:
Advisory council for science, technology and innovation
Prins Willem-Alexanderhof 20
2595 BE The Hague

t. +31 (0) 70 311 0920
e. secretariaat@awti.nl
w. www.awti.nl

impact through application in high-quality products, processes and services. International cooperation is also important to find answers to global social issues and to strengthen the innovation and earning capacity of a knowledge economy such as the Netherlands.

A strong international position of Dutch science is important because the Netherlands wants to be an attractive partner for the international scientific community, civil society organizations and the international business community.[3,4] International top universities, institutes and companies are collaborating more and more closely worldwide, selecting partners that are highly valued by other scientists. The availability of scientific talent (scientists and students) is an important factor in the establishment of R&D-intensive and high-tech companies. That talent is increasingly moving across borders and the Netherlands can attract and retain international talent as an attractive partner. Investments in R&D are also increasingly cross-border. As an attractive partner, the Netherlands can attract R&D investments here and conduct high-level research that requires more funding than is available nationally. Attractiveness as a collaborative partner is often based on reputation and international recognition of the quality of Dutch science.

Quality of science is about qualities of science

The quality of science has to do with values ascribed to certain scientific activities. Therefore, it is better to speak of the qualities of science. For example, the 'intrinsic' value of science refers to the way in which science is practiced. Scientific knowledge is systematically gathered in accordance with the principles of honesty, due care, transparency, independence and responsibility, as included in the Dutch Code of Conduct for Scientific Integrity. The 'extrinsic' value of science is about the impact of scientific knowledge on science, education, society and the economy.[5,6,7,8] This value has to do with the significance that scientific activities have for science itself, through enrichment of the knowledge base; for education, by well-trained students and researchers; and for society and the economy, through the application of knowledge in new products, processes, services and solutions for social issues.

The qualities of science differ between different types of research and types of scientist profiles. In basic scientific research, the quality of science will be more about scientific impact, or the contribution that the scientific results make to the enrichment of the knowledge base. Application-oriented scientific research will rather look at the contribution of the results to applications in new products, processes or services. Other qualities are important for scientists who also teach or often collaborate with the business community, such as the contribution to the development of new educational material and the ability to work on specific questions from the business community. Which qualities are important therefore depends on the totality of activities that scientists carry out and the roles they play.

Moreover, the weighting of the qualities of science is influenced by developments in science and society. An important driving force in this is the development of *open science*,[9,10] because requirements of science practices have changed according to the principles of *open science*. It requires more time and attention from scientists for additional activities.[11] For example, *open science* prescribes that data must be collected

---

[3] AWTI (2017). *STI Diplomacy – Advancing the internationalization of science, technology and innovation* [link].

[4] Ministerie van Onderwijs, Cultuur en Wetenschap (2019). *Nieuwsgierig en Betrokken*. *De waarde van wetenschap* [link].

[5] Ministerie van Financiën (2014). *IBO Wetenschappelijk onderzoek* [link].

[6] Rathenau Instituut (2020). *Balans van de wetenschap 2020*. Den Haag (auteurs: Koens, L. et al.) [link].

[7] CPB (2016). *Kansrijk wetenschapsbeleid*. Den Haag: Centraal Planbureau, p. 51 [link].

[8] KNAW (2018). *Maatschappelijke impact in kaart*, Amsterdam, KNAW [link].

[9] The global movement to make scientific knowledge more accessible, freely available, accessible and reusable through a more open and inclusive way of conducting, publishing and evaluating scientific research.

[10] UNESCO (2021). *UNESCO Recommendation on Open Science* [link].

[11] Rathenau Instituut (2021). *Samen verder met open science – Op weg naar betekenisvolle publieke betrokkenheid bij onderzoek* [link].

and stored according to the FAIR principles.[12] This affects the intrinsic value of science. Other aspects of *open science*, such as more participatory forms of research, have an effect on the extrinsic value of science. An increasing focus on societal challenges in the financing of science leads to more attention for the social impact of science. The rapid spread of information and the emergence of fake news via social media underline the importance of transparency, carefulness, independence and honesty, and thus the quality of scientific practice.

Quantitative evaluation of science gained the upper hand

Evaluating science and research is not a recent development, and dates back to the time when science was institutionalised, around the middle of the 17th century. With the emergence of new structures that organised science, the practice of evaluating science spread to places like laboratories, departments and universities.[13] During the 20th century, government funding for research also became increasingly institutionalised, for example with the establishment of the Foundation for Fundamental Research on Matter (FOM) in 1946, and the predecessor of the Dutch Research Council (NWO), the Netherlands Organization for Pure Scientific Research (ZWO) four years later. From then on, evaluation of science was also used for the selection of research proposals and the allocation of research funding.

For a long time, evaluation of science was based on *peer review*. *Peer review* has the important advantage that other experts in the field (or in a related discipline) can make a substantive assessment of the quality of a scientific article, research proposal, grant application or career. *Peer review* also plays an important role in current scientific assessment, but *peer review* also has a number of disadvantages. Peer reviewers have difficulty valuing high-risk and interdisciplinary research, which puts this type of research at a disadvantage, especially when assessing grant applications.[14] Peer reviewers often disagree with each other, which shows that peer reviews are subjective.[15] Peer reviewers appear to be sensitive to the names of prestigious scientists and institutions[16] and they judge publications by people they are familiar with more positively.[17,18] *Peer review* is therefore sensitive to subjectivity and bias. Despite this, *peer review* is seen as the gold standard for assessing quality in publishing articles and awarding research funding.

Quantitative, in particular bibliometric, indicators were introduced in the 1920s and 1930s. Due to the enormous growth in the number of scientific articles, it became impossible for scientists to keep track of the results of new research. Bibliometric research took off to help scientists gain a better picture of developments in their field. As the use of bibliometric research increased in the 1960s, so did the number of applications.[19] Managers of higher education institutions, for example, saw bibliometrics as an instrument to overcome the aforementioned subjectivity of the peer review system. Incidentally, the question arises to what extent it is a problem that there is subjectivity when peers assess each other on feasibility and consistency. Bibliometric data is also not neutral. Despite this, bibliometric indicators were increasingly used to assess the scientific impact of the applicant in research funding applications. Using bibliometric indicators for the evaluation of the impact of science, however, has a number of drawbacks.

For example, there are major differences between scientific fields and disciplines, which means that bibliometric indicators are not comparable between these fields and disciplines. The scope of disciplines

---

[12] FAIR stands for 'findable, accessible, interoperable and reusable' [link].
[13] Gingras, Y. (2014). *Bibliometrics and Research Evaluation. Uses and Abuses.* MIT Press, p. 56.
[14] Langfeldt, L. (2006). The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation*, Volume 15, Issue 1, April 2006, p. 31–41.
[15] Cole, Cole and Simon (1981). Chance and Consensus in Peer Review. *Science,* Vol 214, Issue 4523 p. 881-886
[16] Langfeldt, L. (2006).
[17] Teplitskiy, M. *et al.* (2018). The sociology of scientific validity: How professional networks shape judgement in peer review. *Research Policy*, Volume 47, Issue 9, November 2018, p. 1825-1841.
[18] Where knowing is defined by proximity in a co-authorship network.
[19] Gingras, Y. (2014). *Bibliometrics and Research Evaluation. Uses and Abuses.* MIT Press, p. 56.

differs and there are differences in publication and citation habits between disciplines. In the humanities, for example, a book is more often published, and English is not always the language of instruction. This is difficult to compare with disciplines in which short research articles are the most commonly used output. It is possible to correct bibliometric indicators for differences between disciplines, but this is not done for indicators such as the journal impact factor and the h-index. Despite this limitation, these two indicators in particular are widely used as core indicators for the evaluation of scientists' careers and research proposals.[20] In addition, both indicators have a number of drawbacks, which make them unsuitable for evaluating and comparing the performance of individual scientists (see box).[21,22]

Disadvantages of the *journal impact factor* and *h-index*

The *journal impact factor* (JIF) is an index based on the number of citations of articles that have appeared in a particular journal.[23] The JIF is used because it would have a predictive value of the impact of scientific publications. Articles appearing in a journal with a high JIF would receive more citations than articles appearing elsewhere. In addition, the bar to publish an article would be higher for journals with a high JIF. An article appearing in such a magazine would certainly be of high quality. However, the JIF has a higher value for disciplines that cite a lot and in which mainly recent articles are cited. The JIF is also influenced by review papers that are cited more often than research articles.[24] In addition, it is unclear how journals calculate their JIF, because journals are not transparent about their calculations. Moreover, scientists may change their behaviour because they are assessed on the JIF, which directs them to publish in a journal with a high impact factor. For example, they choose topics that they think are easier to publish in those kinds of journals. Finally, the JIF relates to the reputation of a journal and is not designed to assess the quality of an individual scientist, although it is oftentimes used for that purpose.

The *h-index*[25] measures the scientific impact of a person's publications over his or her career. The h-index also has flaws, because it is sensitive to age and experience, so that young scientists always have a lower *h-index* than older researchers. Therefore, no fair comparison is possible between scientists at different stages of their careers. In addition, the *h-index* makes it impossible to make comparisons between different scientific disciplines (sometimes the h-index even gives a distorted picture between scientific fields within a discipline), because citation habits differ between scientific disciplines.[26,27]

---

[20] Leydesdorff, L. *et al.* (2016). Professional and citizen bibliometrics: complementarities and ambivalences in the development and use of indicators – a state-of-the-art report. *Scientometrics,* Volume 109, October 2016, p. 2129-2150.

[21] For an overview of the discussion around *journal impact factor* see Larivière, V and C.R. Sugimoto (2019). The Journal Impact Factor: A brief history, critique, and discussion of adverse effects. in: Glänzel, W. *et al.* (Eds.): *Handbook of science and technology indicators*. Springer. 2019; 3–24.

[22] For an overview of the discussion around the *h-index* see for example Bornmann, L. and H.-D. Daniel (2009). The state of h index research. Is the h index the ideal way to measure research performance? *EMBO reports*, 10(1), 2-6. doi:10.1038/embor.2008.233.

[23] The *journal impact factor* of a scientific journal is an index that indicates the annual average number of citations of articles that have appeared in the journal in the previous two years.

[24] An important argument of opponents of the use of the *journal impact factor* in the assessment of applications and careers is the skewed distribution of citations: a small number of articles have a very high number of citations, often including review articles. However, Waltman and Traag (2021) have shown with computer simulations that this statistical argument does not always hold. In certain specific situations, the JIF is a better predictor of scientific impact than the number of citations of an individual article. For example, recently published articles often have few citations. Waltman and Traag argue that the JIF should therefore not be portrayed as a bad measure for statistical reasons. Nevertheless, the authors indicate that the use of the JIF in science evaluation is undesirable because the indicator is sensitive to manipulation and there are problems with transparency and calculation of the indicator values. Waltman, L. and V.A. Traag (2021). Use of the journal impact factor for assessing individual articles: Statistically flawed or not? *F1000 Research,* https://doi.org/10.12688/f1000research.23418.2.

[25] A scientific researcher has index *h* if *h* of the total of N publications has been cited at least *h* times and the other (N-*h*) publications have all been cited less than *h* times.

[26] Larivière, V. en C.R. Sugimoto (2019) p. 34.

[27] Rijcke, de, S. *et al.* (2021). *Halt the h-index*, Blogpost on www.leidenmadtrics.nl, May 19th 2021 [link].

Citation-based indicators can also be manipulated by scientists in an *editor* or *reviewer* role. If a scientist systematically indicates that their articles must be cited, this pollutes the indicators that are based on citations.

Another problem with the focus on the one-sided use of bibliometric indicators in the assessment of scientists is that other qualities of scientists are insufficiently reflected. For example, contributions that scientists make to policy are not taken into account, because there are no useful indicators for this[28], while that contribution to policy is also a part of the impact of science. The scientist who spends more time on other activities such as teaching or sharing knowledge with policymakers and possibly less on publishing scientific articles will by definition be disadvantaged by the one-sided use of bibliometric indicators.

Rankings are popular, but say little

Quantitative indicators, including bibliometric indicators, are increasingly used as yardsticks to compare universities and scientists and to include them in rankings in order of their scores. These university rankings show at a glance who or what scores highest and are therefore widely used. Students and their parents use them to determine where to study and scientists use them to shape their careers. However, they are also used for the development of strategy and policy and for making investment decisions by companies, governments and universities themselves.[29]

However, rankings are problematic because they ignore the complexity of reality. Scientific institutions carry out different activities in different fields of science and in different environments. By placing everything against one and the same yardstick, apples and oranges are being compared. Using university rankings to compare countries is also problematic, because countries differ in the number of universities, in the specialization in scientific fields and disciplines, and in the type of universities (more education vs. research oriented, general vs. technical, etc.). These differences are ignored in a single index or indicator. In addition, the way many of these rankings come about is opaque. The underlying data can be influenced by those who supply the data. Furthermore, a change in the position in rankings says nothing about the change in the performance of institutions and countries. This is because the position is strongly influenced by changes in the numbers of participating institutions and in the methodology.

Growing need to differently evaluate the qualities of science

Since the beginning of this century, there has been growing dissatisfaction with the strong focus on the quantitative assessment of the quality of science and the narrow focus on the results of research in the form of publications and citations. At the same time, more attention is being paid to the impact that scientific results, scientists and knowledge institutions have on education, policy, social issues and innovation and earning capacity.

These developments are the reason for UNL (Association of Universities in the Netherlands, formerly known as VSNU), NFU (Netherlands Federation of University Medical Centres), KNAW (Royal Netherlands Academy of Arts and Sciences), NWO (Dutch Research Council) and ZonMw[30] to publish 'Room for everyone's talent' in 2019.[31] The organizations advocate a new way of ‘recognition and rewards’ in science, in which a number of core domains are distinguished: education, research, impact, leadership and (for UMCs) patient care. First of all, 'the new recognition and rewards' should allow for career diversification. It should therefore be possible to choose a mix of core domains that diversifies the profile

[28] Benedictus, R. *et al.* (2016). ‘Fewer numbers, better science’. *Nature*, volume 538, p. 453–455.
[29] Çakır, M.P. *et al.* (2015). A comparative analysis of global and national university ranking systems. *Scientometrics,* vol. 103(3), p. 813-848.
[30] https://www.zonmw.nl/en/
[31] VSNU, NFU, KNAW, NWO en ZonMw (2019). *Room for everyone’s talent* [link].

of scientists. Second, more attention should be paid to contributions that individual scientists make to collaborative work, shifting the focus from the individual to the team. This creates more appreciation for team science. Thirdly, the intention is to shift the focus from quantitative data, such as the aforementioned bibliometric data, to 'quality'. Quantitative data can indeed be informative about quality, but the publication 'Room for everyone's talent' states that more attention should be paid to content, scientific integrity, creativity and impact, taking into account the context and profile of the scientist. Fourth, there is a focus on the appreciation of activities that contribute to *open science*. Finally, academic leadership should be promoted along the entire academic, hierarchical ladder.

Following the publication 'Room for everyone's talent', UNL has reassessed the university job classification system (UFO) and, together with NWO and KNAW, revised the strategy evaluation protocol (SEP), which is used to assess research units.[32,33] For example, the profile for Data Steward, who ensures the availability and security of data stored online, has been included in the revised UFO. In addition, the adjusted profiles for Assistant Professor, Associate Professor and Full Professor pay more attention to education, impact, open science and leadership. The revised SEP of 2021-2027[34] pays explicit attention to the quality of the research, the social relevance of the research and the future-proofness of the research units. The units are also assessed on the specific aspects: open science, academic culture, HR policy and promotion policy. For the SEP, the research unit writes a narrative argument supported by quantitative indicators that the research unit can choose itself. However, it is not recommended to use the *journal impact factor* or the *h-index*.

Science funders NWO and ZonMw have implemented the principles of the new 'Recognition and Rewards' in their assessment system for selecting research proposals. These financiers use a broader definition of impact, which also includes *open science* and team science. NWO and ZonMw are increasingly using a narrative or evidence-based CV, which consists of:

- an academic profile, in which researchers explain what kind of scientist they are: what is the researcher's research focus, agenda and vision? What have they done to achieve that vision?; and
- a selection of the most relevant 'output' (this is broader than just publications and also includes, for example, scripts and data sets).

Researchers can substantiate the quality of their 'output' on the basis of several indicators, both quantitative and qualitative, as long as they each relate to only one output item. The *journal impact factor*, *h-index* or other indicators and descriptions that refer to a journal, publisher, publication platform or additions and averages of citations are therefore not allowed. It is therefore allowed to say that an article X has appeared in journal Y and that article X has a number of Z citations. However, it is not allowed to say that article X appeared in the leading journal Y with JIF 42, which would suggest that article X, based on the citation score of the journal, is an exceptionally good article. The indicators and descriptions must therefore relate directly to the specific output. NWO is working on a format for the evidence-based CV that offers both applicants and assessors more structure and more clarity about which types of substantiation are and are not allowed. This improves the comparability between applications and streamlines the assessment process.

Broader recognition and appreciation of the different aspects of scientists' work is embraced by many scientists, but there is dissatisfaction with the changes in the measurement of scientific impact. As part of the 'Recognition and Rewards' program, the signatories have announced that they will no longer use the

---

[32] UNL (2021). *Eerste aanpassingen Erkennen & Waarderen in functieprofielen wetenschappers* [link].
[33] VSNU, KNAW, NWO (2020). *Strategy Evaluation Protocol 2021-2027* [link].
[34] The revision was also accompanied by a name change from standard evaluation protocol to strategy evaluation protocol.

journal impact factor and the *h-index* in the assessment of grant applications and academic careers. Critics of these changes believe that too much importance is attached to *peer review*, while *peer review* also has shortcomings (see 'Quantitative evaluation of science gained the upper hand'). Without these indicators, peer reviewers would, according to the critics, be insufficiently able to assess the quality of scientific work. However, the *journal impact factor* and the *h-index* are not suitable for assessing and comparing individual scientists or research proposals. The advantage of *peer review* is that fellow scientists can make a substantive assessment of the quality of a scientific article, a research proposal or a PhD. A descriptive and substantiated CV in which the scientist makes clear what their scientific profile is, which scientific activities and results are important, helps with this. Quantitative indicators can be used in this that provide substantiation but have fewer limitations.

In addition, in the case of qualitative evaluations by peer reviewers, it is important that they can form an opinion on the work of scientists in the same or closely related field on the basis of their knowledge and expertise. Peer reviewers increasingly have to assess the work of other scientists across the boundaries of their own field because scientific domains are broadening and multidisciplinarity is increasing at financiers and research institutions. Bibliometric analyses, for example by machine learning, can be of service to peer reviewers, provided they are used sensibly to assess the qualities of science. Bibliometric analyses can indicate important developments in disciplines, how scientific networks are put together and what the place of individual scientists and institutes is in this. By providing these insights, bibliometrics can strengthen *peer review* and help peer reviewers make a careful assessment.

Joint development of a rich palette of methods and indicators is necessary

Evaluating the qualities of science differently is primarily a cultural change. Cultural changes often take a long time and are met with resistance. It is important to make adjustments step by step, to involve everyone in the science and to keep talking about questions, needs, concerns and possibilities. By experimenting, closely following developments and mapping out the effects, it is possible to learn what works and what doesn't and to make adjustments. The guiding principle here is that there is no single quality of science. To be able to interpret the diversity of qualities of science and scientists, a rich and nuanced palette of quantitative and qualitative methods and indicators is needed. These indicators should suit the characteristics and needs of different disciplines, different core domains of science and different profiles of scientists. Both qualitative and quantitative methods each have their own value and must complement each other. And although the 'Recognition and Rewards' program focuses on universities, universities of applied sciences also need a diverse palette of methods and indicators. Lecturers and researchers at universities of applied sciences also contribute to the qualities of research and education in different ways.

In addition to the universities, university medical centres (UMCs) and research financiers who wrote 'Room for everyone's talent', the Dutch 'Levensbeschouwelijke Universiteiten' (NLU; Philosophy of Life Universities) and 'De Jonge Akademie' (The Young Academy) have published a vision that is in line with the vision of the new 'Recognition and Rewards'.[35,36] At the knowledge institutions, committees have set to work to implement the intentions and recommendations of 'Room for everyone's talent' and to adapt them to the context of their own institution. How different the implementation will ultimately be at the various institutions cannot yet be determined. UNL organizes meetings through the national program 'Recognition and Rewards' to exchange *good practices* in order to achieve convergence at the various institutions.[37]

[35] NLU (2020). *Visie van de NLU commissie 'Erkennen en Waarderen' op de universiteit waar medewerkers het goed kunnen doen en het goed hebben* [link].
[36] De Jonge Akademie (2020). *Good examples merit imitation* [link].
[37] See https://recognitionrewards.nl/wp-content/uploads/2022/11/Programme-plan_RR_ENG.pdf

It is important that the methods and indicators are valid and fit with the objectives of a research fund or program, or of a scientist's profile and career. Further development of these methods and indicators for the core domains of education, leadership, patient care and impact deserves priority. The Dutch knowledge institutions, including universities of applied sciences, must jointly develop this and be able to learn from each other. Even within a knowledge institution, every faculty or research unit does not have to reinvent the wheel if the dialogue about this is well orchestrated. Working together also minimizes differences between knowledge institutions in reassessing recognition and appreciation. Excessive differences could hamper the mobility of scientists and thus their careers in the Netherlands.

The Netherlands is a forerunner in a global movement

The development of other ways of evaluating the qualities of science has been set in motion internationally, and the Netherlands is a forerunner in this. The movement towards *open science*, the growing attention for the social impact of science and the dissatisfaction with the one-sided focus on measuring scientific output is also an issue outside the Netherlands. In 2012, for example, the Declaration on Research Assessment (DORA)[38] was drawn up in San Francisco, which distanced itself from the *journal impact factor*. Meanwhile (as of September 26, 2022) the DORA has been signed by 2633 authorities and 19531 individuals from 159 different countries. In addition to DORA, several other manifestos and principles have been published, including the Leiden Manifesto (2015)[39], The Metric Tide (2015)[40] and the Hong Kong principles (2020).[41] These manifestos point out the misapplication of bibliometric indicators and therefore focus on responsible use of indicators, transparent communication of research results, *open science*, appreciation of different types of research and recognition of all contributions to research and academic activities. All initiatives agree that the commonly used *journal impact factor* and the *h-index* are unsuitable for the assessment of individual scientists.

Europe is taking steps

Actions and programs have now been initiated at the European level, in various European countries and at large international funds for scientific research to develop and apply other ways of evaluating science. For example, the European Commission, Science Europe and the European University Association (EUA) recently launched the agreement 'Agreement on reforming research assessment' (RRA).[42] This treaty contains agreements on how to assess research results in a responsible manner. More than 360 organizations from more than 40 countries have now signed the RRA.[43] NWO, KNAW, UNL and some of the universities and university medical centres have signed the treaty. The RRA calls for a qualitative judgment, in which peer review is central and quantitative indicators are used in a responsible manner. For example, the RRA calls for a move away from journal- and publication-based measures and hardly considers the possible disadvantages of peer review. A balance between quantitative and qualitative measures is necessary to arrive at a suitable instrument for interpreting scientific quality. According to the RRA, quality implies that scientific research is conducted according to transparent research processes and methodologies and that the research data is maintained in such a way that it can be easily reused. Thus, openness and results that are verifiable and reproducible contribute strongly to quality. In this definition, the quality of science is linked to scientific practice.

---

[38] See https://sfdora.org/read/.

[39] Hicks, D. *et al.* (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature.* Issue 520, Page 429-431. https://doi.org/10.1038/520429a.

[40] Research England (2015). *The Metric Tide* [link]. In 2022, The Metric Tide Revisited was launched, a review to see to what extent the recommendations from 2015 have been adopted and whether further recommendations are necessary for the responsible use of indicators in the next round of the Research Excellence Framework.

[41] Moher, D. *et al.* (2020). 'The Hong Kong Principles for assessing researchers: Fostering research integrity'. *PLoS Biology* 18(7): e3000737. https://doi.org/10.1371/journal.pbio.3000737.

[42] See https://www.eua.eu/downloads/news/2022_07_19_rra_agreement_final.pdf.

[43] See https://coara.eu/agreement/signatories/.

The League of European Research Universities (LERU) presented a paper in January 2022 to provide tools for a new way of recognizing and appreciating.[44] According to LERU, too much attention is paid to the debate about bibliometric indicators, without providing sufficient alternatives to assess multidimensional academic careers. This is also an important point of attention for the Dutch initiative (see 'Growing attention to different ways of evaluating the qualities of science'). In line with the RRA mentioned above, LERU called for use of quantitative indicators only to support a qualitative assessment.

In addition to the Netherlands, Finland[45], Norway[46] and the United Kingdom[47] have published *position papers* and launched programs and toolboxes. They all state that quantitative indicators should only serve to support sound qualitative judgement. In the United Kingdom, it is about the need for diversification and inclusiveness in the research field to attract people from different backgrounds in order to meet the upcoming shortage of scientists. The British Ministry for Economy, Energy and Industrial Strategy (BEIS) is taking the initiative, while the initiative in the Netherlands, Finland and Norway has started from science itself. The British government wants to broaden career paths and offer more stability for researchers, making it easier to work temporarily outside of academia. Leadership qualities also receive special attention, just as in the Norwegian and Dutch approach. In addition, the British strategy mentions a narrative CV in which more attention is paid to valuing a wider range of experiences and objectives. In line with this, the UK grant providers signed a letter of intent in 2021 to introduce the narrative CV.[48]

Just like in the Netherlands, the initiatives in Norway and Finland are linked to the movement towards *open science*. Comparable to the Netherlands, the Finnish and Norwegian universities are given scope to draw up their own guidelines, but the documents in Finland and Norway are more detailed. As a result, the national *paper* focuses more on implementation at the level of knowledge institutions.[49] Both Finland and Norway pay a lot of attention to the transparency of the assessment process. In addition, Norway and Finland also offer a national CV and research portfolio template. The Norwegian toolbox goes one step further with an 'automagic' CV: a system in which a researcher can obtain the necessary information in a user-friendly manner and where the CV does not have to be formatted differently for each institution. It is striking that the Dutch core domain of patient care is not included in the Finnish and Norwegian approaches. This probably has to do with how patient care in the Netherlands is closely intertwined with scientific research at University Medical Centres (UMCs) and how the UMCs in the Netherlands are integrated into the university structure.

International research funds endorse the principles

Different evaluation of the qualities of science has also been introduced by various international research funds. Dutch scientists also make use of this: about 5% of the total funding of universities comes from foreign organizations (2018 figures), of which about 80 to 90% comes from the European Framework Programs for Research and Technological Development.[50,51] The European Framework Program follows

[44] LERU (2022). *A Pathway towards Multidimensional Academic Careers* [link].
[45] Working group for responsible evaluation of a researcher (2020). *Good practice in researcher evaluation. Recommendation for the responsible evaluation of a researcher in Finland* [link].
[46] Universities Norway (2021). *NOR-CAM – A toolbox for recognition and rewards in academic careers* [link].
[47] Department for Business, Energy & Industrial Strategy (2021). *R&D People and Culture Strategy* [link].
[48] Alzheimer's Research UK, British Heart Foundation, Cancer Research UK, National Institute for Health Research, Royal Academy of Engineering, UK Research and Innovation and Wellcome Trust (2019). *Funders joint statement: Exploring a shared approach towards a narrative CV* [link].
[49] For example, the Norwegian toolbox has a matrix that indicates which area of competence should be assessed along which axes.
[50] Rathenau Instituut (2020). *Ontwikkeling derde geldstroom en beïnvloeding van wetenschappelijk onderzoek – Een data- en literatuuronderzoek ter beantwoording van de motie-Westerveld*.
[51] The various annual overviews of Dutch universities do not provide any insight into exactly where foreign funding comes from.

the guidelines of the RRA. Another important European funder is the European Research Council (ERC), which stimulates scientific excellence in Europe by allowing researchers in all fields to compete for grants for starting, established and highly experienced researchers. ERC uses an extensive peer review process to assess and select scholarship applications. ERC has signed DORA and the *journal impact factor* is therefore not used in grant proposals. In addition, applicants can name a maximum of ten (five for the Starting Grant) 'outputs'. Bibliometric indicators relevant to the research field in which the researcher operates may be mentioned. The ERC will not sign the RRA, because the ERC considers it unclear what the RRA adds to the peer review method that the ERC follows.[52]

DORA is also signed by several international funds open to scientists from other countries such as Bill & Melinda Gates Foundation, VolkswagenStiftung and Wellcome Trust. The two leading US funds National Science Foundation (NSF) and National Institute of Health (NIH) that also fund foreign scientists (especially NIH) are not signatories to DORA. However, NSF writes that it has adjusted the instructions for grant applications in line with DORA's principles.[53] For example, the header of publications has been adapted to products that include publications, patents, datasets, software and copyrights. The NIH uses a peer review system where peer reviewers assess a *biosketch* of the applicant. The *biosketch* may contain a maximum number of products and contributions. In addition, the *biosketch* contains a paragraph with the most relevant scientific contribution and a reference to an online overview of publications. This could be interpreted as an encouragement for peer reviewers not to fixate on the *journal impact factor* or other indicators.[54]

Effect on the international position of the Netherlands is unknown

Maintaining the strong international position of the Netherlands is important, as argued earlier (see 'Science is a global activity: a strong international position matters'). Critics of the new 'Recognition and Rewards' are concerned that the international position of Dutch science will weaken. They are afraid that the reputation of Dutch science will be damaged if qualities other than scientific impact play a role in the evaluation of science and commonly used impact indicators such as the *journal impact factor* and *h-index* may no longer be used. In basic scientific research in certain fields such as the natural sciences, this would be especially applicable.

University rankings are not very informative to indicate the international position of Dutch science (see 'Rankings are popular but say little'). A commonly used way of interpreting the scientific impact of disciplines is by calculating the publication and citation impact scores of those disciplines as a whole. For example, Dutch scientists are among the most cited scientists in the world. The Netherlands has a weighted citation impact score[55] across all scientific fields that is well above the global average in the period 2017-2020 (citation period up to 2021). The citation score of publications involving international collaboration is highest in the Netherlands. The Netherlands scores high in both the top 1% and the top 10% most cited publications. Only Singapore, Switzerland and the UK score higher. The Netherlands is also one of the top 5 reference countries with the highest citation impact scores in all scientific fields. No

[52] See https://erc.europa.eu/news-events/news/research-excellence-quo-vadis.
[53] See U.S. National Science Foundation | DORA (sfdora.org).
[54] See U.S. National Institutes of Health | DORA (sfdora.org).
[55] The citation impact score indicates how often a publication is cited, compared to other publications from the same field over a year (MNCS). The worldwide average per discipline has been set equal to 1. A citation impact score of 2, for example, means that a publication is cited twice as often as the average publication from the same discipline.

other country is in the top 5 in all areas.[56] This relative position of Dutch science has been stable over the past 10 to 20 years.[57,58]

In addition to the above quantitative indicators, the international position of Dutch science can be determined by looking at the financing of scientists in the Netherlands by European research funds. For example, scientists in the Netherlands often collaborate internationally and they also manage to obtain a relatively large amount of funding from the European Framework Programs and the ERC. The Netherlands is the sixth recipient of funding from Horizon 2020. In no other European country is the available funding from Horizon 2020 per researcher (from the public sector) higher than in the Netherlands. Scientists from the Netherlands are particularly successful in the field of excellent science and research for societal challenges.[59] The Netherlands is in fourth place of the top 10 with countries that received the most ERC grants in the period 2007-2020. Compared to the European average, the Netherlands has scored well in recent years in the domains 'Physical Sciences & Engineering' and 'Social Sciences & Humanities'; the Dutch award percentage is on average 6% higher than in Europe. The award percentage has also risen in the Life Sciences domain over the past two years.[60]

There are as yet no indications that the different evaluation of science and scientists will affect the international position of Dutch science, expressed by the above indicators. It would also be too early for that, because the initiatives to evaluate the qualities of science in a different way are recent and still in full development. The period is too short, because there is quite a bit of time between obtaining funding, conducting research, submitting a scientific article, the final publication and possible citations by peers.

However, there have been other developments in science in recent years, such as open access publishing, that may have had an impact. However, even the flight that *open access* has taken (in 2021, 82% of the peer-reviewed publications of Dutch universities were open access)[61] has had no effect on the international position of Dutch science, measured according to the above indicators. It is unlikely that an increase from 82 to 100% open access will ensure that these effects become visible.

Open access involves a comparable discussion about the effects on the impact of science and scientists, translated into citation impact scores. Critics state that:

- open access journals have a lower prestige, as a result of which other scientists will have difficulty valuing open access articles and journals; and
- in some fields there are few open access options, as a result of which scientific results are less visible and are therefore less cited.

This would also make it unattractive for scientists to work at research institutions where open access is the norm.[62] At the same time, there are studies that show that the citation impact of open access publications is greater and that the *journal impact factor* even increases.[63] There are also studies that

[56] See https://www.rathenau.nl/nl/wetenschap-cijfers/output/publicaties/ontwikkeling-van-het-wetenschappelijk-onderzoeksprofiel-van.
[57] See https://www.rathenau.nl/nl/wetenschap-cijfers/output/publicaties/ontwikkeling-citatie-impact-van-nederlandse-publicaties-wos.
[58] The Humanities are not included, because a relatively small proportion of the citations in these areas appear in Web of Science (WoS), as it is common practice in this field to publish in other ways as well.
[59] See https://www.rathenau.nl/nl/wetenschap-cijfers/nederland-en-horizon-2020.
[60] See https://www.rathenau.nl/nl/wetenschap-cijfers/werking-van-de-wetenschap/excellentie/toekenningen-erc-land-universiteit-en.
[61] See https://www.universiteitenvannederland.nl/nl_NL/nieuws-detail/nieuwsbericht/875-p-nederland-zet-weer-grote-stap-naar-100-open-access-nbsp-p.html.
[62] See https://www.prio.org/publications/11146. See overview in Huang, C. *et al.* (2019). The effect of open access on journal impact factors: A causal analysis of medical journals. *Physica A: Statistical Mechanics and its Applications*, Volume 533, 1 November 2019.
[63] See overview in Momeni, F. *et al.* (2021). What happens when a journal converts to open access? A bibliometric analysis. *Scientometrics*, Volume 126, pp. 9811-9827.

show neither citation advantages nor disadvantages in open access publications.[64] Incidentally, *open access* negatively affects global science. For example, in the 1990s and 2000s, scientists from the Third World could contribute through scientific publications, but often did not read scientific articles. Now a situation has been created in which these scientists can read articles, but not publish them because of the often high publication costs.

Based on these insights, it cannot be concluded that another form of science evaluation threatens the international position or reputation of Dutch science. Moreover, the international movement that is already underway shows that a different way of recognizing and appreciating is inevitable. The Netherlands is at the forefront of this, but not alone. Leading the way offers opportunities to determine the direction. Cooperation in the development of other methods of evaluation is also important from an international perspective. Exchanging knowledge and experience and developing and implementing new approaches in an international context ensures that Dutch science remains in line with the international field and at the same time can continue to give direction to the movement of evaluating the qualities of science in a different way. More attention and appreciation for the various qualities of science may actually make the Netherlands more attractive to new scientific talent, scientific collaboration partners, civil society organizations and the international business community.

Yours sincerely,

dr. E.E.W. Bruins
chairman

P.W.J. Essers
secretary

[64] Liskiewicz, T. *et al.* (2021). Factors affecting the citations of papers in tribology journals. *Scientometrics,* Volume 126, pp. 3321–3336.

Advisory council for science, technology and innovation

The House of Representatives of the Netherlands
Attn. Drs. V.A. Bergkamp
President of the House
PO Box 20018
2500 EA The Hague
The Netherlands

date:	January 11th 2023	appendix:	--
about:	Update Assessing the qualities of science	e-mail:	t.vanviegen@awti.nl
feature:	038/22/sk	Your feature:	--

Dear Mrs. Bergkamp,

Update

The European Research Council (ERC) recently indicated its intention to sign the Agreement on Reforming Research Assessment (RRA). [1] On p.10 of the letter 'Identification of the qualities of science' it was stated earlier: “The ERC will not sign the RRA, because the ERC finds it unclear what the RRA adds to the peer review method that the ERC follows. [2]“ This new development at the ERC underlines once again that the evaluation of scientists is changing internationally. The ERC is an important financier for Dutch scientists and with the announced signing, the ERC will also implement changes that are in line with the Dutch program 'Recognition and Rewards'.

Yours faithfully,

dr. E.E.W. Bruins
chairman

P.W.J. Essers
secretary

cc: Chamber committee Education, Culture and Science

[1] See https://erc.europa.eu/news-events/news/erc-scientific-council-decides-changes-evaluation-forms-and-processes-2024-calls.
[2] See https://erc.europa.eu/news-events/news/research-excellence-quo-vadis.

Visiting address:
Advisory council for science, technology and innovation
Prins Willem-Alexanderhof 20
2595 BE Den Haag

t. +31 (0) 70 311 0920
e. secretariaat@awti.nl
w. www.awti.nl