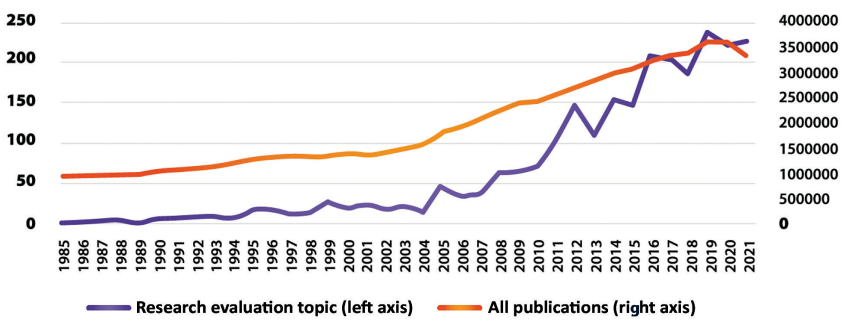
I. Research Evaluation: What, Why, Who?

1. Research evaluation in perspective

Research evaluation is still not a research field per se but rather an incorporated element in the creation of a new, contemporary, and competitive public value, which is often part of other evaluative procedures. Research evaluation seems to be rather a governance than a reflection instrument. However, research evaluation attracted a diversified interest of researchers around the globe in recent years. The total number of papers included in Web of Science Core Collection (WoSCC) mentioning *research evaluation* as a topic grew substantially faster than the database itself (118 % growth for research evaluation papers compared to 30 % growth for all papers in WoSCC for the period 2011–2021).

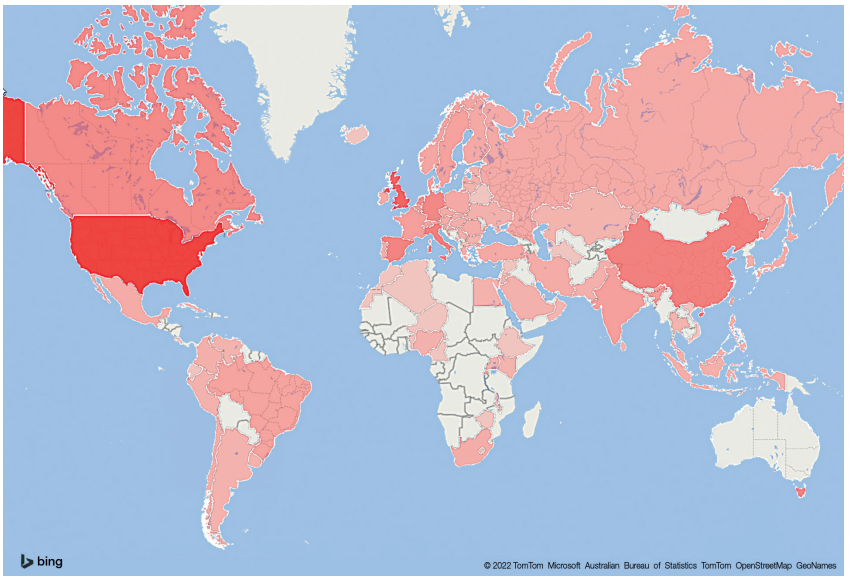


Source: Web of Science Core Collection (accessed February 12, 2022).

Figure 1.1: Growth of interest in publications on "research evaluation" (1985–2021)

The geography of authors publishing on research evaluation expanded with 41 new countries in the last 10 years. Among them are Poland and Bulgaria. In both countries the interest in research evaluation correlated with important policy debates, which lead to the new governance of research and innovation systems in the countries.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb



Source: Web of Science Core Collection (accessed February 12, 2022).

Figure 1.2: Frequency of publications with topic "research evaluation" by country (all time)

The Constitution for Science law in Poland adopted in 2018 significantly changed the evaluation criteria for research performance. As Michal Grabowski, head of department of invertebrate zoology and hydrobiology of the University of Lodz, put it in a recent interview, the law changed the value of an academic unit from "as good as its best scientist" to "as good as its worst scientist" (Zubascu, 2018). Papers by Polish authors constitute 5 % of all papers on the topic published in 2021 in Web of Science. The research policy shifts have different roots – from the pragmatic need to develop a research system which could deliver technology transfer services to local and international companies to the wish to have an instrument for a generational change especially in social sciences.

Fahrenkrog et al. (2002) defines evaluation as a "**systematic and objective process designed to assess the relevance and effectiveness of policies, programmes and projects**". Research evaluation emerged as a policy response to the need to justify budget spending for fundamental and applied research at universities and institutes. Initially resembling more a cost-benefit analysis, later it also became a policy design support tool,

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

systematically gathering information on research performance on national, regional, institutional, or individual level.

2. History

Elements of research evaluation could be seen already at the beginning of the 20th century. As early as 1917, Frances J. Cole and Nelly B. Eels applied a quantitative analysis to comparative anatomy literature from 1543 to 1860. Their work was both descriptive and evaluative in nature. They used a curve to present the rate of document growth over the span of three centuries. They also indicated which aspects of the subject matter attracted researchers' interest in a given period of time (De Bellis, 2009).

The Vannevar Bush's 1945 report to the President of the United States contained a statement that quality control must be left to the internal mechanisms of the research elite using the peer review system. This model was then applied by the US National Science Foundation in 1947, and it was followed by other Western countries.

The first theme registers for scientific publications were created in the 50s and 60s. Soon after, the citation index developed by Eugene Garfield's Institute for Scientific Information was recognised as a way of objectifying research standards. Scientometrics has proven that it is possible to measure specific parameters regardless of some imperfections.

Marjanovic et al. (2009) indicates that one of the earliest studies in the field of evaluation, "The Sources of Invention" (Jewkes et al., 1958), has assessed 61 innovations in different scientific disciplines. This initiative was adopted by the US Department of Defence in 1967. It aimed to provide a justification for the size of the investments made in defence research (Sherwin & Isenson, 1967). Other studies examined back then the return on investment in research. Griliches (1958), for example, has evaluated the social norm for return on investment in hybrid corn-related studies.

Gibbons and Johnston (1974) have studied the role of scientific research in technological innovations and its contribution to industrial research and development. The authors have assessed 30 industrial innovations in Great Britain which include significant technological changes.

In 1968, The National Science Foundation conducted the TRACES study (Illinois Institute of Technology Research) and subsequently expanded this study via the Battelle project (Battelle Laboratories, 1973). It studied how 'non-mission oriented' research had contributed to the practical innova-

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

tions. Battelle is best known for its nuclear research and involvement in the Manhattan project, but throughout the years it established itself as a premier centre for sustainable energy research and innovations.

The top two prolific contemporary authors on research evaluation are Giovanni Abramo and Ciriaco Andrea D'Angelo from the Laboratory for Studies of Research and Technology Transfer (LSRTT), Institute for System Analysis and Computer Science (IASI-CNR), National Research Council of Italy. They have, respectively, 61 and 58 papers (all of which co-authored with Abramo) on the topic in Web of Science Core Collection. The LSRTT school applies economics logic to production of research output, looks for alternatives of established indicators for scientific productivity, and studies the impact of national policies on publication behaviour, i.e. on self-citation behaviour (Abramo et al., 2021).

In research evaluation we find the so-called Hawthorne-like effect (Landsberger, 1958). It directly influences the behaviour of researchers. For instance, if you know you will be evaluated only on the basis of papers published in top journals, you might not publish elsewhere at all (and this could lead to higher turnover of non-tenured professors), and if you receive a bonus based on the number of papers, you may be prolific. If a lesser quality product requiring less efforts (i.e. a monograph of 101 pages published by whatever publishing house) would give you scores more than three times higher than a peer-review article in a Q1 journal, why you should rationally choose to work harder and taking the risk of several rejections? This is exactly the case of the current rules for academic promotion in Bulgaria.

Despite the differences in research evaluation metrics across Europe, there is a certain level of homogeneity of research evaluation systems. Research assessment systems are usually path-dependent – affected by historical, institutional, and political factors. Some countries with intensive research and scientific excellence (such as Netherlands, Austria, Switzerland, or Germany) apply less bibliometrics and more adaptive approaches, while others try to improve their ranking position by applying metrics and showing priority towards publications in English (Ochsner et al., 2018). Typically, research evaluation in post-communist countries is predominantly focused on the quantity rather than quality of publications (Jurajda et al., 2017), which is the case with Poland and Bulgaria as well.

The number of studies on the effects of specific research evaluation measures worldwide has increased significantly in the 1990s (Thomas et al., 2020). The relationships between science, technology, and markets are cru-

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

cial for the market economy. Investments in scientific research are usually associated with high uncertainty – whether the research will lead to an invention of new technology, if the technology could be commercialised, and when these events will occur in time. In order to manage these risks, contemporary mechanisms for evaluating the effectiveness of scientific research have been developed.

Two important trends were observed in the last decade of the 20th century. The first was a shift in the understanding and assessment of research funding: grants perceived as part of the state's responsibility became investments with an expectation of economic efficiency. The second was a rise in the neoliberal approach to the value of knowledge, resulting in pressures to optimise and increase the efficiency[2] and intensity of scientific research. If knowledge is a commodity, then it should be produced by systems, subject to standardisation similar to other commodities (food, pharmaceutical products, cars, etc). Consequently, the mechanisms for evaluating scientific research were calibrated and became a key source of information in the decision-making process with regards to supporting research in the public sphere; public financing is provided on a competitive basis with results being measured on the basis of the generated added economic or social value (Leydesdorff, 2005). In the 1990s, the topic about measuring the impact of a given study attracted the attention of researchers (for example, Mansfield, 1991; Herbertz and Muller-Hill, 1995; Buxton and Hanney, 1996; Martin and Salter, 1996; Dawson et al., 1998; Hanney et al., 2003a, 2003b; Wooding et al., 2004b) and institutions which financed research.

Some changes were applied in the work process with regards to the approach to research evaluation and proof, leading to the creation of the San Francisco Declaration on Research Assessment (DORA), signed in 2013 and supported by 2,552 organisations and 19,126 people (last update: beginning of May 2022). The document offers recommendations targeted at financing organisations, research institutions, publishers, structures, which provide statistics, and researchers. The purpose of these recommendations is to correct the unforeseen effects triggered by the evaluation mechanisms established in the 1990s. In order to make the quality of research output measurable, these mechanisms have been adjusted to transform it into

2 The efficiency shows to what extent the goals set in a given programme have been achieved or whether they are on track of being reached (EC, 2017a).The analysis of effectiveness studies the relationship between the resources invested in a given intervention and the changes achieved (EC, 2017a).

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

quantitative indicators. Since 2013 this approach led to an intense growth in terms of quantity, which critics believe significantly surpasses the increase in the quality of the output.

On the other hand, since the quality of publications should be guaranteed via the peer review mechanisms created by scientific journals, the papers in such journals are privileged as opposed to monographs or publications in local or specialised issues. Apart from that, the quality of research publications is evaluated based on their impact which is calculated using different citation indices the coverage of which is limited due to the contents in databases which serve as the basis for calculations. The indicators also cover a limited amount of time because they set as a premise a relatively high speed of circulation of knowledge which does not account for the different rhythm of development of scientific disciplines.

In order to optimise the evaluation mechanisms for scientific research, the DORA authors recommend to mainly limit the use of metrics related to science journals, to encourage the evaluation of the research itself, not the editions in which it is published, and to use the advantages of online publishing (such as lack of limitations with regards to word count, figures and bibliography) and new indicators for measuring the significance and impact of research. It is noted that research outcomes can vary – new knowledge, data and software, or intellectual property, including well-trained young researchers. The authors believe that the impact of a given study on a specific policy or practice also represents a scientific impact indicator. Mentorship and societal engagement of researchers are other achievements which matter (Curry, 2018), and some of them are being evaluated. For instance, the Plymouth University in United Kingdom introduced Community Research Awards[3] in 2009, being a clear proof of social engagement of researchers.

The DORA group leader is convinced that, despite the restrictive conditions, such an experiment should be conducted in order to introduce real changes in research evaluation and move towards open science, replication of results, and free sharing of knowledge. Open science (scientific research and results to be made available to all inquisitive people, amateurs and professionals) is one of the priorities of the European Commission, and it has been formally introduced in numerous Community documents (European Commission, Directorate-General for Research and Innovation, 2020, 2018). Open science contributed to changes in the business models of

3 Get Involved Awards 2022 – University of Plymouth (last visited April 29, 2022).

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

publishing and accessing the academic research. It democratised the access to the research along with the flourishing of social repositories of academic research such as researchgate.net and academia.edu. Available funding from the European Union programmes increased availability of research to wider audiences (including the start-up community). At the same time predatory publishers applied a complex of marketing techniques to increase citations of the papers published in their open access outlets and to get around the quality controls of the major scientific databases. Various authors doubt the promise of the open science (Böschen et al, 2020) and especially its expected impact on policy making.

Over the last few years, common principles have been commented on and imposed with the aim of resolving part of the existing issues. These principles are not binding, but they correspond to the desire to implement a more independent, transparent, and clear evaluation of research outcome. This trend is observed in the more frequent citation of documents, such as the Leiden Manifesto (2015). It consists of ten principles which are grouped around pre-defined evaluation indicators, taking into consideration the specificities in different fields, using quantitative indicators in support of quality assessment, and timely updating indicators which no longer adequately correspond to the needs for research evaluation.

The efforts towards more effective research assessment are complemented by the Hong Kong principles, which were formulated and endorsed at the 6th World Conference on Research Integrity in 2019 in Hong Kong. They are formulated as follows (World Conferences on Research Integrity, 2019):

- assess responsible research practices (these would include ethical behaviour);
- value added reporting;
- reward the practice of open science (this is particularly important for the access to knowledge by the global south researchers and shortening the time from publication to implementation in business or society);
- acknowledge the broad range of research activities;
- recognise essential other tasks like peer review and mentoring.

As an important step in the same direction, in November 2021 the European Commission announced intentions to outline a framework for research assessment to be applied by all member countries. The focus is put on rewarding ethics, integrity, teamwork, and diversity of outputs in addition to quality and impact (Nature, 2022).

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Organisations are also established with the goal to support responsible evaluation. Such example is INORMS, Research Evaluation Working Group founded in 2001 (https://inorms.net/research-evaluation-group/). Its goal is to encourage interactions between members and sharing of good practices.

3. *Research evaluation: tasks*

Research evaluation is not, and should not be, an end in itself – it is rather a component of the decision-making process governing research at different levels. There are four broad reasons to conduct a research evaluation:

- to increase the accountability of researchers, policymakers, and funding bodies in the eyes of society by making the research team present its outputs and impacts;
- to steer the research process towards studying the outcomes;
- to provide means for advocacy to conduct research or to fund it, based on past outcomes;
- to achieve a better understanding and apply the 'lessons learnt' practice from previous attempts for a research process (Georghiou & Larédo, 2005).

In addition, specific steps are needed to close the gap between the lack of the necessary substantive knowledge and the decision-making process. There are five 'insufficient' or incomplete knowledge categories (gaps) which have a direct impact on the decision-making process:

- difficulties in identifying and interpreting facts (a facts gap);
- difficulties in understanding and grasping certain processes which have a cause-effect relation or create conditions for a series of consequences (natural mechanisms gap);
- difficulties in identifying possible indirect effects (systematic gap);
- difficulties and insecurity with regards to introducing a research product in real manufacturing (technological gap);
- lack of interest which could influence subsequent positive actions (strategic elements gap).

Taking this incomplete knowledge into consideration is a key challenge facing research evaluation. The research evaluation process is characterised by guaranteed quality control on execution; it is conducted by independ-

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

ent experts, research committees, or panels, and it offers a compulsory conclusion and recommendations to the relevant political or financial decision-makers – ministry, heads of executive agencies, donors, etc. This can be treated as a process-oriented evaluation practice.

However, in the last decade evaluation practices (especially those related to large European projects and programmes) are much more oriented towards results – private and public. The evaluation can also be defined as a way of identifying the effect (impact) of particular public activities: scientific and technological, public, and, in some cases, political. This activity can be transposed to a past or future period, and it can be direct and indirect.

The conditions for conducting a research evaluation require two types of process relationships: entry/exit (resources and products) and cause/effect (factors and results).

Research evaluation could be conducted on different levels:

- evaluation of proposals for research projects in terms of quality;
- evaluation of completed national and/or international projects and programme activities;
- evaluation of the research conducted by a given academic unit;
- evaluation of the overall national research system.

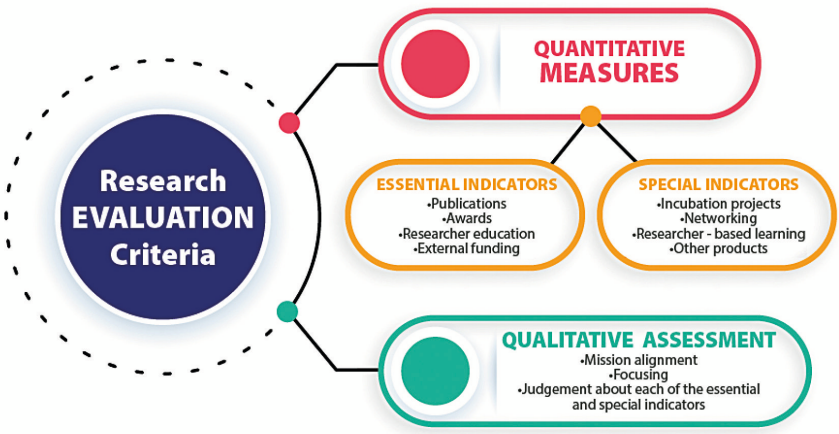
All of these conditions presume a focus on the quality of the research activity. The following elements can be subject to evaluation:

- an economic and/or social effect which comes as a result from the implementation of one or more research programmes, targeted or joint;
- outcomes and effects generated as a result of the implementation of a given research project/programme;
- research methods and their implementation;
- the research and/or technological level (degree of originality) of a given development.

An independent review of the role of metrics in research evaluation and management in the United Kingdom offers a framework for responsible metrics and a set of recommendations (Wilsdon et al., 2015). The study looks at research evaluation indicators in different disciplines and tries to evaluate the negative or unintended effects of metrics on various aspects of research culture.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Evaluation indicators can be quantitative and qualitative and can be applied jointly or independently. The figure shows the specificities of the applicable quantitative and qualitative criteria.



Source: Adapted from Cantu, Bustani, Molina & Moreira (2009).

Figure 1.3: Research Evaluation Criteria

Research evaluation, when properly conducted, improves governance, increases transparency of public funding of higher education and science, contributes to evidence-based decision making with up-to-date information about the quality and impact of the scientific research. Regular and independent research evaluation strengthens the capacity of institutions to conduct strategic planning of their research and in turn increase their competitiveness.

The recommendations of the independent review and analysis of the role of metrics in research evaluation and management in the United Kingdom (Wilsdon et al., 2015) focus on its significance for supporting the effective management of research and on the availability of useful and reliable sources of information to be used in the evaluation process for the purpose of ensuring transparency, avoiding mistakes, carefully using quantitative indicators on the premise that all measures are applied in a coordinated manner. The summarised recommendations from this report (applicable to all research evaluation systems) include:

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

- Choice and application of indicators — they have to be comprehensible for the research community.
- Use of online platforms and tools which can improve access and visibility of scientific research, as well as opportunities for data sharing.
- Expert evaluation of particular activities which, despite some shortcomings and limitations, continues to receive broad support in different disciplines.
- Application of qualitative and quantitative indicators which correspond to individual disciplines and context. (The 'one size fits all' approach is less likely to work in all cases. The unsuitable indicators distort the motivation of a given researcher and the vision for his/her research.)
- Open, transparent and clear structure for research data. There is a growing tendency of journals to request full data-sharing of published papers.
- Complying with common data description, data collection and data disclosure standards.

The evaluation has to be conducted in a way that 'excess' effects are eliminated. For example, when using a lot of frequent evaluation procedures, new initiatives might be destroyed, and a lot of resources could be taken away from the creative process.

Based on studies of numerous research evaluation initiatives and analyses, research evaluation could be understood **as a complex social practice managed by funding agencies or ministries** (Elzinga, 1995). That practice could be subdivided into several elements:

- Social process, ensuring legitimation of policy-making or administrative decision-making, as well as raising transparency of state funds spending on research.
- Setting up expectations, providing a basis for an adequate and effective use of the funds granted (the socio-economic effect is usually applied as a metric – the result of the development).
- Precise steering and correction of workplans and programmes on institutional level as a result of feedback from evaluation results, similar to the peer-review on a paper-level.
- Achieving a rich information infrastructure which could be useful in the decision-making process.

Different 'schools' studying the research evaluation have different specificities, but at the end of the day they show some common characteristics:

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

- Increasing rationality in evaluation where the focus moves from 'formalising a past activity' to 'improving the understanding about the course of action in the future' and setting the focus on 'strategic research' as opposed to 'curiosity research'. This is especially valid in countries where national foresight projects exist. In case the private sector also applies the scenario planning techniques, the research priorities derived from the back-casting contribute to the strategic research agenda.
- Expanding the coverage of the evaluation from 'a problem-specific one' to a 'broader one' or 'systemic one', involving analysis of stakeholders' relationships.
- Expanding the number of stakeholders, who, along with an 'objective external evaluation', offer their own analysis or evaluation.

We could conceptualise all these characteristics of the systemic approach within the context of complex social practice (Elzinga, 1995) as a holistic approach. The holistic approach would see research evaluation in all layers or contexts of social practice – managerial, economic, social, and environmental sustainability. Organisational (including managerial) readiness to foresee trends in science or social, economic, and environmental shocks and its capacity to respond to those shocks should be included as an issue of assessment. The holistic approach requires widening the range of the consulted stakeholders in all phases – from policy formulation to evaluation of results. They should include the users of research results but also the society as a whole – civil society organisations and local unorganised communities. Science communication is also an integral part of the mandate of research organisations. Although researchers typically are not ready to engage in social marketing of their research, it is part of their social responsibility to outreach to the society – be it children at school, who might become the next researchers, or civil society organisations, policy makers, and businesses.

4. Types of research evaluation

The types of research evaluation are, to a large extent, linked to the expectations about its nature, the way it is performed, the tasks carried out, and the manner of addressing the relevant group of stakeholders.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Some authors categorise research evaluation based on the applicable approaches (Worthen et al., 1997):

- Objectives-oriented – focuses on determining the objectives and the degree to which they have been achieved.
- Management-oriented – focuses on determining the information which managers/decision-makers need.
- Consumer-oriented – focuses on providing information to the consumer and on the evaluation of different competing 'products' and services. Consumers here refer to public institutions interested in the new policy design, businesses which need relevant research-intensive product/service, and socially responsible investors, which are accountable to society.
- Expert-oriented – built on the basis of particular experts who determine the quality of the topic which is being evaluated.
- Consultation-oriented and process-oriented – brings together different points of view of the evaluators and compares the pros and cons.
- Participation-oriented – takes into account the variety of viewpoints presented, the values, the criteria, and the needs defined by the stakeholders.

Rossi and Freeman (1999) present the classification of the stakeholder groups which are directly involved or interested in the evaluation process of a given research programme:

- Politicians and decision makers – responsible for deciding the future actions in relation to the programme which is being evaluated.
- Programme sponsors – responsible for financing the evaluation.
- Target participants – entities or units that are at the receiving end of the service being evaluated.
- Programme management – a group which is responsible for the programme.
- Programme staff – a group which delivers the programme.
- Evaluators – a group which conducts the evaluation.
- Programme competitors – groups which compete with the programme.
- Contextual stakeholders – groups in the encirclement of the programme.
- Evaluating community – independent (or second round) evaluators who determine the quality of the evaluation.

Even though individual groups of stakeholders have different influence and perceive the evaluation results in a different way, the stakeholders' expectations become part of the process.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

A wide segment of research assessments is focused on outcomes or results, which are of interest to different target groups. Due to quantitative nature of outcome-oriented evaluations, it creates an illusion for higher objectivity and through quantitative indicators policymakers believe it is easier to prove a given statement. Because of criticism towards the more quantitative approach the system adopts more and more quantifiable proxies. It is quite rarely to find qualitative indicators with a central place in the evaluation system.

Therefore, the outcome-based criteria system is increasingly detailed, and the process evaluation system is not monitored, analysed, and developed. Moreover, it becomes more and more bureaucratised and acquires purely administrative functions. Embedding innovative experimental forms of assessment based on quality will certainly change the general picture of the assessment process and the system will become more flexible, losing some of its rigidity.

According to the European Commission, Directorate-General for Research and Innovation (1997), the evaluation has to be:

- analytical – to apply accepted methodologies for qualitative and quantitative analyses but also participative practices ensuring different viewpoints;
- systematic – to follow a carefully prepared strategy with detailed planning and consistently implemented at different level of system;
- reliable: To evaluate the same data, the application of the same mechanisms must lead to the same results regardless of who the evaluator is. Although subjectivity is inevitable, it should be processed in such way that the result of the evaluation is independent of the participating evaluators, provided they meet certain academic performance criteria.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Source: Adapted from Hong & Boden (2003).

Figure 1.4: Architecture of the programme evaluation system

Although the EC defines three main characteristics of research assessment, and all of them include a process evaluation element, very few assessment systems contain objective criteria for processes evaluation.

The main time-related types of evaluations are *ex ante* – an initial evaluation prior to the implementation of the programme or before the decision for project funding. This is the most frequently used evaluation type. Other types are:

- Mid-term – mid-term evaluation: This could involve ongoing monitoring and feedback which is provided based on the results obtained during the implementation of the project/programme activities. It is an insepar-

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

able part of the evaluation procedure because it provides effective control during the whole project/programme cycle in view of minimising the risk in their implementation.
- *Ex-post* evaluation – being applicated after the completion of a given programme/project (*ex-post* evaluations are rarely and not systematically used. Sometimes they are overdue and implemented only formally). Sometimes *ex-post* evaluations include impact assessments.

In relative terms, the *ex-ante* evaluation is the most clearly defined procedure in the evaluation process. It is widely used in institutions which finance research, as well as in the evaluation of the preliminary plans of research organisations. The criteria used in this type of evaluation are, to a large extent, harmonised, especially on their base level. In specific cases, a second, smaller set of criteria, which reflect the specificity of the particular activity, is included. In general cases the type or impact of the criteria changes, and this is in line with the general balance between applicable criteria.

The mid-term evaluation aims to evaluate the progress in terms of achieving the set goals. It provides an opportunity to make timely changes in order to guarantee that these goals are achieved within the time planned. This type of evaluation provides an opportunity to determine whether:

- the intervention is still aligned with the strategic goals set;
- it is suitable and useful for the key stakeholders;
- it is conducted in an efficient way.

The ex-post evaluation is more difficult to perform, and in a lot of the cases it is conducted as a matter of formality due to the fact that:

- scientific research is not a routine activity with a final limited outcome, thus the quality of the results can be evaluated only at a particular stage;
- scientific research is part of the national innovation ecosystem, and the latter can be evaluated in different aspects;
- only some of the criteria are used in the evaluation – the ones which are applied in the evaluation of the proposals for a given development.

It reports the following types of accompanying activities:

- audit-type evaluation;
- evaluation of policies related to strategic research;

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

- evaluation of the efficiency of functioning of the entire research system (university, research institute);
- impact.

The ex-post and mid-term evaluations are particularly needed in terms of future development of programmes and activities. In recent years (between 2008 and 2018, depending on the programming period) the impact evaluation of part of the activities under the operational programmes of the Structural Funds has been subject to analysis for the majority of new member states. Impact evaluation, however, is based on formalised and, in some cases, unsuitable criteria, and it can be difficult for it to act as a homologue of the *ex-post* evaluation.

The evaluation system in European programmes and projects, to a large extent, applies harmonised basic criteria, and it has created a common framework for their application. Due to the fact that this type of evaluation assesses only intentions and possible outcomes, the degree of uncertainty is very high, and this type of evaluation is more easily digested both by the evaluator and the person being evaluated.

Some evaluation activities of framework programmes of the Community, the programmes of European structures (for example JRC, ECA), can be referred to the last type of evaluation. This evaluation applies the assessment of some typical performance indicators and takes into consideration as well proven effectiveness for the economy or/and society as a result of programme realisation. It is used as an input to outline the framework of programme development in the future.

The types of evaluations, organisation method and relevant activities are systematised in Table 1.1.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Table 1.1: Types of evaluation

Types of evaluation activities	Subjects of evaluation			Expected impact
	Research projects with a clearly defined outcome	**Grants for strategic research programmes**	**Projects and technological and innovation activity**	
Activities for conducting the evaluation process according to subject groups				
Ex ante	A system of evaluators who evaluate the proposals offered, based on a set of criteria	A system of evaluators and consumers evaluating the research quality and the expected economic and public benefits, based on a set of criteria. The expectations of state structures and consumers also impact the evaluation.	A system of experts, including consumers and sponsors/donors who are organised in ad-hoc groups. Addressing the expectations of a given consumer group is of key importance	Ensuring legitimisation and transparency of the public finances provided. A basis for effective use of the funds provided.
Mid term	A system of evaluators who evaluate the progress of the mid-term results, based on a set of criteria.	A system of evaluators, consumers and public structures evaluating the progress in the implementation of a given programme, the quality, and the expected economic and public benefits.	A system of experts, including consumers and sponsors/donors who evaluate the progress and the impact which is the result of the implementation of a given stage of the technological and innovation activity.	Careful monitoring and correction of work plans and programmes as a result of feedback from evaluation results.
Ex post	A system of evaluators who provide a conclusion by comparing the expected results which were originally evaluated and the results achieved in reality	A panel of experts evaluating the degree of achievement with regards to the strategic goals.	A panel of experts and professional evaluators who evaluate the results achieved, based on a comparison with the set expectations from the implementation of a given project.	Achieving a rich information infrastructure.

Source: Authors' own elaboration.

Depending on the subject of evaluation, the evaluation objectives can be very different. We can differentiate the following types of evaluation activities:

- Evaluation of a given research structure – a research institute or a small company with a research profile. This type of activity is performed vis-

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

ibly or away from the public eye; it develops individually and has clearly defined cause-effect relations. The reason for its implementation is because the activity of such units does not just need 'public recognition' but also accountability before the society, in addition to arguments for future financing.

- An evaluation of the general research and technological development is more frequently implemented with the help of international organisations and experts and clearly impacts future policy-related decisions and the introduction of policies, reforms, new structures and programmes promoted on a national level.
- A programme-based evaluation – the most recent evaluation introduced – is aimed at institutionalisation, defining the degree of applicability and usability of the outcome achieved and the effectiveness of its implementation.

The effects of research and technological activity, which can be evaluated, are systematised in Table 1.2.

Table 1.2: Effects from research and technological activity which are subject to evaluation

Main groups of activities financed by the public budget	Direct effect		Indirect effect	
	Short-term	Long-term	Short-term	Long-term
Science	Specific scientific knowledge	Cognitive knowledge	Improved training / research training	Economic/ public benefits
Economy and society	Improved technologies and social cohesion	Better know-how Balanced social relationships	Improved and environmentally friendly productivity Social consensus	Better competitiveness Prosperous society
Policies	Better understanding and implementation of new policies	Resolving social/economic issues	Grasping the genesis of existing issues	Contributing toward reaching a consensus when resolving existing issues

Source: Authors' own elaboration.

Evaluation methods can be examined using a matrix structure in which the elements of the matrix are the data, the type of evaluation, and the

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

analysis applied. The data can be quantitative (statistical and bibliometric) and qualitative – gathered through interviews, statements by focus groups, surveys, etc. The analysis can be different – conducted through peer review panels, based on different studies, technological evaluation, etc.

The indicators used are varied, but a small part of them are basic and are used in almost all types of research evaluations. These are illustrated in Table 1.3.

Table1.3: Types of indicators

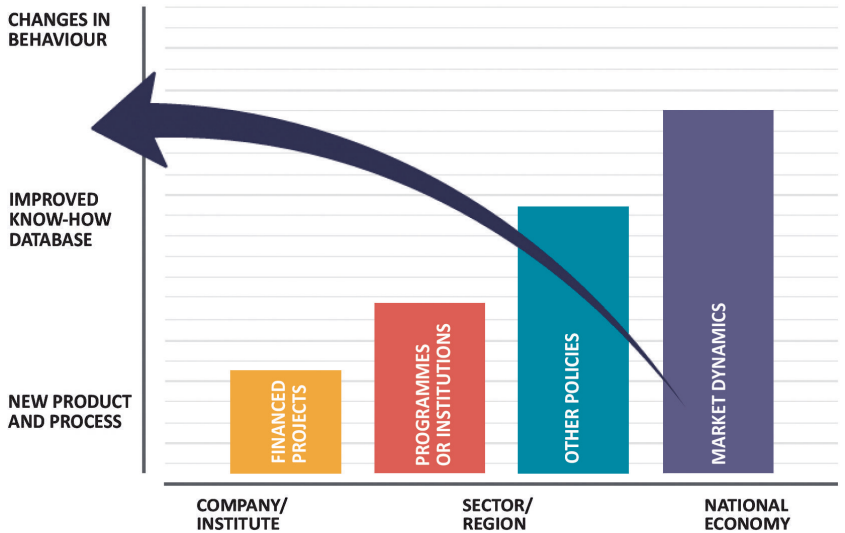
Indicators	Quantity	Quality
Entry	Research staff State of the art of the subject of evaluation	Strategically planned research activities, impact of knowledge absorption
During the activity	Citability of research publications Patents, utility models *Note: Measurable in terms of quantity (number)*	Where the publication is published (presence of an impact factor) What is the type of the patent
Exit	Increased productivity, new workplaces, additional profit, value added Social conflict/issue resolved	Completely new and environmentally friendly products, services, processes Opening new markets and/or market niches, expanding an existing market Entering a global market

Source: Authors' own elaboration.

The first two types of indicators are linked to the research process, while the last type, the exit indicators, are oriented towards finding specific and socio-economic results.

It should be taken into account the fact that, in some cases, evaluation may lead to limitations in the research and technological activity, but at the same time it is still aimed at comprehensive coverage. The comprehensive coverage and the effects of the evaluation activity are illustrated on the diagram of Figure 1.5.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb



Source: Adapted from Accreditation Agency, the Netherlands.

Figure 1.5: Diagram of the effect of technological activity

Figure 1.6 illustrates the five-steps model of the institutional research evaluation framework.

Source: Adapted from Hassanain, Anil & Abdo (2016).

Figure 1.6: Framework concept for institutional research evaluation

If the research in a given region is being evaluated, the most frequently used applicable indicators are as follows:

- gross government expenditures for research as a share of the gross domestic product;

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

- research expenditures provided by the business relative to public expenses;
- workforce employed in the research sector (this includes research organisations and research institutes);
- research expenditures per capita from the working population or per researcher.

The following entry indicators are used for institutional or individual research activity:

- total number of publications in peer-reviewed and indexed journals for which the previous evaluation period is completed;
- total number of citations;
- patents, utility models or technology transferred products in the economy;
- number of PhD candidates who have successfully defended their thesis, for a given period, which is most often determined by using the evaluation methodology.

The following exit indicators for evaluating institutional or personal research are also observed:

- number of publications for the evaluation period in peer-reviewed journals (3–5 years);
- exchange of researchers between research organisations/universities or between research organisations and companies;
- participation/funds attracted under different European programmes;
- participation/funds attracted under national programmes and industry contracts;
- memberships in international research organisations;
- memberships in editorial teams of scientific journals;
- research awards (international and national), etc.

5. Evaluation procedures based on scientometrics

Some authors use mathematical tools for the purpose of creating a model for the growth trend in publications and suggest the term 'scientometrics' for this kind of research (De Bellis, 2009). Scientometrics can be defined as a "quantitative study of science, communication in science and science policy" (Hess, 1997). It has been developing over time. It studies indices

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

for improving the extraction of information from peer-reviewed research publications (usually described as a 'bibliometric' analysis of science) and has gradually expanded to other types of documents and sources of information related to science and technologies.

The scientometric indicators complement and contribute for the standardisation, collection, and analysis of a wide range of activities in the field of science, technologies, and innovation by providing evidence for a selected set of results in science and technologies. There are certain advantages to quantitative approaches in evaluation, using proper statistical data and suitable indicators.

Weinberg (1989) claims that the Board on Physics of the US National Academy of Sciences has applied evaluations to three types of criteria: internal, external, and structural. These have been divided because the inherent indicators include the significance of the research topic, the potential for discovering fundamental laws, the potential for discovering summaries which could be broadly applied, and the attractiveness and maturity of the research. External measures review the potential contribution and stimulation for other sciences, in particular engineering, medicine, and applied sciences, and the contribution towards the national prestige, defence, public education, and international cooperation. The last set of criteria evaluates the need for progress in the discipline and the need to maintain the development in a specific field.

6. *Bibliometric indicators*

Bibliometric indicators provide answers to the following questions:

- How productive is a given research team or an individual researcher (or what is the essence of their/his/her research), what is their level of research competence on an international level?
- Where is the team or the individual researcher positioned in comparison to similar teams in the country and abroad (benchmarking)?

The main limitations to bibliometrics are results which are predominantly applicable to research groups, units and institutions, and which are difficult to apply to innovation activities which have a bigger coverage and do not focus specifically on pure research results. Nevertheless, the increasing capacity of scientific results imputedly intervenes the competitive innovation

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

process via growing intellectual capacity, proved by improved bibliometric indicators (Vutsova, 2009).

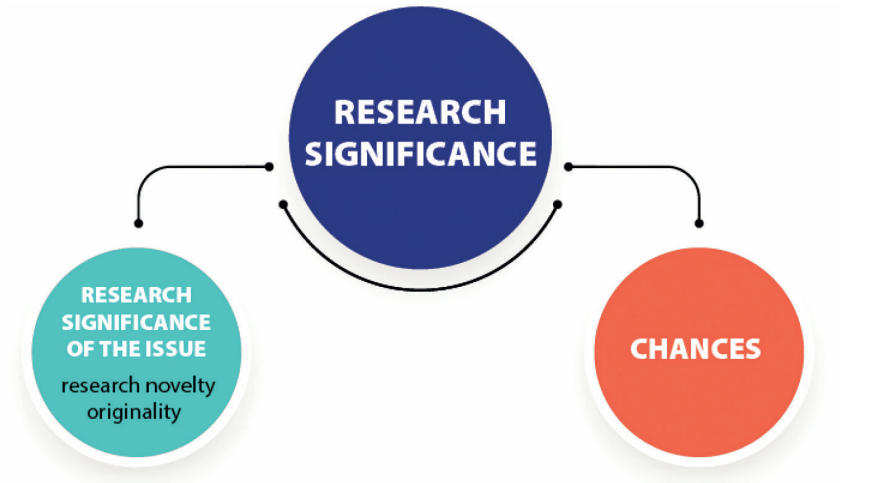
According to Leiden's Manifesto (Hicks et al., 2015) with regards to monitoring the citability of publications, there are often comments about the increasing reliance on concrete data from large databases, namely Web of Science, which has an impact factor that dominates as an indicator in a lot of research evaluation models. Meanwhile, it is not entirely universal. For example, a group of European historians receive a relatively low evaluation because they publish books, not articles in peer-reviewed journals (Hicks & Wouters, 2015). This issue could be avoided by carefully clarifying the specifics in different fields and adding normalising correctives.

Another element which may not be taken into consideration in the impact-factor-based research evaluation is the regional importance of the research despite the fact that Web of Science is undertaking corrective actions in this respect (it includes more journals of regional importance, which are visible in the main data collection, or creates specialised national databases similar to the ones in Norway and Croatia). On the other hand, this is one of the platforms where the biggest amount of information is exchanged in the field of science, and hence access to results is guaranteed.

Problematic fields are mainly those of social sciences and humanities, especially when national/regional issues are examined which would not be of such big interest to international journals. Even in cases a significant global impact has been achieved, as in the case of organisation of the International Philosophy Olympiad (Kolev, 2017), most of the researchers behind it never publish in WoS/Scopus. A balance must be achieved between regional journals included in large databases and monitoring the quality of those which are not. Very often, the evaluation is based on the percentile in which the publication is included.

In the evaluation of research projects, for example, a taxonomic tree of the criteria can be used in view of determining their significance.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

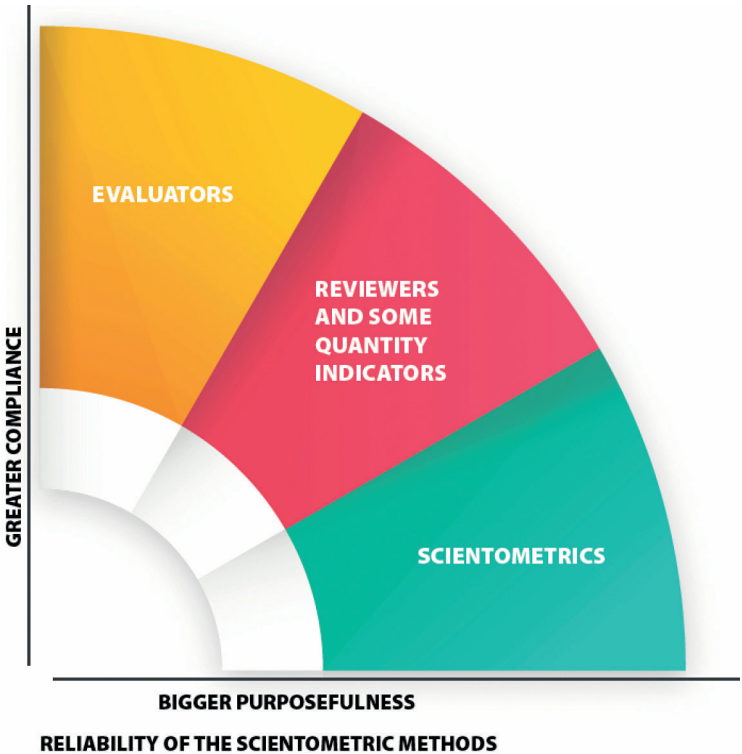


Source: Adapted from German Recotrs' Conference.

Figure 1.7: Taxonomy of scientific research

Scientometrics mainly relies on clearly structured situations which determine the reliability and applicability of its methods. It uses statistical data processing which in some cases creates complications in the analysis. Inaccurate or incorrect conclusions may sometimes be drawn, which significantly differ from the evaluators' opinion.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb



Source: Adapted from Accreditation Agency, the Netherlands.

Figure 1.8: Evaluation methods

Statistical problems which might occur are related to ranking, might be the result of the discrete nature of citation distributions, especially with small samples, and applied a fractional solution (Waltman et al., 2012). In response to such potential threats Bornmann and Williams (2020) suggest guidelines and procedures for the normalisation of percentile ranks based on cumulative frequencies in percentages. University ranking systems are also targeted because of the methodology of assessing research outputs (Fauzi et al., 2018; Bowden, 2000). To tackle these challenges, some authors (Szomszor et al., 2021) believe that new indicators are not necessary, but efforts should be directed towards choosing (a combination of) the proper ones in order to present academic research more adequately. For better decision making a focus on the management and interpretation of results should be put.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

In the evaluation of industrial and applied research the main goal is to determine the contribution towards achieving the company's strategy and objectives. In order to support a given industrial research, the following entry indicators for evaluation should be adhered to:

- the expected effects of the research;
- integrating them with the business strategy of the company;
- the available portfolio of intellectual products;
- monitoring process;
- possible losses due to poor execution compared to the achievements expected.

Apart from that, the results and/or the environment in which industrial research is going to be developed and/or applied are also evaluated through:

- percentage of the turnover generated from innovative products[4];
- percentage of the real innovative products on the market – maintaining a sustainable market;
- the degree of technological innovation of the company;
- granted patents as well as a ratio of the granted patent/applications;
- revenue from the provision of IPR products;
- variety of the technological structure – spin-off, spill over, incubation units, etc.;
- staff involved in R&D activities.

Measuring the effect of industrial research is usually difficult. The Office of Technology Assessment (1986) in the US presents three main reasons in support of their argument as to why measuring the effect of RTD is difficult. Firstly, there are non-economic goals, especially in relation to public and socially desired high-risk investments, such as defence, for example. There is uncertainty in their measurement, for example, progress in healthcare.

In addition, the evaluation criteria are highly contextualised to the national innovation system. In countries where the stock-markets are relatively underdeveloped the role of patents is less-significant compared to US, for instance. Smaller firms in countries with weak law enforcement tend to protect their innovations through classical commercial secrecy.

4 3M for instance applies milestone indicators for a share of revenue generated by new products which are the result of internal research, development, and innovation activities.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

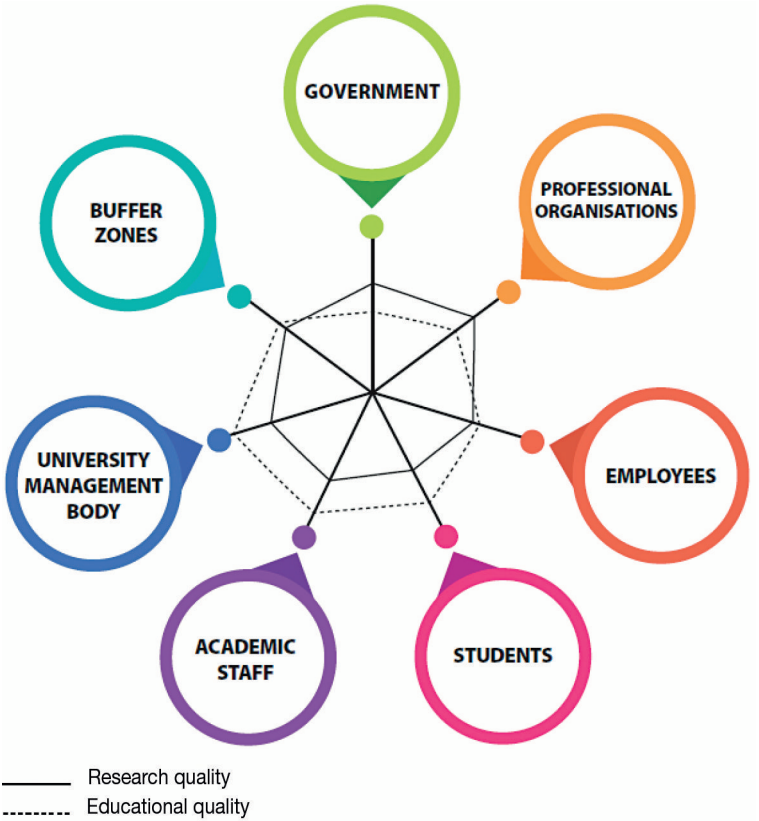
Third, well-known shortcoming of indicators for industrial research is the lack of a transparent ex-post evaluation. One of the reasons for this is also the existence of another regime of company research which takes the competitive conditions on the market into account. Often in relation with this, such researchers are not allowed to publish and disseminate results. At the same time many companies use co-publications with academics as a marketing and legitimisation strategy.

There is another type of evaluation criteria – time-bound relevant indicators. They cover measurable values related to a specific period of time.

If a specific programme is being evaluated (for instance, Cosmic Research), the applicable indicators have to be appropriately selected in order to ensure adequate measurable values. In a lot of cases the indicators are comparable, which allows using an average indicator – for example an average number of citations of one publication by one researcher. In addition, evaluation can be provided for science education and indicators for research management.

In some cases, when the evaluation process covers more than one activity, the different indicators are showcased by presenting all factors influencing the evaluation process. That is how the 'information processing' model is created. The model comprises a set of seven stakeholder target groups or/and interesting parties intervened by the evaluation. This approach is illustrated through the so-called 'radar diagram' (Figure 1.9).

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb



Source: Adapted from Accreditation Agency, the Netherlands.

Figure 1.9: Radar diagram

The indicators which are applied to the management evaluation adhere to a time schedule and address the interests of stakeholders.

The general applicable evaluation methods deal with entry results, exit results, and results from the point of view of the consumers – assessed universities and society. Evaluation methods vary, as is shown below.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

6.1. Evaluation procedures based on expert conclusions

These methods are based on the reviewers' work at the entrance and exit and are focused on:

- evaluating the expectations and, consequently, the research output from the point of view of a research achievement;
- evaluating the result from the point of view of the consumer.

The evaluation conducted by reviewers is most commonly used. It is rooted in tradition and is accepted by the academic body. However, it should be taken into account that there may be deviations due to the lack of balance in the selection of reviewers, resulting in poorly evaluated multidisciplinary studies regardless of the fact that the results are the best exit indicator. That is why this type of expertise is best applied in research-oriented developments. It is accompanied by compulsory collection of adequate data and selection of reviewers.

In the evaluation of specific results through consumers, the issue is made more complicated due to the fact that the latter are directly affected by the execution of a given project and cannot be independent because they know the team that develops the product, and in some cases the final effect impacts more than one project or programme. The effectiveness of integrating a given result, the synergy effect, if there is one, and the extent to which the result obtained is suitable for the intervened system also have an impact in this respect.

6.2. Socio-economic models for research evaluation

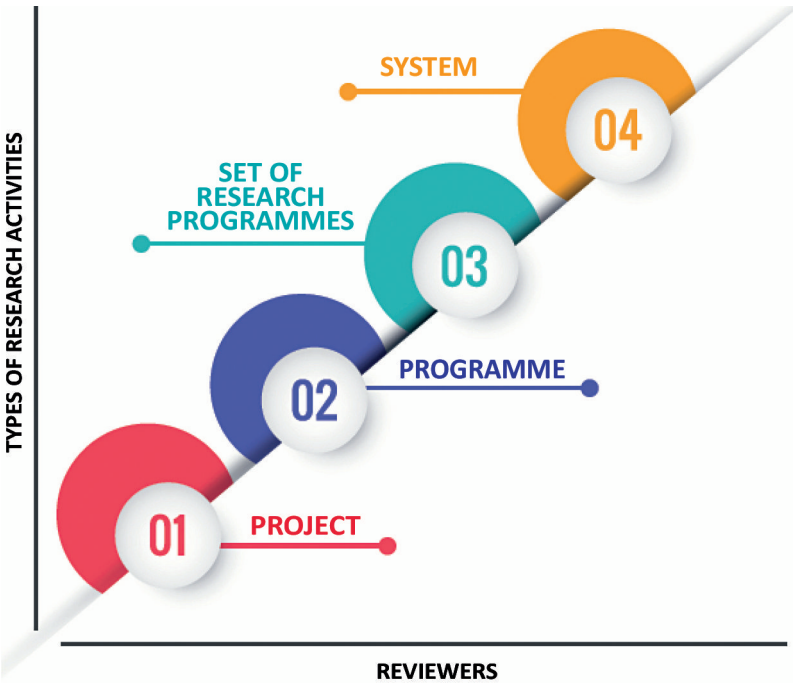
In this model, apart from the traditional scientometric indicators, other methods are used as well – surveys, investigative visits, micro and macroeconomic analyses, comparative analyses, studies of best practices, etc. Nevertheless, an adequate balance between the significance of each one of them has to be sought.

Each of the methods used takes into consideration the objectives of the intentions set in the programme (project) and the development of the programme model, and it applies a realistic approach with regards to the objectives and the designated tasks. For example, the evaluation questions of interim evaluation of Horizon 2020 programme focus exactly on the relevance of the programme, whether it developed as it was expected and

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

if it was adaptive and effective. It includes not only monitoring reports and extensive analysis of the programme itself but also external horizontal studies, data from different EU institutions, input from various stakeholders, and surveys (Interim evaluation of Horizon 2020).

The dynamic of the evaluative process as a series of consecutive actions and in terms of time span is illustrated in Figure 1.10.

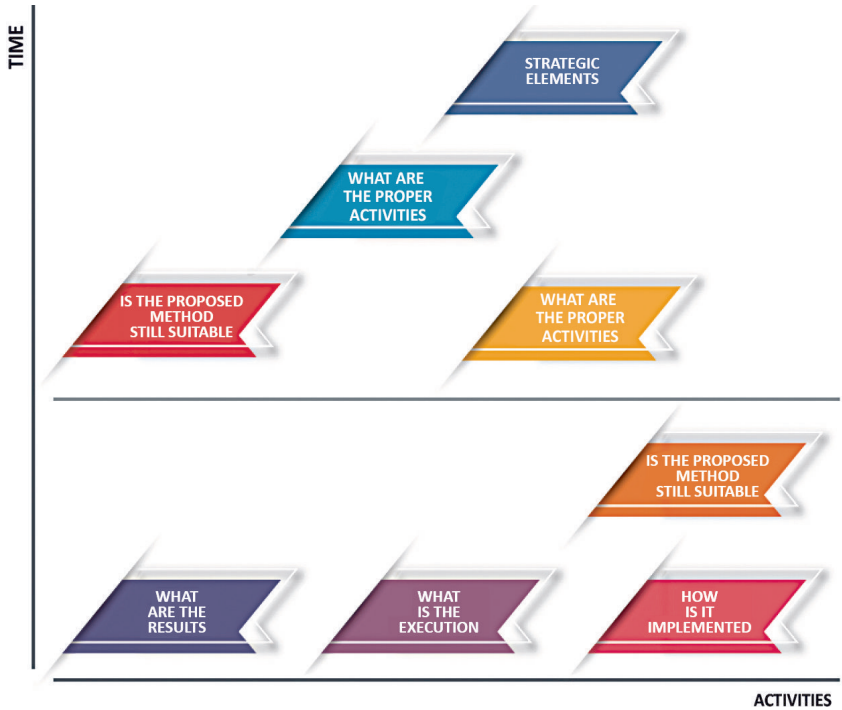


Source: Adapted from Accreditation Agency, the Netherlands.

Figure 1.10: Dynamic of the evaluative process

The peer review evaluation system is practically applied to each of the above-mentioned modules which are subject to evaluation. Figure 1.11. illustrates a full evaluation cycle and includes the following elements:

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb



Source: Adapted from Accreditation Agency, the Netherlands.

Figure 1.11: Levels of activities, subject to evaluation

The specificity of the evaluation process implies the presence of several groups of criteria, for example: to evaluate research projects applying for financing; to evaluate results from completed projects; for institutional evaluation. Each group of criteria contains a specific set of metrics. The majority of them are standardised for different evaluation practices.

6.3. New research evaluation methods

Over time, evaluation methods have evolved by adding new sets of criteria and/or shifting the focus of relevance. The new trends have the ambition to achieve a more complex and truthful assessment. New elements are introduced in the evaluation practice.

The need to evaluate the social impact of science and the aim to achieve a more complete evaluation of research have led to the creation of alternative

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

indicators and approaches. One approach, SCOPE, is suggested by the INORMS Research Evaluation Group. The abbreviation means:

- Start with what you value;
- Context considerations;
- Options for evaluating;
- Probe deeply;
- Evaluate your evaluation.

The supporters of the approach advise first determining the aim of the evaluation and the risks associated to it. The different types of evaluation may affect differently the evaluated, whether we are looking at individuals, institutions, or at national level. Both quantitative and qualitative measures have their imperfections and have to be applied with caution. Still, steps could be taken to mitigate possible negative effects. As for quantitative indicators, they could be used in combination as a 'basket of indicators' and along with qualitative assessment. Among the latter, peer review could be improved through using appropriate experts, more than one and diverse reviewers, and again together with using metrics where appropriate.

An important aspect of the SCOPE model is that the assessment should be conducted in concert with the evaluated individuals or teams. In this way, the evaluator would better understand their aims and the joint interpretation of results would lead to openness and transparency.

Sometimes the evaluation process might have unintended consequences both at institutional and individual levels. Some examples are neglecting activities which aren't measured, discouraging initiative, focusing on the short term, or the academic burden when some academics leave the profession or narrow their publications according to the assessment criteria. That is why it is essential to review or evaluate the evaluation. In this way responsible parties could stay open to adjustments and make sure that the methods they are using are effective (INORMS, 2020).

The aim of some of the additional indicators is to provide additional information which bibliometric indicators cannot. Examples of such evidence regarding research effectiveness are, for example, the number of downloads of a given article or the views and references in social media.

One of the most popular additions to bibliometric indicators is altmetrics. Altmetrics uses indicators for research evaluation based on the activity in online tools and environments (Priem, 2014). Altmetrics.org and Altmetric.com are websites which encourage the use of altmetrics. Altmetric.com aims to popularise and disseminate its products in relation to big academic

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

publishing institutions and financing entities (for example, Taylor & Francis, Wiley, The London School of Economics, and Smithsonian).

In terms of measuring tools, altmetrics are classified based on the function which they offer and the type of users who are interested in the given research outcome. For example, these are categories of different types of altmetrics according to their main functions: discussions, references, readers, reviews, videos and citations.

An evaluation of an article is made by calculating how often it is mentioned in different media platforms, i.e., the popularity of the article is based on how often it is referenced in these sources. In addition to the frequency of referencing, altmetrics uses other measures for viewing, disseminating, and impact indicators.

Even though altmetrics cannot be an alternative to traditional bibliometric indicators, it complements them. The advantage is in the speed in which it gathers and reflects information, something that cannot be said about the other, more frequently used research evaluation indicators. The main limitations to this approach are as follows:

- In case of malicious use of the system, unrealistic results may be generated based on the desire of a specific consumer.
- There is no clear correlation between altmetrics and bibliometrics. The former one includes information from social media such as Facebook and Twitter, which are not academic communities; thus, there is a significant risk that the fundamental research may be neglected.
- The sources of information are not exhaustive.
- There is lack of clarity in the definition and interpretation of a given concept.

Despite these factors, altmetrics is perceived as an area of interest and future research.

Two authors (Herrmannova & Knoth, 2016) have introduced the concept of semanometrics as a new group of research evaluation indicators. They are based on the prerequisite that in order to evaluate a given publication, the full text is needed. The authors believe that a new metric for measuring the impact, which takes the full text of the manuscript into consideration, could be developed by reporting the number of citations and views, and the contribution of the manuscript. They believe that semanometrics has the potential to evaluate to a sufficient degree the quality of research and its contribution.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Another group of authors (Lee, West & Howe, 2017) has established a significant link between the scientific impact and the use of visual information. According to them, the bigger the impact of a given research, the more likely it is to include more diagrams, but on the condition that this number varies a lot in individual areas and is applicable only for some sciences. The aim is to study the organisation of visual information because these information-rich objects are largely ignored in bibliometrics and scientometrics research in comparison to citations and text. The authors introduce the concept of visiometrics in order to discover more interesting and useful applications for their idea.

7. Ethics in research evaluation

National research assessment systems are very different, and they have an impact on the research strategies of research departments. Whitley (2007) highlights that research assessment systems affect the organisation and management of knowledge. The robust assessment systems with high standards, rules, and officially established procedures concerning assessment and publishing of results most definitely influence the research strategies of universities and research organisations in a different way depending on the individual research fields. In addition, the connection between policy and assessment may have an impact on the quality of the research. According to Pleger (2016), assessments are performed in a given political context, and they are influenced by it.

Different researchers believe that the study of ethics with regards to performing a research assessment is an underdeveloped field (Gedutis & Biagetti, 2019). There is a lack of shared understanding as to what ethics are, and the standards for quality and competence are regularly confused with the term 'ethics'. It is not clear what the role of ethics is and who is responsible for implementing ethical practices (Williams, 2016).

Biagetti, Gedutis and Ma (2020) represent research evaluation ethics combining aspects from both research and evaluation good practices. The authors claim that there aren't enough clear guidelines how to establish proper criteria and avoid bias and conservatism in peer review.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Source: Adapted from Biagetti, Gedutis & Ma (2020).

Figure 1.12: Research Evaluation Ethics between Research Ethics and Evaluation Ethics

A lot of authors (Biagetti et al., 2020) note that a number of issues arise during the assessment process, and those are mainly related to the interpretation of bibliometrics.

It is believed that the results-based indicators have to measure the value of research in an objective manner, but the research community believes that they are often a reason for the occurrence of new forms of manipulation. Since the research assessment relies on approximate (proxy) indicators which only measure indirectly what they should actually present (quality, impact, or social significance), it is increasingly difficult to establish to what extent some or all of the indicators are manipulated and to what extent given high-value studies are authentic. According to some authors, the assessment process is still chaotic and the role of the indicators is not quite clear; for example, despite the fact that the impact factor of journals is the most popular indicator, it is not the only one applied and is not uniform for all assessments (Wouters, 2020). There are, however, some dominant trends in the quality assessments which are relevant to citation practices. There is a risk that research which is not measured by specific indicators may be neglected, which threatens the search for knowledge on the part of the universities (Wouters, 2020). The criticism against scientometrics in Eastern Europe is more frequently expressed by scholars who justify in this way the lack of publications which enter in popular databases. Individual research departments within a given research structure are often against the

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

introduction of elementary requirements to PhD candidates (for example, an article published in a journal indexed by Scopus or Web of Science), arguing that the majority of their members do not have such requirements. This poses a serious obstacle for increasing the quality of research and the international convertibility of research and PhD education.

Some authors take into account the fact that the indicator-based assessment creates pressure for active publishing (Fanelli, 2020), but they also remark that the latter does not lead to unlawful actions and does not hinder researchers' integrity.

The desire of journal editors to receive a higher impact factor leads to another specific effect: an artificial increase in the citations through the coordinated efforts of a 'citation cartel' of journals. Such 'citation cartels' have been observed more and more often over the past years (Kojaku et al., 2021). Different researchers direct their efforts at the creation of algorithms and methods for their reporting (Kojaku et al., 2021; Koley & Mishra, 2019; Perez et al., 2019) or at the exclusion of such journals from international databases.

There is a possibility of manipulation via the participation of researchers in editorial boards at international publishing houses, which are oriented towards increasing the number of publications by a given university and, respectively, increasing the number of co-authors in order to improve these results (Biagioli & Lippman, 2020). Such behaviour distorts the objective picture of bibliometrics.

Another issue arises when the expert assessment and the citation analysis contradict each other. If we only rely on the expert assessment, does this not slow down the development of interdisciplinary studies, because the focus is on established and favoured methodologies? On the other hand, we cannot rely only on quantitative indicators. We cannot directly interpret the number of publications or citations, which is normalised for a given field, as an indicator for quality or impact. The high number of citations can be due to the presence of a unique empiricism, an exceptional research with great impact, or it can be the result of repetition of studies or the efforts of citation cartels. A small number of citations, on the other hand, can be the result of a research which is not that interesting or of innovative ideas that are still not recognised (Wouters, 2020) or published in a journal which has a limited reach. An expert assessment is needed in such case. That is why the objective research assessment requires an incredible balance and a careful approach to the above-mentioned activities. In addition, the assessment is closely linked to a political vision and, respectively, specific

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

methods, and it is an instrument which shows us what kind of society we want to build (Mol, 2002; Thurtle & Mitchell, 2002).

Governments often conduct reforms and make decisions regarding financing based on the global rankings of universities (Rouet, 2022). In this case the effect of the strategies for achieving impact through publications may be distorted and conditions for manipulation may be created (gaming opportunities) (Biagioli & Lippman, 2020). Sarah de Rijcke and Tereza Stöckelová claim that the focus of European research policies on 'international publication impact' as a substitute for quality increases the division between the 'international' north and the limited south (Rijcke & Stöckelová, 2020).

The unscrupulous application of assessment indicators has been discussed many times, including in the San Francisco Declaration on Research Assessment, The Metrics Tide, the Leiden Manifesto, etc. There is still, however, a lack of common approach and application of ethical principles with regards to research assessment in the preparation of an assessment and/or the selection of the criteria, regardless of what the effects of the assessment process are going to be (Dahler-Larsen, 2012). Part of the studies, which focus on ethical issues related to the assessment, are dedicated to the assessors' ethics (Morris, 2008; Schwandt, 2015) and study the ethical dilemmas in their professional conduct. At the moment, professional ethics to a great extent focuses on everyday issues related to the individual participants in the process. Ethical issues are examined in the context of interpersonal relationships where the focus is on the issues which occur as a result of the relations between assessors and the other stakeholders (Schwandt, 2018). As regards the independence of assessments, the pressure exercised by the stakeholders is identified as an important ethical challenge (Pleger, 2016). Morris (2015) admits that collecting information about how assessors react to ethical conflicts is of vital importance, whilst also being a delicate endeavour, bearing in mind the defensive position which some studies in this field may lead to.

There is a possibility of expressing preference with regards to gender, race, language, career stage, and the interdisciplinarity (Helmer, 2017; Lee et al., 2013). There is also a possibility of neglect with regards to the use of innovative procedures and platforms in the assessment process (Bornmann, 2011; Horbach & Halffman, 2019) due to the habit of implementing the routine methods or tools or due to unwillingness to try a new work method.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

A socially responsible assessment is one where the rights, dignity, and cultural values of individuals and groups are taken into account. Professional assessors are encouraged to understand and respect the points of view etc. of all stakeholders (Schwandt, 2018). Studies in the field of ethics with regards to assessment should assist assessors, and the latter should use the positive experience of other assessors (Pleger, 2016).

The European Code of Conduct for Research integrity (ALLEA 2017) is a document which postulates the general principles of research ethics, including reliability, honesty, accountability, etc. According to this document, researchers who participate in the assessment process undertake a serious commitment. They have to consider a number of factors such as the presence of a conflict of interests, confidentiality, respecting the rights of authors, etc.

An inseparable part of the integrity of research is the absence of plagiarism, and when it is discovered it should not be neglected, but rather sanctions should be imposed, which would lead to the loss of an academic position. Unfortunately, in Bulgaria the procedure is often suspended without it leading to a direct negative effect for the person responsible for the act of plagiarism. The only exceptions are for people and cases which have become public knowledge (for example politicians), but even that is not guaranteed. A study by Foltynek and Glendinning (2015) shows significant differences among European countries with regards to their understanding of what plagiarism is and to the attitude towards plagiarism, the preparedness how to avoid it, etc.

In a research system where the number of publications is considered an indicator of 'quality' and is a tool used for encouraging career growth and the allocation of grants, 'recycling' a text or self-plagiarism (as a kind of plagiarism) is a way of increasing the results at the expense of other researchers. This raises the question as to what extent the indicators based on results from publications are important assessment criteria for the allocation of work or grants (Horbach & Halffman, 2019). According to some researchers, the solution to the plagiarism issue is to place a focus on quality, not quantity, in the system of criteria (Feenstra, 2021).

According to Helen Simons, plenary lecturer at The Framing Ethics in Impact Evaluation seminar (Barnett & Munslow, 2014), the ethical guidelines or postulates proposed are mainly intentions, and they often relate to the assessment methodology and to the quality of the assessment with regard to a given product rather than focus on whether the research assessment is correct. According to him, and other authors as well, there is

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

a need for ethical norms based on theory, which would guide the assessors' behaviour and choice. Adherence to ethical norms in research assessment is highly dependent on the context of the general level of ethical behaviour in the given country (corruption levels, rule of law, degree of self-regulation in other fields). According to Biagetti and her co-authors Gedutis and Ma (2020), a mixed approach may be applied with regards to resolving the issues of assessment practices.

One of the important questions with regards to the ethics in research assessment relates to its boundaries and the scope of the field. According to Mustajoki and Mustajoki (2017), the identification of ethical issues is achieved in three ways: (a) by identifying the stakeholders (for example, individuals, groups, communities, animals, ecosystems, future generations, etc.), (b) by understanding the rights and responsibilities of the stakeholders, and (c) by defining the options, i.e. searching for a win-win situation or achieving it to the greatest extent possible for the participating stakeholders. Research assessment, be it preliminary or subsequent, concerns important ethical issues. The aim towards a 'common good' in the assessment of a given study means that, in case of a multidisciplinary or interdisciplinary study, each stakeholder involved in the study must be taken into consideration (ESF, 2011).

Some studies show that part of the assessors accept and stand by the claim that ethics do not relate to assessment and that they have never encountered ethical problems in their work (Morris, 2015; Williams, 2016). At the same time, however, it must be taken into account that a well-financed organisation or project can allocate enough financial resources for assessment (especially an interim one or a final internal or external one), while organisations with poor funding do not have this capacity. This would mean that the practice of research assessment is routine in those places where there is funding, not in the places where the assessment is most needed.

Concerning the study, the following four standards, which to a great extent correlate to the ethical principles, must be observed in assessment procedures:

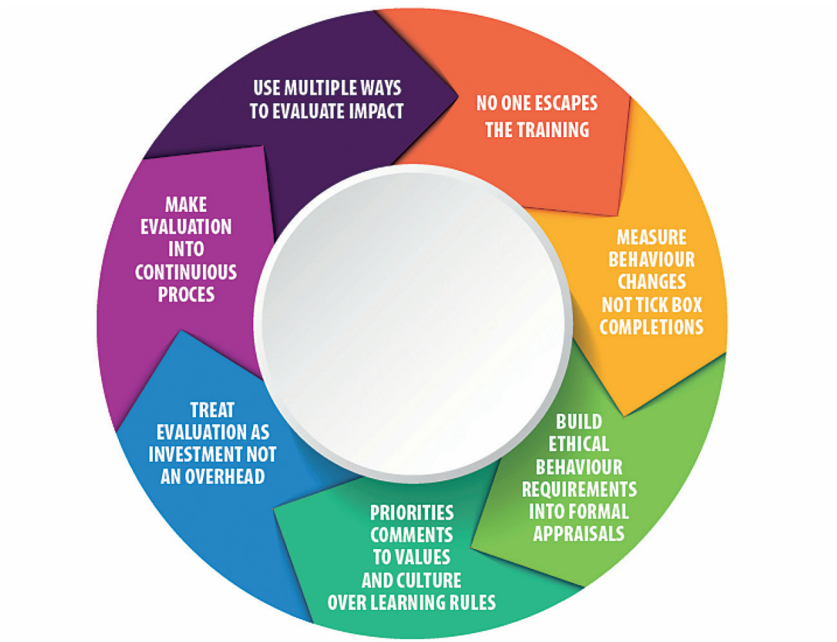
- **Usefulness** – research assessments must address important issues, and the results expected or received must be clear and comprehensible. They should include reasonable recommendations if there is need of such.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

- There should be **realism** with regards to the implementation of a given programme and project, strategic measures, policies, etc. in relation to time and finances.
- **Legitimacy** – the assessments should comply with the respective principles and be institutionally accepted and recognised by the academic community.
- **Accuracy** – the information must be gathered, analysed, reported, and interpreted in an accurate and impartial manner.

The National Science Foundation (USA), NSF, postulates the following four principles in the assessment process: goodwill, trust, professionalism, and confidentiality.

Corporate companies invest a lot in compliance and ethics training. According to Andrew Leigh, there are seven principles of ethics training which underlies its success and might be applied into research evaluation as well.



Source: Adapted from Leigh, 2015.

Figure 1.13: Ethics Evaluation Essentials

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Patton (2014) highlights that studies are something which informs science, while the assessment is something which is targeted at and supports the action itself. In reality, the differentiation between an assessment and a study is rarely that clear, especially with regards to mixed terms about the study of policies, study of the assessment process, or study of the application of results from a given research. The ones in support of the use of experimental methods in the assessment, especially randomised control studies, claim that the knowledge generated is more likely to be precise than approximate, but it is focused on a limited number of issues, and the period for generating it is a lot longer in comparison to the time that is needed to perform applied studies.

Naturally, the assessment process is only one of the factors which contributes to the final result (as studied by Anderson, 2014), but conditions for a distorting effect can be created on all aspects: from the formulation of questions to be the subject of assessment to decisions about the resources, methodologies applied, etc. Adherence to ethical principles supports assessors (and all stakeholders) in their work. Some authors propose a reasonable compromise between the methodological rigor of the assessment process and the assessment itself as a form of knowledge which has to be discussed and used not only by people in positions of power but by the civil society as well. This is quite acceptable when there are a lot of active financial instruments and the specific expectations of each of them are varied.

References

Abramo, G., D'Angelo, C. A., & Grilli, L. (2021). The effects of citation-based research evaluation schemes on self-citation behaviour. *Journal of Informetrics*, 15(4), 101204.

ALLEA – All European Academies (2017). *The European Code of Conduct for Research Integrity*. https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf

Andersen, O. W. (2014). Some thoughts on Development Evaluation Processes. In: B. Befani, C. Barnett & E. Stern (eds). *Rethinking Impact Evaluation for Development. IDS Bulletin*, 45 (6), Brighton: Wiley Blackwell, 77–84.

Åström, F., Hammarfelt, B. & Hansson, J. (2017). Scientific publications as boundary objects: theorising the intersection of classification and research evaluation. Proceedings of the Ninth International Conference on Conceptions of Library and Information Science. Uppsala, Sweden, June, 2016, *Information Research*, 3(1), http://informationr.net/ir/22-1/colis/colis1623.html

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Barnett, C., & Camfield, L. (2016). Ethics in evaluation. *Journal of Development Effectiveness*, 8(4), 528–534, https://doi.org/10.1080/19439342.2016.1244554

Barnett, C., & Munslow, T. (eds) (2014). Framing Ethics in Impact Evaluation: Where Are We? Which Route Should We Take? *IDS Evidence Report* 98, Brighton: IDS, https://www.cdimpact. org/publications/workshop-report-framing-ethics-impact-evaluation-where-are-we-which-route-should-we

Biagetti, M. T. et al. (2020). Ethical Theories in Research Evaluation: An Exploratory Approach. *Scholarly Assessment Reports*, 2(1), 11, https://doi.org/10.29024/sar.19

Biagioli, M., & Lippman, A. (2020). *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. The MIT Press, https://doi.org/10.7551/mitpress/11087.001.0001

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45, 197–245, https://doi.org/10.1002/aris.2011.1440450112

Bornmann, L., & Williams, R. (2020). An evaluation of percentile measures of citation impact, and a proposal for making them better. *Scientometrics*, 124, 1457–1478. DOI: 10.1007/s11192–020–03512–7

Buxton, M., & Hanney, S. (1996). How can payback from health services research be assessed? *Journal of Health Services Research & Policy*, 1(1), 35–43, https://pubmed.ncbi.nlm.nih.gov/10180843/

Constant, N., & Roberts, L. (2017). Narratives as a mode of research evaluation in citizen science: understanding broader science communication impacts. *Journal of Science Communication*, 16 (4), https://doi.org/10.22323/2.16040203

Curry, S. (2018). Let's move beyond the rhetoric: it's time to change how we judge research. *Nature*, 554 (7691), 147–147, https://doi.org/10.1038/d41586-018-01642-w

Dahler-Larsen, P. (2012). *The Evaluation Society*. Stanford, California: Stanford Business Books, https://doi. org/10.2307/j.ctvqsdq12

Daraio, C. (2017). A Framework for the Assessment of Research and Its Impacts. *Journal of Data and Information Science*, 2 (4), 7–42, https://doi.org/10.1515/jdis-2017-0018

Dawson, G., Lucocq, B., Cottrell, R., & Lewison, G. (1998). *Mapping the Landscape: National Biomedical Research Outputs 1988–95. PRISM Policy Report 9*. Wellcome Trust, https://wellcome.ac.uk/sites/default/files/wtd003187_0.pdf

De Bellis, N. (2009). *Bibliometrics and Citation Analysis from the Science Citation Index to Cybermetrics*. Scarecrow Press. DOI: 10.12775/TSB.2012.009

Elzinga, A. (1995). Reflections on research evaluation. *Science & Technology Studies*, 8 (1), 5–23.

ESF (2011). European Science Foundation, Member Organisation Forum European Peer Review Guide Integrating Policies and Practices into Coherent Procedures, https://repository.fteval.at/148/1/2011_European%20Peer%20 Review%20Guide.pdf

European Commission, Directorate-General for Research and Innovation (2017). *Interim evaluation of Horizon 202: Commission staff working document*. Publications Office, https://data.europa.eu/doi/10.2777/78199

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

European Commission, Directorate-General for Research and Innovation. (2018). *The Open Science Policy Platform – Terms of Reference*. https://ec.europa.eu/research/openscience/ pdf/terms_of_reference_ospp.pdf

European Commission, Directorate-General for Research and Innovation (2020). *Progress on Open Science: Towards a Shared Research Knowledge System – Final Report of the Open Science Policy Platform*. Publications Office of the European Union, https://ec.europa.eu/research/openscience/pdf/ec_rtd_ospp-final-report.pdf

Fahrenkrog, G., Polt, W., Rojo, J., Tuebke, A., Zinoecker, K., Kuhlmann, S., et al. (2002). *RTD Evaluation Toolbox – Assessing the Socio-Economic Impact of RTD-Policies*. European Commission, http://repository.fteval.at/id/eprint/272

Fanelli, D. (2020). Pressures to Publish: What Effects Do We See? In M. Biagioli & A. Lippman (eds). *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. Cambridge MA, London: The MIT Press, https://doi.org/10.7551/mitpress/11087.001.0001

Fauzi, M. A., Tan, C. N. L., Daud, M., & Awalludin, M. M. N. (2020). University rankings: A review of methodological flaws. *Issues in Educational Research*, 30 (1), 79–96, http://www.iier.org.au/iier30/fauzi.pdf

Feenstra, R. A., Delgado López-Cózar, E., & Pallarés-Domínguez, D. (2021). Research Misconduct in the Fields of Ethics and Philosophy: Researchers' Perceptions in Spain. *Science and Engineering Ethics*, 27 (1), https://doi.org/10.1007/s11948-021-00278-w

Foltýnek, T., & Glendinning, I. (2015). Impact of policies for plagiarism in higher education across Europe: Results of the project. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 63 (1), 207–216.

Gedutis, A., & Biagetti, M. T. (2019). *Towards Ethical Principles of Research Evaluation in SSH*. RESSH project, https://ressh2019.webs.upv.es/wp-content/uploads/2019/10/ressh_2019_paper_37.pdf

Georghiou, L., & Laredo, P. (2005). *Evaluation of publicly funded research*. Report on Berlin workshop, 26–27 September 2005. Paris: OECD.

Gibbons, M., & Johnston, R. (1974). The roles of science in technological innovation. *Research Policy*, 3 (3), 220–242, https://doi.org/10.1016/0048-7333(74)90008-0

Griliches, Z. (1958). Research Costs and Social Returns: Hybrid Corn and Related Innovations. *Journal of Political Economy*, 66, 419, https://www.journals.uchicago.edu/doi/pdfplus/10.1086/258077

Gutzman, K. E., Bales, M. E., Belter, C. W., Chambers, T., Chan, L., Holmes, K. L., Lu, Y. L., Palmer, L. A., Reznik-Zellen, R. C., Sarli, C. C., Suiter, A. M., & Wheeler, T. R. (2018). Research evaluation support services in biomedical libraries. *Journal of medical library association*, 106 (1), dx.doi.org/10.5195/jmla.2018.205

Hammarfelt, B., & Rushforth, A. D. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*, 26 (3), 169–180, https://doi.org/10.1093/reseval/rvx018

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Hanney, S. R., Gonzalez Block, M. A., Buxton, M., & Kogan, M. D. (2003). The Utilisation of Health Research in Policy-Making: Concepts, Examples and Methods of Assessment. *Health Research Policy and Systems*, 1 (1), 2. DOI: 10.1186/1478–4505–1–2

Helmer, M., Schottdorf, M., Neef, A., & Battaglia, D. (2017). Gender bias in scholarly peer review. *eLife*, *6*, https://doi.org/10.7554/eLife.21718

Herbertz Benno, H., & Müller-Hill, B. (1995). Quality and efficiency of Basic Research in Molecular Biology: A Bibliometric Analysis of Thirteen Excellent Research Institutes. *Research Policy*, 24 (6), 959–979, https://doi.org/10.1016/0048-7333(94)00814-0

Herrmannova, D., & Knoth, P. (2016). Semantometrics: Towards Fulltext-based Research Evaluation. In: *Proceedings of the 16th ACM/IEEE-CS Joint Conference on Digital Libraries, 19–23 Jun 2016, Newark, New Jersey, USA*, 235–236, https://doi.org/10.1145/2910896.2925448

Hess, D. J. (1997). *Science Studies: An advanced introduction*. New York: New York University Press.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520, 429–431. DOI: 10.1038/520429a

Hong, H. D., &Boden, M. (2003) R&D *Programme Evaluation – Theory and Practice*. Aldershot: Ashgate Publishing Limited.

Horbach, S. P. J. M., & Halffman, W. (2019). The extent and causes of academic text recycling or 'self-plagiarism'. *Research Policy*, 48 (2), 492–502. https://doi.org/10.1016/j.respol.2017.09.004

Horbach, S. P. J. M., & Halffman, W. (2019). Journal Peer Review and Editorial Evaluation: Cautious Innovator or Sleepy Giant? *Minerva*, 58 (2), 139–161, https://doi.org/10.1007/s11024-019-09388-z

INORMS Research Evaluation Working Group (2020). *Better decision making through responsible research evaluation*. Research Administration [Briefing Material], https://inorms.net/wp-content/uploads/2020/05/scope.pdf

Jewkes, J., Sawers, D., & Stillerman, R. (1958). *The Sources of Invention*. London: Macmillan.

Joint Research Centre, Institute for Prospective Technological Studies (2002). *RTD Evaluation Toolbox – Assessing the Socio-Economic Impact of RTD-Policies*. European Commission, https://repository.fteval.at/272/1/2002_RTD%20Evaluation%20Toolbox.pdf

Jurajda, Š., Kozubek, S., Münich, D., & Škoda, S. (2017). Scientific publication performance in post-communist countries: still lagging far behind. *Scientometrics*, 112 (1), 315–328, https://doi.org/10.1007/s11192-017-2389-8

Kamenetzky, A., & Hinrichs-Krapels, S. (2020). How do organisations implement research impact assessment (RIA) principles and good practice? A narrative review and exploratory study of four international research funding and administrative organisations. *Health Research Policy and Systems*, 1 (1), https://doi.org/10.1186/s12961-019-0515-1

Kara, H. (2019). The Ethics of Evaluation Research. *Research Ethics Monthly*, https://ahrecs.com/human-research-ethics/the-ethics-of-evaluation-research

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Kojaku, S., Livan, G., & Masuda, N. (2021). Detecting anomalous citation groups in journal networks. *Scientific Reports*, 11, 14524, https://doi.org/10.1038/s41598-021-93572-3

Kolev, I. (2017) International Philosophy Olympiad. In: Peter, M. (ed). *Encyclopedia of Educational Philosophy and Theory*. Singapore: Springer, 1158–1164.

Koley, A., & Mishra, S. (2019). Finding Citation Cartels in Academic Research. EasyChair Preprint no. 1960. https://easychair.org/publications/preprint_open/vM4b

Landsberger, H. A. (1958). *Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry*. Ithaca NY: Cornell University Press.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology,* 64 (1), 2–17, https://doi.org/10.1002/asi.22784

Lee, P., West, J. D., & Howe, B. (2018). Viziometrics: Analyzing Visual Information in the Scientific Literature. *IEEE Transactions on Big Data*, 4 (1), 117–129. DOI: 10.1109/TBDATA.2017.2689038

Leigh, A. (2015, April 13). Does your compliance training make an ethical difference? *Ethical Leadership from Andrew Leigh*, https://www.ethical-leadership.co.uk/compliance-training4/

Leydesdorff, L. (2005). Evaluation of Research and Evolution of Science Indicators. *Current Science*, 89 (9), https://arxiv.org/abs/0911.4298

Mansfield, E. (1991). Academic Research and Industrial Innovation. *Research Policy*, 20 (1), 1–12, https://doi.org/10.1016/0048-7333(91)90080-A

Marjanovic, S., Hanney, S., & Wooding, S. (2009). *A historical reflection on research evaluation studies, their recurrent themes and challenges*. RAND Corporation, https://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR789.pdf

Martin, B., & Salter, B. (1996). *The relationship between public funded research and economic performance*. Science Policy Research Unit.

Milzow1, K., Reinhardt, A., Soderberg, S., & Zinocker, K. (2018). Understanding the use and usability of research evaluation studies. *Research Evaluation*, 28 (1), 94–107, https://doi.org/10.1093/reseval/rvy040

Mol, A. (2002). *The Body Multiple: Ontology in Medical Practice*. Durham, NC: Duke University Press.

Morris, M. (ed) (2008). *Evaluation Ethics for Best Practice. Cases and Commentaries*. New York: The Guilford Press.

Morris, M. (2015) Research on evaluation ethics: reflections and an agenda. In: Brandon, P. (ed). *Research on evaluation: New directions for evaluation*. Hoboken, NJ: Wiley, 31–42.

Morris, M. (2015). Research on evaluation ethics: Reflections and an agenda. *New Directions for Evaluation*, 148, 31–42.

Mustajoki, H., & Mustajoki, A. (2017). *A New Approach to Research Ethics: Using Guided Dialogue to Strengthen Research Communities*. New York: Routledge, https://doi.org/10.4324/9781315545318

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Nigohosyan, D., Vutsova, A. (2018). The 2014–2020 European Regional Development Fund Indicators: The Incomplete Evolution. *Social Indicators Research*, 137, 559–577, https://doi.org/10.1007/s11205-017-1610-8

Ochsner, M., Kulczycki, E., & Gedutis, A. (2018). The Diversity of European Research Evaluation Systems. *International Conference on Science and Technology Indicators*. STI 2018 Conference proceedings, https://hdl.handle.net/1887/65217

OECD Committee on Science and Technology Policy / OECD Working Party on Innovation and Technology Policy (TIP) (2009). *Enhancing Public Research Performance through Evaluation, Impact Assessment and Priority Setting*. OECD, https://www.oecd.org/sti/inno/Enhancing-Public-Research-Performance.pdf

Office of Technology Assessment (1986). Research Funding as an Investment: Can we Measure the Return? Washington DC: US Congress, https://ota.fas.org/reports/8622.pdf

Patton, M. Q. (2014). *Evaluation Flash Cards, Embedding Evaluative Thinking in Organisational Culture*. St Paul, MN: Otto Bremer Foundation.

Penfield, T., Baker, M. J., Scoble, R., & Wykes, M. C. (2014). Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation*, 23 (1), https://doi.org/10.1093/reseval/rvt021

Perez, O., Bar-Ilan, J., Cohen, R., & Schreiber, N. (2019). The Network of Law Reviews: Citation Cartels, Scientific Communities, and Journal Rankings. *Modern Law Review*, 82, 2, 240–268, https://doi.org/ 10.1111/1468–2230.12405

Pleger, L., & Sager, F. (2018). Betterment, undermining, support and distortion: A heuristic model for the analysis of pressure on evaluators. *Evaluation and Program Planning*, 69, 166–172. DOI: 10.1016/j.evalprogplan.2016.002

Research evaluation needs to change with the times [Editorial] (2022). *Nature*, 601, 166, https://doi.org/10.1038/d41586-022-00056-z

Rijcke, S., & Stöckelová, T. (2020). Predatory Publishing and the Imperative of International Productivity: Feeding Off and Feeding Up the Dominant. In M. Biagioli & A. Lippman (eds). *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. Cambridge MA, London: The MIT Press, https://doi.org/10.7551/mitpress/ 11087.001.0001

Rivera, H. (2018). Inappropriate Authorship and Kinship in Research Evaluation. *Journal of Korean Medical Science*, 33 (13), https://doi.org/10.3346/jkms.2018.33.e105

Ronda-Pupo, G. A. (2017). Research evaluation of author's citation-based performance through the relative author superiority index. *Transinformação*, 29 (2), http://dx.doi.org/10.1590/2318-08892017000200006

Rossi, P. H., & Freeman, H. (1999). *Evaluation: A Systematic Approach*. Thousand Oaks CA: Sage Publications.

San Francisco Declaration on Research Assessment (DORA). https://sfdora.org/read/

Rouet, G. (2022) (ed) *Classement des universités. Lectures, publications reçues*. Paris : CNRS éditions.

Schwandt, T. A. (2015). *Evaluation Foundations Revisited: Cultivating a Life of the Mind for Practice*. Stanford: Stanford University Press.

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Schwandt, T. A. (2018). Acting together in determining value: A professional ethical responsibility of evaluators. *Evaluation*, 24 (3), 306–317, https://doi.org/10.1177/1356389018781362

Sherwin, C. W., & Isenson, R. S. (1967). Project Hindsight. *Science*, 156 (3782), 1571–1577. DOI: 10.1109/science.156.3782.1571

Strielkowski, W., Gryshova, I., & Shcherbata, M. (2017). Predatory Publishing and Beall's List: Lessons for the Countries Adapting Novel Research Evaluation Criteria. *Science and education*, 8, 39–43, https://doi.org/10.24195/2414-4665-2017-8-5

Szomszor, M., Adams, J., Fry, R., Gebert, C., Pendlebury, D. A., Potter, R. W. K., & Rogers, G. (2021). Interpreting Bibliometric Data. *Frontiers in Research Metrics and Analytics*, 5:628703. DOI: 10.3389/frma.2020.628703

Thomas, D. A., Nedeva, M., Tirado, M. M., & Jacob, M. (2020). Changing research on research evaluation: A critical literature review to revisit the agenda. *Research Evaluation*, 29 (3), 275–288, https://doi.org/10.1093/reseval/rvaa008

Thomas, R. (2018). *Questioning the Assessment of Research Impact Illusions, Myths and Marginal Sectors*. Springer, https://doi.org/10.1007/978-3-319-95723-4

Thurtle, P., & Mitchell, R. (2002). *Semiotic Flesh: Information and the Human Body*. Seattle: University of Washington Press.

Tusting, K. (2018). The genre regime of research evaluation: contradictory systems of value around academics' writing. *Language and Education*, 32 (6), 477–493, https://doi.org/10.1080/09500782.2018.1505905

Vutsova, A. (2006). *Research Evaluation – Bulgarian Case, Bulgarian Integration into Europe & NATO*. Zvezdi.

Vutsova, A. (2009a). Science and Innovations within the International Cooperation in Europe. *Economic Thought*, 4, 77–89, https://www.ceeol.com/search/article-detail?id=69469

Vutsova, A. (2009b). *Research Policy in the Context of European Integration*. Sofia: Technical University.

Vutsova, A. (2014). *Influence of the EU Instruments on the Institutional Development of Higher Education and Science*. Zvezdi.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., & van Eck, N. J. (2012). The Leiden ranking 2011/2012: data collection, indicators, and interpretation. *Journal of the Association for Information Science and Technology*, 63 (12), 2419–2432. DOI: 10.1002/asi.22708

Weinberg, A. M. (1989). Criteria for Evaluation, a Generation later. In D. Evered & S. Narnett (eds). *Evaluation of Scientific Research*. Ciba Foundation Conference.

Whitley, R. (2007). Changing Governance of the Public Sciences. The Consequences of Establishing Research Evaluation Systems for Knowledge Production in Different Countries and Scientific Field. In R. Whitley & J. Gläser (eds). *The Changing Governance of the Sciences. The Advent of Research Evaluation Systems*. Dordrecht: Springer, 3–27.

Williams, A. E. (2017). Altmetrics: an overview and evaluation. *Online Information Review*, 41 (3), 311–317, https://doi.org/10.1108/OIR-10-2016-0294

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

Williams, L. (2016). Ethics in international development evaluation and research: what is the problem, why does it matter and what can we do about it? *Journal of Development Effectiveness*, 8 (4), 535–552. DOI: 10.1080/19439342.2016.1244700

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. HEFCE. DOI: 10.13140/RG.2.1.4929.1363

Wooding, S., Anton, S., Grant, J., Hanney, S., Hoorens, S., Lierens, A., Shergold, M., & Venema, M. (2004). *The Returns from Arthritis Research. Volume 2: The Case Studies*. RAND Corporation, https://www.rand.org/pubs/technical_reports/TR176.html

World Conferences on Research Integrity (2019). *Hong Kong Principles for assessing researchers*, https://wcrif.org/guidance/hong-kong-principles

Wouters, P. (2020). The Mismeasurement of Quality and Impact. In M. Biagioli & A. Lippman (eds). *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. Cambridge MA, London: The MIT Press, https://doi.org/10.7551/mitpress/11087.001.0001

Zhang, L., Rousseau, R., & Sivertsen, G. (2017). Science deserves to be judged by its contents, not by its wrapping: Revisiting Seglen's work on journal impact and research evaluation. *PLoS ONE,* 12 (3), Article e0174205, https://doi.org/10.1371/journal.pone.0174205

Zubascu, F. (2018) Poland's 'Constitution for Science' gives universities more power and accountability. *ScienceBusiness*, November 20, 2018, https://sciencebusiness.net/news/polands-constitution-science-gives-universities-more-power-and-accountability

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb

https://doi.org/10.5771/9783748937203-19, am 18.09.2024, 14:20:32
Open Access – (cc) BY – https://www.nomos-elibrary.de/agb