

Towards a framework for the appropriate use of bibliometric indicators in research evaluation

Cinzia Daraio^{*}, Juan Gorraiz^{**} and Wolfgang Glänzel^{**}

**daraio@diag.uniroma1.it*
ORCID: 0000-0002-4825-0071
DIAG, Sapienza University of Rome, Italy

***juan.gorraiz@univie.ac.at; wolfgang.glanzel@kuleuven.be*
ORCID: 0000-0002-2414-3212; 0000-0001-7529-5198

Dept Bibliometrics & Publication Strategies, University Vienna, Austria; ECOOM, KU Leuven, Belgium

Conference topic: Practices of evaluation and assessment

Contribution methodological approach: Methodological contributions (incl. indicator development)

Abstract

Recently, research evaluation using quantitative methods has received much criticism, both from a part of the scientific community and from recent initiatives at the European level calling for a rethinking of research evaluation by applying mainly peer-review. We focus on the use of bibliometric indicators in an evaluative context. We sketch a general framework of criteria that should be considered in the use of bibliometric indicators in order to ascertain whether their use, in the specific evaluation context, is appropriate for its intended purpose or not. Are bibliometric indicators always inappropriate, or should they be used with care and skill, with respect to the evaluative problem under consideration? In this paper, we caution against “throwing the baby out with the bathwater” and advocate the idea that bibliometric indicators, even the number of publications and citations received, if used “appropriately” can still be extremely useful for research evaluation.

1. Introduction

Recently the application of science and technology indicators has become the target of heavy criticism for the negative consequences of their use in an evaluative context. Yet, most initiatives aiming at a reform of existing assessment systems do often not clearly delineate the concept of research to be assessed nor do they define the goal, the aggregation level and granularity of the exercise in a satisfactory manner. Therefore, it remains often unclear, if research is regarded as an entire process and being part of broader academic activities or just as the research and its outputs in a narrow sense. Finally, most arguments against the use of bibliometric methods and in favour of peer review are based on phenomena observed at the level of individual scientists, which have already been recognised by the bibliometric community for some time (e.g., Wouters et al., 2013).

The background of the criticism of indicators use can be summarised in brief. Research evaluation using quantitative methods, notably those based on the counting of published papers and citations received by those, has increasingly encountered rejection on a part of the scientific communities (Cronin and Sugimoto 2014, 2015; Gingras 2016, Muller, 2018, Benedictus et al, 2016; Stephan et al. 2017; Zitt 2015). Criticism extends to the spread of the “Publish or Perish culture” (Fanelli 2020), or the proliferation of “desktop bibliometrics” (Katz and Hicks, 1997) for quick, poorly processed and un-validated data on research results (cf. also Glänzel and Schoepflin, 1994) identified as early as three decades ago, and finally the manipulability of the indicators used (Biagioli and Lippman, 2020). Further criticism has been expressed in various declarations and manifestos, which have seen a great proliferation in recent years and which list principles that should be followed in the evaluation of research (cf. Curry et al., 2020, who

compiled fifteen movements that have influenced current debates on responsible research assessment).

More recently, there was another initiative at the European level (European Commission, 2021) calling for a rethinking of current evaluation systems in Europe, citing several existing manifestos on this subject as a basis for reforming the system. In July 2022, an “Agreement on Reforming Research Assessment” (CoARA, 2022) was formulated reporting the main principles introduced in the scoping paper of the European Commission (2021).

In response to these initiatives, while pointing to some weaknesses in their concepts, we argue that indicators themselves should not be condemned and outright rejected since, they are only tools and, at least most of those, are the result of longstanding scientific research conducted by experts in the fields of scientometrics and information science. It is important to point to, to criticise and to prevent their “inappropriate” use. Inappropriate use also implies the application in contexts for which the indicators do not have originally been designed for (Glänzel, 2006). Therefore, it is essential to analyse the suitability of indicators with respect to the specific evaluation problem. To carry out a deep analysis of this, we must be able to have a framework that brings together, combining all aspects regarding such indicators in a proper way, i.e., the main dimensions that need to be taken into account in assessing the appropriateness of the use of the specific indicator, in the evaluation problem under consideration.

The main objective of this paper is, therefore, the use of bibliometric indicators in an evaluative context taking account of recent changes in the concepts and perception of scholarly and non-academic impacts of scientific research. In particular, we sketch a general framework of criteria that should be considered in the use of bibliometric indicators in order to ascertain whether their use, in the specific evaluation context, is appropriate for its intended purpose or not. When peer review is not feasible, does not work, or is inappropriate for the evaluation context, what should be used for research evaluation? Are bibliometric indicators always inappropriate, or should they be used with care and skill, with respect to the evaluative problem under consideration? In this paper, we caution against *throwing out the baby with the bath water* and support the idea that bibliometric indicators, even classical ones such as number of publications and citations received from them, if used in an “appropriate” way can still be extremely useful for research evaluation.

To come up with a constructive criticism of the summary rejection and condemnation of quantitative methods in research evaluation, we will identify some dimensions of assessment that would allow for a more *differentiated* approach, i.e., allow the use of indicators, where this is useful and where their use could facilitate the evaluation process. We will also point to the limitations of their use in the assessment process. At the same time, we will stress that the so-called qualitative methods are not above reproach under all conditions either. In order to do so, we will proceed from some principles regarding the appropriate use of indicators expressed in the EU Scoping Report “*Towards a reform of the research assessment system*”, of November 2021 which states: “The aim is for research and researchers to be evaluated based on their intrinsic merits and performance rather than on the number of publications and where these are published, promoting qualitative judgement with peer-review, supported by a more responsible use of quantitative indicators. The way in which the system is reformed should be *appropriate for each type of assessment* such as research projects, researchers, research units, and research institutions. A reformed system should also be sufficiently flexible to accommodate the diversity of countries, disciplines, research cultures, research maturity levels, the specific missions of institutions, and career paths” (EU 2021, p. 3).

The development of bibliometric methodology and the corresponding tools, the availability and variety of data sources for potential use in the measurement of research performance, and the

concepts and practice in research evaluation have recently undergone severe changes, related to changes in evaluation theories of scholarly activities connected to the new production of knowledge (Gibbons et al., 1994) and the change of knowledge and its interaction with the public in an age of uncertainty (Nowotny et al., 2001). Hence, the way of how knowledge is produced and disseminated has changed and the development of evaluation practices too have influenced this change (Debackere & Glänzel, 2003; Weingart, 2005; Gläser & Whitley 2007). Furthermore, Moed (2017) discusses the use of bibliometric indicators as minimum thresholds and discusses the affordable rate of errors as “*indicators may be imperfect and biased*”. The question of (random or possibly systematic) errors is actually one of the most sensitive issues in any evaluation exercise. This also includes the issue of data quality, and this affects both bibliometric data and expert or survey-based data.

In the following section, we will outline the main elements and dimensions of the proposed framework. This section will be followed by a compilation of important criteria to be met by appropriate indicators. The final section outlines challenging questions and concluding remarks.

2. Important elements and dimensions of the framework

In a quite visionary way, Henk Moed (2017) plotted a framework of an “evaluative informetrics” and its practical application. Three years later, he identified (Moed, 2020, p.4) the key questions to be addressed in the setup of a research assessment study that are: i) unit of assessment (country, institution, research group, individual, research field, international network); ii) dimension of the research process (scientific-scholarly impact, social benefit, multi-disciplinarity, participation in network), iii) purposes and objectives of the assessment (allocate funding, improve performance, increase regional engagement, budget cuts...), iv) characteristics of the units (relevant, general or systemic e.g. a national research community’s orientation towards the international research front, or phase of scientific development).

Proceeding from this, we propose a multidimensional approach for assessing the appropriateness of quantitative indicators in research evaluation by adding one dimension to those proposed by Moed. Thus, important dimensions to consider are as follows.

1. The aggregation level of the analysis

Typical aggregation levels are individual researchers, research projects, departments (micro level), institutions including colleges, universities, hospitals and companies, provinces and regions, and countries. Moed (2015) stressed that the choice of indicators depends on the aggregation level, the purpose and the aspect of the assessment. Consequently, indicators appropriate for one level are not necessarily applicable to another one.

Moed (2020, pp. 2-6) argued that “research output and impact are multi-dimensional concepts; Metrics for individual researchers suggest a *false precision*; university rankings are semi-objective and semi-multidimensional ... and social media-based indicators should at best be used as complementary measures.” He argued, many years before, that the future of research assessment lies in the intelligent combination of “advanced metrics” and “transparent peer review” (Moed, 2007). From both the conceptual-theoretical and the practical viewpoint, the weight of bibliometrics in that combination decreases with the aggregation level. Even at the lowest level (individual researchers, research groups or teams), note that expert opinion and peer review are not always independent and unbiased and may be problematic in emerging topics and interdisciplinary research. Furthermore, peer review is hardly feasible at higher levels of aggregation and not suited for benchmarking because of lacking standards and consistent common criteria. We mention that other “qualitative” but measurable components like recognition, diversity, links to other academic, non-academic activities should find their

way into evaluation, but all these issues justify the (at least supplementary) inclusion of proper quantitative metrics at lower levels as well.

2. The unit of the assessment

This is the question of what or who is to be assessed. This dimension is closely related to the first one but may differ at the same level of aggregation. This dimension also refers to the profile of the assessee and the context in which research is done. Research may require large equipment and personnel or may require different types of input and research output may be manifested in different forms and ways. Sectoral and subject peculiarities play an important part here. Allocating resources based on output measurement is a different situation from comparative performance evaluation and benchmark exercises. This dimension includes the evaluators' expectations regarding the outcomes of the assessment. The different perspectives of ex-post and ex-ante evaluation may just serve as an example. Funding organisations, government and performing organisation may also foster different expectations, which may certainly manifest themselves in the evaluation design.

3. Purpose of the assessment

This is one of the most important issues. The goal of the assessment, the question of why we are carrying out the assessment determines the method and several determinants like the time frame and the respective perspective, the reference standards and the "baseline" to be applied. Promoting scientists, recruiting personnel and proposing researchers for awards would require different perspectives, standard criteria and time frames for the evaluation – just to mention some examples.

4. The context of the assessment

This dimension refers to the context, environment and circumstances into which the evaluation task has been embedded and, as such, this is a general dimension of the framework. This also refers to the other components of the assessment, e.g. if universities are the unit of the assessment, the countries of the universities may determine the context.

5. Elements of the research process in the assessment

Moed (2017) has identified several components of the research process that need to be considered in any assessment. He distinguishes between *input*, *output*, *process* and *impact*. He already distinguished between academic (scientific-scholarly) and non-academic (e.g., societal, economic) impacts. Moed (2017) provided a (still non-exhaustive) list of various research outputs and impacts ranging from scholarly to educational, technological, economic social and cultural ones.

6. Stakeholders' engagement in research

Generally, by stakeholders we mean all interested parties and those who can influence or be influenced by research and research evaluation activities. This dimension is relevant to several aspects, including delineating the impact of research activity and assessing the desired and undesired effects or consequences of research and research evaluation.

As a matter of course, these dimensions are strongly inter-dependent and have overlapping components. Each dimension implies several possible approaches and nearly excludes others. And this has strong implications on the following topic.

3. Important criteria for the appropriate use of indicators

The Multidimensional Research assessment Matrix elaborated by the *Expert Group on Assessment of University-Based Research* (AUBR, 2010) and Moed (2017) on p. 46 ff listed several measures, methods and indicators in use for the quantification and measurement of research performance and its various impacts. These indicators already extend to the new, alternative metrics, designed to depict important aspects of non-academic outputs and impacts as well. While Moed provides concrete indications and specific applications in respect of such indicators, including a discussion of their strength and limitations, the AUBR matrix discusses methods in more general terms while also providing indications of purpose and possible use. Beyond any doubt, these compilations provide a useful framework and overview of the contexts in which (bibliometric and other) indicators can be used in research assessment, but the question of how to build and apply appropriate and reliable metrics in an evaluative context cannot be answered by selecting a number of potential indicators. Even a properly designed and theoretically sound metric could harm and cause damage if applied in inappropriate contexts. So the title of this section should actually read as “designing reliable metrics and using those in an appropriate way”.

The combination of qualitative and quantitative assessment tools requires the integration and harmonisation of differently structured information and data. In this context, we wish to recall the sketch of a standardised approach for this (see Figure 1) according to Daraio and Glänzel (2016).

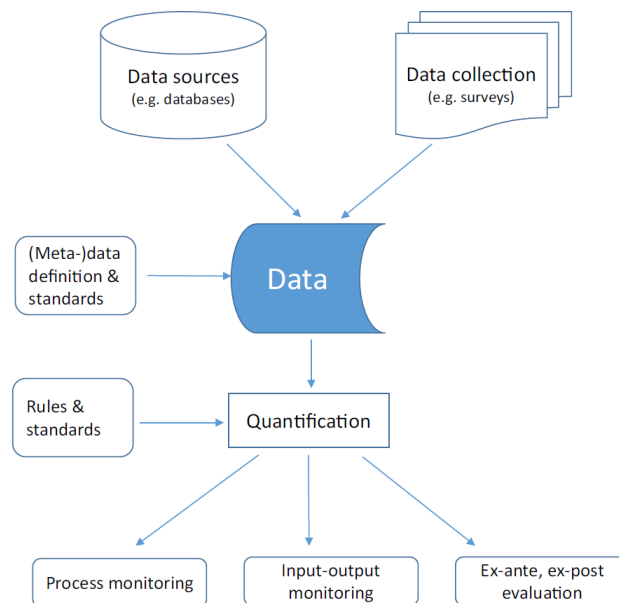


Figure 1. Sketch of data integration in use for different purposes with interference points for standardisation. Source: Daraio and Glänzel (2016, p.392)

Independently of their sources, *quantifiable data* must meet several basic demands for possible bibliometric use, particularly:

- Indicators can only be as good as the underlying data sources allow (“quality requirement”).
- The data sources must provide a basis for comparison and benchmarking exercises (“commensurability”).
- The data sources must allow replicability of results under the same conditions of data retrieval and collection at any time (“validatability”).

Bookstein (1997) identified three (out of other) demons to measurement that “frustrate our efforts to gain understanding by empirical investigation” and are challenges to quantitative approaches. In particular, he noted randomness, fuzziness and ambiguity. This applies to both concepts and methodology, including the indicator design. In addition, *indicators* that are applied to research assessment must meet several important preconditions, as they must be

- valid (measures what it is designed for and intended to measure)
- meaningful (significance of measurement)
- reliable (statistically and in terms of coverage and availability of underlying data)
- robust (insensitive to negligible changes in the system)
- normalisable (some indicators do not meet this criterion, cf. h-index)
- standardisable (for reasons of comparability and replicability).

Once reliable indicators have been built, the use of responsible metrics requires the choice of appropriate indicators according to the dimensions outlined above. As the dimension responds to the *conceptual framework*, the definition and choice of corresponding metrics is, if properly done, given by the parameters of this framework. This implies the “sample size” based on the unit of assessment, the disciplinary peculiarities in scholarly and scientific communication, e.g. in terms of required infrastructure, specific outputs, publication and citation behaviour, etc. Consequently, a framework is needed to orient users of indicators in using them “appropriately”, choosing those in line with their “fitness for purpose” and deciding what level of errors one can afford depending on the evaluation problem and context. The use needs to be accompanied by strict guidelines for interpretation, including the discussion of caveats and limitations arising from *methodological pitfalls* and *data quality issues*. The assessment may relate to the (post and ante) evaluation of research projects and results, or to the complete process (ranging from project planning over funding, conducting research, communication of results and findings to the implementation of research output and its transformation to science, technology, economy and society). Therefore, there is a need to balance between quantitative and qualitative methods depending on this very purpose and its requirements. The integration of “qualitative” components (like reward, recognition, diversity, and links to other academic, and non-academic activities) is hereby of paramount importance. On the metric part, composite measures should be avoided mainly because of the *time-variant nature of the underlying sources*, the *possible interdependence of components* the *arbitrariness of weights* chosen for the components, which may result in different outcomes when even slightly altered.

4. A preliminary outline of our framework

Figure 2 outlines the main dimensions of our framework represented by an optical prism. The three basic dimensions of our framework, the basis of our prism in Figure 3, from which to begin are: the unit to be evaluated (*whom we are assessing*), the research process to be evaluated considering its boundaries (*what we are assessing*), and the main goal of the assessment (*why we are doing the assessment*). We then have two important dimensions that allow us to specify *where*, *when*, and most importantly, *how* the assessment is carried out. They are the level of aggregation and the context of the evaluation, which constitute the two sides of our framework. Finally, we have the dimension that completes our framework represented by all *stakeholders* interested in the evaluation and its impacts and effects (consequences). Our framework aims to apply some kind of spectral decomposition of the complex assessment task represented by light entering the prism. If it works correctly, the prism should provide a proper evaluation spectrum for the unit under study.

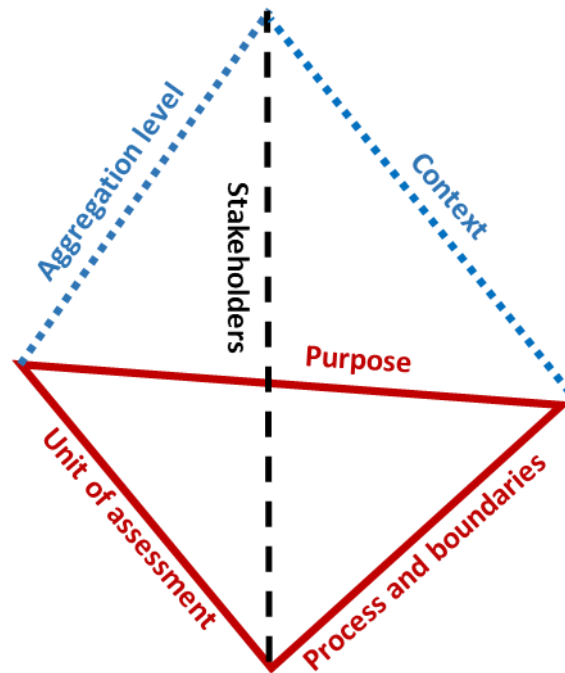


Figure 2. Main dimensions of our Prismatic Framework

5. Discussion and conclusions

A framework is needed to orient users of indicators in appropriately using them, depending on the evaluation problem and context. Our framework may be helpful to

- The choice of tools and indicators according to their “fitness for purpose”, and
- the decision of what margin of errors one could afford in the current evaluation exercise.

In this context, the framework sketched in this paper should be used to build a *checklist* for possible indicators to be used for the assessment task in the light of its dimensions and all known parameters, work to be done on a list of several indicators.

The respective bibliometric application needs to be accompanied by strict guidelines for interpretation, including the discussion of caveats and limitations arising from methodological pitfalls, data quality issues, and the choice of chosen parameters (Robinson et al., 2024).

The shift towards evaluating research through a broader lens, as suggested by the “beyond publications, beyond citations” motto, raises crucial questions about what manifestations of research outputs and their measurements might fairly capture a researcher’s scholarly and non-academic impacts. If we move beyond traditional metrics, such as journal publications and citation counts, what alternative research outputs could be considered valuable, and how might new evaluation criteria affect the academic landscape? One significant challenge is ensuring these new indicators do not inadvertently replicate the problems associated with citation-based evaluations. For instance, would introducing alternative forms of research output lead to the same proliferation issues seen with publications? The current oversaturation of published academic papers – many of which are little or not at all noticed and read – demonstrates a system where quantity is often set over quality. Could new metrics similarly encourage researchers to upload large quantities of work to achieve visibility without substantive impact? To address this, limiting the number of publications researchers can submit for evaluation has been suggested as a solution to the issue of quantity over quality. However, such limits must be implemented carefully to avoid disadvantaging early-career researchers or those in fields where

frequent publishing is essential for knowledge dissemination. Is a cap on publications feasible, and would it effectively curb the volume of research output, or might it simply shift focus to other potentially exploitable metrics?

Transparency of tools and methods and reproducibility of results furthermore form necessary requirements, but these can be met, if the framework is properly implemented.

In this context, frequent interaction with client, user, scientists, and stakeholders (if applicable) can support and enhance the meaningfulness of the results, their accuracy and the level of error margins.

To ensure meaningfulness and to meet the complexity of the evaluation tasks, well-defined bundles of valid indicators should be applied. Bibliometricians should, however, resist the temptation to build composite indicators, such as linear combination of indicators using arbitrary weights in combining possibly interdependent measures, as those indicators may cause unpredictable and inestimable error rates, force multi-dimensionality into linearity with the inevitable loss of information, transparency and reproducibility of results. Due to these properties, composite indicators may contradict the main goals of the above framework and should, therefore, be used with utmost care.

The concept of “narrative bibliometrics”, as described in recent literature (Torres-Salinas et al., 2024), proposes a more nuanced approach to interpreting research impact by focusing on contextualized storytelling around citation patterns. Could this method help address the limitations of traditional bibliometric data, which often fails to capture the true influence of a work? Narrative bibliometrics presents an opportunity to interpret data within a meaningful framework, but integrating this approach into current systems may prove difficult, particularly given the rapidly evolving landscape of Artificial-Intelligence (AI) aided tools that are reshaping data analysis and interpretation.

However, while this approach might address certain limitations of traditional metrics, it also brings significant challenges. One key concern is that the inherent *objectivity* associated with bibliometric analysis could be compromised by the *subjective* nature of narrative interpretation. By embedding a narrative, evaluators introduce elements that could vary widely depending on personal perspectives, which risks may undermine neutrality and objectivity that bibliometrics traditionally aims to uphold.

As already stressed earlier, fair assessment should be based on an intelligent combination of “advanced metrics” and “transparent peer review” (Moed, 2007). We have discussed several requirements and conditions to be met by advanced metrics. On *the qualitative part*, similar strong requirements and criteria apply as in the metric component. Transparency and reproducibility often remain a challenge for qualitative assessments, and a certain extent of subjectivity can never be avoided either. In practice, biases in terms of arbitrariness, fuzziness and ambiguity experienced in quantitative measurement (cf. Bookstein, 1997) may emerge from traditional qualitative assessment too. Therefore, we conclude, at the bottom line that the sound balance between quantitative and qualitative methods depends on the above-mentioned dimensions and the integration of the corresponding *qualitative* components (e.g., reward, recognition, diversity, links to other academic and non-academic activities) adding supplementary and otherwise not or hardly quantifiable information is hereby of paramount importance.

Finally, the potential role of AI in enhancing bibliometric assessments offers both promise and caution. AI could streamline complex data analysis and highlight meaningful research

contributions more effectively, but there is also the risk of over-reliance on automated systems, potentially resulting in biases or oversights in evaluation. In summary, rethinking evaluation systems to include diverse research outputs and innovative metrics is a promising yet complex task. Balancing the desire for comprehensive assessments with the risk of creating new, unintended issues requires a careful, collaborative approach from academic institutions, researchers, and funding bodies.

Open science practices

The present study is an opinion paper discussing the role and correct use of quantitative indicators in research assessment exercises. The authors have not used any data for this.

Author contributions

All authors contributed equally to the paper.

Competing interests

Authors declare that they have no competing interests.

Acknowledgement

Figure 1 has been reproduced from Daraio & Glänzel (2016) with permission by the publisher.

References

- AUBR (2010), *Assessing Europe's University-Based Research – Expert Group on Assessment of University-Based Research*. (2010). Research Policy. European Commission. doi:10.2777/80193
- Biagioli M, Lippman A eds. (2020), *Gaming the metrics: Misconduct and manipulation in academic research*, MIT Press.
- Bookstein A. (1997), Informetric distributions. III. Ambiguity and randomness. *JASIS*, 48(1), 2–10.
- CoARA (2022), *Coalition for Advancing Research Assessment*. (accessible at: https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf)
- Cronin, B., Sugimoto, C. R. (2014). *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*. Cambridge, MA: MIT Press. DOI: <https://doi.org/10.7551/mitpress/9445.001.0001>.
- Cronin, B., Sugimoto, C. R. (Eds.). (2015). *Scholarly metrics under the microscope: from citation analysis to academic auditing*. Medford: Information Today.
- Curry, S., de Rijcke, S., Hatch, A. et al. (2020), *The changing role of funders in responsible research assessment: progress, obstacles and the way ahead*. Working Paper. Research on Research Institute (RoRI) <https://doi.org/10.6084/m9.figshare.13227914.v1>
- Daraio, C., Glänzel, W. (2016). Grand challenges in data integration—State of the art and future perspectives: An introduction. *Scientometrics*, 108(1), 391-400.
- Daraio, C., Glänzel, W. (2020). Selected essays of Henk F. Moed. *Evaluative Informetrics: The Art of Metrics-Based Research Assessment: Festschrift in Honour of Henk F. Moed*, 15-67.
- EU (2021), “*Towards a reform of the research assessment system*”, EU Scoping Report. ISBN 978-92-76-43463-4. Accessible at: <https://op.europa.eu/en/publication-detail/-/publication/36ebb96c-50c5-11ec-91ac-01aa75ed71a1/language-en>.
- Fanelli, D. (2020). Pressures to publish: What effects do we see? In M. Biagioli and A. Lippman (Eds.), *Gaming the metrics: Misconduct and manipulation in academic research* (pp. 111–122).
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. Los Angeles, CA: Sage.
- Gingras, Y. (2016). *Bibliometrics and research evaluation: Uses and abuses*. MIT Press.

- Glänzel, W. (2006), *The perspective shift in bibliometrics and its consequences*. Accessible at <https://de.slideshare.net/inscit2006/the-perspective-shift-in-bibliometrics-and-its-consequences>
- Glänzel, W., Debackere, K. (2003), *On the opportunities and limitations in using bibliometric indicators in a policy relevant context*, In: R. Ball (Ed.), *Bibliometric Analysis in Science and Research: Applications, Benefits and Limitations*, Forschungszentrum Jülich (Germany), 225–236.
- Glänzel, W., Schoepflin, U. (1994), Little scientometrics, big scientometrics ... and beyond? *Scientometrics*, 30(2–3), 375–384.
- Gläser, J., Whitley, R. (Eds.). (2007). *The changing governance of the sciences: The advent of research evaluation systems*. Germany: Springer.
- Katz, J. S. A. H., Hicks, D. (1997). Desktop scientometrics. *Scientometrics*, 38(1), 141-153.
- Moed, H. F. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy*, 34(8), 575-583.
- Moed, H. F., Halevi, G. (2015). Multidimensional assessment of scholarly research impact. *Journal of the Association for Information Science and Technology*, 66(10), 1988-2002.
- Moed, H. F. (2017). *Applied evaluative informetrics*. Berlin: Springer International Publishing.
- Moed, H. F. (2020). Appropriate use of metrics in research assessment of autonomous academic institutions. *Scholarly assessment reports*, 2(1): 1. DOI: <https://doi.org/10.29024/sar.8>
- Nowotny, H., Scott, P., Gibbons, M. (2001). *Re-thinking science: Knowledge and the public in an age of uncertainty* (p. 12). Cambridge: Polity.
- Robinson-Garcia, N., Vargas-Quesada, B., Torres-Salinas, D., Chinchilla-Rodríguez, Z., & Gorraiz, J. (2024). Errors of measurement in scientometrics: classification schemes and document types in citation and publication rankings. *Scientometrics*, 1-21.
- Stephan, P., Veugelers, R., Wang, J. (2017). Blinkered by bibliometrics. *Nature*, 544, 411–412. DOI: <https://doi.org/10.1038/544411a>
- Torres-Salinas, D., Orduña-Malea, E., Delgado-Vázquez, Á., Gorraiz, J., & Arroyo-Machado, W. (2024). Foundations of Narrative Bibliometrics. *Journal of Informetrics*, 18(3), 101546.
- Weingart, P. (2005), Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62(1), 117–131
- Wilsdon, J., et al. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. DOI: <https://doi.org/10.4135/9781473978782>
- Wouters, P., Glänzel, W., Gläser, J., Rafols, I. (2013). The dilemmas of performance indicators of individual researchers—An urgent debate in bibliometrics. *ISSI Newsletter*, 9(3), 48-53.
- Zitt, M. (2015). The excesses of research evaluation: The proper use of bibliometrics. *Journal of the Association for Information Science and Technology*, 66(10), 2171–2176.