Scientometrics (2020) 124:1411–1419
https://doi.org/10.1007/s11192-020-03516-3

The implicit preference of bibliometrics for basic research

Paul Donner[1] · **Ulrich Schmoch**[2]

Received: 8 January 2020 / Published online: 25 May 2020

Abstract
By individually associating articles to basic or applied research, it is shown that basic articles are cited more frequently than applied ones. Dividing the subject categories of the Web of Science into a basic and an applied part, the mean field-normalization rate is referred to the applied or basic part depending on the research orientation of the paper analysed. By this approach, a distinct difference of the citations for the applied and basic parts of most subject categories is found. However, differences of the citation scores of applied and basic research organisations are found as well, but are less clear. The explanation is that applied and basic research organisations generally publish a mix of basic and applied articles. In consequence, the standard normalization without distinction of basic and applied papers is generally sufficient for the bibliometric assessment of research organisations.

Keywords Basic orientation · Applied orientation · Preference for basic research · Normalization by research level

Introduction

Bibliometrics is the quantitative analysis of scientific publications. In addition to simple publication numbers, it also uses citations as a measure of impact. In practice, citations are also used as a proxy for research quality of research institutions, e.g., in the Research Excellence Framework (REF) in the United Kingdom.

To analyse citations appropriately, many scholars have developed a set of standards, in particular the CWTS group in Leiden (De Bruin et al. 1993; Moed et al. 1995; Braun and Glänzel 1990; Schubert and Braun 1986; Vinkler 1986). One of the most popular is the so-called "crown indicator". This indicator is defined as:

$$CI = CPP/FCSm$$

✉ Ulrich Schmoch
us@isi.fhg.de

Paul Donner
Donner@dzhw.eu

[1] Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Berlin, Germany

[2] Fraunhofer ISI, Karlsruhe, Germany

where CPP is the mean citation rate of the papers (CPP) of a research unit under study, and FCSm is the mean citation rate of the scientific fields into which these papers were classified. In practice, these fields are typically the subject categories of the Web of Science (WoS) or the subject areas of the Scopus All Science Journals Classification (ASJC), both of which are assigned at the level of whole journals.

This measure takes into account that the citation scores often differ considerably by scientific field, e.g. biotechnology compared to mathematics, which makes the citations of publications in different fields incomparable. Therefore, the mean citation rates of a research unit are normalized by the mean citation rates of their respective fields. This approach is able to solve a major methodological problem of citation analysis.

Recently, there has been some controversy concerning how to calculate field normalization appropriately (Opthof and Leydesdorff 2010; Van Raan et al. 2010; Lundberg 2007; Waltman et al. 2011a, b). However, the basic concept of normalizing the citation rates of a research unit by relating them to the typical citation rates of a reference set of papers from the same discipline (and publication year) is not contested. In particular, CWTS suggested a new way of calculating the crown indicator (Waltman et al. 2011a, b). In this paper, the new version of calculation is used.

Underlying the concept of field normalization is the implicit concept that a higher citation rate reflects a higher impact or a higher performance, but that the different levels of citations by field have to be taken into account. The type (articles, letters, notes, proceedings, reviews) and year of publication are seen as the only additional relevant factors that may influence the citation score.

However, even at the very beginning of bibliometrics, some authors already noted that papers in basic research are cited more frequently than those in applied research (Garfield 1979; Lindsey 1978). The reason for this phenomenon is evident: Articles in basic research are relevant for many scientists working in the respective discipline and often even outside that discipline as well, e.g. basic findings in physics may be interesting for electrical engineering. Articles in applied research have a more limited audience, i.e. to researchers working on similar applications. Using the standard bibliometric indicators implies a preference for basic research. In fact, Opthof (2011) and van Eck et al. (2013) found such citation impact differences within medical fields between basic and clinical research topics. Research units with a strong orientation towards basic research obtain higher citation scores than those with a more applied orientation. There is no epistemic justification for rating basic research higher than applied research. In addition, applied research often involves very complex topics at a very high level (see, e.g. Schmoch et al. 2019). Experienced bibliometricians are aware of the different citation levels of applied and basic articles, but most users of bibliometric assessments do not know that these differences exist and assume that highly cited publications have a high quality independent of their applied or basic orientation.

Narin et al. (1976) developped a classification of medical journals into 4 research levels and applied it to about 900 biomedical journals. This scheme was expanded to physical sciences by . The number of journals classified by research level was enlarged to more than 4000 journals, but still covers only a part of all journals in the WoS or Scopus. This classification was used to analyse different aspects such as collaboration or link to industry, but not as to the citation level. A major shortcoming is that the approach is refers to a limited set of journals.

Boyack et al. (2014) developed on this basis an approach to automatically classify basic and applied articles based on characteristics of publications, i.e., all articles in databases

Table 1 Research levels suggested by Narin et al. (1976) and Carpenter et al. (1988) and used by Boyack et al. (2014)

Research level	Biomedical definition	Non-biomedical definiton
1	Clinical observation	Applied technology
2	Clinical mix	Engineering-technological mix
3	Applied research	Applied research
4	Basic research	Basic scientific research

such as WoS or Scopus are individually classified by research level, all articles are covered. The authors used Narin's four research levels, shown in Table 1. A disadvantage of this classification is the unclear definition of the levels 1 to 3. E. g., the level 2 "mix" may be useful for journals encompassing articles of different orientation, but for individual articles as in the approach of Boyack, this category is less useful. Furthermore, the difference between applied technology and applied research is vague. In addition, the citation scores of the levels 1 to 3 using the Boyack approach prove to be quite similar. Therefore, we reduce the scheme to the two levels "applied" (levels 1 to 3) and "basic" (level 4). This method makes it possible to distinguish basic and applied articles within fields and thus to introduce normalization that distinguishes applied and basic publications.

A possible objection against this simplification could be that, e.g., the distribution of levels in WoS is largely equivalent, i.e., each level is linked to about one quarter of the articles. This is primarily due to articles in medicine which represent a large part of WoS and where the share of basic research (level 4) is only 10%. Most papers are oriented on clinical issues. Also in engineering, the share of basic research is 9%. But in most other areas, the share of basic and applied research (levels 1 to 3) is largely equivalent. E. g., in chemistry the share of basic research is 42%, in physics 37%. In addition, the publications with level 3, applied research, dominate within the levels 1 to 3 in fields of natural sciences. Thus, the aggregation of the levels 1 to 3 to one unit "applied research" appears to be justified.

In this article, we explore in more detail whether the distinction between basic and applied research has a relevant impact on citation scores, in particular with regard to the analysis of research institutions.

Distribution of basic and applied research by subject categories

To some extent, the Web of Science's subject categories already consider the basic-applied distinction, as there are categories for applied microbiology, chemistry, mathematics, physics and psychology. However, there are no categories explicitly for basic research.

To assign articles in WoS to basic and applied research, we used the published version of Boyack and colleagues' model to classify all WoS records of papers automatically using their titles and abstracts. This paper explores whether this distinction has a relevant impact on citation scores. We assume that the classification model of Boyack et al. (2014) is sufficiently accurate, at least on the level of larger sets of publications, but we have not yet independently verified its performance. For this alternative analysis, we use in the formula for the field normalized citation rate for FSMm, not the mean citation rate for the subject category in total, but we divide each subject category in an applied and basic part, thus we double the number of scientific fields, e.g., we divide the field "Optics" into "Applied optics" and "Optics, basic research". With the standard normalisation without division of

Table 2 Average citation counts in selected WoS categories (articles, 2005–2015). *Source*: WoS, own search

Web of Science subject category	Applied		Basic		Total	
	Avg. cit.	Pub.	Avg. cit.	Pub.	Avg. cit.	Pub.
Cardiac and cardiovascular systems	7.1	138,729	9.1	8,576	7.2	147,305
Chemistry, multidisciplinary	8.2	183,812	9.5	223,511	8.9	407,308
Chemistry, physical	8.5	261,363	7.7	195,227	8.2	456,588
Clinical neurology	5.8	183,212	7.5	18,956	5.9	202,168
Materials science, multidisciplinary	6.0	477,704	9.1	121,024	6.6	598,724
Nanoscience and nanotechnology	9.0	149,758	11.7	50,638	9.7	200,391
Optics	4.8	163,959	5.3	55,937	4.9	219,894
Physics, condensed matter	6.2	152,543	6.6	118,386	6.4	270,925
Surgery	4.3	279,772	3.9	2,549	4.3	282,320

the subject categories, we have a mix of applied and basic articles implying an implicit preference for basic articles in a citation analysis. The division of subject categories is the only realistic approach of a clear division between applied and basic. A clear division between applied and basic journals is not possible, as many journals include both types of papers such as the journals Physical Letters or Electrical Engineering.

As a first test, we looked at nine WoS subject categories and selected the three medical categories studied by Van Eck et al. (2013), and six categories in which the German research organisations Fraunhofer and Max-Planck, which are analysed in more detail below, have substantial numbers of publications. We included articles published between 2005 and 2015. We divided each category into applied and basic publications, and calculated the referring citation scores with a three-year citation window. The results are presented in Table 2. For seven of the nine cases, we find higher citation scores for basic articles than for applied ones. In most cases, this difference is 20% or more. A lower value of 6% is only found in "Physics, condensed matter", where it is difficult to distinguish basic and applied research.

In "Physical chemistry" and "Surgery", the citation scores for applied articles are higher than for basic ones. Physical chemistry is generally a basic field, but there is a close relation to material sciences, in particular new materials, e.g. the American Chemical Society says:

"Physical chemistry is the study of how matter behaves on a molecular and atomic level and how chemical reactions occur. Based on their analyses, physical chemists may develop new theories, such as how complex structures are formed. Physical chemists often work closely with materials scientists to research and develop potential uses for new materials."[1]

A closer analysis shows that about 40% of all papers in WoS classified in physical chemistry deal with nanotechnology which generally achieves high citation scores. It can be assumed that the focus on new materials implies a higher average citation score of applied than that of basic papers.

[1] https://www.acs.org/content/acs/en/careers/college-to-career/areas-of-chemistry/physical-chemistry.html.

Table 3 Crown Index (mean field-normalized citation rate) for selected German research organizations. *Source*: WoS, own searches

	Fraunhofer Society	Max Planck Society	Universities
Without research level normalization			
2005	1.06	1.69	1.19
2010	1.12	1.75	1.27
2015	1.38	1.70	1.25
With research level normalization			
2005	1.06	1.66	1.18
2010	1.14	1.70	1.26
2015	1.42	1.64	1.25

Surgery, on the other hand, is a primarily applied field. Only 1% of the papers are associated to basic research, thus a much lower share than in medicine in geneal and in consequence, the number of applied articles is much higher than that of basic ones. In this specific case, the applied articles may be more interesting for the relevant community.

To sum up, basic articles are generally cited more often than applied ones and the difference is relevant with 20% and more.

Comparison of a basic and an applied research organization in the same subject category

In Germany, there are two major non-university research organizations with very different missions. The institutes of the Max Planck Society (MPG) "conduct basic research in the natural sciences, life sciences, and humanities ".[2] The Fraunhofer Society (FhG), on the other hand, is an application-oriented research organization. An important part of its work is contract research for private industry, which accounts for about one third of its revenue.[3] Applying normalization by research level to compare these two organizations, we expect the scores for the Max Planck Society to decrease and those for the Fraunhofer Society to increase.

As a general finding, the Fraunhofer values do improve slightly, whereas the Max Planck ones decrease, but the Fraunhofer values remain below the Max Planck ones, see Table 3. The major reason for the decline in the difference between FhG and MPG scores is the improvement of the Fraunhofer values between 2005 and 2015, while the MPG values remain stable over time.

Looking at German universities as a reference, we find only minor changes from introducing normalization by research level. The reason is that universities conduct both applied and basic research to a similar degree, so that the higher values of applied publications are compensated by the lower values of basic publications.

[2] https://www.mpg.de/11761628/profile-visions. accessed 19 Oct. 2019.

[3] https://www.fraunhofer.de/en/about-fraunhofer/profile-structure/facts-and-figures/finances/contract-research-revenue.html. accessed 19 Oct. 2019.

Table 4 Crown Index (mean field-normalized citation rate) for selected German research institutes, 2015. *Source*: WoS, own searches

	Standard	With research level normalization
Winners		
Fraunhofer-Institut für Solare Energiesysteme (ISE)	2.39	2.52
Europäische Atomgemeinschaft	1.72	1.81
Deutsches Konsortium für Translationale Krebsforschung (DKTK)	2.49	2.56
INM—Leibniz-Institut für Neue Materialien gGmbH	1.87	1.94
Robert-Bosch-Krankenhaus	1.76	1.82
Losers		
Max-Planck-Institut für molekulare Zellbiologie und Genetik	2.07	1.93
Max-Planck-Institut für Pflanzenzüchtungsforschung	1.92	1.77
Max-Planck-Institut für Physik (Werner-Heisenberg-Institut)	2.57	2.42
Max-Planck-Institut für Entwicklungsbiologie	2.08	1.92
European Molecular Biology Laboratory (EMBL)	4.96	4.38

The considerations for Max-Planck, Fraunhofer and universities can also be applied on the level of research institutes. An arbitrary selection of German research institutes leads to the Crown Indices shown in Table 4.

Again, more applied research institutes obtain higher indices with research level normalization. The values increase by between 3.4 and 5.4%. In particular, Fraunhofer ISE is assigned better values. The indices of the more basic research institutes decrease by between 5.8 and 11.7%.

Thus, the differences between the standard indices and the indices with research level normalization are higher for research institutes than for whole research organizations, but they are still modest. However, the changes for some specific institutes are limited but substantial. The reason for this difference between institutes and organizations, or between large and smaller units, is that the spectrum of publications is broader and less distinctly oriented towards applied or basic research for organizations or larger units.

In the case of the Max Planck Society and the Fraunhofer Society, the comparison between the two is highly artificial as their fields of activity are quite different, as documented in Table 5.

The Fraunhofer Society focuses on applied categories such as "Electrical Engineering", "Telecommunications" or "Computer Science"; the Max-Planck Society concentrates on basic categories such as "Astronomy", "Physical Chemistry" or "Particles Physics". Thus, the indices in Table 3 mean that Fraunhofer and Max-Planck are both above the worldwide averages in their specific fields, and that Max-Planck achieves relatively higher indices within its specific communities.

A more meaningful comparison can be made for the activities in "Materials Science, Multidisciplinary", where both organizations have relevant activities. This comparison reveals that the indices for Fraunhofer increase and those for Max Planck decrease by applying research level normalization. However, the differences are so moderate that the ranking does not change, see Table 6. The higher position of Fraunhofer in 2015 is the result of generally higher citation rates and not of the research level normalization.

Table 5 Publications of the Fraunhofer Society and the Max Planck Society in the Web of Science in the 15 most relevant categories, 2015 (shares in percent of all publications). *Source*: WoS, own searches

Fraunhofer society	Share	Max-planck society	Share
Engineering electrical electronic	26.6	Astronomy astrophysics	14.6
Physics applied	15.8	Materials science multidisciplinary	8.0
Optics	13.5	Multidisciplinary sciences	6.8
Materials science multidisciplinary	11.8	Physics applied	6.7
Energy fuels	7.9	Chemistry physical	6.0
Telecommunications	6.8	Chemistry multidisciplinary	5.8
Computer science theory methods	6.2	Biochemistry molecular biology	5.8
Computer science information systems	5.8	Neurosciences	5.2
Computer science artificial intelligence	3.9	Physics multidisciplinary	5.2
Nanoscience nanotechnology	3.9	Physics particles fields	4.1
Computer science software engineering	3.4	Physics condensed matter	4.0
Chemistry physical	3.3	Cell biology	3.9
Computer science interdisciplinary applications	3.3	Physics atomic molecular chemical	3.3
Engineering industrial	3.0	Optics	3.2
Physics condensed matter	3.0	Physics fluids plasmas	2.9

Table 6 Crown Index (mean field-normalized citation rate) of the Max Planck and the Fraunhofer society for the category "Material Science, Multidisciplinary". *Source*: WoS, own searches

	Max Planck society	Fraunhofer society
Without research level normalization		
2013	1.25	1.53
2014	1.20	1.65
2015	2.06	1.62
With research level normalization		
2013	1.34	1.46
2014	1.29	1.57
2015	2.19	1.54

This outcome may not be in line with the expectations, as Fraunhofer and Max Planck have clearly different missions. However, a more detailed analysis reveals that, e.g. in 2015, most publications of Fraunhofer were applied, but 11% were basic. Max Planck has a higher share of basic publications with 36%, but also a substantial share of applied publications with 64%. The basic orientation of these organizations does not mean that all their publications are either basic or applied. Rather, there is a mix of both types with different structures. This mix explains why the differences between the standard indices and the research level normalized indices for organizations and also institutes are smaller than might be assumed.

Nevertheless, normalization by research level does provide instructive insights into the structure of subject categories and relevant differences for some organizations and institutes. Therefore, it may be interesting to include the research level of publications in the Web of Science or in Scopus. Boyack et al. have shared the codes of their approach

as a Python program on the software platform Github. The program is a multinomial logistic regression classifier, trained to classify articles using four research levels based on the titles and abstracts.

A possible alternative is to use journal-normalized instead of field-normalized citation rates, as suggested by Grupp et al. (2001). As most journals have a clear orientation towards either applied or basic research, the indices increase for applied organizations in a similar way as by applying research level normalisation. This approach may be easier to implement in citation analysis than the research level normalization.

Conclusions

Assigning the research levels "basic" or "applied" to the features of individual articles makes it possible to analyse whether applied or basic articles achieve different citation rates on average. The analysis of subject categories in the WoS and of different types of research organizations and institutes reveals that organizations' basic publications have higher citation rates than applied ones within subject categories and that institutes and organizations with a distinct basic research orientation achieve higher citation rates than those with a distinctly applied one. In consequence, the normalization of citation rates using research levels leads to higher citation indices for applied organizations/institutes and lower indices for basic ones. However, the differences between normalization with and without research levels are only relevant for the different applied and basic parts of the subject categories/scientific fields, and are less pronounced for organizations/institutes. The explanation for this finding is that most organizations/institutes publish a mix of applied and basic articles, despite their clear overall orientation towards basic or applied research. Thus, a major result of the analysis is that for analysing the performance of organisations/institutes, a differentiation of the normalisation into applied and basic is not essential.

Nevertheless, the analysis by research levels can provide new instructive insights into the publication strategies of organizations/institutes, or the development of emerging fields from basic towards more applied research.

In any case, it would be useful to include the feature "research level" into publication databases such as WoS or Scopus in order to achieve citation indices that more accurately reflect a basic or applied research orientation.

Acknowledgements Certain data included in this paper are derived from the Science Citation Index Expanded (SCIE), the Social Science Citation Index (SSCI), the Arts and Humanities Citation Index (AHCI), and the Index to Social Sciences & Humanities Proceedings (ISSHP) prepared by Clarivate Analytics, Philadelphia, Pennsylvania, USA.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Boyack, K. W., Patek, M., Ungar, L. H., Yoon, P., & Klavans, R. (2014). Classification of individual articles from all of science by research level. *Journal of Informetrics, 8*(1), 1–12.

Braun, T., & Glänzel, W. (1990). United Germany: The new scientific superpower? *Scientometrics, 19*(5–6), 513–521.

Carpenter, M. P., Gibb, F., Harris, M., Irvine, J., Martin, B. R., & Narin, F. (1988). Bibliometric profiles for British academic institutions: An experiment to develop research output indicators. *Scientometrics*, *14*(3–4), 213–233.

De Bruin, R. E., Kint, A., Luwel, M., & Moed, H. F. (1993). A study of research evaluation and planning: The University of Ghent. *Research Evaluation, 3*(1), 25–41.

Garfield, E. (1979). *Citation indexing—Its theory and application in science, technology and humanities*. New York: Wiley.

Grupp, H., Schmoch, U., & Hinze, S. (2001). International alignment and scientific regard as macro-indicators for international comparisons of publications. *Scientometrics, 51*(2), 359–380.

Lindsey, D. (1978). *The scientific publication in social science*. San Francisco: Jossey-Bass.

Lundberg, J. (2007). Lifting the crown—Citation *z*-score. *Journal of Informetrics, 1*(2), 145–154.

Mittermaier, B., Holzke, C., Tunger, D., Meier, A., Glänzel, W., Thijs, B. & Chi, P.-S. (2018). Erfassung und Analyse bibliometrischer Indikatoren für den PFI-Monitoringbericht 2019. http://www.hdl.handle.net/2128/21780. Accessed Jan 07, 2020.

Moed, H. F. (2009). New developments in the use of citation analysis in research evaluation. *Archivum Immunologiae et Therapiae Experimentalis, 57*, 13–18.

Moed, H. F. (2010). CWTS crown indicator measures citation impact of a research group's publication oeuvre. *Journal of Informetrics, 4*(3), 436–438.

Moed, H. F., De Bruin, R. E., & Van Leeuwen, T. N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics, 33*(3), 381–422.

Narin, F., Pinski, G., & Gee, H. H. (1976). Structure of the biomedical literature. *Journal of the American Society for Information Science*, *27*(1), 25–45.

Opthof, T. (2011). Differences in citation frequency of clinical and basic science papers in cardiovascular research. *Medical & Biological Engineering & Computing, 49*, 613–621.

Opthof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics, 4*(3), 423–430.

Schmoch, U., Beckert, B., & Schaper-Rinkel, P. (2019). Impact assessment of a support programme of science-based emerging technologies. *Scientometrics*, *118*(3), 1141–1161.

Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics, 9*(5–6), 281–291.

Tijssen, R. J. W., Visser, M. S., & Van Leeuwen, T. N. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics, 54*(3), 381–397.

Van Eck, N. J., Waltman, L., van Raan, A. F., Klautz, R. J., & Peul, W. C. (2013). Citation analysis may severely underestimate the impact of clinical research as compared to basic research. *PLoS ONE, 8*(4), e62395.

Van Raan, A. F. J., Van Leeuwen, T. N., Visser, M. S., Van Eck, N. J., & Waltman, L. (2010). Rivals for the crown: Reply to Opthof and Leydesdorff. *Journal of Informetrics, 4*(3), 431–435.

Vinkler, P. (1986). Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics, 10*(3–4), 157–177.

Waltman, L., Jan Van Eck, N., Van Leeuwen, T. N., Visser, M. S., & Van Raan, A. F. J. (2011a). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics, 5*(1), 37–47.

Waltman, L., Van Eck, N. J., Van Leeuwen, T. N., Visser, M. S., & Van Raan, A. F. J. (2011b). Towards a new crown indicator: An empirical analysis. *Scientometrics, 87*, 467–481.