5th International Conference on Advanced Research Methods and Analytics (CARMA2023)
Universidad de Sevilla, Sevilla, 2023
DOI: http://dx.doi.org/10.4995/CARMA2023.2023.16413

Account for variation by field in publication: bibliometric databases' analysis in a Portuguese Higher Education Institution

Cátia Malheiros[1], Conceição Gomes [1], Filipa Campos [1,2], Sofia Eurico[1]
[1]CiTUR, School of Tourism and Maritime Technology, Polytechnic of Leiria, Portugal.
[2]CICF, Research Centre for Accounting and Taxation, School of Management, IPCA, Barcelos, Portugal

Abstract

This study seeks to examine the variation by field in publication practices in a Portuguese Higher Education Institution (HEI), where both research in Social Sciences and in Hard Sciences is conducted. The intention is to raise the issue of the suitability of bibliometrics for the Professors/researchers' evaluation considering their areas of research, as well as understanding the sort of use they make of these instruments. Different Bibliometric Databases were managed to analyze the use given to them by all the researchers in this HEI and to find out the main differences in its use according to the researched field of study. These results might represent a valuable source of information for HEIs in the process of finding the balance between the different procedures and format for the evaluation of researchers, to identify their in/ability to proficiently use these tools and to study the suitability of each tool to different profiles.

Keywords: *Evaluation; Higher Education; Bibliometrics; Social Sciences; Hard Sciences*

1. Introduction

In a recent past, bibliometrics have been assuming a crucial role in the evaluation process of Higher Education Institutions' professors/researchers, both in an individual basis as well as in a collective one, positioning the institution according to its information science results. However, the rapid evolution of bibliometric science and its close liaison to the evaluation of researchers does not make of the former expert users, or even interested ones in this method. Their evaluation was once a task led by peers and data are now "increasingly used to govern science" (Hicks et al., 2015:429) by Institutions, regardless the researchers' will or expertise in using these tools and the effectiveness of the service of bibliometric support research in libraries.

The concept of Bibliometrics has been used since 1969, when Pritchard (1969: 348) defined it as the "application of mathematics and statistical methods to books and other media of communication" and it dates back to the early 19th century, when the impact factor was firstly described by Eugene Garfield (1955) and when Tibor Braun launched the first dedicated journal Scientometrics in 1978 (Springer, 2023). Later on, and according to Furner, (2014:146), bibliometrics was described as being "about what people (authors, readers, etc.) do with documents (books, journal articles, web pages, tweets, etc.), for what reasons, and with what effects"

Introduction to Bibliometrics for the Evaluation of Scientific Information happens later on, on the threshold of the 21st century and reliance to its use is among much of the scientific community (Sugimoto & Larivière, 2018). The use of various research indicators should be done responsibly and ensure that these ones are not detrimental to the scientific community and that research is measured productively, supporting, rather than destroying, the scientific system (Sugimoto & Larivière, 2018). As Hicks et al. (2015:430) acknowledge and alert "Across the world, universities have become obsessed with their position in global rankings (…) even when such lists are based on what are, in our view, inaccurate data and arbitrary indicators". Moreover, the account for variation by field in publication is a concern that must be attended in order to avoid inequities and biases in the evaluation process (Nederhof, 2006).

According to the Leiden Manifesto for Research Metrics (Hicks et al., 2015) there are 10 ten principles to guide research evaluation, and the sixth one is directly linked with this topic. This study will therefore try to verify if the predominant area of study of the different researchers in a Portuguese HEI could be related to their presence and proficiency in the use of the metric databases at their disposal, leading to the following research question: What is the influence of the scientific field studied in the use of different bibliometric platforms by researchers to account for their publications?

2. Account for variation by field in publication

As clearly explained by Leiden (Hicks et al., 2015), quantitative metrics may not reflect with the same precision and justice the production of researchers from different areas of knowledge. Whether in the arts, social sciences or other areas, the studies that result from them have very specific nature and characteristics, which often do not match with publications that are plausible for bibliometric measurement. Working mainly on content, distancing themselves from quantitative instruments for the analysis of results, it is the social sciences that most resort to the use of qualitative methodologies, having often been kept away from publications in the 1st and 2nd quartiles and even recognizing a tendency to difficulty in being accepted in indexed journals for this very reason. According to College & James (2015:62), "the diverse nature of research at the institution as well as in the field should be highlighted, and appropriate denominators and indicators requested". This same idea is reinforced by Coombs & Peters (2017:8) about the Leiden manifesto, when saying that "the field normalization can be responsible for strongly influencing the result of the quantitative assessment, even more than the actual performance of the field".

The availability and willingness to use these platforms is often less among social science researchers who often publish their work in formats other than articles, which are the most easily measurable and accepted format for indexed publication. Some studies have been conducted in order to understand research output performance of social scientists as far as bibliometrics are concerned (Thanuskodi, S., 2017; Glänzel & Schoepflin,1999). For this study different bibliometric databases were considered with the purpose of gathering comprehensive research activity. Either researchers' unique digital identifier, as ORCID (Lehmann-Haupt, 2022), which allows the research to be guided by the individual, or bibliometrics databases, as Scopus (Scopus, 2023) which in turn guide research by the output results of researchers, have been used with the aim of enlarging the scope of this study.

3. Bibliometric databases' analysis in a Portuguese Higher Education Institution: methodological procedures

A Portuguese HEI which offers Bachelor's and Master's degrees in the field of Social Sciences (hereafter referred to as SSs) and that of Hard Sciences (hereafter referred to as HSs) and, therefore, having researchers from one area and the other equally, was chosen for the study. The Institution hosts two Research Units, one linked to SSs and one to HSs. The research units will be referred to as Group A (social sciences) and Group B (hard sciences).

All professors, working full time in this institution, teach and research simultaneously, and there is no separation of careers. From the 148 professors, only full members of the research units mentioned before and simultaneously working full time have been

considered, namely 35 (53%) in relation to the SSs one and 31 (47%) to the HSs one. The remain 15 professors are members of other research units not connected with the studied HEI.

Five different databases were explored during three months (December, 2022 to February 2023), namely: ORCID (ORCID, 2023), Scopus (Scopus, 2023), Web of Science (WOS) (Web of Science, 23), Dimensions (Dimensions, 2023), Google Scholar (Google Scholar, 2023) and 3 categories that were common in all the platforms: Articles in journals; Book chapters and Conference papers. Besides these 3 categories, some more were identified in ORCID and Google Scholar such as books, books edition, posters and patents. The information for each researcher was searched by name in all the mentioned databases and in ORCID. Sometimes, due to the difficulty in finding them by the name, it was necessary to add the institution or research centre to which they belong to.

Analysis and processing of data was achieved through excel and statistical package for social sciences (SPSS) software. Several descriptive statistics' measures were used such as median and variables were explored to analyse their normality and the existence of outliers through Kolmogorov-Smirnov test and Box-plot. As significance level of Kolmogorov-Smirnov test was <0.001, null hypothesis - the population is normally distributed - was rejected. Thus, as there is not normality, non-parametric tests were the option. Both research centres were compared based on publication practices using Mann-Whitney test (Pestana & Gageiro, 2014).

4. Results and discussion

4.1. Main differences in the use of database platforms according to the researched study field

By analysing the different database platforms, it is clear that the rate of publication in the SSs and HSs varies and Figure 1 shows the large discrepancy between the number of papers produced by researchers in Group A and those in Group B. It should be noted that ORCID and Google Scholar databases are those that bring together the largest number of publications by both groups. Considering that ORCID is the database that gathers the largest number of papers from both Group A and Group B, this base was selected to compare the two research groups in terms of quantity of publications.

Publication practices were compared between Group A and Group B researchers, using Mann-Whitney test. Null hypothesis (H0) was formulated: the distribution of total publications of *ORCID/Scopus/Web of Sciences/Google Scholar/Dimensions* is the same across group A and Group B. These hypothesis were rejected. Regarding these results, the difference between Group A and Group B is evident. The number of publications differs significantly between research centres, being necessary to determine which type of research

Group has the most publications in the analysed databases. Then, the median of the total publications was calculated in each Group for each database. The results show that the range of values of the median of publications of HSs is 41 to 22 and for SSs is 5 to 0. The difference between the groups was clarified, being the HSs researchers the ones with the highest number of publications.

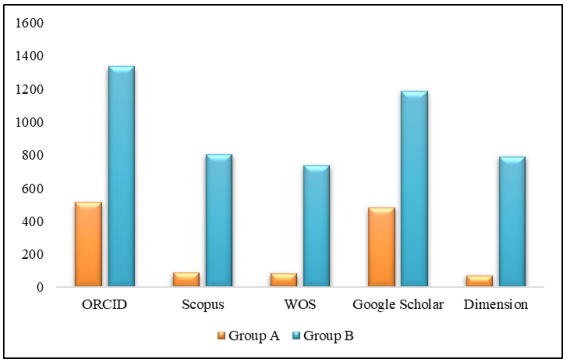

Figure 1 – Publications by database and research centre

While the normality was tested the variables were explored, and some outliers were found. This means that, in all the population, some researchers show up. They have so many publications that they become outliers, standing out from the overall pattern of the researchers considered for the study.

4.2. The variation by field regarding publication categories

Considering that ORCID is the most reliable database in terms of authorship, the typologies of documents in it were analysed in order to compare the differences in publication categories between research fields. Observing Table 1, in ORCID 72% of the documents were produced by Group B researchers and the predominance of publications in this group are articles in journals and posters. In the case of Group A, the typology with greater expression is the conference papers. Notice that there are neither patents in the case of Group A, nor edition of books in the case of Group B (Table 1).

ORCID database publication practices were compared between categories through Mann-Whitney test. Null hypothesis was formulated: the distribution of the articles in journals/books/book chapters/conference papers in ORCID is the same across the groups. For articles in journals and conferences papers this hypothesis was rejected. Moreover, the median of articles in journals highlights in Group B with the value of 26 against a value of 3 in Group A. In addition, the opposite happens in conferences papers, where Group A stands out with a median of 9 facing Group B which has a median of 2. With regard to books and book chapters H0 is not rejected, meaning that the distribution is similar between both groups.

Table 1. Publication categories in *ORCID*

ORCID	Group A	Group B	Total	Group A	Group B
Articles in journals	165	797	962	17%	83%
Books	28	40	68	3%	4%
Book chapters	67	78	145	7%	8%
Posters	3	351	354	0%	36%
Conference papers	240	54	294	25%	6%
Book's edition	10	3	13	1%	0%
Patents	0	12	12	0%	1%
Total	513	1335	1848	28%	72%

Furthermore, HSs researchers record the highest number of articles in journals. However, SSs researchers have the highest number of conference papers. These researchers have different profiles according to the field that they belong to, being undeniable the higher number of the publications of HSs' researchers.

5. Conclusions, limitations and further research

From the obtained results, it is clear that HSs' researchers present more publications than those from the SSs and that, in both areas, some publish more than others within the same group. There are also differences in the typology of documents produced by each group. Journal articles are very high in Group B and they also publish many posters and have patents. On the other hand, conference papers are higher in group A. These results corroborate the research question of the study, that inquires the influence of the scientific area studied and its relationship with publication performance in different bibliometric platforms.

As for the studied platforms, ORCID presents itself as the one with the larger number and diversity of documents and both ORCID and Google Scholar gather the greatest diversity of documents. Scopus, WOS and Dimension only consider three types of documents, disregarding other works for the indicators.

A careful reflection of the obtained results, and in the light of the theoretical framework, highlights that a proper and legitimate diversity of the different scientific areas requires a correct and dignified treatment of scientific production, regardless of its nature or format. Scientific production demands, in its essence, quality and accuracy and these should prevail over formatting standards, style and methodologies that are imposed and that, ultimately, do

not dignify the research product *per si*. The scientific production of the SSs ends up having less expression and visibility in the platforms. Their lower presence may not be an indicator of lower productivity, but perhaps of the lack of systematisation and registration of production in these instruments or the inadequacy of the parameters and formats required equally for all areas, not looking at their specificities.

Moreover, the danger of the commercialising of science by imposing practices for measuring results that do not always match the nobility, breadth and diversity of types of studies of scientific production may result from an incorrect use of these instruments. If they are not seen as an auxiliary measuring mechanism, instead of a prevailing instrument to validate scientific production, we may move towards a pathogenic culture that, according to Mendon (1942), results from an imperative logic of publication as a way of belonging to the community, and that is frequently cause for fraudulent behaviour.

As for the limitations of the present study, in fact they somehow enhance future research. For instance, in some cases, difficulty in identifying the researcher in the different databases, due to the absence of a profile or publication or even the presence of more than one profile for the some researcher, leads to the need of a future study that compares different databases and author identifiers and recognizes weaknesses and advantages among them.

This study compares 4 databases and 1 author identifier, for a singular HEI and the development of similar studies, but in a broader context, including different Portuguese HEIs and even expanding it to a worldwide context, would be advisable. Further research should also consider the implementation of new instruments for database assessment and better performance of the analysis, through more advanced artificial intelligence.

Funding: This work is financed by national funds through FCT - Foundation for Science and Technology, IP, within the scope of the reference project UIDB/04470/2020.

References

Colledge, L. & James, C. (2015). "A "basket of metrics"—the best support for understanding journal merit", *European Science Editing*, *41*(3), 61-65.

Coombs, S. K. & Peters, I. (2017). The Leiden Manifesto Under Review: What Libraries Can Learn From It. *Digital Library Perspectives*. https://libereurope.eu/wp-content/uploads/2020/11/DLP-Paper.pdf

Dimensions. (2023). *Linked research data from idea to impact*. Digital Science & Research Solutions Inc. https://www.dimensions.ai/

Furner, J. (2014). The Ethics of Evaluative Bibliometrics. In CRONIN, Blaise; SUGIMOTO, Cassidy R., eds. (2014). Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact. Cambridge: MIT, 85-107.

Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, *122*(3159), 108-111. http://www.garfield.library.upenn.edu/papers/science_v122v3159p108y1955.html

Glaènzel, W., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing and Management, 35*(1), 31-44. https://www.sciencedirect.com/science/article/pii/S0306457398000284

Google Scholar (2023). Google Scholar. https://scholar.google.com/

Hicks, D. & Wouters, P. (2015). The Leiden Manifesto for research metrics. *Nature*, 520, 429-431.

Lehmann-Haupt, J. (2022). *ORCID's Frist decade: From Startup to Sustainability*. ORCID. https://info.orcid.org/wp-content/uploads/2022/11/R2_Orcid-10th-Ann-Booklet-FOR_WEB.pdf

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities, *Scientometrics,* 66(1), 81-100.

ORCID. (2023). *ORCID – Connecting research and researchers*. https://orcid.org/

Pestana, M. & Gageiro, J. (2014), Análise de Dados Para Ciências Sociais, 6th ed., Sílabo.

Pritchard, A. (1969). Statistical Bibliography or Bibliometrics. *Journal of Documentation, 25*(4), 348-349.

Scopus (2023). *Start exploring - Discover the most reliable, relevant, up-to-date research*. Elsevier B.V. https://www.scopus.com/search/form.uri?display=basic#basic

Springer (2023). *Scientometrics - An International Journal for all Quantitative Aspects of the Science of Science, Communication in Science and Science Policy*. Springer Nature Switzerland AG. https://www.springer.com/journal/11192/editors.

Sugimoto, C.R. & Larivière, V. (2018). *Measuring research: what everyone needs to know*. New York: Oxford University Press.

Thanuskodi, S. (2017). Journal of Social Sciences: A Bibliometric Study. *Journal of Social Sciences, 24*(2), 77-80. https://www.tandfonline.com/doi/abs/10.1080/09718923.2010.11892847

Web of Science. (2023). *Web of Science Platform - Training Resources*. Clarivate Accelerating innovation. https://clarivate.com/webofsciencegroup/support/wos/