



Philosophers' appraisals of bibliometric indicators and their use in evaluation: from recognition to knee-jerk rejection

Ramón A. Feenstra¹ · Emilio Delgado López-Cózar²

Received: 25 September 2021 / Accepted: 3 January 2022
© The Author(s) 2022

Abstract

The knowledge and stance of researchers regarding bibliometric indicators is a field of study that has gained weight in recent decades. In this paper we address this issue for the little explored areas of philosophy and ethics, and applied to a context, in this case Spain, where bibliometric indicators are widely used in evaluation processes. The study combines data from a self-administered questionnaire completed by 201 researchers and from 14 in-depth interviews with researchers selected according to their affiliation, professional category, gender and area of knowledge. The survey data suggest that researchers do not consider bibliometric indicators a preferred criterion of quality, while there is a fairly high self-perception of awareness of a number of indicators. The qualitative data points to a generalised perception of a certain rejection of the specific use of indicators, with four main positions being observed: (1) disqualification of the logic of metrics, (2) scepticism about the possibility of assessing quality with quantitative methods, (3) complaints about the incorporation of methods that are considered to belong to other disciplines, and (4) criticism of the consequences that this generates in the discipline of philosophy.

Keywords Bibliometric indicators · Philosophy · Spain · Research quality · Evaluation

Introduction

The generalisation of the use of bibliometric indicators for scientific evaluation has been a constant for a wide range of contexts and disciplines (Butler, 2003, 2007; Geuna & Martin, 2003; Hammarfelt & Rushforth, 2017; Moed, 2005; Whitley, 2007). And although it initially seemed that the humanities would be barred from adopting bibliometric indicators because of their epistemological differences, their peculiarities in terms of their research practices, publication habits and citation processes (Garfield, 1980; Nederhof, 2006), today these techniques are also widely used for this field. It is recognised, however, that the use

✉ Ramón A. Feenstra
feenstra@uji.es

Emilio Delgado López-Cózar
edelgado@ugr.es

¹ Universitat Jaume I de Castelló, Castelló, Spain

² Facultad de Comunicación y Documentación. Universidad de Granada, Granada, Spain

of metrics for research evaluation can be difficult or problematic for this particular field (Hammarfelt & Haddow, 2018; Hug et al., 2014; Thelwall & Kousha, 2015). Its stronger national focus, higher use of vernacular languages, predilection for publishing books, single author approach, and higher number of publications addressed to a non-scholarly public may make the use of this type of techniques inefficient or give rise to imbalances (Borrego & Urbano, 2006; Hicks, 2004; Nederhof, 2006; Ochsner et al., 2013). The wide and diverse range of disciplines and identities (Hammarfelt, 2017; Laudel & Gläser, 2006; Thelwall & Delgado, 2015) is also deemed to make it difficult to measure their quality in a quantitative way (Hammarfelt & Haddow, 2018; Hug et al., 2014; Ochsner et al., 2017). An added difficulty is the poor coverage of the humanities in citation databases (Web of Science and Scopus) (Hicks, 1999, 2004; Archambault et al., 2006; Martín-Martín et al., 2018, 2021). Finally, another central topic regarding the use of bibliometric indicators in the evaluation of the humanities also warns us of the effects it can have on the idiosyncrasies of the discipline (Hammarfelt, 2017; Hicks et al., 2015).

Given the widespread use of bibliometric indicators in evaluation, several studies have been undertaken on researchers' perceptions of them, the uses they make of them, and their assessments of their strengths and weaknesses. Such studies have been extended to different fields (Buela-Casal & Zich, 2012), some of which are focused on the life sciences (Aksnes & Rip, 2009), biomedicine and sociology (Hargens & Schuman, 1990), medicine and physics (Derrick & Gillespie, 2013) or biomedicine and economics (Hammarfelt & Rushforth, 2017). A few far more recent studies deal with social sciences (Haddow & Hammarfelt, 2019) or the humanities (Bayer et al., 2019; Hammarfelt & Haddow, 2018) and, within this field, more specifically with areas such as literature and art history (Hug et al., 2014).

The geographical scope of the studies conducted on the reception of bibliometric indicators in the humanities has been both transnational and national. Cross-national studies (Galleron et al., 2017) include work that focuses specifically on the representations of young researchers as well as the perceptions of humanities researchers (Jamali et al., 2020; Nicholas et al., 2020). On the other hand, there is information available for specific countries such as Switzerland, Australia or Russia (Hug et al., 2014; Hammarfelt & Haddow, 2018; Narayan et al., 2018; Griněv, 2021). For the case of Spain, which is the focus of this work, some studies centred on the humanities have been conducted either for the entire country (Giménez-Toledo, 2016) or for a particular Spanish institution (Bayer et al., 2019). Also for this context as a whole, there is research focused specifically on the perception of young people (Rodríguez-Bravo & Nicholas, 2018, 2020).

Studies that focus on analysing the perceptions of researchers in the humanities have highlighted a number of aspects. For example, attention has been drawn to researchers' criticism of the use of methods considered to belong to other disciplines, reservations about the use of quantification of research quality or the lack of consensus on the quality criteria shared between different areas (Bayer et al., 2019; Galleron et al., 2017; Hammarfelt & Haddow, 2018; Hug et al., 2014). There is also a widespread feeling of frustration on the part of humanities researchers (Giménez-Toledo, 2016; Hammarfelt & Haddow, 2018), while others report that humanities researchers show a certain lack of knowledge about how journals improve their performance on metrics (Narayan et al., 2018).

However, despite all this abundant literature, much remains unknown about the views on the use of bibliometric indicators in the evaluation of the humanities and even more so in specific areas such as philosophy and ethics. Apart from some notable data from the study by Hammarfelt and Haddow (2018), who warned specifically of the relevance of expanding this type of approach, there are not many studies that examine how experts in

philosophy and ethics perceive bibliometrics for the scientific evaluation of their field. The present study attempts to fill this gap by looking exclusively at these areas. This research gains added relevance if we bear in mind that we are undertaking our study in a country that is paradigmatic in the use of metrics for the evaluation of its scientists in all branches (Jiménez-Contreras et al., 2002; Jiménez-Contreras et al., 2003; Butler, 2004; Fernández et al., 2006, 2011; Osuna et al., 2011; Delgado-López-Cózar, 2010; Molas-Gallart, 2012; Derrick & Pavone, 2013; Marini, 2018; Cañibano et al., 2018; Rodríguez-Bravo & Nicholass, 2018; Bautista-Puig et al., 2020).

In short, with this study our aim is to answer some key questions: What are the bibliometric indicators that researchers in ethics and philosophy in Spain know and value the most? What are their main concerns or worries? Do they consider that metrics are capable of measuring the quality of their publications and, indirectly, their research?

Methodology

The study combines a survey and 14 in-depth interviews in order to gather both quantitative and qualitative data.

Self-administered questionnaire

The study population comprised university researchers working in the knowledge areas of philosophy and ethics in Spain. We identified the members of this academic community through a systematic search of the websites of Spanish universities and research centres. In the vast majority of institutions we were able to identify affiliated members and researchers, with only four exceptions. Through these inquiries we identified 541 members of researchers, of whom 521 worked in universities and 20 in the research centre CSIC; 44 universities (37 public and 7 private) took part in the study and responses were received from all but three institutions.

The data was collected using Google Forms. The survey went through a previous pilot testing process in which 9 researchers from four different categories participated at the beginning of 2019 (2 research fellows, 3 lectures, 2 senior lectures and 2 professors). Once this validation process was completed, the survey was launched and remained open for responses between February and June 2019. On 25 February, a message was sent to the institutional email address of the 541 university teachers and researchers identified, followed by two reminders, two and four weeks after the initial contact. We also approached the main scientific societies and associations for Spanish philosophy professionals—the Spanish Association for Ethics and Political Philosophy (AEEFP), the Academic Society of Philosophy (SAF) and the Spanish Philosophy Network (REF)—requesting their collaboration. In May, these organisations contacted their members by email to encourage them to participate in the survey. The survey was closed on 14 June 2019.

The survey included 22 questions, of which 21 were multiple choice questions and 1 an open-ended question. The questions were divided into four main sections: (1) information search behaviour, (2) communication practices on pay-to-publish and open access (Feenstra & Delgado-López-Cózar, 2021b), (3) ethics in scientific publication (Feenstra et al., 2021), and (4) scientific evaluation and metrics, the latter being the main focus of the present study.

In order to obtain a general idea of the role attributed by our researchers to bibliometric indicators, we asked (Q1) about the criteria that would best reflect the quality of a scientific publication. This question was formulated in such a way as to include different criteria so that we could contrast the hierarchical ranking and the position of bibliometric indicators in the participants' answers. We expressly offered two options: the journal impact indicators, which are the indicators preferentially used in Spain to evaluate publications and scientists (Feenstra & Delgado-López-Cozar, 2021a), and citation counts in general. Since journal-level metrics are based on citations, we could capture not only the degree of specific valuation of the journal impact indicators and the citation counts of a publication, but also the degree of agreement between the two values.

Q1 Please rate which criteria best reflect the quality of a publication (1 not at all, 5 very much).

- Adherence to the editorial standards for the presentation of scientific publications
- Journal impact indicators in which it is published
- Media mentions (press, radio, TV)
- Mentions in blogs and social media
- Likes and recommendations
- The number of citations received
- The opinion of scientists as measured by surveys
- Presence in bibliographic repertories and databases
- The economic repercussion and impact
- Social and cultural repercussion and impact
- Peer review of the publication
- The views and downloads received

Furthermore, the results of this question are compared with those obtained in item 2 of the survey on publication practices (Feenstra & Delgado-López-Cózar, 2021b), which specifically asked about the criteria used when selecting the journal or publisher. This is an aspect that allows us to contrast the reliability of the data.

The other question included in our survey refers to awareness and appraisal of bibliometric indicators. In this case, we have focused mainly on the journal impact indicators, given that they are the most widely used in Spain for scientific evaluation at all levels (BOE, 2015). The possible answers also include other indicators such as those that measure the influence of authors' publications or their dissemination and public attention. The intention behind including this variety of indices is to investigate the consistency of the participants' knowledge as regards the indicators currently promoted in these areas.

Q2 Please rate the bibliometric indicators you are aware of in terms of their relevance (1 not at all, 5 very much).

- H-index
- Journal Impact Factor (JCR)
- Citescore (Scopus)
- SJR (Scopus)
- Altmetric Score
- RG Score

Table 1 Demographics of the survey of Spanish university teaching staff and researchers in philosophy and ethics

Field	Total population	Respondents	Average response rate (%)
Philosophy	380	115	30.2
Ethics	161	86	53.4
Total	541	201	37.1

SNIP

As mentioned before, the questionnaire also included an open-ended question where respondents could freely express their opinions on the object of study. In the survey data collection process, 201 responses were obtained, representing 37.1% of the population surveyed. Responses were received from researchers from 41 universities (35 public and 6 private) as well as from the CSIC research centre. Table 1 shows the distribution by knowledge area. The open-ended question received a total of 60 detailed responses, of which 24 answers focused on bibliometric indicators and their use in the areas of philosophy and ethics.

Interviews

The interviews took place in September and October 2019. The 14 interviewees were selected according to the criteria of affiliation, professional category, gender and disciplinary area in order to guarantee the widest possible range of profiles. Seven of the interviewees were men and seven were women, and seven worked in the field of ethics and seven in philosophy. The profiles of the interviewees included three research fellows, two lecturers, five senior lecturers and four full professors.

The interviewees were affiliated to the universities of Barcelona, Castellón, Complutense (Madrid), Granada, Murcia, Valencia, Zaragoza and the Basque Country, as well as the Institute of Philosophy at the CSIC. The semi-structured interviews lasted an average of 35.30 min; the shortest was 14.14 min and the longest, 59.10 min. The interviews were conducted individually by telephone and always by the same researcher. The conversations were recorded and later transcribed for analysis. Questions were posed on five central issues: (1) document genre and preferred language of publication, (2) peer review, (3) paying to publish and open access, (4) perceptions of research misconduct, and (5) bibliometric indicators and the Spanish assessment system. Specifically, for the present study, a generic question was asked about their vision on the bibliometrics and their use in the assessment of research in philosophy and ethics.

The qualitative data collected from the interviews and from the open-ended question in the survey was then subjected to a content analysis, first classifying the verbatim according to the main sections of the study, and then moving on to a new coding and classification according to the topics raised. The most significant verbatims are included in the results section using the acronyms of Table 2.

Table 2 List of acronyms

<i>Career stage</i>	
Research Fellow	Rf
Post-doctoral researcher	Pdr
Lecturer	L
Senior lecturer	Sl
Professor	P
<i>Source of information</i>	
Open section of the survey	
Interview	

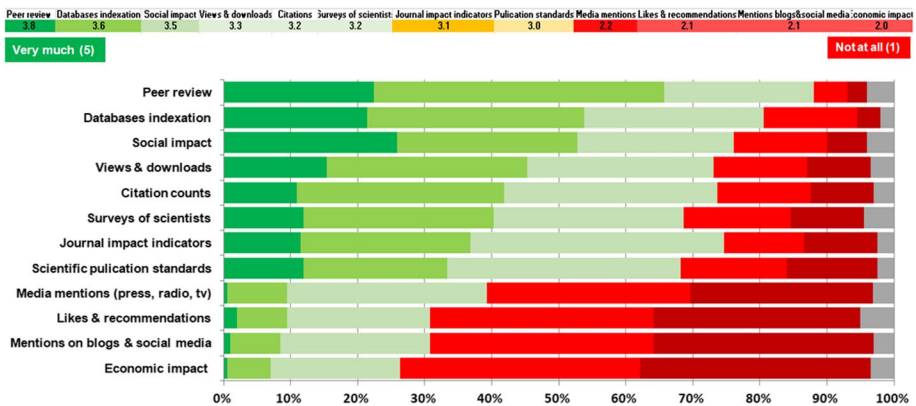


Fig. 1 Evaluation of the criteria reflecting the quality of a publication by Philosophy and Moral Philosophy teachers and researchers at Spanish universities

Results

The perception of the relevance of bibliometric indicators as regards the publication and its potential quality

When researchers were asked about the criteria they took into account when measuring the quality of a publication, they ranked the bibliometric indicators in secondary positions: the number of citations in fifth place (3.2 on a scale of 1 to 5) and the journal impact indicators in which it is published in seventh place (3.1 on the same scale) among the possible options (Fig. 1). The scores for the bibliometric indicators are very similar to each other, which reflects a very high degree of reliability of the responses obtained. It should also be noted that, although these quantitative criteria are not the preferred ones, they do obtain a slightly positive score. More researchers rate bibliometric indices highly or fairly highly than those who rate them poorly or not at all. In contrast, it is very significant that the most important criterion for estimating the quality of a publication is peer review (Fig. 1). It is precisely this qualitative and subjective criterion, based on the value judgements made by researchers, that philosophers consider their favourite.

If we contrast these data with those obtained in a specific study on publication practices (Feenstra & Delgado-López-Cózar, 2021b), we can see the consistency of the responses of

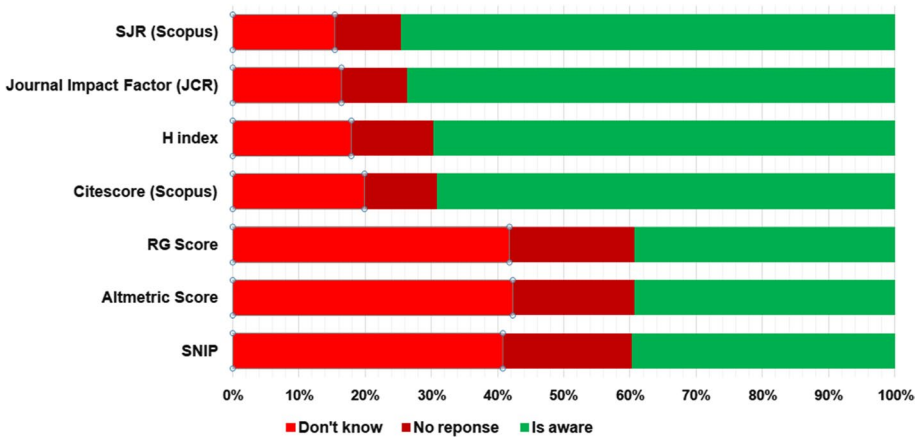


Fig. 2 Level of awareness of various bibliometric indicators by Philosophy and Moral Philosophy teachers and researchers at Spanish universities

this research community, since “impact on number of citations” is not the preferred criterion for researchers when making a decision on the journal in which to publish their work (with a score of 3.5 on a scale of 1 to 5). In fact, this criterion is ranked sixth on the list provided. Aspects such as prestige, subject orientation, not having to pay to publish, quality and visibility all score higher.

It is therefore clear that journal impact indicators are not a preferred criterion for researchers, although the general assessment of these indices is more positive than negative. This is interesting, bearing in mind that the very promotion of researchers depends to a large extent on their publishing a certain number of articles in high-impact journals (Feenstra & Delgado-López-Cózar, 2021a).

Awareness and assessment of bibliometric indicators

Looking at the awareness that researchers in philosophy and ethics claim to have of some bibliometric indicators, it is worth noting that this data is higher than might perhaps be expected for a discipline that traditionally has not used quantitative indices such as the ones reviewed here. The data in our study show an unequal degree of knowledge of the indicators and clearly reflect the existence of two groups of indicators (Figs. 2 and 3). On the one hand, there are four indicators (*SJR*, *Journal Impact Factor*, *H-index* and *Citescore*) that are valued by 70% or more of the respondents, while less than 20% are unaware of them. The high awareness of *SJR* is particularly striking if we compare it, for example, with the results extracted from Springer, where up to 52% were unaware of this indicator, while *Impact Factor* and *h-index* were only unknown to 5% and 10% of the participants, respectively (Penny, 2016).

Moreover, our study shows that there are three indicators that are openly unknown by respondents (only 40% value them and 40% are not even aware of their existence), namely, *RG Score* (*ResearchGate*), *Almetric Score* and *SNIP*. The lack of awareness of the latter does not seem strange if we compare it with other studies that observe, for example, a lower degree of use or knowledge of indicators such as *SNIP* (Rousseau & Rousseau, 2017) or the *Almetric Score* (Aung et al., 2019; Lemke et al., 2021; Thelwall, 2018).

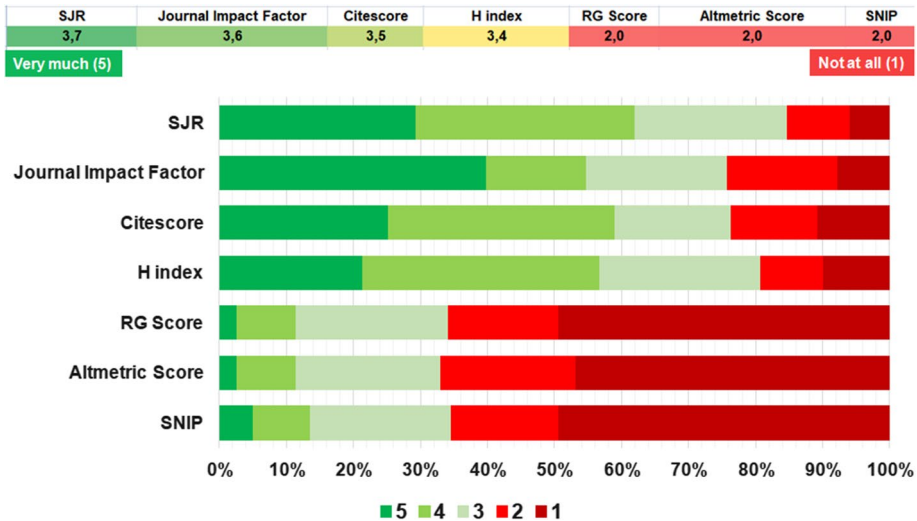


Fig. 3 Assessment of the relevance of various bibliometric indicators by Philosophy and Moral Philosophy teachers and researchers at Spanish universities

When asked about the evaluation of these indicators, it is worth noting that none of them received a very high rating from the researchers. However, four indicators were rated positively by respondents: *SJR*, *Journal Impact Factor*, *Citescore* and *H-index*, with values almost half a point higher than those obtained in question 1 (Fig. 1). The highest rated indicator is *SJR* (3, 7 on a scale of 1 to 5), which is generated from the Scopus database. The fact that the *SJR* calculates the impact of Humanities journals, in general, and of a good number of Philosophy journals (446)¹ probably explains this inclination. It should be remembered that the *Journal Impact Factor* (*JIF*) is not calculated for Humanities journals, although some Philosophy journals (no more than 25)—those included in the Social Sciences edition—may have impact indices (Sivertsen, 2014; van den Akker, 2016).² It is significant that two indicators for measuring the impact of journals (*SJR* and *JIF*) are associated with or derived from the Web of Science and Scopus databases, which are the most well-known, reputable and widely used by the scientific community. The reason may be linked to the use that the Spanish performance evaluation agencies make of them (BOE, 2015). This conclusion coincides with that observed by Hammarfelt and Rushforth (2017) for the fields of biomedicine and economics, where *JIFs* and *h-indices* are preferred, in their opinion, largely because “they are well-established tools of evaluation within some disciplinary communities” (Hammarfelt & Rushforth, 2017, 178). Other recent studies show how, in general, for fields such as social sciences, technology and medicine *WoS*, *Scopus* and *Google Scholar* gain greater importance as indexing databases (Wijewickrema, 2021). They also highlight the fact that for the field of social sciences, the *Scopus* database

¹ See <https://www.scimagojr.com/journalrank.php?category=1211&area=1200&type=j&wos=true>.

² On the other hand, it should be noted that the new *Journal Citation Indicator* (*JCI*) has been introduced in 2021, which introduces new features in this respect with the inclusion of journals from the Arts & Humanities Citation Index (*AHCI*) and *ESCI*.

is the most important, while WoS and Google Scholar rank second and third in terms of importance.

Finally, it should be noted that the data on the assessment of the bibliometric indicators lead us to observe that, although researchers say they do not use the number of citations received as a priority criterion for defining the quality of a publication, they do exhibit a good knowledge and a slightly positive assessment of the two journal impact indicators most commonly used in scientific evaluation in Spain (SJR and JCR). These data provide us with some key clues as to how researchers evaluate and use the metrics, which should be complemented with qualitative material that delves deeper into their perceptions.

Stances of philosophy and ethics researchers as regards the use of bibliometric indicators

The number of responses received in the open section of the questionnaire seems to suggest that the use of bibliometric indicators is perceived as a crucial issue. Of the 60 responses received, a total of 24 focus on bibliometric indicators and their use in the evaluation of researchers. In general, the qualitative material of our study (survey and interviews) show that the stance of researchers as regards the use of metrics for scientific evaluation reaches a high level of rejection that can even become or mutate into a certain level of *viscerality*. In this section we group together some of the most prominent stances in an attempt to collect a large number of voices due to their high level of exposition.

Positive assessment: an exception in the qualitative part of the study

It should be noted that in the qualitative instrument as a whole, only two voices are clearly positive regarding the use of bibliometric indicators and one voice can be classified as rather neutral. Thus, one participant in the study describes “adopting metrics that cannot be controlled only by one’s buddies” as a clear attempt to put an end to an academic system defined as patronage-based (Rf.1-survey). Another considers that there is no real problem with a scientific evaluation model based on journal citation metrics, stating that:

I honestly do not think that excessive attention to quality indices is the problem in the area of moral philosophy, quite the contrary: many people still do not publish (and do not even try to) in quality journals that are well positioned in JCR or SCImago. (L.1-survey)

Finally, another participant pointed out that although the system could be improved, it is not really problematic.

Overall I think the (metrics-based) evaluation system leaves a lot to be desired, but it is better than having no system at all. (L.9-survey)

It is noteworthy that all three voices come from researchers who are in the early stages of their careers. This is significant given that studies that extract data on the position of young people seem to note a greater predilection (or at least a greater use and knowledge) of metrics by this group. This is, for example, borne out by multidisciplinary studies (Nicholas et al. 2020b), as well as those focused on the humanities, where Hammarfelt and Haddow noted that academics, and especially those with a 5–15 year track record, “were the most frequent users of metrics” (Hammarfelt & Haddow, 2018, 931).

Critical voices: a majority and with disparate arguments

In general, philosophy and ethics researchers are mostly critical of bibliometric indicators, practically identified with journal-level metrics, when they express their opinions qualitatively. The arguments are varied, but we can group them into four central ideas: (1) disqualification of the logic of metrics, (2) scepticism about the possibility of evaluating quality with quantitative methods, (3) complaints about the incorporation of methods that are considered to belong to other disciplines, and (4) criticism of the consequences for the discipline of philosophy. It is relevant to note, however, that the critical voices are mainly expressed with regard to the specific use of metrics in the Spanish context, as can be seen in points 2, 3 and 4 below.

Disqualification of the logic of metrics First, there are researchers who criticise the citation counting system in general. These comments are in line with the literature that is critical of the limitations of bibliometrics (e.g. Baum, 2011; Moed, 2005; Moed & Van Leeuwen, 1996). Some voices indicate, for example:

Indices are a mere convention. (S1.15-survey)

(...) I think quartiles are like a soap bubble. (P.3-interview)

The difference between high impact journals and low impact journals is what, twenty citations? ... (S1.3-interview)

The "ranking" systems (e.g. SJR) that change every year are a disaster. (P.8-survey).

The quality of philosophical research as something not "quantifiable" The predominant argument of philosophy and ethics researchers is focused on pointing out the inadequacies of using "quantitative" indices to measure the "quality" of philosophical research. In this respect, there is a generalised rejection. The statements stand out for their categorical nature and for focusing on the specific use of bibliometric indicators by Spanish evaluation agencies. In this sense, we can read, for example:

The root of the problem is quantitative evaluation. This is already well diagnosed even in the field of experimental sciences. There is the famous California DORA declaration [sic], where biologists, physicists and mathematicians, journal editors, etc. said that decisions regarding professional and research careers (promotions, recruitment, etc.) could not be made based solely on quantitative indicators, which is precisely what we are doing in Spain. We are evaluating careers on the basis of publications (...) in indexed journals, which are evaluated as the famous quartiles.

In other words, the evaluators do not actually read the texts. Qualitative evaluation has been eliminated from the entire university system. (...) and that is the big problem we have. (...) It would be an important change to recover quality (P.1-interview)

It is supposed to assess the quality of research, but in practice it only takes the quantity into account, so in time everyone ends up slipping through the net. (S1.21-survey) Being asked for Q1 journals is a purely quantitative criterion that does not reflect the researcher's worth. To do things properly, what they should do is to actually read the person's articles, and not look at whether it is Q1 or Q2. Because, the fact that you don't have the talent to know which journal to choose... It's a question of knowing how to find your way around, or knowing how to manage, or having friends in certain journals. (L.2-interview)

Impact is a joke that has nothing to do with academic quality. Impact means that a journal has been cited forty times in a year... and what does that mean? (...) for example, there is no Italian or Spanish journal (in philosophy) that is Q1 and there are some very good journals. (SI.3-interview)

(...) quality indices and soft soap (with respect), which are pure window dressing. The quality of a shoe has to do with the quality of its materials, how its structure is designed, how its different parts are sewn and adapted to the whole. It will never depend on the shop window where it is displayed. (SI.2-survey)

The proliferation of quantitative indices in no way guarantees the quality of what is published. (Pdr.1-survey)

Research in certain areas, such as Philosophy, is not going to be compatible with submitting it to “objective” criteria of productivity and scientific quality. (L.15-survey)
Content should be more important than figures. (Pdr.2-survey)

What journal rankings (JCR) do is measure impact, the impact by the citations they receive (...) this idea that quality is based on the number of citations received seems to me to be a random criterion (...) The exclusive quantitative criterion of impact seems to me to be reductionist and counterproductive (...) (SI.5-interview)

True quality has nothing to do with what is understood in academia as quality criteria. The current system of *acreditación*³ and *sexenios*⁴ does not improve quality. (L.13-survey)

As in the case of French universities, it is essential that the Spanish university system, and in particular the accreditation agencies, abandon the current criteria of scientificity imposed by Anglo-American and German publishing holdings. (L.9-survey)

The evaluation of the quality of knowledge, in any field of expertise, must be carried out in a qualitative manner by peers who are experts in the field and do it publicly. 2) Quality assessment should not be confused with intellectual censorship, but should be compatible with respect for ontological, epistemological and ethical-political pluralism, especially in the humanities and social sciences. (P.6-s)

It is clear that Kant today would have problems to have *sexenios* or to be *accredited* (i.e. to be promoted). (SI.6-survey)

The criticism of the quantitative measurement of quality coincides with other studies. Some of those focused on the perception of researchers in the field of the life sciences have noted that scientists in this branch also discuss this possibility and point out how “some emphasised that citations only reflect intellectual influence among the academic community and do not represent a balanced indicator of quality” (Aksnes & Rip, 2009, 905).

Within the humanities and social sciences, resistance to the quantification of quality seems to be even stronger (Bayer et al., 2019; Hammarfelt & Haddow, 2018; Hug et al., 2014; Ochsner et al., 2017). Moreover, studies examining the perceptions of young researchers note that ‘arts and humanities and social science ECRs rated reliance on quantifiable metrics lowest’ (Nicholas et al., 2020a, 7). Likewise, within the field of philosophy

³ *Acreditación* is a tenure review for promotion.

⁴ *Sexenio* is an individual bonus based on the recognition of research performance assessed by a committee every six years.

itself, there have also been reports of resistance to this phenomenon of quantification of research results. In this regard, Hug, Ochsner and Daniel pointed, by way of example, to the unease of Australian philosophers in relation to the journal ranking in the Excellence in Research for Australia (ERA), who considered that ‘those standards cannot be given simple, mechanical, or quantitative expression’ for disciplines such as philosophy (Hug et al., 2014, 46). Similar concern was expressed in 2020 in an open letter written by the Institute of Philosophy of the Russian Academy of Sciences concerning the use of a metric-driven assessment system proposed by the Ministry of Science and Higher Education of the Russian Federation.⁵

The participants in our study seem to agree with the assessment of the discord generated by using quantitative methods. As other authors have pointed out for the humanities (Hug et al., 2014; Ochsner et al., 2017), there are strong disagreements about what research quality means and how (and even whether it is possible) to measure it.

Colonisation of philosophy by numbers The rest of the interventions collected mention aspects linked to the specific use of metrics in the evaluation of philosophy. In this sense, what is considered to be a clash between disciplines appears on a recurrent basis. Here, the arguments appear in two directions: (a) the problems derived from measuring philosophy with the parameters of other sciences, and (b) the need to look for evaluation criteria specific to the scientific branch of philosophy. This type of criticism ties in with the warnings that point both to the specificities of the humanities (Hicks, 2004; Nederhof, 2006) as well as to the plural range of disciplines and identities that compose it (Hammarfelt, 2017; Laudel & Gläser, 2006; Thelwall & Delgado, 2015).

The researchers themselves put it this way:

There are specialities in Philosophy that cannot be equated with the Sciences – there are no patents, etc. It is another type of research, but still RESEARCH. (SI.13-survey)

The application of the research criteria of the experimental sciences to philosophy is detrimental. (SI.19-survey)

Humanistic publications cannot be assessed on the basis of the criteria applied to the sciences, where work is done in a different way. The decisive criterion should be internal (the objective quality of the content of publications), not external (journal quality indicators, impact indices, etc.). (SI.9-survey)

(...) I believe that the evaluation of articles and journals in philosophy should not be measured in the same way as scientific ones; their dimension is very different and so are their objectives. This is due to the fact that positivism predominates today. The humanities, without losing their quest for rigorousness, must be faithful to their own idiosyncrasy. And it is this idiosyncrasy that should mark the quality indices. This is why it is important to find specific rather than generalist evaluation scales. (L.4-survey)

It is necessary to limit the importance of impact indices and allow the studies themselves to be read and criticised by the authors, taking into account, in turn, the characteristics of our area of study. (L.8-survey)

The most appropriate scientific criteria must be sought for each speciality. (SI.11-survey)

⁵ See https://iphras.ru/pismo_06_02_2020.htm.

It has something of a whirlwind that is alien to the demands of quality work in philosophy. (Pdr.3-survey)

An agreement is needed to unify the evaluation criteria in each subject area. (P.1-survey)

Criticism also extends to areas within philosophy itself. Thus, several voices consider that the current use of bibliometric indicators is prioritising (and promoting) some areas over others. In other words, it is believed that a clash between specialisations in philosophy is being encouraged (Sl.10-survey, Sl.7-survey, P.3-interview).

The effects on the internal dynamics of philosophical science A final aspect widely commented on by researchers derives from the consequences that they observe regarding the effects that the specific use of bibliometric indicators has on their discipline. This idea ties in with a whole specific literature on the effects of evaluation policies (Aagaard et al., 2015; de Rijcke et al., 2016; Hicks et al., 2015; Wouters, 2014; Wouters et al., 2015). In this case, researchers from the fields of philosophy and ethics in Spain show that the evaluation system affects research behaviour in a significant way: the transformation of research agendas, the modification of publication practices (document type and language of publication), the neglect of teaching work, a perception of increasing research misconduct and, finally, a negative effect on mental health. Although to a lesser extent, other consequences included increased research productivity and enhanced transparency and impartiality in academic selection processes. The large number of effects observed and the specificity of the topic have led to this issue being researched in a specific study on research misconduct (Feenstra et al., 2021) and on the general consequences of the Spanish evaluation system (Feenstra & Delgado-López-Cózar, 2021a).

Discussion/Conclusions

This study has shown that bibliometric indicators are not a preferred criterion for researchers when it comes to defining what is considered to be the quality of a publication and, indirectly, of research. However, despite the fact that the indicator of the number of citations is not regarded as a preferred criterion, we have also been able to see how the researchers give it a slightly positive evaluation on a scale of 1 to 5. The evaluations are very similar (3.2 for the definition of quality and 3.5 for the journal selection criterion), which would indicate a fairly consistent judgement and would also confirm the reliability of the data from the quantitative material.

These data do not reflect a decided rejection of the citation measurement indicator but, on the other hand, they seem to show that the penetration of bibliometric indicators among Spanish philosophers is rather modest (at least for the time being); indeed, the researchers themselves place it below other criteria in their decision-making and preferences. It is also worth contextualising these data in an evaluation system that establishes the need for researchers to publish a minimum number of articles in high impact journals in order to have a chance of promotion in their academic careers (for example, 20 publications in Q1 or Q2 JCR or Q1 SJR in order to aspire to a professorship in ethics with the highest likelihood of success. See ANECA, 2019). Moreover, Spanish universities themselves often encourage and monitor the bibliometric results of their researchers (Bautista-Puig et al., 2020). Despite this, it does not seem to condition the researchers' decision-making, or at least that is how they express it at the moment.

Likewise, when asked about their opinion of the indicators, they show a strikingly good awareness of them, and especially of the journal impact indicators. Although it may be surprising that researchers in philosophy and ethics show this degree of knowledge, these data coincide to a certain extent with those provided in some of the few studies that exist on these fields. In this sense, when asked about the use of metrics in evaluation or self-promotion in applications and CVs among different areas of the humanities, Hammarfelt and Haddow (2018) found that the “areas of ‘philosophy, ethics, and religion’ averaged above the rest of the areas with almost 37%, compared to an average of 32% or 24% in art”. These data led them to conclude that “The importance of (international) journal publishing in these fields, especially in philosophy and ethics, is one possible explanation for the relatively frequent use of metrics” (Hammarfelt & Haddow, 2018, 928). However, it is also worth noting that our survey data were focused on the researcher’s awareness and specifically the self-perception of their relevance for reflecting the quality of a publication. Our responses do not allow us to assess the exact extent of this knowledge, which is an attractive issue to be addressed in future studies. Moreover, it would also be interesting to conduct future research that analyses to what extent and at what level Spanish philosophers regularly monitor and make use of bibliometric indicators in the Web of Science and Scopus databases. In this case, our data show that those indicators used in the evaluation are the best known, which seems logical, bearing in mind that their progression in their academic career depends on this. Furthermore, the survey shows that the indicators used by the national evaluation system are not only the best identified but also the most highly valued. In this sense, the SJR is the indicator with the highest rating (3.7/5), which is understandable, moreover, due to its greater coverage of the journals in this field. These data as a whole can be interpreted, as we have pointed out earlier, within the Spanish evaluative culture that, by prioritising these indicators, has generated a social and academic environment in which even researchers in the fields of the arts—such as Philosophy and Ethics—are aware of these indicators and know of their relevance in their immediate environment. On the other hand, it cannot be ruled out that these responses are to some extent conditioned by an effect of wishing to continue with the requirements of the evaluation system.

The study has also made it possible to observe aspects highlighted in the qualitative part that contextualise the information from the survey and allow more in-depth analyses. In this section, it is worth highlighting the generalised rejection of the indicators, although this rejection is expressed more with regard to the specific use of these indicators. It is not so much the bibliometric indicator that generates rejection as its preponderance in the Spanish system for evaluating philosophy, the imbalances it generates and its effects on the discipline. Opinions that reflect strictly upon the indicators are very scarce, with the metric-driven evaluation process being the one that attracts the most attention. Even in the moments of the study where participants were free to express themselves on any subject (such as the open-ended question of the survey), this was the predominant issue. It cannot be ruled out that some voices with more pro-indicator views may not have expressed their opinion during this study, or that in the future these positions may become more significant in these areas, especially among younger people. Future panel studies would therefore allow us to assess this evolution. However, the general trend perceived in this study even shows a certain degree of *viscerality* in the qualitative setting of the research. The arguments that are outlined—such as complaints about the use of methods that are considered to belong to other disciplines or scepticism about the possibility of assessing quality quantitatively—are arguments that have already been observed by other studies for different branches of the humanities (Hammarfelt, 2017; Thelwall & Delgado, 2015; Thelwall & Kousha, 2015). Thus, philosophy and ethics researchers in Spain state that qualitative

evaluation is the most appropriate method for the evaluation of their field, a stance that we know is generally expressed for the humanities (Galleron et al., 2017; Hammarfelt & Had-dow, 2018). It is also significant that when asked in the survey about which is the best quality criterion, peer review was the one most highly valued by researchers, with the qualitative criterion based on reading and evaluation being precisely the one that is recognised as specific to the discipline.

Our study allows us to observe, in short, how the essential origin of criticism or frustration is linked to the quantitative way of assessing philosophical quality and what is seen as a colonisation by the latest metrics. Once again, the debate arises about the complexity of measuring the quality of a branch of the humanities in numerical terms and the call for a set of its own consensus criteria, which is a goal that has long proved elusive (Hug et al., 2014; Ochsner et al., 2017),

Acknowledgements The authors would like to thank Daniel Pallarés-Domínguez for his help in the data collection process and Mark Andrews for the English translation. We would also like to thank the three main Spanish philosophy and ethics associations, Asociación Española de Ética y Filosofía Política (AEEFP), Sociedad Académica de Filosofía (SAF) and Red Española de Filosofía (REF), for their collaboration during the research process and their endorsement of the data collection report.

Authors' contributions Conceptualisation: RF and EDLC; Methodology: EDLC; Formal analysis and investigation: EDLC; Writing—original draft preparation: RF; Writing—review and editing: EDLC and RF.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. No funding was received for conducting this study.

Availability of data and material Open data report available (in Spanish) at: <http://repositori.uji.es/xmlui/handle/10234/189924>.

Declarations

Conflict of interest The authors have no financial or proprietary interests in any material discussed in this article.

Ethical statement Prior to answering the survey, participants were informed about the nature of the study and that their anonymity would be preserved; they were only asked to state their affiliation, area of knowledge and academic status. Verbal informed consent was obtained from the participants in the interviews. An anonymised summary of the data gathered during the research was shared with the participants and published in an open report. This report has been endorsed by the three main Spanish philosophy and ethics associations: SAF, AEEFP and REF.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aagaard, K., Bloch, C., & Schneider, J. W. (2015). Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation*, 24(2), 106–117. <https://doi.org/10.1093/reseval/rvv003>
- Aksnes, D. W., & Rip, A. (2009). Researchers' perceptions of citations. *Research Policy*, 38(6), 895–905. <https://doi.org/10.1016/j.respol.2009.02.001>
- ANECA. (2019). *Méritos evaluables para la acreditación nacional para el acceso a los cuerpos docentes universitarios: cátedra de universidad*. Ciencias Sociales. Madrid. Retrieved 18 July 2021, from: <https://www.educacionyfp.gob.es/dam/jcr:d23eb60f-fa48-4f25-a6d3-342d3056f3f3/criterios-sociales-juridicas.pdf>
- Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, 68(3), 329–342. <https://doi.org/10.1007/s11192-006-0115-z>
- Aung, H. H., Zheng, H., Erdt, M., Aw, A. S., Sin, S.-C.J., & Theng, Y.-L. (2019). Investigating familiarity and usage of traditional metrics and altmetrics. *Journal of the Association for Information Science and Technology*, 70(8), 872–887. <https://doi.org/10.1002/asi.24162>
- Baum, J. A. (2011). Free-riding on power laws: Questioning the validity of the impact factor as a measure of research quality in organization studies. *Organization*, 18(4), 449–466. <https://doi.org/10.1177/1350508411403531>
- Bautista-Puig, N., Moreno Lorente, L., & Sanz-Casado, E. (2020). Proposed methodology for measuring the effectiveness of policies designed to further research. *Research Evaluation*. <https://doi.org/10.1093/reseval/rvaa021>
- Bayer, F., Gorraiz, J., Gumpenberger, C., Iturbide, A., Iribarren-Maestro, I., & Reding, S. (2019). Investigating SSH research and publication practices in disciplinary and institutional contexts. A survey-based comparative approach in two universities. *Frontiers in Research Metrics and Analytics*, 4, 1. <https://doi.org/10.3389/frma.2019.00001>.
- BOE. (2015). Ministerio de Educación, Cultura y Deporte, 17/6/2015, p. 50319–50337 <https://www.boe.es/buscar/doc.php?id=BOE-A-2015-6705>.
- Borrego, Á., & Urbano, C. (2006). La evaluación de revistas científicas en Ciencias Sociales y Humanidades. *Información, cultura y sociedad*, 14, 11–27. <https://doi.org/10.34096/ics.i14.886>.
- Buela-Casal, G., & Zych, I. (2012). What do the scientists think about the impact factor? *Scientometrics*, 92(2), 281–292. <https://doi.org/10.1007/s11192-012-0676-y>
- Butler, L. (2004). What happens when funding is linked to publication counts? In *Handbook of quantitative science and technology research* (pp. 389–405). Springer. https://doi.org/10.1007/1-4020-2755-9_18.
- Butler, L. (2003). Modifying publication practices in response to funding formulas. *Research Evaluation*, 12(1), 39–46. <https://doi.org/10.3152/147154403781776780>
- Butler, L. (2007). Assessing university research: A plea for a balanced approach. *Science and Public Policy*, 34(8), 565–574. <https://doi.org/10.3152/030234207X254404>
- Cañibano, C., Vilardeil, I., Corona, C., & Benito-Amat, C. (2018). The evaluation of research excellence and the dynamics of knowledge production in the humanities: The case of history in Spain. *Science and Public Policy*, 45(6), 775–789. <https://doi.org/10.1093/scipol/scy025>
- Delgado-López-Cózar, E. (2010). Claroscuros de la evaluación científica en España. *Medes: Medicina en Español*, 4, 25–29.
- Delgado-López-Cózar, E.; Feenstra, R.A.; Pallarés-Domínguez, D. (2020). *Investigación en Ética y Filosofía en España*. Asociación Española de Ética y Filosofía Política, Sociedad Académica de Filosofía, Red Española de Filosofía. Retrieved 13 May 2021, from: <http://hdl.handle.net/10234/189924>
- Derrick, G. E., & Gillespie, J. (2013). 'A number you just can't get away from': Characteristics of adoption and the social construction of metrics use by researchers. In Hinze, S. & Lottman, A. (Eds.) *Proceedings of the 18th international conference on science and technology indicators* (pp. 104–116)
- Derrick, G. E., & Pavone, V. (2013). Democratising research evaluation: Achieving greater public engagement with bibliometrics-informed peer review. *Science and Public Policy*, 40(5), 563–575. <https://doi.org/10.1093/scipol/sct007>
- Feenstra, R., 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), 1–21. <https://doi.org/10.1007/s11948-021-00278-w>
- Feenstra, R. A., & López-Cózar, E. D. (2021a). The footprint of a metrics-based research evaluation system on Spanish philosophical scholarship: An analysis of researchers perceptions. *ArXiv*, 2103.11987. (<https://arxiv.org/abs/2103.11987>).

- Feenstra, R. A., & López-Cózar, E. D. (2021b). Spanish philosophers perceptions of pay to publish and open access: books versus journals, more than a financial dilemma. *Learned Publishing*, 1–12. <https://doi.org/10.1002/leap.1426> (online first)
- Fernández Esquinas, M., Díaz Catalán, C., & Ramos Vielba, I. (2011). Evaluación y política científica en España: el origen y la implantación de las prácticas de evaluación científica en el sistema público de I+D (1975–1994). In: González de la Fe, T., & López Peláez, A. (eds) *Innovación, conocimiento científico y cambio social: ensayos de sociología ibérica de la ciencia y la tecnología* (pp. 93–130). Centro de Investigaciones Sociológicas.
- Fernández Esquinas, M., Pérez Yruela, M., & Merchán Hernández, C. (2006). El sistema de incentivos y recompensas en la ciencia pública española. In J. Sebastián & E. Muñoz (Eds.), *Radiografía de la investigación pública en España* (pp. 148–206). Biblioteca Nueva.
- Galleron, I., Ochsner, M., Spaapen, J., & Williams, G. (2017). Valorizing SSH research: Towards a new approach to evaluate SSH research's value for society. *f1eal Journal for Research and Technology Policy Evaluation*, 44, 35–41. <https://doi.org/10.22163/fteval.2017.274>.
- Garfield, E. (1980). Is information retrieval in the arts and humanities inherently different from that in science? The effect that ISI's citation index for the arts and humanities is expected to have on future scholarship. *The Library Quarterly*, 50(1), 40–57. <https://doi.org/10.1086/629874>
- Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva*, 41(4), 277–304. <https://doi.org/10.1023/B:MINE.0000005155.70870.bd>
- Giménez-Toledo, E. (2016). *El malestar de los investigadores ante su evaluación*. Iberoamericana.
- Griněv, A. V., Bylieva, D. S., & Lobatyuk, V. V. (2021). Russian university teachers' perceptions of scientometrics. *Publications*, 9(22), 1–16. <https://doi.org/10.3390/publications9020022>
- Haddow, G., & Hammarfelt, B. (2019). Quality, impact, and quantification: Indicators and metrics use by social scientists. *Journal of the Association for Information Science and Technology*, 70(1), 16–26. <https://doi.org/10.1002/asi.24097>
- Hammarfelt, B. (2017). Four claims on research assessment and metric use in the humanities. *Bulletin of the Association for Information Science and Technology*, 43(5), 33–38. <https://doi.org/10.1002/bul2.2017.1720430508>
- Hammarfelt, B., & Haddow, G. (2018). Conflicting measures and values: How humanities scholars in Australia and Sweden use and react to bibliometric indicators. *Journal of the Association for Information Science and Technology*, 69(7), 924–935. <https://doi.org/10.1002/asi.24043>
- 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>
- Hargens, L. L., & Schuman, H. (1990). Citation counts and social comparisons: Scientists' use and evaluation of citation index data. *Social Science Research*, 19, 205–222. [https://doi.org/10.1016/0049-089x\(90\)90006-5](https://doi.org/10.1016/0049-089x(90)90006-5)
- Hicks, D. (2004). The four literatures of social sciences. In Moed, H. Glänzel, W. & Schmoch, U. (eds.), *The handbook of quantitative science and technology research* (pp. 473–496). Kluwer. https://doi.org/10.1007/1-4020-2755-9_22.
- Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, 44, 193–215. <https://doi.org/10.1007/bf02457380>
- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature News*, 520(7548), 429. <https://doi.org/10.1038/520429a>
- Hug, S. E., Ochsner, M., & Daniel, H. D. (2014). A framework to explore and develop criteria for assessing research quality in the humanities. *International Journal of Education Law and Policy*, 10(1), 55–68. https://doi.org/10.1007/978-3-658-05969-9_13
- Jamali, H. R., Nicholas, D., Herman, E., Boukacem-Zeghmouri, C., Abrizah, A., Rodríguez-Bravo, B., Xu, J., Swigon, M., Polezhaeva, T., & Watkinson, A. (2020). National comparisons of early career researchers' scholarly communication attitudes and behaviours. *Learned Publishing*, 33(4), 370–384. <https://doi.org/10.1002/leap.1313>.
- Jiménez-Contreras, E., de Moya Anegón, F., & Delgado-López-Cózar, E. (2003). The evolution of research activity in Spain: The impact of the National Commission for the Evaluation of Research Activity (CNEAI). *Research Policy*, 32(1), 123–142. [https://doi.org/10.1016/S0048-7333\(02\)00008-2](https://doi.org/10.1016/S0048-7333(02)00008-2)
- Jiménez-Contreras, E., Delgado-López-Cózar, E., Ruiz-Pérez, R., & Fernández, V. M. (2002). Impact-factor rewards affect Spanish research. *Nature*, 417(6892), 898–898. <https://doi.org/10.1038/417898b>
- Laudel, G., & Gläser, J. (2006). Tensions between evaluations and communication practices. *Journal of Higher Education Policy and Management*, 28(3), 289–295. <https://doi.org/10.1080/13600800600980130>

- Lemke, S., Mazarakis, A., & Peters, I. (2021). Conjoint analysis of researchers' hidden preferences for bibliometrics, Altmetrics, and usage metrics. *Journal of the Association for Information Science and Technology*, 72, 777–792. <https://doi.org/10.1002/asi.24445>
- Marini, G. (2018). Tools of individual evaluation and prestige recognition in Spain: How sexenio 'mints the golden coin of authority.' *European Journal of Higher Education*, 8(2), 201–214. <https://doi.org/10.1080/21568235.2018.1428649>
- Martín-Martín, A., Orduña-Malea, E., & Delgado López-Cózar, E. (2018). Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: A multidisciplinary comparison. *Scientometrics*, 116, 2175–2188. <https://doi.org/10.1007/s11192-018-2820-9>
- Martín-Martín, A., Thelwall, M., Orduña-Malea, E., & Delgado-López-Cózar, E. (2021). Google scholar, microsoft academic, scopus, dimensions, web of science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, 126, 871–906. <https://doi.org/10.1007/s11192-020-03690-4>
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Springer.
- Moed, H. F., & Van Leeuwen, T. N. (1996). Impact Factors Can Mislead. *Nature*, 381, 186. <https://doi.org/10.1038/381186a0>
- Molas-Gallart, J. (2012). Research Governance and the Role of Evaluation. A Comparative Study. *American Journal of Evaluation*, 33(4), 583–598. <https://doi.org/10.1177/1098214012450938>.
- Narayan, B., Luca, E. J., Tiffen, B., England, A., Booth, M., & Boateng, H. (2018). Scholarly communication practices in humanities and social sciences: A study of researchers' attitudes and awareness of open access. *Open Information Science*, 2(1), 168–180. <https://doi.org/10.1515/opis-2018-0013>
- Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1), 81–100. <https://doi.org/10.1007/s11192-006-0007-2>
- Nicholas, D., Herman, E., Jamali, H. R., Abrizah, A., Boukacem-Zeghmouri, C., Xu, J., Rodríguez-Bravo, B., Watkinson, A., Polezhaeva, T., & Świgoń, M. (2020). Millennial researchers in a metric-driven scholarly world: An international study. *Research Evaluation*, 29(3), 263–274. <https://doi.org/10.1093/reseval/rvaa004>
- Nicholas, D., Watkinson, A., Abrizah, A., Rodríguez-Bravo, B., Boukacem-Zeghmouri, C., Xu, J., Świgoń, M., & Herman, E. (2020). Does the scholarly communication system satisfy the beliefs and aspirations of new researchers? Summarizing the Harbingers Research. *Learned Publishing*, 33(2), 132–141. <https://doi.org/10.1002/leap.1284>
- Ochsner, M., Hug, S. E., & Daniel, H.-D. (2013). Four Types of research in the humanities: setting the stage for research quality criteria in the humanities. *Research Evaluation*, 22(2), 79–92. <https://doi.org/10.1093/reseval/rvs039>
- Ochsner, M., Hug, S., & Galleron, I. (2017). The future of research assessment in the humanities: Bottom-up assessment procedures. *Palgrave Communications*, 3(1), 1–12. <https://doi.org/10.1057/palcomms.2017.20>
- Osuna, C., Cruz-Castro, L., & Sanz-Menéndez, L. (2011). Overturning some assumptions about the effects of evaluation systems on publication performance. *Scientometrics*, 86(3), 575–592. <https://doi.org/10.1007/s11192-010-0312-7>
- Penny, D. (2016). What matters where? Cultural and Geographical Factors in Science. <https://doi.org/10.6084/m9.figshare.3969012.v1>
- Rijcke, S. D., Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—a literature review. *Research Evaluation*, 25(2), 161–169. <https://doi.org/10.1093/reseval/rvv038>
- Rodríguez-Bravo, B., & Nicholas, D. (2018). Reputación y comunicación científica: Investigadores españoles en el inicio de su carrera. *El Profesional De La Información*, 28(2), e280203. <https://doi.org/10.3145/epi.2019.mar.03>
- Rodríguez-Bravo, B., & Nicholas, D. (2020). Descubrir, leer, publicar, compartir y monitorizar el progreso: Comportamiento de los investigadores junior españoles. *El Profesional De La Información*, 29(5), 1–16. <https://doi.org/10.3145/epi.2020.sep.03>
- Rousseau, S., & Rousseau, R. (2017). Being metric-wise: Heterogeneity in bibliometric knowledge. *El Profesional De La Información*, 26(3), 480–487. <https://doi.org/10.3145/epi.2017.may.14>
- Sivertsen, G. (2014). Scholarly publication patterns in the social sciences and humanities and their coverage in Scopus and Web of Science. In *Proceedings of the science and technology indicators conference* (pp. 598–604). Universiteit Leiden.
- Thelwall, M., & Kousha, K. (2015). Web indicators for research evaluation. Part 1: Citations and links to academic articles from the Web. *El Profesional de la Información*, 24(5), 587–606. <https://doi.org/10.3145/epi.2015.sep.08>.

- Thelwall, M. (2018). Altmetric prevalence in the social sciences, arts and humanities: Where are the online discussions? *Journal of Altmetrics*, 1(1), 4. <https://doi.org/10.29024/joa.6>.
- Thelwall, M., & Delgado, M. M. (2015). Arts and humanities research evaluation: No metrics please, just data. *Journal of Documentation*. *Journal of Documentation*, 71(4), 817–833. <https://doi.org/10.1108/JD-02-2015-002>
- van den Akker, W. (2016). Yes we should; research assessment in the humanities. In Ochsner, M., Hug, S. E & Daniel H.-D. (Eds.), *Research Assessment in the Humanities* (pp. 23–29). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_3.
- Whitley, R. (2007). Changing governance of the public sciences. In Whitley, R. & Gläser, J. (Eds.), *The changing governance of the sciences* (pp. 3–27). Springer. https://doi.org/10.1007/978-1-4020-6746-4_1.
- Wijewickrema, M. (2021). Authors' perception on abstracting and indexing databases in different subject domains. *Scientometrics*, 126(4), 3063–3089. <https://doi.org/10.1007/s11192-021-03896-0>
- Wouters, P. (2014). The citation: From culture to infrastructure. In Cronin, B., & Sugimoto C. R. (Eds.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact* (pp. 47–66). MIT press. <https://doi.org/10.7551/mitpress/9445.003.0006>.
- Wouters, P., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S., Rushforth, A., & Franssen T. (2015). *The metric tide: Literature review. Supplementary report I to the independent review of the role of metrics in research assessment and management*. HEFCE.