

Quality, impact and quantification: Indicators and metrics use by social scientists

Gaby Haddow*

Dept. of Information Studies, School of Media, Communication and Creative Arts

Curtin University

GPO Box U1987, Perth, Western Australia 6845, Australia

+61 8 9266 2707

G.Haddow@curtin.edu.au

Björn Hammarfelt

Swedish School of Library and Information Science

University of Borås,

Allégatan 1, Borås, 501 90, Sweden

+46 33-435 40 00

bjorn.hammarfelt@hb.se

*corresponding author

This is the peer reviewed version of the following article: Haddow, G. and Hammarfelt, B. 2018. Quality, impact and quantification: Indicators and metrics use by social scientists. *Journal of the Association for Information Science and Technology*. 70 (1): pp. 16-26, which has been published in final form at 10.1002/asi.24097. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving at <http://olabout.wiley.com/WileyCDA/Section/id-828039.html>

Quality, impact and quantification: Indicators and metrics use by social scientists

Abstract

The use of indicators and metrics for research evaluation purposes is well-documented, however less is known about their use by individual scholars. With a focus on the social sciences, this paper contributes to the existing literature on indicators and metrics use in fields with diverse publication practices. Scholars in Australia and Sweden were asked about their use and reasons for using metrics. A total of 581 completed surveys were analyzed to generate descriptive statistics, with textual analysis performed on comments provided to open questions. While just under half of the participant group had used metrics, the Australians reported use in twice the proportion of their Swedish peers. Institutional policies and processes were frequently associated with use, and the scholars' comments suggest a high level of awareness of some metrics as well as strategic behavior in demonstrating research performance. There is also evidence of tensions between scholars' research evaluation environment and their disciplinary values and publication practices.

Introduction

While demonstrating the quality of scholarly work is not a recent phenomenon, the widespread use of indicators and metrics to assess research has 'proliferated' over the last two decades (Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015). Ranging from citations to article downloads, the Impact Factor to ranked journal lists, the growth of indicators and metrics has run parallel with an increasing focus on research evaluation in the higher education sector (Box, 2010; Hicks et al., 2015). Dahler-Larsen (referring to Koselleck) makes the point that the components of evaluation systems are "in constant tension with the reality they seek to describe" (2012, p. 213). For many social scientists, the reality is that indicators and metrics used in research evaluation are unlikely to fully reflect the quality of their work. This tension at the individual level, and the potential of a "more results-oriented vocabulary" (Dahler-Larsen, 2014, p. 981) of academic environments to influence metrics use, is the focus of our study.

History and an abundance of literature demonstrate that bibliometrics are inadequate measures of research quality and impact for many fields in the social sciences. Yet a range of metrics are often components of research evaluation, at all levels in the higher education sector. When considered alongside peer review, quantifiable criteria for research evaluation have clear advantages in terms of the time required to gather data and the perception of numeric comparability. However, this perception is frequently founded on misunderstandings or misinterpretations of the metrics available and these metrics are better suited to fields for which the main form of research output is the journal article.

In 2004, Diana Hicks (p. 474) described the publications of social sciences as a "messy set of literature". Focusing on the limitations of bibliometrics, Hicks explained the problems for four categories of published outputs in the social sciences: journal articles, books, national literatures, and non-scholarly literature. Later research identifies specific types of social science literature, such as book chapters, conference papers, magazines, reports, cases and theses (van Leeuwen, van Wijk, & Wouters,

2016). While these literatures each play an important role in scholarly communication for social scientists, they were, and remain, poorly indexed or not indexed at all in the main bibliometric tools, Web of Science and Scopus. As a result, effective evaluation of the published output of social scientists is more complex than for the sciences and this presents a challenge for scholars wishing to show evidence of the value of their work.

Yet, very little is known about if and why individual social scientists use metrics, and how they employ specific metrics in their work. The perception and use of metrics has been studied in medicine and science fields (Aksnes & Rip, 2009; Derrick & Gillespie, 2013), while others have included the social sciences in their samples (Buela-Casal & Zych, 2012; Hargens & Schuman, 1990). With the exception of a recent paper (Ma & Ladisch, 2016), multidisciplinary studies have focused on a single bibliometric indicator. Our approach is narrower, in that social scientists form the study population, but broader in its inclusion of a range of indicators and metrics. This paper explores: (1) the extent of indicator and metrics use; (2) the purpose and context of use; and (3) features of the research evaluation environment that appear to be associated with use. An online questionnaire gathered data from social scientists working in Australia and Sweden, which presented the opportunity to explore indicator and metrics use by scholars from different national academic cultures and languages, and with different research evaluation systems in place.

Background

Bibliometrics, indicators and research evaluation in the social sciences

The diverse publishing practices of the social sciences limits the use of indicators and metrics in the social sciences. Books, book chapters, reports, conference papers, and a range of others are not uncommon publication forms (Engels et al., 2012; Gumpenberger et al., 2016; Huang & Chang, 2008; Laudel & Gläser, 2006; Lindholm-Romantschuk & Warner, 1996; Verleysen & Weeren, 2016) and only a tiny proportion of these are covered by Web of Science and Scopus, the most important citation indexes. Over the last decade, the potential for alternative metrics to play a role in evaluating a range of publication types has been proposed, however, the new metrics are yet to prove themselves as credible tools (Wouters et al., 2015).

Thus, citation counts and related indicators are either unavailable or compromised as research evaluation criteria for much of the social science journal literature. Some fields, particularly psychology and economics, are exceptions and are more widely indexed (Butler, 2008; Hicks, 2004; Huang & Chang, 2008). Given that the vast majority (95.45%) of social science articles indexed by Web of Science are in English (Gingras & Mosbah-Natanson, 2010), it is not surprising that these fields tend to publish in highly regarded English-language international journals (Verleysen & Weeren, 2016).

Writing for national audiences in the national language is common to many social science fields (Hicks, 2004; Sivertsen, 2016), although there appears to be a move away from publishing in non-English languages (Engels, Ossenblok, & Spruyt, 2012; Gumpenberger, Sorz, Wieland, & Gorraiz, 2016; Hammarfelt & de Rijcke, 2015). In the Nordic context, Olsson and Sheridan (2012) raised concerns that research has been overlooked when published in the national language. There are a number of possible explanations for this trend, such as perceptions that non-English language

journals are less scholarly (Hicks & Wang, 2011), and that English language articles have higher impact (van Leeuwen, 2013); each feeding into research evaluation criteria.

Like indexing coverage, publication in languages other than English is more common to some social science fields, such as law (de Jong, van Arensbergen, Daemen, van der Meulen, & van den Besselaar, 2011). The importance of national language publications has been acknowledged in the development of journal lists for research evaluation, and while the lists include titles that are not indexed by major citation databases, they are not without their own set of problems. Criticisms include flawed methods for their development and the Impact Factor-like assignment of a journal rank on individual articles (Ferrara & Bonaccorsi, 2016; Rafols, Leydesdorff, O'Hare, Nightingale, & Stirling, 2012; Tourish & Willmott, 2015). It is worth noting that scholars in smaller English-speaking countries, like Australia, can also be affected by indexing coverage and ranked journal lists (Genoni & Haddow, 2009).

Despite the limitations of bibliometrics and other indicators, scholars across all disciplines are inclined to use them. Wouters (2014, p. 56) notes the “ambivalent attitudes with respect to performance and citation indicators” evident in previous studies, however, research into the use and perceptions of the Impact Factor (Buela-Casal & Zych, 2012) and citations (Hargens & Schuman, 1990) found that social scientists use these citation indicators and, with some qualifications, regard them as reflecting quality. A more recent study (Ma & Ladisch, 2016) touches on issues that are of interest in this paper, such as scholars’ strategies relating to perceptions of indicators, including alternative metrics. The preliminary findings suggest the same ambivalent attitude to metrics, but they are used by scholars nevertheless.

Dahler-Larsen (2012, p. 14) describes evaluation as having “the potential, to facilitate a new view of customary practice that breaks with existing habits and convictions” and the effects of different evaluation systems can be wide-ranging (de Rijcke, Wouters, Rushforth, Franssen, & Hammarfelt, 2016; Lewis & Ross, 2011). Clearly, institutional policies and processes that include incentives (or disincentives) to comply are highly likely to influence individual responses (Aagaard, Bloch, & Schneider, 2015; van Dalen & Henkens, 2012).

Research evaluation systems in Australia and Sweden

In Australia, a new national research evaluation system was trialed in 2010 (Australian Research Council, 2017). The Excellence in Research for Australia (ERA) was accompanied by a ranked journal list, introducing quality as a measure for journal publications. By 2012, the ranking of journals had been abandoned, although the list of approved titles remains as a criterion of eligibility for assessment. Quality is assessed through peer review for all social science fields, with the exception of psychology, for which citation-based metrics are applied. However, in the development of the ERA, metrics were proposed as an indicator for all fields (Butler, 2008) and they have been the subject of research and discussion in Australia for much longer (Bourke, 1994; Royle & Over, 1994).

Books, book chapters, conference papers and creative works are eligible for assessment, but journal articles comprise the greatest proportion of assessed social sciences submissions (Turner & Brass, 2014). The ERA Journal List is a key tool

(<http://www.arc.gov.au/era-2018>). Titles are classified with between one and three Field of Research codes (Australian Bureau of Statistics, 2008; Haddow, 2015), and this classification is responsible for allocating research outputs to fields. It is significant that although outcomes of an ERA assessment - ratings of Fields of Research by institution - have not been linked to funding to date, achieving high ratings is an important driver of research at Australian universities (Knott, 2015).

Sweden introduced a new system for research funding allocation to higher education institutions in 2009, which is currently under revision. Unlike the Australian model, it assesses research across all fields using citations and number of publications as measures (Hammarfelt, Nelhans, Eklund, & Åström, 2016). This model has influenced funding allocation at an institutional level, with most Swedish institutions using bibliometric indicators to distribute funds. Some universities have implemented the Norwegian model, which incorporates both quality, through a two-tiered journal and publisher list, and quantity. Like the national system, this model “emphasizes international and peer reviewed publications” (Hammarfelt & de Rijcke, 2015, p. 66), and although English as language of publication is not a specified criterion, the international component of the model means that English is likely to be the language of publication for many journals.

There are numerous terms, such as ‘indicators’ and ‘metrics’, used interchangeably in the literature of bibliometrics and research evaluation, and it is beyond the scope of this paper to attempt to differentiate between them. Unless we are discussing a specific measure, the term ‘metrics’ is used here to encompass the range of tools used in the evaluation of research, whether they are citation-based metrics or other indicators, such as the Norwegian ‘publication indicator’ (Aagaard et al., 2015), the now-defunct ranked ERA Journal List (Australian Research Council, 2017), and the ABS Academic Journal Guide (Association of Business Schools, 2015).

Research design

This paper is reporting on a subset of a broader study involving social science and humanities (SSH) scholars in Australia and Sweden. The humanities subset was explored separately and is reported elsewhere (authors, in press). Due to the large SSH populations in Australia and Sweden (17,840 and 7,791, respectively), an online questionnaire was selected as the most effective research instrument. A survey developed for arts and humanities scholars at a Swedish university (Hammarfelt & de Rijcke, 2015) provided a base from which questions were added, modified and omitted for the SSH study.

Demographic questions for country, academic position, time in academia and gender, allowed for the testing of associations across variables. For identifying research field, we used the research classification schemes relevant to Sweden, the revised Field of Science and Technology Classification (OECD, Organisation for Economic Co-operation and Development, 2007), and Australia, the Australian and New Zealand Standard Research Classification (Australian Bureau of Statistics, 2008). An extensive list of metrics was developed as response options and three open questions were designed to collect qualitative data about metrics use and changes in publication practices. (See Appendix 1 for an abridged version of the survey). Approval to conduct the research was gained from Curtin University’s Human Research Ethics Committee.

In order to attract a good response rate several strategies were implemented, such as keeping the survey as brief as possible and distributing the initial invitation to scholars with 'legitimate authority' (academic line managers) over the study population (Dilman, 2006, p. 20). The variety of university organizational structures and unit titles, as well as the time required to identify SSH academic areas meant that units with the title 'health science' were excluded. It is likely that some SSH-related sub-units (eg., psychology and social work) were therefore omitted from the invitation to participate. Individually addressed email invitations were sent to 134 Australian and 58 Swedish line managers, at 38 and 18 universities, respectively, requesting that they distribute the invitation. After four weeks and reminder emails, the survey had gathered only 503 responses. Using what was essentially a random sample of universities - those from which no responses were received - invitations to all SSH scholars (approximately 6,000) at the non-responding universities with email addresses available were distributed. The universities were representative of their country's higher education sector and included research-intensive universities, regional universities, and more teaching oriented institutions. This second recruitment round doubled the responses to 1189. Of these, 160 responses without university data were excluded, leaving a total of 1029 completed surveys from SSH scholars. The responses from social scientists accounted for 581 of these.

Like Hicks' (2004) description of its literature, a definition of social science is a little messy. For the purposes of this study, social science fields were defined by the OECD classification scheme, which comprises eight fields: Economics & Business, Educational Sciences, Law, Media & Communications, Political Science, Psychology, Social & Economic Geography, and Sociology, as well as an overarching 'Social Science' and 'Other' classification. A total of 581 participants had noted one or more of these classifications: 249 (42.9%) Australians and 332 (57.1%) Swedish scholars. As a proportion of all social scientists in Australia (11,049) and Sweden (5,228), our respondent sample represents approximately 2.25% and 6.35% of the total population, respectively (Turner & Brass, 2014; Statistics Sweden, 2017).

The quantitative data analysis was performed using the survey software's analysis functions and Excel to produce descriptive statistics and to explore some cross tabulations. Comments made in response to the open questions were exported into documents in order to conduct manual counts and to extract quotes relating to identified themes.

In the findings section, we first discuss the demographic profile of the participant group, followed by the results for extent of use and types of metrics used. We then discuss the purpose and context of use, and features of the research evaluation environment that appear to be associated with use. The results are reported for the full set of responses (581) and, in some cases the eight fields (477).

Demographic profile of the participant group

Swedish social scientists participated in the study in higher numbers (57.1%) than their Australian peers (42.9%), and the survey appeared to attract scholars in more senior positions, with over 45% at the professor and associate professor levels. In terms of academic age, the highest number of participants was in the 5-10 year range (25.3%), followed closely by researchers who had worked in academia for over 20

years (23.4%). Females participated in slightly higher numbers (50.1%) than males (47.7%), with some scholars choosing not to identify with a gender. English was the preferred language of publication, with only 58 (17.4%) Swedish scholars indicating they preferred their national language.

For the analysis of discrete fields, 477 of the 581 scholars had indicated their research area as one or more of the eight fields (522 instances of fields). Another 49 noted the general Social Sciences and 56 noted the Other Social Sciences classifications only. These broad fields were not analyzed separately. Economics & Business was the largest field with 139 participants, almost double the size of Educational Sciences (79) and Law (72). Scholars identifying with the remaining fields are: Sociology – 65; Political Science – 53; Psychology – 43; Media & Communications – 41; and Social & Economic Geography - 30.

When the demographic data are compared by country (Fig. 1), minor differences are apparent for gender, with stronger variations evident for position and academic age. The much higher number of Australians at the lecturer position potentially relates to the way the title is used in Sweden, which is frequently associated with teaching rather than research positions. It is possible the survey was regarded as less relevant to these academics. The much higher number of Australians with an academic age of over 20 years is less explicable. Due to the low numbers, postdoctoral, doctoral student, and research assistant responses are not presented in the Figure, however, a relatively high proportion of the Swedish respondents were doctoral students (14.8%), compared to the Australian group (1.2%). This may be a product of distribution lists used in the recruitment of respondents. It may also relate to the trend in Sweden for PhD students to publish during or as part of their doctoral work – perhaps considering themselves ‘researchers’ – whereas thesis by publication is less common in Australia.

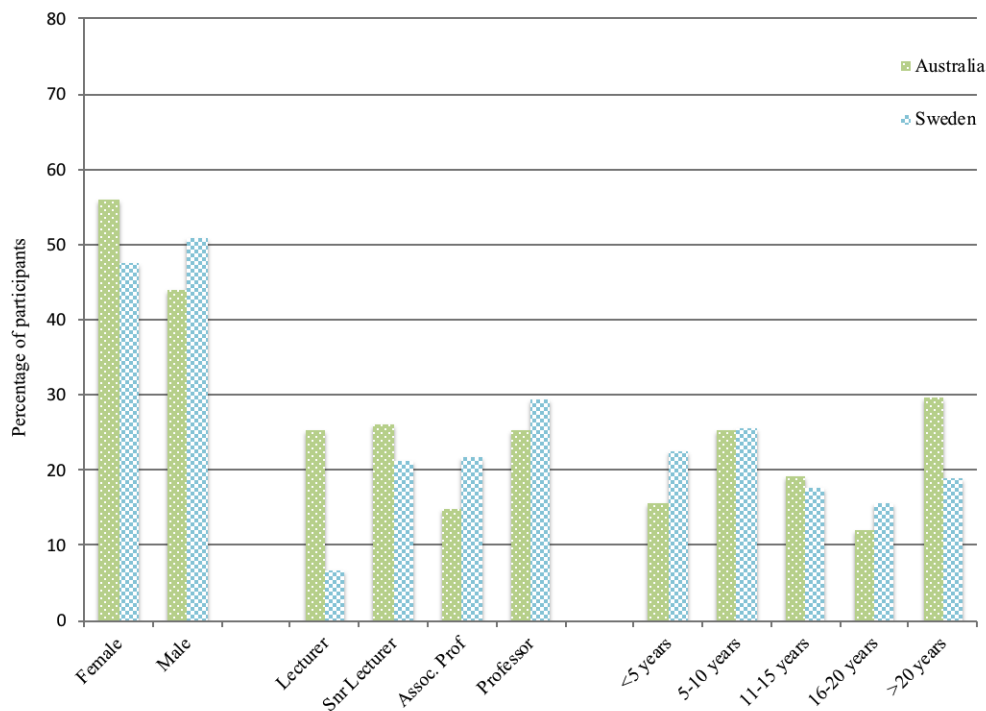


FIG. 1 Demographic profile of participating social scientists by country (Australia $n=249$, Sweden $n=332$)

Very few of the participants did not respond to the demographic questions, however the number of responses to later questions varied and this is indicated with n in the results below.

Findings

Metrics use: Extent and types of tools used

When asked if they used one or more of a range of metrics in CVs, promotion and grant applications almost half (47.2%) of the social science participants ($n=517$) responded affirmatively. However, a major difference in metrics use was found when the countries were compared, with 67.5% of the Australian group ($n=231$) and 31.1% of the Swedish scholars ($n=286$) reporting they had used one or more of the metrics listed.

We examined demographic factors to identify any possible associations with use and found that professors and associate professors reported using metrics in higher than expected numbers. Scholars with an academic age of less than five years reported much lower use than the other ages, 17.6% compared with a range of 45.9% - 57.1%, which likely reflects their limited opportunity to submit grant applications and apply for promotion.

When use was analyzed for field and country (Figure 2), Law scholars reported the lowest use, although the majority of Australian Law participants reported using metrics; as did all Australian social science fields. Australian Psychology and Sociology scholars used metrics in lower proportions than the overall use reported by Australians. The exclusion of health science units from the recruitment process almost certainly resulted in fewer psychology scholars responding to the survey, and those who did are potentially more aligned with qualitative approaches. In relation to the Sociology scholars' use, our findings are higher than those reported previously for citations use (Hargens & Schuman, 1990). On the basis of previous studies, Economics scholars would be expected to report high use and we speculated that the inclusion of Management in the classification influenced our results. To test this, we examined the ANZSRC fields selected by the Economics & Business scholars and found that approximately 60% of the cohort identified with fields outside of Economics. Overall, the difference between countries clearly exceeds those concerning fields. At the same time there are indications that some fields – like Law – are less inclined to use metrics.

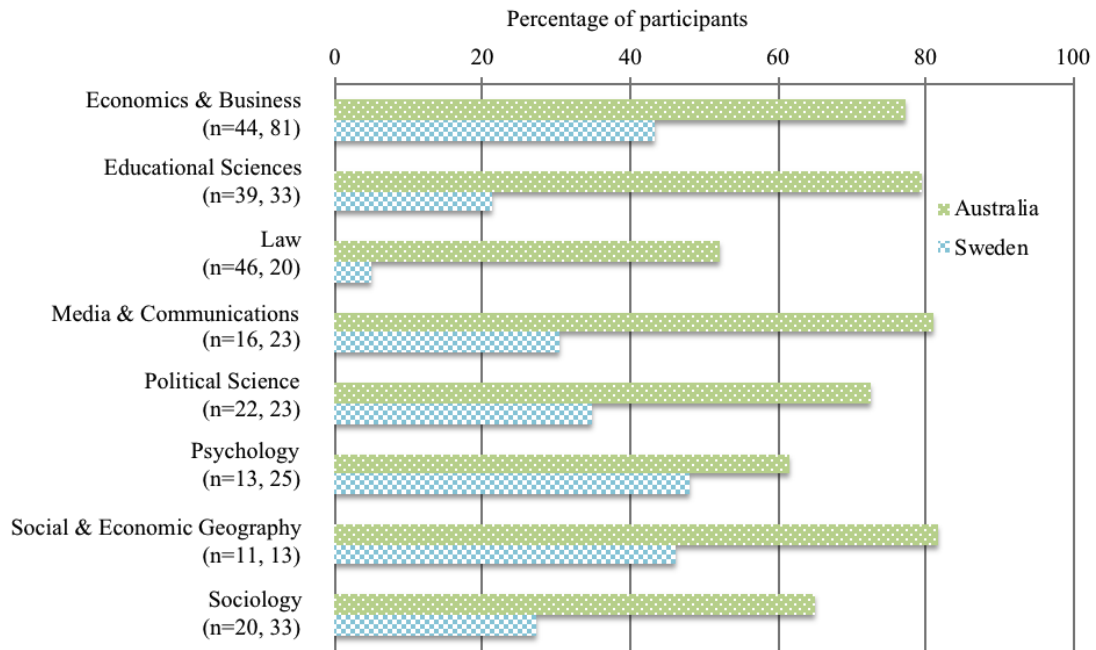


FIG. 2. Use of metrics by field and country

An open question about which metrics were used and for what purpose provided data about specific tools. Researchers who responded to this question had all reported using metrics in CVs, promotion and/or grant applications. Manual counting was performed to identify the metrics mentioned most often and these are presented as a percentage of the number of responses to the question for the Australian and Swedish groups (Figure 3).

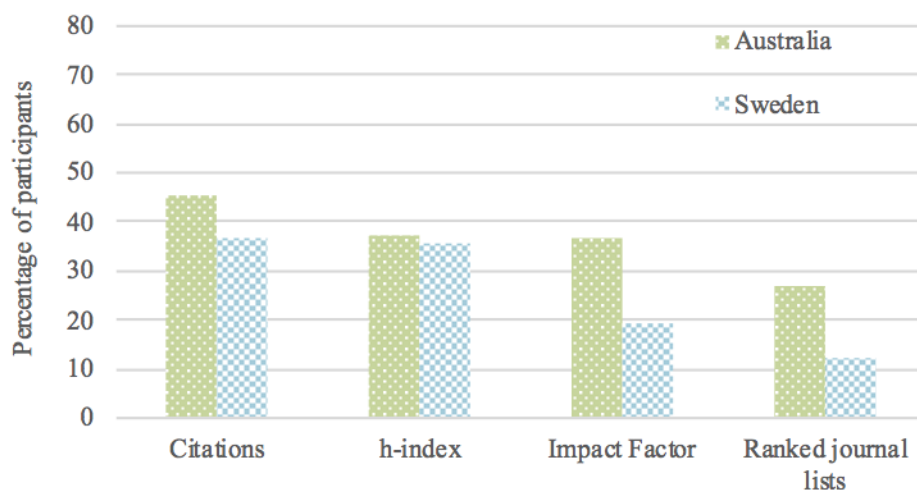


FIG 3. Metrics used by country (Australia $n=148$, Sweden $n=82$)

Although there is only a slight difference between the Australian and Swedish groups' reported use of citations and the *h*-index, the Australians mentioned the Impact Factor and ranked journal lists in much higher proportions. The poor coverage on non-English language journals by Web of Science means that very few Swedish-language

journals will have an Impact Factor. Indeed, the *Source Publication List* for the Social Sciences Citation Index lists only two journals in Swedish (2017, http://mjl.clarivate.com/publist_ssci.pdf).

The use of ranked journal lists is another point of difference. Although the ERA ranked journal list was discontinued in 2012 a good number of Australians mentioned it, along with the Australian Business Deans Council journals list. In contrast, only four Swedish participants mentioned the Norwegian list. Other lists, such as the Financial Times top 50, SCImago Journal Rank, and the UK business journal rankings were also noted by a few researchers. Library holdings, downloads and altmetrics were listed only once. Some scholars were less specific and recorded the names of sources, such as social media platforms, Google Scholar, Web of Science and Scopus, without an associated metric. Google Scholar was the most frequently recorded source by scholars from both countries.

In Figure 4, the four metrics presented in Figure 3 are broken down for the social science fields. Psychology scholars mentioned citations and the *h*-index in much higher proportions than any other field. Citations were used by the majority of Political Science scholars, as was the Impact Factor by researchers in Social & Economic Geography. Law scholars most frequently used journal lists and the Sociology field exhibited relatively even use of the citation-based tools. Somewhat surprising is the finding that only a quarter of the Economics & Business scholars noted ranked journal lists, as this field tends to use discipline based lists (Hammarfelt & Rushforth, 2017).

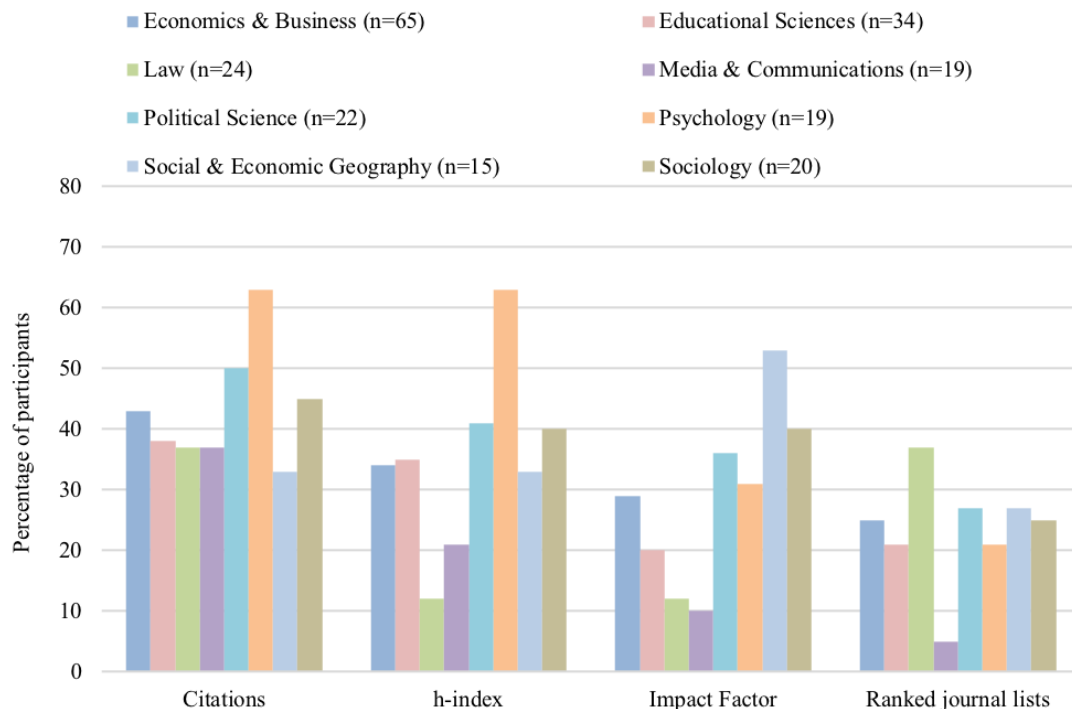


FIG. 4. Metrics used by field

Metrics use: Purpose and context

We identified the purpose and context in which metrics were used from 227 responses to the open questions. Of these, 149 were Australian and 78 Swedish. Scholars from both countries referred to the examples of use (CVs, grant and promotion applications) included in the survey, some providing context around that use. For instance, an Australian used Google Scholar citations *'to embed in a narrative for grant applications'*, and another had used the *h-index*, *'heavily qualified'*, but *'necessary given the scientific lean of the promotion panel'*.

Additional purposes of use were noted, the most frequent being annual (performance) reviews and reporting to the institution. Citations and the Impact Factor were used by an Australian to show *'I'm hitting institutional targets'*, and another Australian had used metrics *'to satisfy the uni that I am following their directions and requirements'*.

Many social scientists described their use of metrics with terms that attribute a value and we identified several recurrent terms - quality, impact, and quantify. Quality was associated with all the main metrics presented in Figure 3 and generally related to publications, for example *'to highlight the quality of publications'* and *'h-index which is best correlated with quality in the current publication climate'*. One participant had used the ERA journal list for a promotion application and wrote *'shorthand for quality even though it isn't'*.

Impact was associated with the scholars' research and was employed in a less descriptive manner, such as *'to demonstrate impact'* and *'to indicate the impact of my work'*. However, the comments of a number of scholars hinted at specific uses, such as *'to demonstrate international impact'*, and altmetrics had been used to demonstrate *'wider impact'*. When using the term quantify it was in a similarly limited form as impact, although a Swedish scholar noted using the Impact Factor and *h-index* to *'quantify the spread'* of their work. Scholars also used metrics to benchmark or compare with others in their field, to indicate reputation and esteem, to provide evidence of scholarly activity, *'capacity for institutional mobility'*, and to *'sell'* their work.

Evidence of a rather more directive environment can be seen in the comments of two Swedish scholars; one stating they were *'forced to use an h-index'* and the other had used citations *'because counting them was required'*. In the same vein, an Australian wrote *'we are mandated to use'* Google Scholar, Scopus and the Journal Citation Reports. Another Australian indirectly referred to their institution as an important driver in the use of citation-based indicators, with: *'to ward off any criticisms that ... articles are not in A* or A ranked journals'*.

Mentions of specific alternative metrics were few; instead the social media platforms that generate metrics were referred to. An Australian scholar's comment demonstrates a high level of awareness of different sources of metrics, listing ResearchGate and Academia *'stats because they can be better than Google Scholar and Scopus for my discipline'*.

Factors associated with the use of metrics

Scholars commented on metrics, publishing trends and expectations of their institution or broader research environment in the open questions. In some instances, it was difficult to disentangle the different aspects of academic lives to identify

influencing factors in the use of metrics, however two interconnected matters emerged strongly: an increased focus on journal publications and research policy at institutional and national levels.

The emphasis on journal publications was associated with ranked journal lists and indexing by the main citation databases, which often equates to English language, international journals. Australian participants were particularly candid about ranked journal lists, while their Swedish peers commented more frequently on international journals and the indexing coverage of journals, although a few referred to the Norwegian and the ABS journal lists.

A change in publication practices was evident across all social science fields, even Law, for which *'journal metrics were not important a decade ago'* and *'bibliometrics and rankings is [sic] getting increasingly important'*. In general, scholars noted a *'decline in the importance of books'*, and in Sweden a shift *'from monographs in Swedish to articles in international journals'* and *'increased importance on publishing in English'*. On this point, only three of the Swedish scholars who reported using metrics indicated that their preferred language of publication was Swedish.

Early career researchers (ECRs) in both countries remarked on the pressure to publish with specific metrics in mind, illustrated by: *'in the past, just having journal articles as an ECR was important, now it appears that the journals need to be well known'* and *'I am trying to combat the shift towards international publication by also writing in Swedish, but it is hard when you are at an early stage in your career: you need those Norwegian listed journal publications'*.

As the findings in the previous section indicate, metrics are used in CVs, promotion and grant applications and this use could be a matter of individual choice. However, there was evidence that grant agency requirements, at least in some cases, were a factor and the research environment was heavily criticised for the *'neurotic attention to ranking indicators in recent years'* and *'an obsession with quantifiable research outputs'*. In Australia, this was expressed most often in association with the continued use of the ERA ranked journal list by institutions to reward (or penalise) scholars for their research outputs, articulated in the following comment:

The insidious ranking of academic journals in Australia ... the Faculty uses the repudiated ARC ranking of academic journals to dispense financial rewards for research and conference attendance, and to reduce teaching time. The whole matter creates gross perverse incentives.

Many social scientists wrote of their institutions' policies and performance measures relating to preferred research outputs (journal articles) and the *'value'* and *'relevance'* attributed to different publication types. For example, Australian scholars wrote: *'journal articles have increased in value in terms of promotion'* and *'my uni no longer sees books or book chapters as relevant or even counts them among outputs'*. An awareness of the limitation of metrics being used for performance measurement was also noted: *'The University appears obsessed with citation indexes, and metrics of all sorts, as proxies for 'quality' research'* and *'I now aim for journals formerly ranked as 'A' under the officially defunct ERA journal ranking system. This is what my university uses to evaluate performance in the absence of any other useful measure.'*

Swedish scholars wrote of similar institutional policies, again with a focus on journal articles as the preferred publication. Their comments included: *‘institution policy is to reward publication in journals that are listed in Web of Science or at least on the Norwegian list’*; *‘more and more international, peer-reviewed articles due to policy for allocation of state research funding to web-of-science [sic] articles’*; and *‘nowadays much more focus on indexing in Scopus/WoS due to bibliometric evaluation of performance’*.

Discussion

It has been widely acknowledged that the scholarly communication practices of many social science fields are largely incompatible with applying metrics. Yet, a substantial number of social scientists in our study, particularly in Australia, have used these tools. Our data do not provide an explanation of why Swedish social scientists are less inclined to use metrics, however we suggest two possibilities. Firstly, the difference between how each country has implemented their national research evaluation model is notable. In Australia, the creation of the ranked journal list to assess all article outputs in the ERA has attracted a great deal of discussion and debate (Bennett, Genoni & Haddow, 2011; Knott, 2015), which, as this study shows, remains a contentious issue today. On the other hand, although the Swedish national research evaluation model uses citations data to assess all disciplines, the model is not fully integrated into all institution’s research performance systems and therefore there is less expectation that scholars use metrics for other purposes. A second, and perhaps a more fruitful line of thought, is that language of publication is a factor. Swedish participants noted a shift to more English language (journal) publishing, which lends weight to an argument that due to the coverage of the major citation databases, metrics were largely unavailable in the past and therefore not habitually used. The emphasis on English language evident in the Swedish responses suggests, however, that their use of metrics is likely to increase in the future.

When considering metrics use at field level, we acknowledge the low number of responses limits the authority of our findings. Nevertheless, they support previous work relating to citation coverage and publication practices for Psychology, which used metrics (citations and *h*-index) in the highest proportions, and Law, which used citation-based tools (*h*-index and Impact Factor) least (Butler, 2008; Verleysen & Weeren, 2016). It seems likely that Law’s much higher use of ranked journal lists is the flip-side of the low metrics use for the field. The research also confirms one of the few previous studies of citations use in the social sciences, with Sociology scholars using all the citation-based tools in relatively high numbers (45%) (Buela-Casal & Zych, 2012).

Terms that pervade the research evaluation vocabulary (quality, impact and quantify) appeared frequently in the scholars’ comments, and the context in which they were applied suggest pragmatism and compliance in relation to metrics use. Metrics were used in promotions and grants applications to match the scholars’ perceptions of assessors’ preferences and to demonstrate research achievements more widely. Performance reviews and meeting institutional targets were mentioned, and at times scholars’ responses were expressed as a defense of their work. Some chose to use a particular tool because it provided a ‘better’ metric for the field, while others sought to contend the prevailing policy by presenting alternative measures to highlight the value of their work. These actions and reactions by social scientists suggest a high

level of awareness of some metrics and strategic behavior in demonstrating research performance.

A cause and effect cycle may exist between scholars' use of metrics and the emphasis on journal publications by institutions. As Naylor wrote in 2001 (in relation to the Research Assessment Exercise), the journal article is 'king', and these publications are most able to provide quantifiable assessments in an environment that is focused on a "results-oriented vocabulary" (Dahler-Larsen, 2014, p. 981). In this environment, scholars need to be aware of the tools that will identify 'quality' journals and that may also mean using other metrics to support performance reviews, grant and promotion applications.

Burrows wondered if metrics were "experienced as oppressively" as he had suggested in his paper (2012, p. 369). The frank comments about institutional policies in relation to what was valued as a research output, together with the metrics relevant to those outputs, indicate that for approximately 15% of social scientists in this study the answer is yes. These scholars appear to be experiencing significant tensions in their academic life, evident in the Law participants' remarks about changes in their field and also in observations about the shift away from books to journal articles. As previous research has shown (Laudel & Gläser, 2006; Wouters, 2014), an ambivalent attitude towards using metrics is not uncommon, and Laudel and Gläser (p. 293) describe the tension emanating from "a message from the exogenous evaluation that directly contradicts the endogenous valuation of communication channels by their disciplines". In this study, fields that traditionally produce books are seeing the value of those research outputs diminishing, which may in turn lead to the breaking of 'existing habits and convictions' (Dahler-Larsen, 2012, p. 14). Tensions between intradisciplinary criteria of quality and external evaluation systems' measures was also found when studying humanities scholars and their attitudes towards metrics (authors, in press). Other tensions can be seen for young researchers, who appeared vexed by some expectations for academic achievement. This finding evokes Malsch & Tessier's (2015, p. 85) autoethnographic research about ranked journal lists in accounting, in which the term 'identity fragmentation' is used to describe junior researchers' experience of being driven "professionally and intellectually in contradictory directions".

Our study's findings are based on a very small proportion of social scientists working in Australia and Sweden, resulting in low numbers of participants in the eight fields. This affects the degree to which we can make strong claims about the use and attitudes to metrics in the social sciences. Furthermore, the study's topic may have attracted scholars who felt strongly about research evaluation, which would explain the many forthright comments provided. In relation to the social science field, the OECD classification created an element of ambiguity as to the research area with which participants identified. If a similar study was to be conducted, using free text for the field question may produce more precise results in order to compare differences across the social sciences.

Conclusion

Many social scientists use metrics, despite the literature and expert opinion that warns against it in most fields. However, there appears to be an awareness, at least for some, of the limitations of these tools in ascribing quality. While Australian social scientists

use metrics in much higher proportions than their Swedish peers, the valuing of journal articles in international publications in Sweden is likely to result in much higher use of metrics in the future.

The importance of institutional policies regarding incentives and disincentives in performance management and academic promotion came through clearly in the data. In both countries there is an increasing focus on journal articles and it appeared to be associated with indicators such as the ERA ranked journal list and citation-based indicators. This was accompanied by ambivalence towards metrics, with some scholars expressing resistance, disbelief or frustration with the research evaluation policies that are in place. Both institutional policies and those relating to research evaluation at a national level were mentioned and in many cases it is not possible to separate one from the other due to the expected trickle-down effect within a national higher education system. While, a number of scholars referred to specific metrics that are components of their respective national research evaluation system, the strong focus on institutional issues suggests several interconnected factors are associated with the use of metrics by social scientists in Australia and Sweden.

Finally, it is useful to consider how nations and institutions can alleviate the less positive experiences that appear to be associated with research evaluation and the metrics used to assess research. Institutions are likely to adopt aspects of a national model and they are also driven by factors such as university rankings. However, a balance needs to be struck between effective research policy and recognition of the non-metric contribution made by the social sciences. Adjustments are possible, and the introduction of engagement and impact assessment in Australia, in 2018, will provide social scientists with alternative methods to demonstrate the value of their work. Sweden has also mooted the use of ‘societal impact’ to assess research, which in turn may create more opportunities for publications in the national language.

Acknowledgement

Gaby Haddow would like to acknowledge Curtin University’s study program support for enabling her role in this research. Björn Hammarfelt’s contribution was supported by *Riksbankens Jubileumsfond: (The Swedish Foundation for the Social Sciences and Humanities)* (grant number SGO14-1153:1). We also wish to thank the reviewers for their generous comments and constructive feedback on our manuscript. Finally, we would like to express our gratitude to all the Australian and Swedish scholars who participated in our survey.

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.
- Aksnes, D.W., & Rip, A. (2009). Researchers’ perceptions of citations. *Research Policy*, 38(6), 895–905.
- Australian Bureau of Statistics (2008). *Australian and New Zealand Standard Research Classification (ANZSRC)*. Retrieved November 3, 2017, from

<http://www.abs.gov.au/ausstats/abs@.nsf/0/6BB427AB9696C225CA2574180004463E>.

Australian Research Council (2017). *Excellence in Research for Australia*. Retrieved November 3, 2017, from <http://www.arc.gov.au/excellence-research-australia>.

Bennett, D., Genoni, P., & Haddow, G. (2011). FoR codes pendulum: Publishing choices within Australian research assessment. *Australian Universities' Review*, 53(2), 88-98.

Box, S. (2010). Performance-based funding for public research in tertiary education institutions: Country experiences. *OECD, Performance-based funding for public research in tertiary education institutions: Workshop proceedings*. Retrieved November 3, 2017, from <http://dx.doi.org/10.1787/9789264094611-6-en>.

Buela-Casal, G., & Zych, I. (2012). What do scientists think about the Impact Factor? *Scientometrics*, 92(2), 281-292.

Bourke, P. (1994, May 19-25). Quantitative research indicators: Citations treatment 'deficient'. *Campus Review*, p. 9.

Burrows, R. (2012). Living with the *h*-index: Metric assemblages in the contemporary academy. *The Sociological Review*, 60(2), 355-372.

Butler, L. (2008). Using a balanced approach to bibliometrics: Quantitative performance measures in the Australian Research Quality Framework. *Ethics in Science and Environmental Politics*, 8, 83-92.

Dahler-Larsen, P. (2012). *The evaluation society*. Stanford, CA: Stanford University Press.

Dahler-Larsen, P. (2014). Constitutive effects of performance indicators: Getting beyond unintended consequences. *Public Management Review*, 16(7), 969-986.

de Jong, S.P.L., van Arensbergen, P., Daemen, F., van der Meulen, B., & van den Besselaar, P. (2011). Evaluation of research in context: An approach and two cases. *Research Evaluation*, 20(1), 61-72.

de Rijcke, S., 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.

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 S. Hinze & A. Lottman (Eds), *Proceedings of the 18th International Conference on Science and Technology Indicators* (pp. 104-116). Berlin.

Dilman, D.A. (2006). *Mail and internet surveys: The tailored design method*, (2nd ed.). Hoboken, NJ.: Wiley.

- Engels, T.C., Ossenblok, T.L., & Spruyt, E.H. (2012). Changing publication patterns in the social sciences and humanities, 2000–2009. *Scientometrics*, 93(2), 373–390.
- Ferrara, A., & Bonaccorsi, A. (2016). How robust is journal rating in humanities and social sciences? Evidence from a large-scale, multi-method exercise. *Research Evaluation*, 25(3), 279-291.
- Genoni, P., & Haddow, G. (2009). ERA and the ranking of Australian humanities journals. *Australian Humanities Review*, 46, 7-26.
- Gingras, Y., & Mosbah-Natanson, S. (2010). Where are the social sciences produced. *World social science report: Knowledge divides* (pp. 149-53). UNESCO Publishing and International Social Science Council, Paris, France. Retrieved November 5, 2017, from <http://unesdoc.unesco.org/images/0018/001883/188333e.pdf#page=120>.
- Gläser, J. (2007). The social orders of research evaluation systems. In R. Whitley & J. Gläser (Eds), *The changing governance of science: The advent of research evaluation systems* (pp. 245-266). Dordrecht, The Netherlands: Springer.
- Gumpenberger, D., Sorz, J., Wieland, M., & Gorraiz, J. (2016). Humanities and social sciences in the bibliometric spotlight: Research output analysis at the University of Vienna and considerations for increasing visibility. *Research Evaluation*, 25(3), 271-278.
- Haddow, G. (2015). Research classification and the social sciences and humanities in Australia: (Mis)matching organizational unit contribution and the impact of collaboration. *Research Evaluation*, 24(3), 325-339.
- Hammarfelt, B., & de Rijcke, S. (2015) Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the Faculty of Arts at Uppsala University. *Research Evaluation*, 24(1), 63–77.
- Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities. *Research Evaluation*, 25(3), 292-305.
- 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.
- Hargens, L.L., & Schuman, H. (1990). Citation counts and social comparisons: Scientists' use and evaluation of citation index data. *Social Science Research*, 19(3), 205–221.
- Hicks, D. (2004), The four literatures of social science. In H.F. Moed, W. Glänzel, & U. Schmoch (Eds), *Handbook of quantitative science and technology research*

- (pp. 473–496). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Hicks, D., & Wang, J. (2011). Coverage and overlap of the new social sciences and humanities journal lists. *Journal of the American Society for Information Science and Technology*, 62(2), 284-294.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., and Rafols, I. (2015). Leiden Manifesto. *Nature*, 520, 429-431.
- Huang, M., & Chang, Y. (2008). Characteristics of research output in social sciences and humanities: From a research evaluation perspective. *Journal of the American Society for Information Science and Technology*, 59(11), 1819-1828.
- Knott, M. (2015, November 17). Universities questioned over alleged ‘gaming’ of research rankings. *Sydney Morning Herald*. Retrieved November 5, 2017, from <http://www.smh.com.au/national/education/universities-questioned-over-alleged-gaming-of-research-rankings-20151117-gl0yva.html>.
- Laudel, G., & Gläser, J. (2006). Tensions between evaluations and communication practices. *Journal of Higher Education Policy and Management*, 28(3), 289-295.
- Lewis, J.M., & Ross, S. (2011). Research funding systems in Australia, New Zealand and the UK: Policy settings and perceived effects. *Policy & Politics*, 39(3), 379–398.
- Lindholm-Romantschuk, Y., & Warner, J. (1996). The role of monographs in scholarly communication: An empirical study of philosophy, sociology and economics. *Journal of Documentation*, 52(4), 389-404.
- Ma, L., & Ladisch, M. (2016). Scholarly communication and practices in the world of metrics: An exploratory study. *Proceedings of the Association for Information Science and Technology*, 53(1), 1-4.
- Malsch, B., & Tessier, S. (2015). Journal ranking effects on junior academics: Identity fragmentation and politicization. *Critical Perspectives on Accounting*, 26, 84-98.
- Naylor, B. (2001). The king is in the altogether? *Serials*, 14(2), 149-153.
- Olsson, A., & Sheridan, V. (2012). A case study of Swedish scholars’ experiences with and perceptions of the use of English in academic publishing. *Written Communication*, 29(1), 33-54.
- Organisation for Economic Co-operation and Development (2007), *Revised Field of Science and Technology (FOS) Classification in the Frascati Manual*, OECD, Paris. Retrieved November 5, 2017, from <https://www.oecd.org/science/inno/38235147.pdf>.
- Rafols, I., Leydesdorff, L., O’Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy*, 41(7) 1262-1282.

- Royle, P., & Over, R. (1994). The use of bibliometric indicators to measure the research productivity of Australian academics. *Australian Academic & Research Libraries*, 25(2), 77–88.
- Sivertsen, G. (2016). Patterns of internationalization and criteria for research assessment in the social sciences and humanities. *Scientometrics*, 107(2), 357-368.
- Statistics Sweden (2017). *Employees in higher education 2016*. Swedish Higher Education Authority. Retrieved November 5, 2017, from <https://www.scb.se/publication/31450>.
- Tourish, D., & Willmott, H. (2015). In defiance of folly: Journal rankings, mindless measures and the ABS Guide. *Critical Perspectives on Accounting*, 26, 37-46.
- Turner, G., & Brass, K. (2014), *Mapping the humanities, arts and social sciences in Australia*. Academy of the Humanities, Canberra, ACT. Retrieved November 5, 2017, from <https://www.humanities.org.au/issue-item/mapping-humanities-arts-social-sciences-australia>.
- van Dalen, H.P., & Henkens, K. (2012). Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. *Journal of the American Society for Information Science and Technology*, 63(7), 1282-1293.
- van Leeuwen, T. (2013). Bibliometric research evaluations, Web of Science and the social sciences and humanities: A problematic relationship? *Bibliometrie – Praxis und Forschung*, Band 2. Retrieved November 5, 2017, from <http://www.bibliometrie-pf.de/article/view/173>.
- van Leeuwen, T.N., van Wijk, E., & Wouters, P.F. (2016). Bibliometric analysis of output and impact based on CRIS data: A case study on the registered output of a Dutch university. *Scientometrics*, 106, 1-16.
- Verleysen, F.T., & Weeren, A. (2016). Clustering by publication patterns of senior authors in the social sciences and humanities. *Journal of Informetrics*, 10(1), 254-272.
- Wouters, P. (2014). The citation: From culture to infrastructure. In B. Cronin & C.R. Sugimoto (Eds), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact* (pp. 47-66). Cambridge, MA: MIT Press.
- Wouters, P. et al. (2015). *The metric tide: Literature review* (Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management), HEFCE. DOI: 10.13140/RG.2.1.5066.3520.

Appendix 1: Survey

1. In which country are you an academic?

Australia|Sweden

2. What is the name of your university? (free text)

3. What is your academic position?

Professor|Associate Professor (Reader)|Assistant Professor/Senior Lecturer|Lecturer|Postdoc|Doctoral student|Research assistant|Other

4. Gender

Female|Male|Do not identify with any of the above|Do not wish to specify

5. Please select up to three research classification terms from both of the lists below that best describes your field of research.

OECD classification (whole list)

ANZSRC Fields of Research (whole list)

6. What is your preferred language when publishing?

English|Swedish|German|Dutch|French|Spanish|Other (please specify)

7. Researchers in your field most often publish: (Rank first three)

Book chapters
Conference papers
Creative works
Journal articles
Monographs

8. On average, how many of the following types of research-related outputs do you publish each year? Please use whole numbers regardless of whether your outputs are sole or co-authored works.

Blog posts|Book chapters|Book reviews|Creative works|Conference papers (published in full)|Edited books|Monographs|Newspaper articles|Peer reviewed journal articles|Peer review reports, e.g. article reviews|Policy reports|Research proposals|Trade, professional journal articles|Tweets|Wikipages |Other, please specify

9. How important are the following factors to you when choosing a publication channel? (*Extremely important* | *Very important* | *Somewhat important* | *Not important* | *Not familiar with this*)

Peer review processes|Quality of peer review|Reputation of publishing channel|Open access|Publication speed|Maximum outreach to target audience|International visibility|National visibility|Department policy|Institutional policy|National policy|Requirements of funding agencies|Indexing by international database (eg Scopus, Web of Science)|Journal Impact Factor|Scimago Journal Rank|Journal ranking lists|

Inclusion in Norwegian list (SWE)|Inclusion in ERA Journal List (AUS)|Field of Research code assignment(AUS)

10. Are there other factors that are important to your selection of publication channels? (please specify)

11. How familiar are you with the following tools for evaluating research?

Citations (Web of Science, Scopus, Google Scholar)|Impact factor|Scimago Journal Rank|h-index|Norwegian list (SWE)|ERA journal list (AUS)| Other journal rankings, ratings|Other journal, publisher rankings|Altmetric services (Impact Story|Altmetrics.com)|Social platforms(Research Gate, Academia.edu, etc)

12. Have you used any of these evaluation tools in your CV, promotion applications, grant applications, etc.?

Yes|No

13. If yes, please list the tools you have used and for what purpose

14. Please comment on any changes in publication practice in your field over the last ten years

15. Please comment on any changes in your own publication practice over the last ten years