Original articles

A "basket of metrics"—the best support for understanding journal merit

Lisa Colledge, Chris James
Elsevier, Amsterdam, The Netherlands; l.colledge@elsevier.com

Abstract

Aim: To survey opinion of the assertion that useful metric-based input requires a "basket of metrics" to allow more varied and nuanced insights into merit than is possible by using one metric alone.

Methods: A poll was conducted to survey opinions (N=204; average response rate=61%) within the international research community on using usage metrics in merit systems.

Results: "Research is best quantified using multiple criteria" was selected by most (40%) respondents as the reason that usage metrics are valuable, and 95% of respondents indicated that they would be likely or very likely to use usage metrics in their assessments of research merit, if they had access to them. There was a similar degree of preference for simple and sophisticated usage metrics confirming that one size does not fit all, and that a one-metric approach to merit is insufficient.

Conclusion: This survey demonstrates a clear willingness and a real appetite to use a "basket of metrics" to broaden the ways in which research merit can be detected and demonstrated.

Keywords

Metrics, research assessment, impact factor, journal metrics, basket of metrics

Introduction

A group of publishers, editors and editorial organizations of scholarly journals, including EASE, state at the outset of The San Francisco Declaration on Research Assessment (DORA)[1] that, "There is a pressing need to improve the ways in which the output of scientific research is evaluated by funding agencies, academic institutions, and other parties". They go on to note that, "The Journal Impact Factor is frequently used as the primary parameter with which to compare the scientific output of individuals and institutions", which is a misuse of the original intention of the journal-level calculation.

The impact factor was first postulated in 1955[2] as a means to assist librarians in collection management. It is a useful metric in this regard but has since been used for purposes for which it is not suitable, for example as the primary means "to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions"[1]. It is not inherently incorrect to derive a measure of an article's or a researcher's performance based on where they are published; editors are recognised experts and their decision to publish an article in a journal associated with their own academic standing rightly lends credence to that article and its authors. The problems derive from assuming that the actual merit of an article and its authors equates to the impact factor of the journal in which it is published[3]. There are many instances of well-cited articles in journals with low impact factors, and uncited articles in journals with high impact factors.

Editors want their journals to be excellent, but what is considered "excellent" varies. Editors of international journals will find it important that their contributors are globally distributed, whereas editors of applied journals may focus on securing articles and readers from the corporate as well as academic sector. Journal excellence is multi-faceted and cannot be captured by any single metric. Useful metric-based input requires a "basket of metrics" to allow more varied and nuanced insights into performance than is possible by using the impact factor, or any one metric, alone[6]. Figure 1 illustrates a vision of the basket of metrics for journals, that encompasses the:

- Community that contributes to its content, that is, its editor(s), board and authors. A journal could promote their geographical and sector spread, and might also showcase metrics about their performance as individuals such as the *h*-index.

- Counts and characteristics of contributions. It may be important to focus on particular types of outputs, such as conference proceedings, raw research data or medical guidelines, or on the funders that have supported the published research.

- Consumption of content. A journal can be well cited, and citation-based measures such as impact factor, source-normalized impact per paper (SNIP)[7] and SCImago journal rank (SJR)[8] will remain important. There is also interest in metrics that provide a more immediate measure, and the inclusion of usage metrics in the "basket" can address this. Furthermore, if a journal's content is targeted at an international audience that is distributed across multiple sectors – corporate, government, and health as well as academic – then this should be apparent.

- Esteem: a journal's academic authority and reputation. Altmetrics draw on the referencing and conversations in scholarly tools, such as Mendeley and CiteULike.

- Impact outside the academic world. Journals that attract and publish content that leads to improvements in quality of life are recognised by media coverage and mentions in policy documents, for instance.

Journal performance also reflects the individual articles, and the researchers—authors, readers, reviews, and editors—associated with them. The basket accommodates this by offering metrics that apply to multiple entities, and not only to journals. For example, the metrics could be calculated for sets of related journals, or pieces of a journal such as a
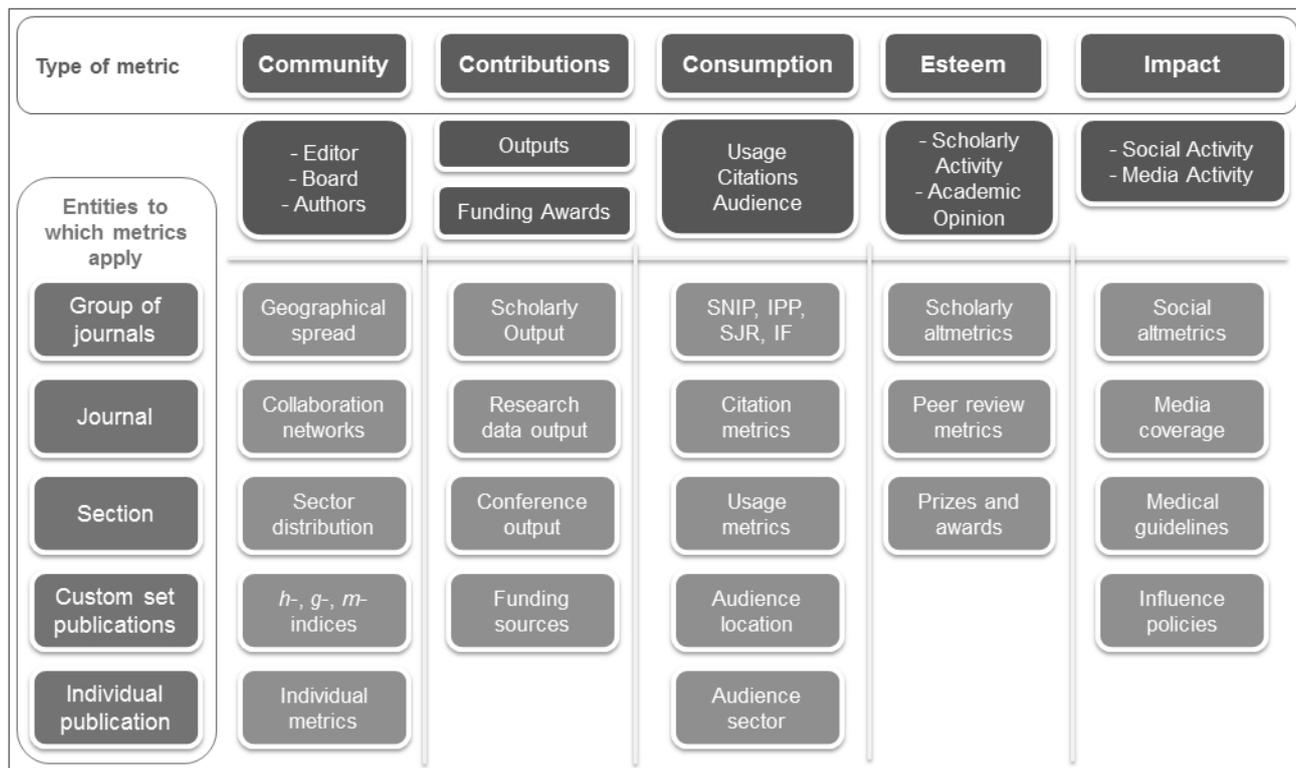
Type of metric	**Community**	**Contributions**	**Consumption**	**Esteem**	**Impact**
Entities to which metrics apply	- Editor - Board - Authors	Outputs Funding Awards	Usage Citations Audience	- Scholarly Activity - Academic Opinion	- Social Activity - Media Activity
Group of journals	Geographical spread	Scholarly Output	SNIP, IPP, SJR, IF	Scholarly altmetrics	Social altmetrics
Journal	Collaboration networks	Research data output	Citation metrics	Peer review metrics	Media coverage
Section	Sector distribution	Conference output	Usage metrics	Prizes and awards	Medical guidelines
Custom set publications	h-, g-, m-indices	Funding sources	Audience location		Influence policies
Individual publication	Individual metrics		Audience sector		

Figure 1: A "basket of metrics" for understanding journal performance.

thematic section, case studies or proceedings content. They could even be calculated for sets of articles that a reader has collected from several different journals — a personal "virtual special issue".

There is no single correct way to apply this basket to a journal. A major benefit of this approach is that different journals can apply it in different ways, and any one journal might use it variously in different situations; for instance, editors of an arts and humanities journal and a chemistry journal may well have different preferences.

This article reports on the reaction of the international research community to metrics based on usage data. Usage data are created when a user makes a request to an online service to view scholarly information. Elsevier has recently started to implement usage metrics in merit tools such as My Research Dashboard and SciVal[9]. We have previously published examples of how usage metrics help to build a comprehensive understanding of performance[10]. Usage metrics are particularly interesting because:

1. Research excellence comes in many shapes and forms.
2. Research is complex and is best quantified by multiple criteria.
3. Usage is one of the earliest indicators of interest in research.
4. Non-publishing, and hence non-citing or cited, users are estimated to constitute one-third of the research community[11]. This group includes undergraduate and graduate students, as well as researchers in the corporate sector. In effect, their engagement is hidden from citation data, but usage data make their interests visible.
5. Research may not be published with the intention of

being cited. For example, clinical research is aimed at practitioners working with patients, who tend not to publish themselves but who read voraciously.

This article shares the results of a poll to survey opinions about using usage metrics alongside output and citation metrics in merit systems[12].

Methods

A webinar, entitled, "See the bigger picture with usage metrics", was held on 25 March 2015[9]. An invitation was emailed directly to a mailing list of over 50 000 researchers, librarians, senior management and faculty members globally, including subscribers to Elsevier's *Research Trends* newsletter and attendees of previous Elsevier research intelligence webinars. The seminar was also advertised on Elsevier.com and promoted on social media, and was free to anyone who wished to register, including Elsevier staff; responses by Elsevier staff have been removed in the statistics reported in this article.

204 participants joined from all over the world; the best-represented countries were the United States (N=53; 26 %), United Kingdom (N=27; 13%), the Russian Federation (N=24; 12%), Italy (N=16; 8%) and Canada (N=14; 7%). Attendees were asked to state their positions, and these ranged from visiting professors to vice-deans and librarians to research development coordinators.

The webinar was based on our article giving practical examples of the benefits of using usage metrics[10]. The survey was composed of five questions that were distributed throughout the webinar[9], and to which respondents could select from pre-defined options. Responses were recorded by the webinar software. The questions were:

- Question 1a: How often do you use usage metrics? Response options are shown in Figure 2.
- Question 1b: For those of you that said that you used usage metrics never or infrequently, why? Response options are shown in Figure 3.
- Question 2: Which of these statements is the most important reason for you to use usage metrics? The five response options have previously been published[10].
- Question 3: Which usage metric would be most useful to you? The options had been previously explained as follows:
  - Views Count calculates the total views that an entity has received.
  - Views per Publication corrects for differing article outputs of entities. It divides Views Count by the metric Scholarly Output.
  - Field-Weighted Views Impact corrects for the different levels of activity that can be associated with articles in distinct disciplines, of different types, and published at varying times. If this metric is 1, viewing activity is average compared to similar outputs globally; if it is more than 1 then the count of views is above average; and below 1 indicates a below average count of views.
- Question 4: If you had access to them, how likely would you be to include usage metrics in your evaluation of research? Respondents could select from "not likely", "likely" or "very likely".

Results

Out of 204 external participants, between 122 and 129 (response rate 60%-63%) responded, depending on the question.

The frequency of metrics usage is shown in Figure 2. 129 (63%) external participants responded to the question, "How often do you currently use usage metrics?" More than one third (N=45; 34%) of respondents reported using them monthly or even weekly, and more than one fifth (30, 23%) never used them. Those who selected "never" or "infrequently" were asked to select a reason for this, and Figure 3 shows that the most common obstacles were lack of access (N=21; 46%) or lack of awareness (N=19; 41%).

125 (61%) external participants responded to the question, "Which of these statements is the most important reason for you to use usage metrics?" "Research is best quantified using multiple criteria" was selected by most (N=50; 40%) respondents, followed by: "publications may be used (viewed) but not cited" (N=24; 19%); "reflects the engagement of the entire research and student community" (N=21; 17%); "viewing is an early indication of interest" (N=16; 13%); and "research excellence comes in many shapes and forms" (N=14; 11%). We also asked

respondents to rate their second and third most important reasons, and opinion was equally divided across the five options (results not shown).

122 (60%) external participants responded to the question, "Which usage metric would be most useful to you?" Figure 4 shows that almost half (N=60; 49%) of respondents selected Field-Weighted Views Impact as the most useful usage metric available, with over a third (N=43; 35%) selecting Views per Publication and a sixth (N=19;16%) selecting Views Count.

At the end of the presentation, 123 (60%) external participants responded to the question, "If you had access to them, how likely would you be to include usage metrics in your evaluation of research?" The majority (N=117; 95%) of respondents indicated that they would be likely (N=63; 51%) or very likely (N=54; 44%) to use usage metrics in their assessments of research merit, if they had access to them (results not shown).

Discussion

The availability of usage metrics to benchmark research is relatively novel. Respondents who said that they infrequently or never used usage metrics were offered a range of options as to why, drawn from common concerns that we have heard anecdotally and during user testing. Only 1 respondent (2%) selected ease of manipulation (Figure 3); this was surprising since usage data are often perceived as relatively easy to manipulate, despite the clear guidelines of the industry standard COUNTER (Counting Online Usage of NeTworked Electronic Resources) which aim to reduce this undesirable practice[13].

Field-Weighted Views Impact was selected by 49% of respondents as the most valuable usage metric, in line with the popularity of its citation-based equivalent, Field-Weighted Citation Impact (data from SciVal; not shown). User feedback indicates that these weighted metrics are seen as convenient in that they inherently compensate for differences between: behaviour of researchers working in distinct disciplines; modes of consumption of various article-types such as original research articles versus case studies; and awareness and uptake of outputs with different years of publication.
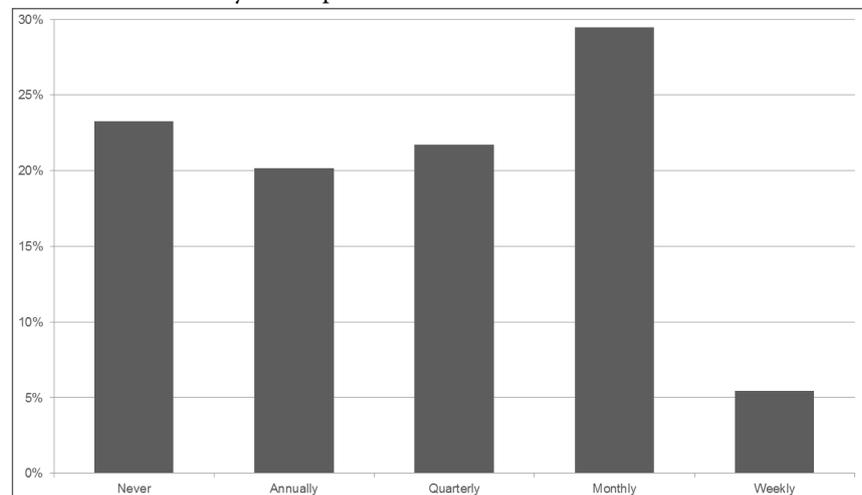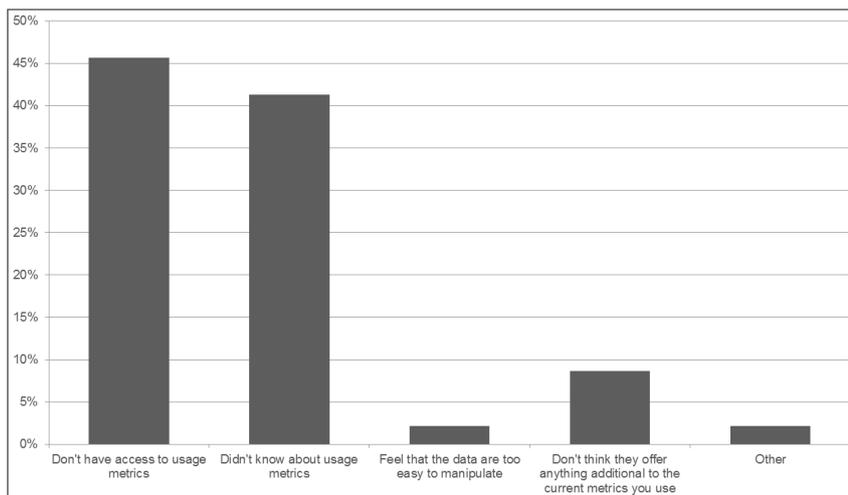

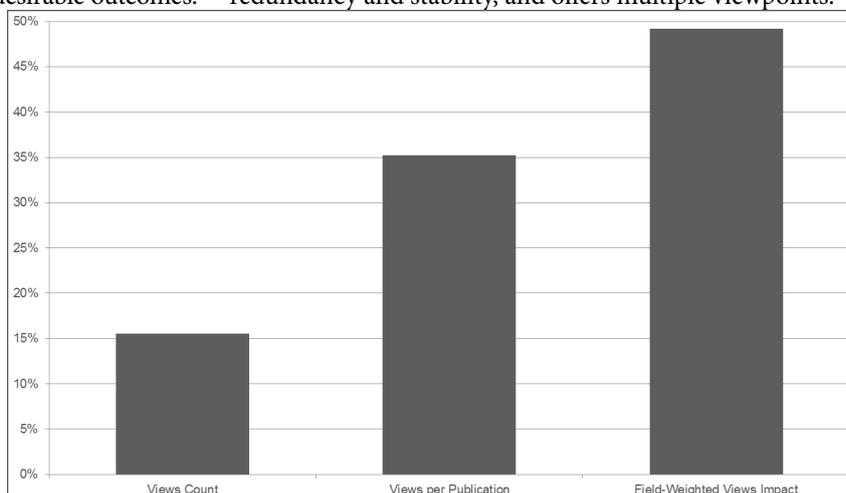**Figure 2: Current usage of usage metrics (N=129; 63%).**

Figure 3: Reasons for not yet using usage metrics, for those who responded that they "never" or "infrequently" use them (N=49; 24%).

Fifty one percent of respondents selected one of the more mathematically simple metrics as the most useful to them: Views Count and Views per Publication were selected by 16% and 35% of respondents, respectively. This may be because weighted metrics, like all metrics, also have weaknesses: the weightings complicate the method, making them more difficult to validate, and their relative value does not convey an impression of absolute magnitude. The similar degree of preference for simple and sophisticated (weighted) metrics is confirmation that a one-metric approach to merit is insufficient. A basket of metrics approach offers each user a choice of a primary metric that reflects their particular interest, while encouraging awareness and use of additional measures that reflect different needs.

A measure of an article's or a researcher's performance can be derived from where they are published, but problems arise from assuming that the actual merit of an article and its authors equates to the merit, most often the Impact Factor, of the journal in which it is published[3]. Such misuse of the Impact Factor has led to two common criticisms of research metrics, and unfortunately often used as a reason to advocate ignoring them completely. Firstly, any metric can be used inappropriately, which drives undesirable outcomes. Researchers have been shown to change their behaviour in response to the main evaluation criteria applied by the UKs national assessment exercise, for example by "salami slicing" to increase their output when article counts were used as the principal metric of evaluation[4]. Secondly, metrics can be manipulated; the construction of the Impact Factor makes it especially open to abuse[5].

There is best practice, however, that addresses the common concerns of using research metrics, and promotes their appropriate use in understanding performance. With a little common sense, the research community can benefit from the responsible use of research metrics to contribute to the understanding of not only journal, but also article and researcher merit, and indeed the merit of any entity[6]:

•   All merit systems should be based on multiple types of measurement and never on one alone. Research metrics are an essential part of any merit system, but are insufficient on their own, regardless of how sophisticated the metrics system is. It would be foolish for a journal editor to decide whether to accept a submission based only on metrics about its topicality or the past performance of its author(s), for instance, although these could be useful when used together with the editor's judgement and experience. This approach makes it much more difficult for any type of assessment to be abused – if metrics are being manipulated it would be apparent from the qualitative assessment, and vice versa.

•   The metrics-based aspect of understanding merit must always, without exception, draw on at least two types of measurement from a "basket of metrics". Every metric has weaknesses, but employing multiple metrics enables them to be complemented by the strengths of others. This approach also makes manipulation more difficult, and therefore less likely. There is not, and will never be, one single metric that encompasses all aspects of excellent performance, no matter how well or carefully designed. Trying to reduce excellent performance to one metric will almost certainly drive unbalanced, undesirable behaviour as those being evaluated optimise their performance according to that metric. But how can researchers optimise across three or five different metrics, except by doing genuinely better research?

These best practices ensure that no single approach or metric fully carries the load, so that a merit system does not have a single point of failure. Rather, the system has built-in redundancy and stability, and offers multiple viewpoints.

Figure 4: Most useful usage metric (N=122; 60%).

The main driver selected in the survey for using usage metrics was that research excellence is a multi-faceted phenomenon that is best quantified using multiple criteria. This was selected by 40% of respondents and, together with the strong support indicated by 95% of respondents for incorporating usage metrics in understanding research performance, demonstrates a clear willingness and a real appetite to use a "basket of metrics" to broaden the ways in which research excellence can be detected and demonstrated. Indeed, industry articles call for the use of a broad range of criteria when assessing researchers' track records and when considering which research to publish[14].

Members of the research community often seem to feel that the approach to measuring research merit is outside their influence and that there is little point speaking up. The Snowball Metrics initiative[15], in which research institutions agree on a consistent message and speak with one voice to influence the way that institutional performance is measured by suppliers and funders, has proved that change can come from within. The community can change the status quo by supporting and using the basket of metrics in understanding and communicating the performance of not only journals, but also of articles, and researchers.

The results in this report have previously been presented by the authors on their website. Elsevier Connect, posted May 2015. Available from: http://www.elsevier.com/connect/how-you-are-using-usage-data-to-measure-research-impact-or-whats-stopping-you

References

1 The San Francisco Declaration on Research Assessment. Available from: http://am.ascb.org/dora/.
2 E. Garfield. Citation indexes for science. *Science* 1955:122(3159), 108-111.
3 E. Garfield. How can Impact Factors be improved? *British Medical Journal* 1996: 313(7054):411-413.
4 H.F. Moed. UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics* 2008:74(1):153-161. doi: 10.1007/s11192-008-0108-1.
5 N.C. Romano Jr. Journal Self-Citation V: Coercive Journal Self-Citation – Manipulations to Increase Impact Factors May Do More Harm than Good in the Long Run. *Communications of the Association for Information Systems* 2009: 25 (article 5):41-56.
6 L. Colledge. Elsevier's response to HEFCE's call for evidence: independent review of the role of metrics in research assessment. Available from: http://bit.ly/hefceresponse.
7 H.F. Moed. Measuring contextual citation impact of scientific journals. *Journal of Informetrics* 2010: 4(3):265-277. doi: 10.1016/j.joi.2010.01.002.
8 B. González-Pereira, V. P. Guerrero-Bote, and F. Moya-Anegón. A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics* 2010:4(3):379-391. doi: 10.1016/j.joi.2010.03.002.
9 L. Colledge. Usage Guidebook. Available from: http://bit.ly/usageguidebook.
10 L. Colledge and C. James. 5 ways usage metrics can help you see the bigger picture. *Elsevier Connect*, posted on 23 March 2015. Available from: http://www.elsevier.com/connect/5-ways-usage-metrics-can-help-you-see-the-bigger-picture.
11 S. Haustein. *Readership Metrics* in *Beyond Bibliometrics: Harnessing Multidimensional indicators of Scholarly Impact*, ed. B. Cronin and C.R. Sugimoto. 2014:ch.17:327-344.
12 L. Colledge and C. James. How you are using usage data to measure research impact – or what's stopping you? *Elsevier Connect*, posted May 2015. Available from: http://www.elsevier.com/connect/how-you-are-using-usage-data-to-measure-research-impact-or-whats-stopping-you.
13 Guidelines from Project COUNTER (Counting Online Usage of NeTworked Electronic Resources). Available from: projectcounter.org/r4/COPR4.pdf, page 25.
14 The findings of a series of engagement activities exploring the culture of scientific research in the UK. 2014. Available from: http://nuffieldbioethics.org/project/research-culture/.
15 Snowball Metrics. Available from: www.snowballmetrics.com.