



Time to Modernize Our Metrics

By Seth C. Murray, CSSA President, Seth.Murray@ag.tamu.edu

| June 19, 2023



Photo courtesy of Adobe Stock/treenabeena.

Goodhart's Law was an observation made in the social sciences that posits when a measure becomes a target for performance, it can lose its effectiveness as a measure. This is because people and organizations will change their behavior to meet the target,

rather than pursuing the underlying goal. This phenomenon is similarly referred to as “gaming the system” or “teaching to the test.”

CSSA members throughout academia, government, and industry should be familiar with Goodhart’s law. In academia, we’ve come to fetishize grants, peer-reviewed publications and citations, sharing admiration for those who achieve high metrics across these areas and regarding the metrics as import for one’s research reputation and the discoveries we care about. Similarly, for large companies, as quarterly reporting has become more accessible, investors have increasingly become focused on commercialization and short-term profits, which can direct research towards areas of immediate gains and away from long-term goals and innovations. Neither trend bodes well for advancing solutions to our grand challenges within agriculture, which requires long-term efforts.

As a further example, in both academia and industry, some organizations have gone through periods of patenting everything they could as a metric of performance to be counted; yet many patents remain unused. Combined with William Bruce Cameron’s observation that “not everything that can be counted counts, and not everything that counts can be counted,” the importance of metrics is worth reflecting on for both our scientific societies and within our own professional careers. Should we focus our time on improving our metrics, or should we focus our time on improving agriculture and solving difficult problems?

Should we focus our time on improving our metrics, or should we focus our time on improving agriculture and solving difficult problems?

What Do the Metrics of Revered Scientific Leaders Tell Us?

In a cross-university graduate course on Scientific Career Success and Leadership experimentally taught this year, students presented on a scientific leader they admired. We, of course, had Drs. Norman Borlaug and Rachel Carson, and we were also pleased to learn more about Kevin Folta (outstanding in social media outreach and communication in our field), Jane Goodall, Maurice Hilleman (inventor of the MMR and more than 40 other vaccines), and Paul Jackson (an outstanding horticultural professor at Louisiana Tech, who has made a difference to his students). Through this journey, the major discovery the class had was that none of the individuals the students chose as scientific leaders were selected based on the traditional scientific metrics we measure and celebrate. In fact, many of them had what some would consider to be poor scientific metrics. For example, Norman Borlaug had relatively few publications, none particularly that well cited, and many were based in policy as opposed to original research findings; nevertheless his global achievements are indisputable.

While obvious in some ways, this was a discovery because some students perceived that their career value came from a narrow set of publishing metrics that focused on

authoring articles in *Cell*, *Nature*, or *Science* (CNS); this is a message science is sending. Instead, we found unifying themes of scientific leaders selected included that they forged ahead in new areas and approaches, generally being the first to do so; they focused on making positive impacts for other people, wildlife, or the environment; finally, they communicated directly with the public beyond their scientific communities and not just in CNS publications.

Output vs. Impact

There is a continuum from research activities that we conduct to the outputs, outcomes, and eventually impacts made; these words are deliberate. It remains increasingly common, at least in academia, to focus efforts on outputs—things like publications, presentations, graduate students trained, grants written, etc. Outcomes from these outputs we know are more important, and some are easily measured—citations on each publication, attendees at each presentation, graduate students' first jobs, grant submissions funded, or awards received. Yet often our important outcomes are more difficult to measure or obtain data for—farmers growing a cultivar released, labs using a new protocol, etc.

The most important and most difficult to measure end of the activities we conduct are impacts. Impacts often take a comparatively long time to come to fruition. How much more profit did a cultivar generate for farmers, or how did it increase human health? How has a research program changed the research field or improved people's lives? Often the outcomes and impacts we most value are poorly aligned with the shorter-term outputs we report, measure, and evaluate as important. If we had a crystal ball, we might better divine which outputs will become impactful, but we often just don't know. I suspect the increased focus on outputs in academia is at least partially in

response to outside skepticism of the value of taxpayer-funded research and partially due to the ease of reporting output metrics.

Impact is a word you might be most familiar with in reference to “impact factor” (IF) of a journal (JIF, trademarked by Clarivate). The JIF is calculated by taking the citations of all articles in a journal in a year as the numerator and dividing by the total number of articles published in that journal over the two previous years. I would argue that JIF measures outcomes and not impacts, but “outcome factor” (OF) does not have the same ring to it, I suppose. It’s also notable that JIF is a metric of the journal and not of the individual articles (i.e., yours), which it contains. For evaluating an individual scientist, total citations or the h-index (h is equal to the h number of papers that have each been cited at least h times) are better metrics over the JIF of the journal they submit to. Still citations and h-index can be affected or even gamed, such as through unnecessary self-citations.

Looking at the Impact of CSSA Journals

There are many great reasons to publish in our non-profit CSSA journals, from helping to support member activities you value, to enhanced visibility to your target audience, to impact on agricultural policy as discussed in prior President’s Messages. Yet, the most frequent concern members report is that our CSSA journals have a lower JIF than our members (as well as their administrators and promotion and tenure committees) would like. Within most journals, there exists articles cited 0 times in the first two years with a long tail having just a few very highly cited articles (Berg, 2016; Larivière et al., 2016; Figure 1—using *Crop Science* only as an example). One common proposed solution is that we become even more exclusive; rejecting more articles unlikely to be cited (assuming we know) reduces the JIF denominator of 0 cited articles. However,

we also like our students to be able to publish their high quality work in the highest visibility journal, even if work that is highly specialized, incremental, or validating, which may not obtain high citations. Anything that increases rejection based on divining JIF would likely exclude specialty crops from publishing work in our journals. A reduced publication volume helps neither our members' careers nor our Society's budget. Another proposed solution is more open access or review articles, which indeed could help as they are more likely to be highly cited (Figure 1), which brings JIF up through the numerator.

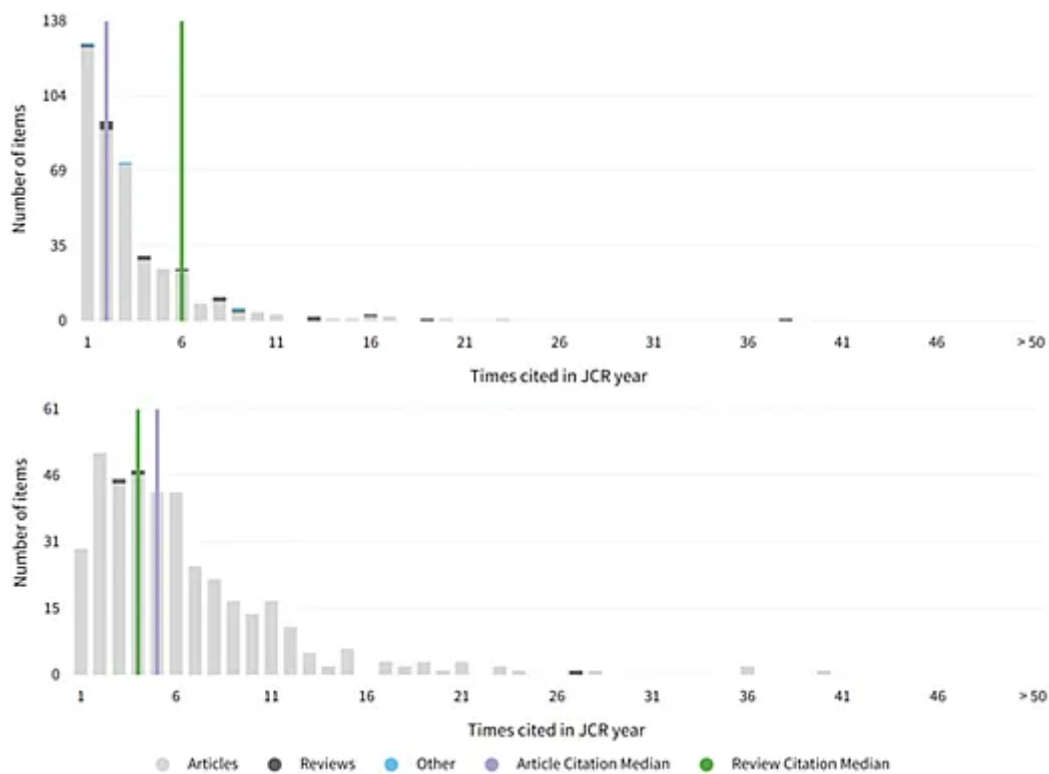


Figure 1. Top: Crop Science's 2021 citation distribution. The journal is ranked 25th among agronomy journals and has a 2021 journal impact factor (JIF) of 2.76. Bottom: Field Crops Research's 2021 citation distribution. This journal is ranked 8th among agronomy journals with a 2021 JIF of 6.15. Information from Clarivate Journal Citation Reports (accessed Apr. 24, 2023).

I think our disciplines also have structural reasons for a lower JIF. Agricultural research is often regionally and species specific, comparatively under-resourced with few publishing scientists, and takes more than two years to incorporate findings from previous research. The recent addition of the Declaration on Research Assessment (DORA) including altmetrics and article views may be more useful than JIF to show article-specific value. We can increase JIF through improving the use of the science we publish as well as the visibility and access to our science (this being a goal of our Wiley agreement). I've observed firsthand the incredible efforts and thoughtfulness our staff, editors, board, reviewers, and authors all put into ethically improving all aspects of our publications. In contrast, some predatory journals have found it is possible to game JIF by encouraging unnecessary within-journal citations; rules have increasingly been put in place to minimize this trick.

Anyone who has published in our journals knows that the articles published are trustworthy and well edited because the review process is difficult (but fair), especially compared with what might be expected from JIF alone (which continues to increase for most CSSA journals). It is reasonable to wonder why JIF is considered as a metric of importance when it hurts many individual researcher as well as our field, and we know it is not representative of the quality of our journals. Inflated impressions of any metric, such as JIF, can lead to distorted priorities and even unethical behavior. When a measure becomes a target for performance, it loses its effectiveness as a measure for what we really care about.

Three Questions for You

I leave you with three major questions to reflect on. What do you think are good metrics by which to be evaluated in your work? What do you want to have accomplished by the end of your career? My guess is that your target

accomplishments are impacts and not outputs. So finally, does the metric you first specified measure progress towards the accomplishments you seek, and if not, is it a good metric to use for evaluating the work of others?

References

Berg, J. (2016, Aug. 4). Journal impact factors—fitting citation distribution curves. Sciencehound. <https://www.science.org/content/blog-post/journal-impact-factors—fitting-citation-distribution-curves>

Larivière, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., Pulverer, B., ... & Curry, S. (2016). A simple proposal for the publication of journal citation distributions. BioRxiv, 062109. <https://doi.org/10.1101/062109> bioRxiv

Text © . The authors. CC BY-NC-ND 4.0. Except where otherwise noted, images are subject to copyright. Any reuse without express permission from the copyright owner is prohibited.