**White Paper**

*May 2019*

**Article Citations are an Insufficient Building Block for Measuring Research Impact in the Twenty-First Century**

**Richard Wynne**

[**https://www.orcid.org/0000-0002-9217-0407**](https://www.orcid.org/0000-0002-9217-0407)

[**https://www.rescognito.com/0000-0002-9217-0407**](https://www.rescognito.com/0000-0002-9217-0407)

**Background**

According to the US National Science Board, annual global expenditure on research was estimated to exceed $1.9 trillion in 2018[[1]](#endnote-1). Public and private research funders therefore need the best possible tools to measure the outcomes of such a massive recurring investment. Global technological, social and ecological problems need urgent solutions and even a 1% improvement would result in significant beneficial economic and social impact. By more effectively measuring research outcomes, funders can direct future investments to the most fruitful ideas, and to the most productive teams and individuals.

Assessing research outcomes is more easily said than done~~!~~ Funders have developed a variety of techniques to optimize their investment decisions including: peer review of research proposals, literature review, financial overview, and expert panels. However, fundamental to almost all quantitative research evaluation measures is the publication of research findings in recognized peer review journals and counting of subsequent citations. In other words, the impact of research is signaled by publication in journals that have a track record of high citation, as measured by the Journal Impact Factor[[2]](#endnote-2) (JIF).

The limitations of the JIF as a measure of individual research quality have been much discussed[[3]](#endnote-3). Alternative approaches such as the *h*-index[[4]](#endnote-4) focus on citations of individual researchers rather than overall journal performance. The JIF is also criticized because it symbiotically reinforces and underpins an elite journal structure resulting in the “Matthew Effect” where highly cited authors attract more citations. The “publish or perish” culture is said to encourage “goal displacement” as researchers excessively focus on the metrics (article citations) at the possible expense of the underlying goal (research quality). In some cases, researchers are even awarded cash bonuses of as much as $165,000 if they publish in highly cited journals such as Nature[[5]](#endnote-5).

However, despite such long-standing and evident concerns about how it is calculated and used, the JIF remains an influential metric when assessing research investment choices and researcher career advancement. In the research community, garnering article citations therefore continues to be an animating force. From a researcher point of view, more citations equate to greater reputation and funding opportunities.

Article citations are the common building blocks for almost all quantitative measurements of research quality[[6]](#endnote-6). The emergence of Artificial Intelligence (AI) as a tool[[7]](#endnote-7) for research analysis also relies heavily article citations as its raw data input for connecting articles and ideas.

By default, article citations are therefore the ubiquitous currency for quantifying and signaling the research quality and researcher reputation. They are the building blocks from which the edifices of the JIF and *h*-index are built. They are also the *de facto* raw materials for next generation AI research analytics.

In view of the growing demands placed on article citations as a ubiquitous metric, now is a good time to re-examine some of their limitations and to ask: can we improve?

**Imprecision – Article citations lack specificity in author identity attribution**

A typical manuscript citation includes multiple authors with each author identified by a text string (e.g. J. Smith, S. Brown). In the parochial world of twentieth century scholarly print publication this was acceptable since the reader likely knew which “J. Smith” was the expert in their field. In the world of electronic publication with global authorship and machine-reading applications, text string-based identification is the equivalent of using a carrier pigeon for mobile communications.

Thankfully ORCID iDs have emerged as a free, persistent and unambiguous mechanism to identify individual authors using a 16-digit number[[8]](#endnote-8). At time of writing, approximately 6 million ORCID iDs had been issued. By embedding authenticated ORCID iDs into the scholarly workflow, the activities and outputs of individual researchers can be consistently credited and tracked. An “authenticated” ORCID iD is superior to a hand-entered ORCID iD because it has been verified by its owner via the ORCID API (Application Programming Interface).

ORCID iDs therefore have the potential to help solve the problem of unreliable text-based author names. However, capturing authenticated ORCID iDs at the front-end of the publication process has proven problematic. Based on data generated from the Crossref API[[9]](#endnote-9) it’s clear that only handful of publishers (e.g. PLOS, American Chemical Society, Hindawi, PeerJ) are depositing authenticated ORCID iDs in their CrossRef metadata, and almost zero manuscripts contain an ORCID iD for *all* the authors of multi-authored manuscripts.

Publishers’ reluctance to capture validated ORCID iDs for all authors early in the publication process is understandable because of the additional administrative burden that it places on authors and journal editorial offices. In many cases authors many not appreciate the benefit of obtaining an ORCID iD until months or years after the manuscript has been published. While every effort should continue to be made to collect ORCID iDs at the start of the publication process, it would be pragmatic and beneficial to have a mechanism for authors to claim and verify their contributions after publication and to allow third parties to validate such claims.

**Imprecision – Article citations blur individual accountability and recognition**

The number of authors per manuscript is steadily increasing. Using Scopus data, the STM association reports that the average number of scholarly article authors increased from 3.2 in 1996 to 4.4 in 2015[[10]](#endnote-10). In certain fields, such as high-energy physics, there are a growing number of “hyper-authored” papers with more than 50 or even 1,000 contributors.

Our existing system of researcher evaluation relies on article citation counting as a measure of research quality. From an author’s point of view, one citation to a manuscript with 1,000 authors is considered to have equivalent impact as a citation to single-authored manuscript. This cannot be justified since the former citation provides a lot less useful information about the contribution of an individual researcher than the latter.

When one manuscript cites another it creates a useful, one-to-one relationship between the two *documents* and the ideas and findings they contain. However, for the purpose of research evaluation, article-level citations overlook potentially more granular and useful connections between the individual contributors. In an example where one article with 5 authors cites another article with 5 authors, that single manuscript citation represents 25 (5X5) possible channels of single author recognition (including self-citation).

To better understand the value and impact of researcher contribution, it would be beneficial to also know exactly which author(s) is/are citing exactly which other authors’ work. Such author-to-author citation would reveal specifically which author’s work is being valued by the citing manuscript and would be a more accurate signal of researcher contribution. It would also enable better human and machine-based analysis of knowledge development and patterns of knowledge development over time.

**Imprecision – Article citations don’t identify specific author contributions**

When viewing a typical author list for multi-authored articles, the specific contributions of each individual author are usually opaque and governed by implied rules. Some research disciplines attach more importance to the first, last or second authors. Author order significance also varies by culture.

Today’s article citation format does not explicitly communicate which author was responsible for specific activities such as: data curation, methodology, or software development. If citations are being used to measure individual researcher impact is it sensible that there is no explicit recognition of what work the individual actually contributed?

A handful of journals (e.g. PLOS, Cell) are working to address this problem by adopting the CRediT taxonomy[[11]](#endnote-11) to identify specific author contributions. The CRediT taxonomy identifies 14 author contribution roles that are applied at the individual author level. This additional data, especially when associated with a validated ORCID iD, greatly enriches the value of the citation as tool to signal research quality because it unambiguously identifies the individual and the work they undertook with respect to the article.



Example use of CRediT on a PLOS One manuscript

In due course CRediT and other domain-specific taxonomies will allow recognition of a much broader range of manuscript contributions. This will also make it possible to legitimately include additional “authors” (e.g. software developers, clinical trials administrators, etc.) who are currently excluded because they do not fit the traditional author definition.

Regrettably the vast majority of journals and preprint servers still do not collect information about author contributions or provide it in their published citation metadata. As with the collection of authenticated ORCID iDs, authors and journal offices do not wish to slow down the publication process with the additional administration that would be required to collect these data. In most cases it’s more realistic and pragmatic to expect that the enrichment of citation metadata with contribution taxonomies will occur after publication. This means there will be a need for a mechanism to validate such assertions about contribution after publication.

**Imprecision – Article citations miss many useful research contributions and outputs**

According to the Declaration on Research Assessment (DORA)[[12]](#endnote-12) : “*the Journal Impact Factor has a number of well-documented deficiencies as a tool for research assessment “* and *“the outputs from scientific research are many and varied, including: research articles reporting new knowledge, data, reagents, and software; intellectual property; and highly trained young scientists”* .

By definition traditional article citations only recognize article-related contributions and omit other potentially valuable activities such as: mentoring, teaching, teamwork and public advocacy[[13]](#endnote-13). Transparent and open recognition of such contributions would help indicate the future potential of a researcher rather than simply document historic track record.

Metrics such as the JIF that rely exclusively on journal citations as their input miss recognition of many beneficial activities. An improved tool for measuring research contributions would include a credible way to acknowledge and assess such other activities.

**Imprecision – Article citations don’t give a reason for citing**

The unstated assumption of today’s impact metrics is that more citations correlate to higher positive impact. In an age of paper-based publication this rule of thumb was pragmatic and good enough. However, powered by twenty-first century technology and standards can we improve?

Citations would be enhanced if we know exactly why a source was cited. The proposed CiTO, Citation Typing Ontology[[14]](#endnote-14) suggests a way to characterize the purpose of citations. Of the 23 suggested terms 10 are “factual” (such as “uses data from”), 6 are “positive” (such as “confirms”), 5 are “negative” (such as “corrects”) and 2 are “neutral” (such as “discusses”). Obviously, an article that receives “corrects” citations means something different than one that receives “confirms” citations, even if the citation counts are the same.

The benefit of adding meaning to citations is not a new idea. In the legal field a process known as Shepardizing involves consulting Shepard's database (LexisNexis) to see if earlier cases have been overturned or reaffirmed by later citing cases. Historically, scholarly citations have not been structured in this manner, but It may be possible to use natural language processing or human processes to categorize scientific literature citations into broad groups such as “confirming”, “mentioning” and “refuting”[[15]](#endnote-15). If successful, this would enable a form of Shepardizing for the scientific literature.

While neither of these initiatives has yet achieved broad adoption in scholarly publication, they illustrate that citations are not homogeneous. Metrics such as the JIF that treat citations as homogeneous overlook their nature and consequently provide a sub-optimal assessment of research impact. In the future, if the author’s underlying purpose for citing an earlier work can be accurately captured this would improve the efficiency and fairness of research evaluation.

**Imprecision – Opacity means that article citations can be easily gamed**

In his book about the formula for success[[16]](#endnote-16), the academic Albert-László Barabási estimates that each scholarly journal citation is worth approximately $100,000 in research funding. When the economic incentives are so great it’s not surprising that even the most honest researchers look for ways to optimize article citations.

There are numerous known techniques for increasing citation counts and Journal Impact Factors. For example, by their nature review and summary articles attract above average citation counts, so journals and individuals can easily boost their averages by increasing the proportion or review articles that they publish.

Another technique is for editors and journals to suggest or mandate the citation of articles from their own journals, as illustrated in the Tweet reproduced here.



There are also cases of outright fraud where “citation rings” are used to fraudulently boost citation counts and impact factors[[17]](#endnote-17) by a group of cooperating journals and editors.

Any system used to measure performance will inevitably be gamed. Even with the advent of Blockchain technology, it would be prohibitively expensive and cumbersome to implement workflows that totally eliminate the possibility of gaming. A more practical approach is to envision a recognition strategy that increases transparency and accountability so that automated tools can better highlight and adjust for inevitable gaming behavior. For example, where an identifiable author repeatedly cites themselves or a pair of identifiable individuals reciprocate recognition without any third-party validation, such behavior patterns could be used to “discount” the assessment of impact.

In this context, the problem with article citations is that as a broad-based measure of impact they are susceptible to gaming because they are insufficiently transparent and granular.

**Article Citations - Inadequate building blocks**

In summary, article citations indicate a possible connection between any of number of ambiguously-identified authors to an unspecified number of ambiguously-identified authors of the cited manuscript, for an unspecified reason, in recognition of unspecified contributions, and for motivations that are not transparent; while totally excluding non-manuscript related contributions. On this basis we generate and use derivative measures of “impact” that are reported as accurate to three decimal places[[18]](#endnote-18) and use the result to evaluate careers of millions of researchers and direct billions of dollars in annual investment.

And yet despite their limitations, article citations are countable metadata, and so we count them. They are the *de facto* basis of purportedly high-precision measures of impact and the raw materials of AI analysis. But to improve their use as a signal of research quality we need to know much more about their provenance context: “*provenance is both more and less than metadata. It involves the origin and history of something (somebody) and documentation of the chain of evidence custody and relationships to other entities*”[[19]](#endnote-19).

Today, article citations can be defended as being the worst approach for measuring impact, except for all the other available methods. As such, citations will continue to play an important and effective role in research output evaluation. However, the emergence of new standards and global electronic connectivity means that it will be possible to re-think and improve how research quality is signaled in a more transparent and open manner. Some of these approaches, such as using ORCID iDs and contribution taxonomies, are already underway. Others will emerge from research institutes and entrepreneurial organizations that are ready to think beyond article citations.

**Beyond article citations – what would an improved system look like?**

An improved research evaluation system would:

* Be fair and respectful of researchers
* Go beyond article citations as a homogeneous unit of impact
* Be founded on open and persistent identifiers such as ORCID and DOI (Digital Object Identifiers) to minimize ambiguity
* Provide granular recognition and accountability for a comprehensive range of contributors
* Provide recognition for a broader range of contributions than are currently recognized (e.g. mentoring, social impact, teaching, teamwork, etc.)
* Be transparent and open
* Offer a clear chain of evidence with improved “provenance context”
* Promote and recognize teamwork and research groups
* Provide tangible, material benefits for individual researchers

**Article Citations – Time to change**

Historically, relationships between personally known authors, reviewers and editors could form the basis of trust in research objects such as journal articles, but this no longer holds true. Research objects, increasingly exhibit the characteristics of pure data objects that are “read” by a global, impersonal readership of humans and machines.

The current system of research evaluation is sub-optimal and needs updating to conform to twenty-first century standards, technologies and needs. The potential societal and economic benefits arising from improving research outcomes generated from $1.9 trillion in annual spending are enormous. It’s time to look beyond the article citation as the primary quantitative signal of research quality.

1. https://www.nsf.gov/statistics/2018/nsb20181/report/sections/research-and-development-u-s-trends-and-international-comparisons/cross-national-comparisons-of-r-d-performance [↑](#endnote-ref-1)
2. <https://clarivate.com/essays/impact-factor/> [↑](#endnote-ref-2)
3. https://doi.org/10.1371/journal.pbio.2004089 [↑](#endnote-ref-3)
4. The H Index is widely used by Google Scholar and RELX services. [↑](#endnote-ref-4)
5. <https://www.timeshighereducation.com/news/chinese-scholars-win-ps127000-for-papers-in-top-journals#survey-answer> [↑](#endnote-ref-5)
6. With the exception of “altmetrics” that count social media impact on platforms such as FaceBook and Twitter. [↑](#endnote-ref-6)
7. Such as Meta https://www.meta.com, a Chan Zuckerberg Foundation initiative. [↑](#endnote-ref-7)
8. <https://www.orcid.org> [↑](#endnote-ref-8)
9. The [www.crossref.org](http://www.crossref.org) API can be used to query publisher deposited metadata including the authenticated status of ORCID iDs. [↑](#endnote-ref-9)
10. <https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf> - Page 118 [↑](#endnote-ref-10)
11. <https://casrai.org/credit/> [↑](#endnote-ref-11)
12. <https://sfdora.org/read/> [↑](#endnote-ref-12)
13. <https://doi.org/10.1371/journal.pbio.2004089> [↑](#endnote-ref-13)
14. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2903725/> [↑](#endnote-ref-14)
15. [www.scite.ai](http://www.scite.ai) [↑](#endnote-ref-15)
16. “The Formula: the universal laws of success” Albert-László Barabási [↑](#endnote-ref-16)
17. <https://scholarlykitchen.sspnet.org/2017/03/09/citation-cartel-or-editor-gone-rogue/> [↑](#endnote-ref-17)
18. <https://support.clarivate.com/ScientificandAcademicResearch/s/article/Journal-Citation-Reports-Notes-on-Impact-Factors-Taken-to-Three-Decimal-Points?language=en_US> [↑](#endnote-ref-18)
19. “Theories of infometrics and scholarly communication” Cassidy R Sugitomo https://books.google.co.uk/books?id=UYulCwAAQBAJ&printsec=frontcover#v=onepage&q&f=false [↑](#endnote-ref-19)