Citations are a good way to determine the quality of research

Martin Caon¹ · Jamie Trapp² · Clive Baldock³

Published online: 9 November 2020

© Australasian College of Physical Scientists and Engineers in Medicine 2020

Introduction and overview: Clive Baldock, moderator

From the time that Henry Oldenburg created the Philosophical Transactions of the Royal Society of London in the seventeenth century, scientists have endeavoured to share the results of their research efforts to a wider audience by publishing in scientific journals [1]. Since that time, many new journals have been founded that have given authors many options for publishing. This has been particularly significant in recent years, due to the growth of the internet and the availability of online and open access journals [2, 3]. With such growth has come challenges. A significant issue for active research scientists is in keeping up-to-date with published research is due to the large body of research outputs, resulting in only a fraction of the literature being read and subsequently cited [4]. Due to many published papers going unnoticed or unread, many articles either never get cited or are only self-cited by the authors of the paper. When a paper is published, it has no immediate impact. However, many would contend that the number of times a paper has been subsequently cited is a good indicator of the impact of the paper and therefore the quality of the research [5].

In this Topical Debate, the previous and current Editors in Chief of this journal engage in a spirited and timely debate regarding citations and whether they are a good way to determine the quality of research.¹

Arguing for the proposition is Dr. Martin Caon, previous Journal Editor and from Flinders University. Dr Caon was born and grew up in Adelaide. His PhD in Medical

Physics was undertaken at the University of South Australia and the Women's and Children's Hospital, Adelaide. This came after an undergraduate degree in Physics and Zoology, with honours in Zoology from Adelaide University, a Diploma in Education, and after a Master's Degree in Science Education from Curtin University. Martin worked as a secondary school teacher in Mount Gambier and Adelaide for ten years before moving to Flinders University in 1988. There, the somewhat unusual undergraduate combination of Physics and Zoology proved useful in the teaching of physical science and anatomy and physiology to health sciences students.

While at Flinders University, his research interests were in the teaching of science to nursing students and in calculating, using Monte Carlo computer methods, the radiation dose to children from CT scanning. This latter work involved constructing anatomically correct mathematical models of child anatomy from medical images, initially using manual and then semi-automatic methods. Commencing in 2003, Martin was Associate Editor of this journal for three years and then Editor in Chief for twelve years. Martin continues as an Associate Editor, but retired from the Editor in Chief position and from Flinders University at the end of 2018. He has spent his retirement playing saxophone, hockey, tennis, bushwalking, running and spending time at his hobby farm and on matters concerning the Friends of the Vulkathunha-Gammon Ranges National Park, of which he is president. He has also completed the recently published 3rd edition of his book: Examination Questions and Answers in Basic Anatomy and Physiology: 2900 multiple choice questions, which has been published by Springer. Additional writing projects have been an article describing the Friends' work of monitoring *Mogurnda clivicola*, a land-locked fish found in the Vulkathunha-Gammon Ranges National Park; and researching and writing the 115-year history of the Forestville Hockey Club.

✉ Clive Baldock
cbaldock@uow.edu.au

¹ Clarence Park (retired), Adelaide, Australia

² School of Chemistry, Physics and Mechanical Engineering, Queensland University of Technology, Level 4 O Block, Garden's Point, Brisbane, QLD 4001, Australia

³ Research and Innovation Division, University of Wollongong, Wollongong, NSW 2522, Australia

¹ Contributors to Topical Debates are selected for their knowledge and expertise. Their position for or against a proposition may or may not reflect their personal opinions.

Arguing against the Proposition is Jamie Trapp who is an Associate Professor at Queensland University of Technology (QUT). After completing his PhD at QUT in 2003 he moved to London and worked as a postdoctoral research scientist at the Institute of Cancer Research at the Royal Marsden Hospital. In 2004 he commenced an academic position at Royal Melbourne Institute of Technology, and in 2007 moved back to QUT into a similar position. In 2005 he commenced as an Associate Editor for this journal, in 2016 moved to a Deputy Editor role, and at the end of 2018 became Editor in Chief after Dr Caon's retirement. He has published over 100 peer reviewed articles in various topics associated with radiation measurement and medical physics.

Opening statement—Martin Caon

How can one determine the quality of a quantum of research? A quantum in this case is an article published in a peer-reviewed scientific journal. One supported by a constituted National or International Society of Scientists sharing a common interest in some field of endeavour, such as this one.

Being published in a journal means that the manuscript has navigated the obstacle course of peer review and been adjudged to be sound, novel and of publishable quality by some members of the academic community. The historical gold standard to assess published articles has been: Be an expert in the field, read a hard copy and assess it in the light of your own expertise. Unfortunately, this method was only available to experts with access to the journal. Fortunately, science advances.

Electronic publication has provided novel, measurable ways to access research. This ability to measure is important. William Thomson said: “When you can measure what you are speaking about, and express it in numbers, you know something about it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind” [6]. Citation metrics mean that the assessment of research quality has left behind the dark ages of meagre and unsatisfactory kinds of knowledge. By expressing quality as a number, we know that we know something about it.

What metrics are available to us? The “number of reads” is a measure of scientific interest. However “Number of Downloads” is a better metric as it implies that the interest is sufficiently deep to require a sustained effort to understand the published work. The NoD is directly proportional to the quality of the article. Contrarily it may be argued that some scientists download rubbish articles to make a point, but people soon tire of reading nonsense, and will return to quality. In fact, the NoD may even underestimate interest as pdf copies electronically sent to colleagues, saves them downloading it anew.

A still better metric is the Number of Citations (NoC) made to published work. Scientists that cite an article are a subset of those that download it. Citing scientists have read the article, found it relevant and useful, and wish to associate their own work with it. Such scientists are themselves researchers who have the expertise to recognise important work. They are part of the community of cognoscenti. Their number can be counted. Now we can say that five scientists have noticed this work, or that 55 scientists approve of the work. We have an improved gold standard. Previously, few people knew how many experts had assessed an article. Its quality was a secret. Now everyone, including the authors, want to know.

The record of quality is permanent. The NoC stays on the record and increases with time making the importance of the work more apparent. Hence the NoC is flexible and responsive to changes in knowledge. It is unlikely that modern day scientists will overlook the work, or the work will be forgotten, as has happened in the past.

The metrification of quality allows for better metrics to be introduced. For example the Relative Citation Ratio measures quality at the article level [7, 8]. This metric addresses the widely-acknowledged limitations of metrics such as raw

citation count. RCR measures impact while also accounting for field of study and time of publication. This may be done by means of “expected citation rates”, derived from all citations to all contributions of the same type, published in the same year in the same ISI journal.

Just as cobalt machines have been relegated to the basement by modern linacs, x-ray film has been supplanted by digital radiography, CT scanning is now helical, and just as few medical physicists now know what a typewriter is, so the modern assessment of quality has improved and supplanted the historical method. The assessment of quality has been democratised because now Scientists can “vote” on the quality of a manuscript by citing it. Previously quality was assessed subjectively by individual scientists reading the article. Now quality assessment is a transparent collegial exercise in consensus by expert researchers and no longer at the whim of individual scientists.

Opening Statement—Jamie Trapp

The modern concept of citation indexing in research was introduced by Eugene Garfield in 1955, based on the earlier legal citation database *Shepard's Citations* which dates back to 1873 [9]. The purpose of a legal citation database is to ensure that precedents set in legal cases can be found easily and used in later court decisions. Garfield argued that a similar citation database for academic research can be a useful tool to more easily find earlier research in a topic so that it can be built upon by researchers, rather than spending time rediscovering the same, and also that criticisms of earlier research are more easily available to inform future directions. In Garfield's paper, the term Impact Factor is introduced relating to the number of citations an article or journal receives.

Garfield's argument that the number of times a particular item/discovery/methodology is subsequently used to build later research [9] holds true, but simply counting all citations also includes the number of times an item has been criticised. For example, Radicchi showed that papers which have been criticised via Comments are cited more than non-commented papers, and more likely to be among the most cited of a journal [10]. Although a paper which is commented and criticised is not necessarily of low quality (and the criticism itself is often debunked by author responses), there are certainly some instances where criticism is valid, yet this very act of criticising a research paper contributes to the ‘impact’ and therefore ‘quality’ via the citation.

It has recently been shown that 74% of citations occur in the Introduction and Discussion sections of papers [11]. A proper Introduction to a paper puts the present research into context in terms of recent progress in the topic and gaps in the knowledge that need to be filled, and likewise a good

Discussion puts the results of the research into context. However, when a paper is cited in an Introduction or Discussion section one has to consider whether it truly informs the present research, or is just listed as a convention where other work is acknowledged, e.g. authors, particularly inexperienced authors, might feel compelled to have a long introduction citing many articles in order to follow the format of other papers they have seen. If these cited papers don't directly lead to the work being presented, they are essentially just saying ‘these other people worked on something similar to us, but not quite the same’.

Papers which are truly influential, such as introducing a new process, procedure, or discovery are more likely to be cited in the Methodology, as this is where they influence future work. Therefore, one must ask the question of whether citation metrics should be based on all citations, or only the 26% which do not occur in Introductions or Discussions.

The above arguments can be taken further when considering the Matthew effect [12], which is very eloquently described with an example given by Ioannidis [13]. In short, the Matthew effect in relation to citations is the concept that a paper with many citations will gather even more citations than, say, a comparable similar paper with fewer or no citations (a self-propagating cycle). Ioannidis describes the situation of over simplified research tools, however the concept can generally be expanded to other situations such as inexperienced authors who might see certain papers being cited in Introductions, which then lead them to think that they should cite the same papers in their own Introductions as some type of unwritten convention in the field (even if not directly relevant), thus the citation cycle is established.

Finally, lack of citations does not mean a piece of work is necessarily of lower quality. The number of citations can be influenced by numerous factors [14], including age of the article (even Nobel prize winning papers started with 0 citations), discipline norms, access [15], self-citations (and self-citation schemes) [16], and even reviewer-coerced citations [17].

Rebuttal—Martin Caon

Implicit in the statement that “citations are a good way to determine the quality of research” is that the citations are “appropriate” (and appropriately placed within the article). A spurious citation will leave the reader wondering at the quality of the article being read and perhaps to its dismissal as unworthy of study. An article that is cited in the Introduction as reporting a novel result cannot be said to be less worthy of citation than one cited in the Methods section for, say, the kilovoltage selected on the machine. If inappropriate, the snowballing of citations, because “everyone is citing

it”, is noticed and dismissed by clever researchers—and that should include all of us.

Peer criticism is an essential part of improving the quality of research and by attracting criticism, the shortcomings of research are uncovered and avoided in future. This improves quality. So while a critical citation may not flag the cited paper as a quality one, it contributes to future quality.

My esteemed colleague contends that lack of citation does not imply lack of quality. I agree, but that is not the proposition we are debating.

Rebuttal—Jamie Trapp

I acknowledge the excellent point of my esteemed colleague that the number of reads and number of downloads of an article is a good measure of its scientific interest. Likewise, the use of normalised citation metrics such as relative citation ratio and others including relative citation impact, relative citation rate, etc. can allow for variations between different research fields.

However, the premise of citation counting is that the total number of citations received is the sole indicator of the quality or impact of a piece of research, regardless of whether the citation is positive, negative, or just there to fill in the introduction and bibliography.

Citation counting is an easy thing to measure, originating from the time when bibliometrics started (when tables of contents and bibliographies were the easiest data to analyse with the tools at the time). To make citation counting appear more meaningful (or perhaps hide the flaws of the system), contemporary research managers and organisations have chosen to use words like ‘research quality’ and ‘impact’ to disguise the fact that it is merely a ‘total citation count’, thereby surreptitiously giving a new, different, definition to these words when the consumer assumes the older, traditional meaning. This is classic case of cognitive bias (imagine, for example, if ‘quality’ suddenly changed to mean ‘word count’, then very long articles would be considered more important than succinct articles, and important works in currently high ranking journals with strict word limits, of say 1500 words, would suffer).

I agree with my esteemed colleague in his point that times change and I would take this a little further. Surely, with the development of data mining techniques a better measure of quality and impact can now be found—there is no longer the need to stick with the ‘easy’ solution.

Note: To support this debate and provide data for the future, a separate article has been produced in this Issue [18]. The article is intentionally of obvious and deliberate poor quality (in the traditional sense of the word). The purpose of the other article is to act as an experiment within itself, to see if a poor article gains citations how this would be considered for

‘quality’ and ‘impact’ in the contemporary sense used in this debate.

References

- Guédon J-C (2001) In oldenburg’s long shadow: librarians, research scientists, publishers and the control of scientific publishing. Association of Research Libraries, Washington, DC
- Larivière V, Gingras Y, Archambault É (2009) The decline in the concentration of citations, 1900–2007. *J Am Soc Inf Sci Tec* 60:858–862
- Armato S, Baldock C, Orton CG (2015) Hybrid gold is the most appropriate open-access modality for journals like Medical Physics. *Med Phys* 42(1):1–4
- Meho LI (2007) The rise and rise of citation analysis. *Phys World* 20:32–36
- Biagioli M, Lippman A (2020) Metrics and the new ecologies of academic misconduct. In: Biagioli M, Lippman A (eds) *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. MIT Press, Cambridge MA
- Lord Kelvin (1883) Lecture to the institution of civil engineers https://todayinsci.com/K/Kelvin_Lord/KelvinLord-MeasureQuote500px.htm. Accessed 11 August 2020
- Hutchins BI, Yuan X et al (2016) Relative citation ratio (RCR): a new metric that uses citation rates to measure influence at the article level. *PLoS Biol* 14(9):e1002541
- Purkayastha A, Palmaro E et al (2019) Comparison of two article-level, field-independent citation metrics: Field-Weighted Citation Impact (FWCI) and Relative Citation Ratio (RCR). *J Informetrics* 13:635–642
- Garfield E (1955) Citation indexes for science a new dimension in documentation through association of ideas. *Science* 122(3159):108–111
- Radicchi F (2012) In science “there is no bad publicity”: papers criticized in comments have high scientific impact. *Sci Reports* 2:815
- Gerlach M, P-C et al (2019) Large-scale analysis of micro-level citation patterns reveals nuanced selection criteria. *Nat Human Behav* 3:568–575
- Baldock C, Schreiner LJ, Orton, (2017) Famous medical physicists often get more credit for discoveries due to their fame than less prominent scientists who may have contributed as much or earlier to these developments. *Med Phys* 44(4):1209–1211
- Ioannidis JPA (2018) Massive citations to misleading methods and research tools: Matthew effect, quotation error and citation copying. *Eur J Epidemiol* 33:1021–1023
- Tahamtan I, Afshar AS, Ahamdzadeh K (2016) Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics* 107:1195–1225
- De Groote SL (2008) Citation patterns of online and print journals in the digital age. *J Med Library Association* 96(4):362–369
- Ioannidis JPA, Baas J et al (2019) A standardized citation metrics author database annotated for scientific field. *PLoS Biol* 17(8):e3000384
- Wren JD, Valencia A, Kelso J (2019) Reviewer-coerced citation: case report, update on journal policy and suggestions for future prevention. *Bioinformatics* 35(18):3217–3218
- Trapp JV (2020) Citations equals research quality? If you agree then don’t cite this stupid, totally terrible article. *Phys Eng Sci Med*. <https://doi.org/10.1007/s13246-020-00942-8>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.