Hydrobiologia (2013) 702:1–4
DOI 10.1007/s10750-012-1392-6

OPINION PAPER

The propagation and dispersal of misinformation in ecology: Is there a relationship between citation accuracy and journal impact factor?

D. C. Drake · B. Maritz · S. M. Jacobs · C. J. Crous · A. Engelbrecht · A. Etale · M. J. Fourie · D. G. Furniss · S. L. Scott · S. Parusnath · D. R. Tye

Received: 24 July 2012 / Revised: 5 October 2012 / Accepted: 5 November 2012 / Published online: 19 November 2012

Research almost always builds upon existing, peer-reviewed literature. This is how we understand the development of ideas in our disciplines, formulate meaningful new questions, and advance our knowledge. But several of the authors of this article have noted disturbingly inaccurate reporting of our study findings when cited by others. We have also noted copying and pasting of phrases and citations from published articles in a way that distorts or even completely changes the meaning of the source. For us, this calls into question the quality and meaning of scientific advancement. It is tempting to believe that citation practices are worse in lower-impact journals, and that we can more easily rely on the quality of the logic and ideas presented in more prestigious journals. But we have noticed inaccurate citations in some high-impact journals as well. These observations led us to attempt to quantify the extent of inaccurate citation in the recently published ecological literature, to compare citation accuracy with journal impact factor (IF) and also compare our results to a previous study of citation accuracy in journals with IF > 1 (Todd et al., 2007).

Publication is one of the primary measures of success in the competitive field of academics. This, perhaps unavoidably, encourages the stretching of research to maximise the quantity of publications produced (Lawrence, 2003), can discourage transparency when it comes to sharing of information on methods (Anderson et al., 2007) and can seem to reward minor misconduct such as poor citation practices. The resulting unintentional or negligent spread of misinformation undermines scientific advancement and individual's professional credibility over the long term. Financial incentives for publishing houses may also contribute to the problem of poor citation practices. Scientific writing (to our surprise, at least), is not only our primary means of communication with the wider scientific community, but is also a product with a market value (Bergstrom and Bergstrom, 2006). Growing financial incentives (e.g. library subscription fees) are at least partially responsible for the recent proliferation of new journals in ecology, and more broadly in science. Since 1980, the number of ecology-related journals listed by the Institute for Scientific Information (ISI) has doubled (currently 129), with for-profit or joint for-profit/non-profit publishers accounting for most of this increase.

A concern over the perceived growth of unethical practices in science is implicit in many opinion pieces

Handling editor: Sidinei Magela Thomaz

D. C. Drake (✉) · B. Maritz · A. Etale · D. G. Furniss · S. L. Scott · S. Parusnath · D. R. Tye
School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Private Bag 3, WITS 2050, Johannesburg, South Africa
e-mail: Deanne.Drake@wits.ac.za

S. M. Jacobs · C. J. Crous · A. Engelbrecht · M. J. Fourie
Department of Conservation Ecology and Entomology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

(e.g. Lawrence, 2007; Anderson et al., 2007), but few studies show unequivocal evidence of a shift in ethics in publication. The worst examples of scientific misconduct involve fabrication or falsification of data (quantified by Fanelli, 2009). A recent analysis of publication in biomedical and life-science research showed a tenfold increase in the percent of retracted articles for reasons of fraud since 1975 (Fang et al., 2012). This may, however, reflect greater scrutiny, greater frequency of misconduct, or, most likely, both. Poor citation practices are clearly a less odious type of scientific misconduct and have been analysed and criticized in a number of contexts. Bibliographic errors (errors in citations such as incorrect journal titles), for example, occurred in 45 % of cases examined (Harper, 2001). An examination of citation behaviour in ethnobotany found that the theoretical contributions made by an author were often not cited, possibly due to the fact that the citer has not read the original article or at least only read the article superficially (Ramos et al., 2012). Simkin and Roychowdhury (2003), using highly cited papers in physics, quantified the propagation of bibliographic errors assuming that copied errors meant that the original article was not read. Based on this they estimated that only about 20 % of cited articles were actually read by the citer. This estimate is shockingly low—but some grey area does exist in that there are perhaps few papers that are read very, very thoroughly—i.e. it is easy to read parts of papers that are of particular relevance to one's own work. It is also, admittedly, easy to misconstrue the finer points of a paper when one is looking for support of one's own ideas.

Quantification of citation accuracy and its relationship to

We selected 33 journal titles from those listed under the 'ecology' heading by the ISI in 2011 and two general science journals that contain ecological articles. Impact factors ranged from ~0.1 to ~31.4. Three or four papers published between 2009 and 2012 were randomly chosen from each of these journals, and one key citation reporting an important premise for the paper was chosen from the Introduction of each paper for evaluation. Self-citations and statements referring to multiple citations were excluded. A total of 124 statements were evaluated. Six to 15 statements were assigned to each author for accuracy checking. We assigned ten of the citations to two checkers for consistency of judgements within the group. The citation checkers were responsible for locating the originally cited source and categorising the accuracy of the statement (1 = completely supported/accurate; 2 = partially supported; 3 = completely unsupported, and 4 = source material could not be located). The latter statements were unveriable and were subsequently removed from the calculations of accuracy. Empty citations (copy and pasting of ideas and citations, frequently out of context) were noted when evident. This was a judgement call, but in general word-for-word duplication of phrases more than five or six words in length were noted as 'empty'. Empty citations were also ranked for accuracy. To minimise bias, the authors, journals and impact factors were not provided to the citation checkers in most cases. Chi-squared tests for goodness of fit were used to compare frequencies of citations in IF categories determined here, and to compare our results to those of Todd et al. (2007). Overall and journal-specic average accuracies were calculated assuming that a score of 1 was 100 % accurate, 2 was 50 % accurate and 3 was 0 % accurate.

Outcomes

Only 67 of the 124 citations (54.0 %) examined here were judged completely supported, 33 (26.6 %) were partially supported, 14 (11.3 %) were totally unsupported, and ten (8.1 %) were unveriable. Seven of the ten unveriable sources were from obscure sources not available at either of the well-resourced universities (with access to most electronic journals in ecology) with which the authors of this article are associated, one was cited as 'in press', and two were from unobtainable internet sources. Seven of the ten unveriable sources were cited in journals with IF < 1. Empty citations were noted in ten of the 114 verifiable citations. This strongly suggests that the authors did not read the original source in at least ~8 % of their introductory citations.

Journal IF was strongly related to accuracy (Table 1; Fig. 1). Differences between accuracy based on IF were highly significant (Table 1, $\chi^2 = 39.8$, $v = 4$, $P < 0.001$). Verifiable citations in journals with IF > 5.0 were 91.4 % accurate while verifiable

Table 1 Average accuracy calculated for a range of journals by IF

Impact factor	<1	1–5	>5
Average % accuracy	49.0	74.6	91.5

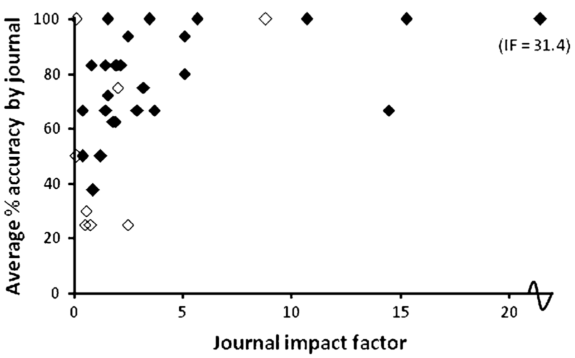

Fig. 1 Each diamond represents the average citation accuracy rating for one journal (unless there was only 1 value for a journal). Solid diamonds represent averages for journals in which all citations were located, open diamonds represent average values for journals in which one or two sources were unverifiable

citations in journals with an IF of <1 were only 49 % accurate (Table 1). We note that in many cases average accuracy in low-IF journals was high, and that the average citation accuracy in lower-IF journals was far more variable than that in high-IF journals (Fig. 1).

Consistency of judgements within the group was lower than expected, suggesting that personal interpretation of 'accuracy' can vary considerably. Only six of the ten citations checked by two people received the same accuracy rating from each. This highlights the need for unambiguous writing in science.

Are our results consistent with previous studies?

A similar study on citation accuracy in ecology journals was published by Todd et al., 2007 (based on analysis of citations in papers published mostly in early 2006, Table 2). We were not aware of this study until our reviews were essentially completed. Similar methods were used in the two studies, although the analysis by Todd et al. was limited to journals with IF > 1 and no unverifiable citations were reported. The proportion of citations in journals with IF > 1 that

Table 2 Citations in accuracy categories

	Todd et al. (2007) IF > 1 % of papers	This study 2012 IF > 1	This study 2012 all IF
Completely supported	76.1	68.9[a]	58.8[a]
Partly supported	11.1	22.2[a]	28.9[a]
Totally unsupported	7.2	8.9[a]	12.3[a]
Unverifiable	Not reported	4.3	8.1
Empty	5.6	5.4	8.0
n	306	92	124

[a] Values from this study calculated by excluding unverifiable sources (for direct comparison to Todd et al. 2007)

were completely supported was 76.1 % in the 2007 study and 68.9 % here (Table 2). These rates are not significantly different, showing no convincing change in citation accuracy over this relatively short period of time. This may bear re-examination in the future.

Our estimate of citation without reading differed dramatically from Simkin and Roychowdhury's 2003 estimate. Our estimate was based on apparent and fairly obvious copying and pasting of ideas and citations, leading us to conclude, conservatively, that 8 % of original papers were not read. Simkin and Roychowdhury (2003) estimated that only 20 % of cited papers are actually read by the citer (thus 80 % are not read) by assuming that any propagated (copied and pasted) errors in the actual citations meant that papers were not read. Technically, both estimates could be correct, but it is our opinion and hope that the actual rate is closer to 8 % than 80 %.

Conclusions

Writing scientific papers is usually slow, hard work and it is not possible to know everything. Our access to information (literature) and ideas has probably outpaced our ability to use this information. Moreover, the proliferation of journals and papers and the increasing availability of information make it increasingly difficult to remain up-to-date with the literature. Most of us struggle with citation accuracy at some point in our careers, perhaps especially early in our careers when we are less familiar with the ideas that underlie our disciplines or when we branch into new areas of research. Our susceptibility to rushing and carelessness may also increase when we are pressured by the demands of promotion or the tenure process. Only about half of the citations examined here were judged completely accurate. This is discouraging, but not unexpected given the general agreement with the results of Todd et al. (2007). The significantly higher accuracy of citations in high-IF journals quantified here, however, lends credibility to the IF system, i.e. articles published in high-IF journals are, on average, of better quality and more reliable in respect to foundational literature. Other authors have made articulate arguments that metrics such as IF and H′ factor (which was only introduced 7 years ago; Hirsch, 2005) have led to or encouraged an increase in scientific misconduct (e.g. Lawrence, 2007; Anderson et al., 2007) but the link is difficult to quantify. We remain ambivalent about IF and other currently popular metrics (see also Todd and Ladle 2008), even though IF does appear to reflect citation accuracy and, by extension, quality of research.

An analysis such as this should go beyond describing problems and offer some potential solutions. Given that the advancement of any science must rely on the desire of practitioners to produce high quality, useful work, we do not favour a 'legislative' approach. There are, however, several potential self-regulatory steps that we can take as individuals in addition to being careful with our own citation practices: Academic supervisors can discuss the importance of reading original papers and the ethical problems involved in poor citation practices with their graduate students and in classes. Reviewers may choose to check one citation in each manuscript for accuracy and automatically reject with possibility to resubmit if it is not accurate. Citation accuracy checks may also be included as part of an internal manuscript review process within an organisation or laboratory. The use of plagiarism detection software as part of the review process might help alleviate inaccurate citing. Last, we believe that a number of the statements selected for this exercise were established ideas (e.g. 'Water is critical for life on earth.') and did not require citation. Curbing such unnecessary citation may help contribute to improvement of citation practices. Ultimately the pervasiveness of poor citation practices and more serious scientific misconduct will be determined by the culture that we ourselves build in the workplace and the classroom.

References

Anderson, M. S., E. A. Ronning, R. De Vries & B. C. Martinson, 2007. The perverse effects of competition on scientists' work and relationships. Science and Engineering Ethics 13: 437–461.

Bergstrom, C. T., & T. C. Bergstrom, 2006. The economics of ecology journals. Frontiers in Ecology and the Environment 4: 488–495.

Fanelli, D., 2009. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS ONE 4: e5738. doi:10.1371/journal.pone.0005738.

Fang, F. C., R. G. Steen & A. Casadevall, 2012. Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences. doi:10.1073/pnas.1212247109.

Harper, J. A., 2001. Citation inaccuracy in a scientific journal. A continuing issue. Science and Technology Libraries 20: 39–44.

Hirsch, J. E., 2005. An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences 102: 16569–16572.

Lawrence, P. A., 2003. The politics of publication. Nature 422: 259–261.

Lawrence, P. A., 2007. The mismeasurement of science. Current Biology 17: 583–585.

Ramos, M. A., J. G. Melo & U. P. Albuquerque. 2012. Citation behaviour in popular scientific papers: what is behind obscure citations? The case of ethnobotany. Scientometrics. pp. 1–9, doi:10.1007/s11192-012-0662-4.

Simkin, M. V., & V. P. Roychowdhury, 2003. Read before you cite! Complex Systems 14: 269–274.

Todd, P. A., & R. J. Ladle, 2008. Hidden dangers of a 'citation culture'. Ethics in Science and Environmental Politics 8: 13–16.

Todd, P. A., D. C. J. Yeo, D. Li & R. J. Ladle, 2007. Citing practices in ecology: can we believe our own words? Oikos 116: 1599–1601.