Erratum Citation and Accuracy in the Publication Record

Marshall Thomsen, Department of Physics and Astronomy. Eastern Michigan University, USA

Christopher Aubin, Department of Physics and Astronomy. Eastern Michigan University, USA

Barbara Hall, Department of Physics and Astronomy. Eastern Michigan University, USA

Matthew Knee, Department of Physics and Astronomy. Eastern Michigan University, USA

Keywords: Erratum, Citation, Publication

Technological advances have greatly influenced the ways in which science is communicated. However, the refereed journal remains an important element of the system, providing a permanent record of information with some quality control over the scientific content. In trying to keep abreast of recent developments in a field or when entering a field of study for the first time, scientists often rely on the refereed journal as their primary information source. Thus accuracy of the written record becomes a significant issue.

While much has been written about the publication process in general, (1) we will focus on a small piece of the process that lends itself to accumulation of basic statistical information and, we hope, provides some insight into other broader aspects of publication. In particular we will look at physics papers that have an erratum associated with them and study how these papers are cited in subsequent literature. There are several issues we will examine. If an erratum is written, how likely is it that those who have read the original paper also will have read the erratum? If a corrected paper is cited, how likely is it that the authors who cited the paper also cited the erratum? Is it misleading to cite the original paper but not the erratum? Do authors typically cite their own errata?

Some of these questions have been addressed before. For instance a 1990 study of retracted medical papers showed that retractions tended to reduce, but not eliminate, citation rates (2). A 1995 study of errata in physics journals showed that when corrected papers are cited, most often the corresponding erratum is not cited (3). The authors of the study commented at the time that part of this citation problem was associated with the logistical issue of locating an erratum. It is much easier to search the publication record backward in time by studying citations. Moving forward in time to locate errata requires scanning journal contents or using an index (such as the *Science Citation Index*). The authors speculated that as more journals were provided in an electronic format, locating errata would be easier since the original paper presumably would be linked electronically to the erratum.

The American Physical Society now has a large collection of its journals available online via subscription. All of their recent online papers that have an associated erratum have a link to that erratum. We thus undertook a new study to determine if this electronic linking has improved the

Corresponding author: Marshall Thomsen, Department of Physics and Astronomy, Eastern Michigan University, Ypsilanti, MI, 48197, 734-487-8794 (voice), 734-487-0989 (fax), phy_thomsen@online.emich.edu.

citation rate of errata. Curiously, we find that, if anything, the citation rate for errata seems to have decreased since the introduction of the electronic format.

Study Design

Our study consisted of an examination of the citations of 14 papers from *Physical Review Letters* in 1995 and 1996 that had errata making nontrivial corrections. We included papers with calculational errors requiring replotting graphs or reproducing tables, papers in which derivations needed modifications, papers in which data needed to be reanalyzed due to misinterpretation, etc. We excluded papers in which simple typographical errors were corrected or acknowledgments of related work were added. The goal was to focus on papers in which there was a scientific error of substance being corrected by the erratum. At the same time, none of the errata reported on here amount to a complete retraction of a paper. For clarity in discussions below, we refer to these 14 papers as the *primary* papers.

We selected our primary papers from amongst the first papers to become available in the *Physical Review Letters* online collection. Hence the primary papers all have electronic links to their errata, and they have all been available in the literature for several years (thus increasing their chances of being cited).

Physical Review Letters is one of the most selective physics journals, containing papers describing some of the most recent and significant advances across all physics disciplines. We focussed on these papers since they are higher profile and hence likely to produce a greater set of citation data. In contrast, the 1995 study of errata in physics journals included papers from both *Physical Review Letters* and *Physical Review B*, the latter being a more specialized journal. That study showed that papers in *Physical Review Letters* tend to be cited two to three times as often as papers in *Physical Review B.* The 1995 study also showed the citation rate for errata in *Physical Review Letters* was substantially higher than that for *Physical Review B.* Thus our present study focuses on a journal with a *relatively* high erratum citation rate.

We attempted to identify all papers that had cited the primary papers and/or their associated erratum, using the *Science Citation Index* as our main tool. We located 507 papers citing the primary papers and/or their errata. We refer to this collection of papers as *secondary* papers. It is interesting to note that a small portion of these secondary papers cited only the erratum and not the corresponding primary paper. As a spot check on the accuracy of *Science Citation Index*, we used the citation list provided by the online version of *Physical Review Letters*. It should be noted that the journals used in this citation index are much more limited in scope than those used to assemble the *Science Citation Index*, listing citations by only American Physical Society journals. We selected three primary papers from our list that, according to the *Science Citation Index*, had no citations to their erratum. We verified this finding with all available listings on the more limited *Physical Review Letters* citation data base and also confirmed that all 21 secondary papers appearing on this database also appeared on the *Science Citation Index* data base. That is, we discovered no evidence that *Science Citation Index* was omitting papers appropriate for our secondary category.

Results and Discussion

The collection of secondary papers was divided into two categories. The first category contained those papers in which there was an overlap between the authors of the secondary paper and those of the cited primary paper. The second category consisted of those secondary papers in which there was *not* any overlap of authorship with the cited primary paper. The purpose of this division was to address separately the questions of how often authors cite their own errata and how often independent authors cite errata. The cases with overlapping authors will be considered first.

Table 1 shows data for authors citing their own errata. We exclude from consideration in the secondary paper data set those papers published prior to the appearance in print of the erratum. We are left with 59 secondary papers that could have cited an erratum. Of these, 25 (42 percent) actually did cite the erratum. The reason for the remaining 58 percent of the secondary papers not including the erratum citation is not clear. One possibility is that the author of the primary paper and erratum chose not to cite the erratum. Another possibility is that the person or persons of the secondary paper who took the most responsibility for writing that paper were not among the authors of the primary paper. In this case, it would be possible for the

Paper Identification Number	Potential Erratum Citations	Actual Erratum Citations
1	1	1
2	5	0
3	2	0
4	4	2
5	4	3
6	3	2
7	0	0
8	5	2
9	3	1
10	1	0
11	5	0
12	3	0
13	23	14
14	0	0
Total	59	25

Table 1. Analysis of citations by one or more authors of the original (corrected) paper. Potential erratum citations represent the total number of papers citing the original paper, its erratum, or both. Actual citations represent the number of times the erratum was cited. Only papers appearing after the publication date of the erratum were considered in columns 2 and 3.

writer of the secondary paper to be unaware of the existence of the erratum. However, assuming the erratum author read through the secondary paper prior to publication, then either that author chose not to add the erratum citation to the list or overlooked the absence of the erratum in the references. We will return to this issue later.

Table 2 shows data for secondary papers sharing no authors in common with the cited primary paper. We exclude from the secondary paper data set those papers that did not appear in print at least one year after the publication date of the erratum. This is to ensure that the authors of the secondary paper had the chance to see the erratum at the time they were writing their own paper. After reducing the data set as described, 355 secondary papers remained. Of these, just 59 (17 percent) cited the erratum. The 1995 study of 9 primary papers in *Physical Review Letters* had a citation rate of 39 percent (51 of 131) when a similar approach to data analysis was used. While there are obviously statistical fluctuations associated with this sampling, it is worth noting that only 4 of the 14 primary papers in the present study had an erratum citation rate exceeding the 39 percent average from the previous study. It is thus safe to conclude that the advent of electronic journals has not had the desired impact on erratum citation.

We now return to the issue of the extent to which it is a problem that errata are not generally being cited. There are three fundamental questions. First, does the reader of the secondary paper need to be aware of the erratum? Second, will a reader discover an erratum based on information provided by the authors of a secondary paper? Third, whose responsibility is it to locate the erratum?

We will examine the first question in the context of the errata discussed here: those providing substantive corrections. The 1995 study of erratum citations showed that a little more than half of the primary papers examined were cited “in passing” in the secondary reference. In these cases, the secondary authors were primarily acknowledging the work of others in the field rather than laying down specific ground work for their own paper. These citations typically occur in the introductory section. The remaining citations to the primary papers

Paper Identification Number	Potential Erratum Citations	Actual Erratum Citations
1	6	0
2	4	0
3	22	0
4	13	8
5	8	2
6	15	7
7	3	1
8	6	6
9	2	0
10	2	2
11	8	0
12	17	1
13	248	32
14	1	0
Total	355	59

Table 2. Analysis of citations not involving authors of the original (corrected) paper. Potential erratum citations represent the total number of papers citing the original paper, its erratum, or both. Actual citations represent the number of times the erratum was cited. Only papers appearing one year or more after the publication date of the erratum were considered for columns 2 and 3.

indicated that the authors of the secondary paper were using one or more results or ideas from the primary paper to support their own work. This latter group of citations raises the erratum citation question in a direct way. Even if the erratum did not have any direct bearing on the portion of the primary paper that was drawn upon, citing the erratum is still significant in that it indicates that the secondary authors are aware of its existence and took it into account (if necessary) in preparing their paper. Furthermore, a reader of the secondary paper who is inclined to investigate the topic more thoroughly can be misled if unaware of the existence of the erratum.

Returning to the citations "in passing," there are typically two motivations for providing such a citation. First, one may wish to pay tribute to predecessors in a particular field. Second, one may wish to direct the reader to papers with relevant background information. Papers cited for the second reason also should have their corresponding errata cited as a service to the reader.

We now consider the second question: Will a reader discover an erratum based on information provided by the authors of a secondary paper? Obviously, if the authors have cited the erratum, the answer is yes. If the authors have not cited the erratum, then there are a number of ways in which the reader may discover the erratum. For instance, the reader may look up the primary paper electronically and discover a link to the erratum. This is constrained by the fact that not everyone has access to journals in electronic form and not all journals are available in this format. When using journals in printed format, the reader must rely on techniques such as searching the journal index for errata or using a more comprehensive index such as the *Science Citation Index*. Otherwise, the erratum might only be discovered by chance while browsing through an issue.

Perhaps authors of the secondary papers assume that the interested reader will be able locate errata on their own. Short of taking a survey, we can only speculate as to whether this is the rationale for authors not citing errata. However, given the fact that this citation problem predates the electronic journal format, it is unlikely that most authors are consciously electing not to cite an erratum on these grounds. It is possible, however, that this rationale may explain the *drop* in the erratum citation rate between the 1995 study and the present study.

This brings us to our final question: Who is responsible for locating the erratum? It is reasonable to view a reference to a paper as a recommendation of a source to consult for further information. In making that recommendation, an author thus has some responsibility to ensure that it is a sound recommendation. However, a reader of a secondary source who is making an in depth study that requires consulting cited references also bears some responsibility for seeking out relevant errata. While it is difficult to say who has the greater responsibility, neither side can be removed from the equation.

It is worth noting that the secondary author is somewhat more likely to be aware of the erratum than the reader of the secondary paper, because often one cites papers written by people with whom one has had some direct or indirect association or by people whose work one has followed closely. This correlation of course is particularly true in the case of a secondary author also being a primary author. This observation coupled with the fact that erratum citation is not routine even when there is an overlap between primary and secondary authors leads us to speculate that secondary authors are not always citing errata even when they are aware of their existence. Why is this the case? One possible argument is that some perceive there is a stigma associated with publishing an erratum and hence they prefer not to call attention to it. Arguably, however, publishing an erratum is a sign of both integrity and attention to detail. It is likely most physicists who have done any significant amount of research have encountered papers that should have had errata but the authors chose not to write one. Clearly there is more damage to the field by uncorrected papers than by those properly corrected. The irony is that if one takes the time to do the right thing—to write the erratum—it is not clear how many people are going to read it.

Conclusions

We conclude as the previous study did with the hope that eventually the conversion of printed journals into electronic databases will resolve the erratum citation problem. In particular, if we reach a point where all journals are in a dynamic electronic database that is updated with appropriate links as errata are written and electronic access is as pervasive as printed access, then it becomes unnecessary to cite errata. While many physics journals are headed in this direction, it is not clear if and when all

will get there. Particularly problematic is the task of going through older journals and converting them to electronic format. In the meantime, citing errata will continue to be an important part of the service provided by authors in their reference sections.

Even if the erratum citation problem is resolved, the fact that it has existed raises more general questions concerning the integrity of the publication record. Specifically, is the accepted norm that authors *do* have a responsibility to cite errata or is the expectation that the reader is responsible for locating them? More generally, is this problem a sign of pervasive sloppy practices in publication or is it merely a situation of ill-defined responsibility? The answers to these questions will become clearer only after more discussion within the scientific community.

Paper Number	Paper Reference	Erratum Reference
1	**74:**694	**75:**355
2	**74:**1839	**76:**4097
3	**74:**4101	**76:**4293
4	**75:**1447	**75:**3781
5	**75:**394	**75:**1874
6	**75:**3549	**77:**2345
7	**75:**4413	**76:**3242
8	**76:**014	**76:**2826
9	**76:**1031	**77:**4278
10	**76:**2848	**77:**5148
11	**76:**3955	**78:**3227
12	**77:**127	**78:**3587
13	**77:**3865	**78:**1396
14	**77:**4066	**78:**162

The table above provides references to the papers used in this study. All are from Physical Review Letters, *published by the American Physical Society. The format is volume:beginning page.*

Acknowledgments

It is a pleasure to acknowledge support from the Graduate School, the College of Arts and Sciences, and the Department of Physics and Astronomy, all at Eastern Michigan University.

Bibliography

1. See, for instance, LaFollette MC. Stealing into print: fraud, plagiarism, and misconduct in scientific publishing. Berkeley (CA): University of California Press; 1992.
2. Pfeifer MP, Snodgrass GL. The continued use of retracted, invalid scientific literature. Journal of the American Medical Association 1990 March 9; 263(10): 1420-1423.
3. Thomsen M, Resnik D. The effectiveness of the erratum in avoiding error propagation in physics. Science and Engineering Ethics 1995; 1(3):231-240.