

What Fraction of Literature References are Incorrect?

HELMUT A. ABT

Kitt Peak National Observatory, National Optical Astronomy Observatories,¹ Box 26732, Tucson, Arizona 85726-6732
 Electronic mail: apj@noao.edu

Received 1991 October 16; accepted 1991 December 20

ABSTRACT. From a systematic study of 1009 references in *The Astrophysical Journal* it was found that 12.2% had errors. Only 0.4% of the referenced papers could not be found at all; another 3.0% were found by searching in volume, annual, and five-year indices and the remaining correctly and incorrectly referenced papers were found right away. Another 8.3% of the references have errors in the first authors' names or in the journal names, and volume and page numbers such that they could be misplaced in the *Science Citation Index* (SCI). However, the compilers of SCI match all citations against a computerized file of the source papers and correct some of the citations, so that only 3.6% of the citations are missing or displaced in SCI.

1. INTRODUCTION

References at the ends of scientific papers serve two purposes. The main purpose is to locate relevant publications whose contents are needed to support the scientific arguments or to provide data. For that purpose the journal names, volumes, and page numbers are generally sufficient information to find the source papers. Information about the authors' names and publication years is superfluous except that it fixes in our minds the identification of the researchers and publication years when the research was done. Paper titles, which cannot be published in most journals because of the space they take, often tell us whether we really wish to consult the source papers or not.

The second purpose of references is to provide the database for citation studies, primarily through the *Science Citation Index* (hereafter referred to as SCI). That information is useful for searching for papers on specific topics and for citation studies of individual authors. For that purpose the references need to have the first authors' names as well as the paper locations. For instance, if an author's name is given incorrectly in a reference as "C. Anderson" when it is really "C. Andersson," the incorrect citation would be found 107 pages away from the correct citations in the 1980–1984 SCI. Most people will not think of looking at the wrong spellings when they count citations. Another example is that if one cites "H. Schmidt" when the paper is really by "H. U. Schmidt," one will find that there are 23 other authors listed between those two in the same five-year index and, again, citation counts may be incorrect.² References should be accurate for use in the citation database.

We have all found errors in reference lists but we can usually decipher them. Most of the errors are typos, e. g., to list a page as 431 instead of 413, or 121 instead of 1121. In looking at the SCI we find that for nearly all papers having 75 or more citations, there are one or more citations listed above or below with different page numbers that probably refer to the same paper. (But it could be that the author does have a paper on page 413 as well as on 431!) Citations with the wrong volume numbers or journal

names are found further away, but then it is not clear whether those refer to the same papers.

How frequent are reference errors? The answer to that is the purpose of this study.

Reference lists in journals and books are the only parts that are not reviewed systematically. Referees and editors³ do not generally have the time to look up each reference, which average 35 per paper in *The Astrophysical Journal*, and the manuscript editors do not have the available libraries. Thus reference lists will continue to be the weakest part of papers until our scientific literature is transferred to computer memory with automated search techniques.

2. DATA AND RESULTS

I looked up 1009 references published in the 1991 January 1 issue of *The Astrophysical Journal*, Parts 1 and 2. Perhaps the results are typical of other astronomical journals, although the results do depend upon the care shown by individual authors. Not included in this search, of course, were papers "in press." To maximize the number of references searched with minimum effort, I searched for only those references to eight journals: A&A, A&A Suppl., AJ, ApJ, ApJ Suppl., MNRAS, Nature, and PASP. I doubt that the errors made with those journals are more or less frequent than for less frequently cited journals or monographs. Monographs generally do not have volume numbers; otherwise they have just as many opportunities for author and typesetting errors.

I found that 123 of the 1009 references, or 12.2%, had errors of some kind. The breakdown is given in Table 1.

We see that only four references, or 0.4%, could not be found at all. For those the volume and page numbers were both incorrect, and perhaps even the journal names. Under those circumstances it is unclear to which papers the authors intended to refer. But that does not seem to be a high error rate.

For 30 references, or 3.0%, I was able to locate the likely paper by using the volume, annual, or five-year indices. The most frequent errors were in page numbers (2.0%) or volume numbers (0.8%). Most of the errors looked like typos in writing, typing, or typesetting the ref-

¹Operated by the Association of Universities for Research in Astronomy, Inc., under contract with the National Science Foundation.

²I have been aware of this frequent error and when I made citation counts, I generally looked for alternate uses of initials to recover more of the citations, e.g., A. R. Sandage as well as A. Sandage.

³To my knowledge the only astronomical editor who looked up each reference in his journal—and found many errors to which he called the authors' attention—was William P. Bidelman when he edited these *Publications*.

TABLE 1
Summary of Errors

	No.	Misplaced or missing in SCI
Not found (0.4%)		
Wrong volume and wrong page	4	3
Found with the use of indexes (3.0%)		
Wrong name	1	1
Wrong journal	1	0
Wrong volume	8	7
Wrong page	20	18
Incorrect but found (8.9%)		
Misspelled names ^a	7	0
Missing authors' names ^a	4	0
Inverted first/last names ^a	2	0
Inappropriate "et al."	2	0
Inverted order of authors ^a	2	0
Wrong years	4	4
Missing initials	61	2
Wrong initials	11	1
Extra initials	4	0

^aNot including the first author.

erences. One had the wrong journal (ApJ instead of ApJ Suppl.) and one had the wrong author's name. The source papers were recovered with an extra effort. However this does not justify the authors putting the interested readers to this extra trouble.

For 90 references (8.9%) there were errors that did not interfere with finding the source papers, but some of those could cause those citations to be misplaced in the SCI. The majority (7.5%) were due to errors in initials. Some (1.4%) misspelled authors' names or omitted names or inverted first and last names or used an inappropriate "et al." when the full set of names should have been given. And some (0.4%) stated the wrong years. The total of these (9.3%) is more than the quoted 8.9% because some references had more than one error.

Among the errors listed in Table 1 we isolated those involving only the first authors' names and paper locations. Those number 84 references. Thus 8.3% of the references

could be misplaced slightly or substantially in the SCI.

As a check I looked up all the incorrect references in the SCI (1991 January–February issue). Only three references (0.3%) were not found. An additional 33 references (3.3%) were misplaced slightly (i.e., they were next to the correct locations) or more distantly in SCI. The breakdown for these two groups is listed in the last column of Table 1. The remaining 86 references (8.5%) were corrected and found in SCI for the following reason: SCI employs a code entitled KEYSAVE that, for each published paper entered, records the first author's truncated name, volume, page, and year. Each citation is then matched against that file and if slight differences occur (e.g., missing initials, wrong year), the citations are corrected before inclusion in the SCI. Therefore the 12.2% author errors result in only 3.6% misplacements or omissions in SCI.

As mentioned above, the errors are author dependent. Among the 37 papers scanned, four had no errors and six had 20% or more incorrect references. For the latter the average was eight errors per paper. The record was a paper in which 71% of the references were incorrect, of which 14% required searches of indexes.

What can be done to reduce these errors? The ultimate solution is to have our literature in computer memory so that each reference can be checked against the literature for accuracy. Until that is available authors should be encouraged to be more careful. It is never permissible to copy a reference from another paper without looking up the referenced paper. Some authors have been embarrassed when the references they copied are found to be incorrect so that their laziness is revealed to all. And the "It doesn't matter" attitude shown by authors with up to 71% errors should not be condoned. Although it is easy to expect referees and editors to look up each reference, that is often not practical, but individual errors should be called to the authors' attention.

I am avoiding adding any references to this paper!

I am indebted to Dr. Eugene Garfield of the Institute for Scientific Information for two careful readings of the manuscript and suggestions for corrections.

