



REFERENCE ERROR IN BOOK MANUSCRIPT FROM LIPI: HOW GOOD OUR SCIENTISTS ARE IN COMPOSING REFERENCES

Martinus Helmiawan

LIPI Press, Lembaga Ilmu Pengetahuan Indonesia

Correspondence: martian319@yahoo.com

Diajukan: 18-06-2019; Direview: 20-01-2020; Diterima: 08-03-2020; Direvisi: 20-03-2020

ABSTRAK

Daftar referensi dari 13 manuskrip buku yang ditulis oleh penulis dari LIPI dan dimasukkan untuk proses penerbitan di LIPI Press pada 2018 diperiksa. Sebanyak 879 entri daftar referensi ditemukan. Seluruh entri tersebut diperiksa keakuratannya. Pemeriksaan dilakukan dengan cara menentukan terlebih dahulu tipe kesalahan dan gaya referensi yang digunakan untuk mengecek kesalahan. Setiap entri ditelusuri untuk menemukan kesalahan. Kesalahan yang ditemukan kemudian dicatat dan dikategorikan berdasarkan tipe kesalahan yang ditentukan. Hasilnya, terdapat 100 entri yang bersih dari kesalahan, dan sisanya 779 entri terdapat kesalahan. Total kesalahan ada 3.651 entri, yang terdiri dari 1.576 kesalahan tanda baca; 396 kesalahan kapitalisasi huruf; 338 kesalahan penggunaan *italic*; 206 kesalahan ejaan dan pemilihan diksi; 126 kesalahan sintaksis; 642 kesalahan penggunaan spasi; 46 kesalahan terkait penambahan informasi yang tidak perlu; dan 321 kesalahan terkait tidak lengkapnya informasi yang diperlukan. Dari hasil tersebut, ditemukan rata-rata kesalahan per-referensi sebesar 4,15. Studi ini menunjukkan bahwa tingginya tingkat kesalahan yang ditunjukkan disebabkan oleh kelalaian penulis sendiri dalam menyusun referensi.

ABSTRACT

Reference lists from 13 book manuscripts that were submitted to LIPI Press in 2018 written by the writers from LIPI were examined. In total, there were 879 reference list entries, and the accuracy of each citation was examined. The examination was conducted by firstly determining the type of error and reference style used for cross-checking. Afterwards, each entry was thoroughly checked for errors. Found errors were grouped into each type of error. The result, only 100 entries were error-free, and from the 779 entries that contain errors, 3,651 errors were found. The errors were categorized into 1,576 punctuation errors; 396 capitalization errors; 338 italicization errors, 206 spelling and word choice errors; 126 syntax errors; 642 spacing errors; 46 extraneous information errors; and 321 missing data errors. From that result, error rate of 4.15 was achieved. This paper concluded that the overall huge error rate found shows the negligence of the writers in composing references.

Keywords: Reference errors; Reference style; Books; Researchers; Reference management software

1. INTRODUCTION

One of the most important parts of scientific writing is citation and its corresponding reference list. Citing previous studies is crucial as a context-maker for current research, while the quality of reference list can be interpreted as an extension of a scientist's level of knowledge. However, it is a problem when this section of scientific publication is often the most neglected in the writing process, thus resulting in a publication with low quality referencing (Santini, 2018; Taylor, 2002). This includes errors from punctuation and spelling to inconsistencies in style used. The reason for that problem is that to understand and implement citation style is a difficult task (Homol, 2014). Nowadays, Reference Management Software (RMS) is projected to ease the hardship of going through complicated reference style and putting the correct entry in reference list. However, Stevens (2016) found that errors still occur

even when RMS is used. Therefore, there is an indication that the problem stems from internal factor (the writers themselves) rather than external factor (the reference style or the need of software). This paper put an effort to further analyze the internal factor, which is somewhat overlooked by other studies that are generally more focused on the use of RMS (Brahmi & Gall, 2006; Kessler & Van Ullen, 2005, 2006; Kratochvil, 2017).

LIPI, as a national scientific institution in Indonesia, is the home of scientists who have produced a huge amount of scientific publications in regular basis, from scientific articles to books. Books in particular give the biggest point in the career of scientists in Indonesia. Therefore, books that are published by LIPI scientists should have been the main showcase of knowledge and science in Indonesia. This paper, therefore, examined whether scientific book writers from LIPI have put a well-composed reference list in their ready-for-publication manuscript or not by detecting errors in reference lists. Also, most of the studies on reference used journal articles as the basis of their study. Similar study done for books is still rare.

2. LITERATURE REVIEW

There have been several previous researches on citation and reference list. Some have put emphasis on identifying and counting errors in references. Karabulut (2017) utilized The Cited Reference Search function of Thomson Reuters' Web of Science database (formerly the Institute for Scientific Information's Web of Knowledge database) to identify erroneous citations. The study found that incorrect citations adversely affected the impact factor of the AJR by 0.065 in 2012 and by 0.123 in 2013. In the same fashion, Teixeira, et al. (2013) surveyed ecology journals indexed in the Web of Science and calculated the appropriateness of citations of review papers. The result is that reviews were significantly more often cited than regular articles, with an addition of 22% of citations were inaccurate, and another 15% unfairly gave credit to the review authors for other scientists' ideas.

Meanwhile, Stevens (2016) assessed learning approach to citation error by doing in-class activity related to referencing to students, while noting that it may be possible that the persistence of error-ridden citation is caused by more than just carelessness. Similarly, Melles, & Unsworth (2015) have done an extensive studies on the behavior of postgraduate students in practicing reference management. Others studies specifically note the performance of references management software. Brahmi and Gall (2006) compared citation format in EndNote version 7 and Reference Manager version 11 with the citation format for references found in the instructions to authors from the most significant medical literature. Homol (2014) tested the accuracy of citations generated by web-based citation tools, such as EBSCO Discovery Service's Cite tool, EndNote Basic, RefWorks, and Zotero. The study found that none of the programs is capable of generating an error-free citation. Strikingly similar study was done by Kratochvil (2017), analyzing the accuracy of citation generated by EndNote, Mendeley, RefWorks and Zotero, while pointing out that several mistakes were caused by technical limitations of the reference managers. Kessler & Ullen (2005, 2006) have done extensive studies on the performance of free or inexpensive Web-based tools that have been developed to create citations and format bibliographies. The result is that some knowledge of proper citation formats is necessary to use these programs effectively as the error rate per citation is quite high. In the follow up study to that (Ullen & Kessler, 2012), there is an improvement in the error rate, but the actual number is still unacceptably high.

This paper adopted the method used in the previous studies to assess the accuracy of reference lists from LIPI scientists, whilst analyzed the comparison of error type and error rate, and drew conclusion from that analysis.

3. METHOD

To measure the accuracy of reference lists generated by LIPI scientists, this paper examined reference list in book manuscripts submitted to LIPI Press in 2018. Revised manuscripts after first submission were not counted. Manuscripts were collected from the database in LIPI Press' server. In total, there were 13 book manuscripts that were written by author(s) from LIPI, and submitted in 2018. All manuscripts were in .docx format.

Table 1. List of Manuscripts and Its Number of Reference List Entries as Study Objects

No.	Title	Department	Entries
1	<i>Science, Technology & Society (STS): Bidang Agrikultural di Indonesia, Sejarah Problematika dan Prospek</i>	Pusat Penelitian Kemasyarakatan dan Kebudayaan	175
2	<i>Sistem Pengukur Intersepsi Curah Hujan untuk Pemantauan Komponen Siklus Hidrologi yang Terlalaikan</i>	Pusat Penelitian Metrologi	205
3	<i>Sistem Pemantauan Pengelolaan dan Distribusi Air Minum untuk PDAM Bangka Barat</i>	Pusat Penelitian Metrologi	34
4	<i>Status Keanekaragaman Hayati Flora Indonesia</i>	Pusat Penelitian Biologi	54
5	<i>Rekayasa Benefisiensi Pemrosesan Mineral untuk Peningkatan Nilai Tambah Sumber Daya Marginal</i>	Pusat Penelitian Geoteknologi	48
6	<i>Pengelolaan Kebun Raya Daerah: Antara Harapan dan Kenyataan</i>	Pusat Penelitian Teknologi Tepat Guna	14
7	<i>Keanekaragaman Tumbuhan Pulau Sempu dan Ekosistemnya</i>	Balai Konservasi Tumbuhan Purwodadi	121
8	<i>Jenis-Jenis Karang di Perairan Teluk Ambon</i>	Pusat Penelitian Laut Dalam	17
9	<i>Strategi dan Rencana Aksi Konservasi 12 Jenis Pohon Langka Indonesia 2018-2028</i>	Pusat Penelitian Biologi	26
10	<i>Teknik Penyusunan Prosedur Kerja dengan Metode Integrasi Business Process Improvement dan Risk Based Thinking</i>	Pusat Penelitian Standar Mutu dan Teknologi Pengujian	66
11	<i>Usada: A Book About Traditional Balinese Medicinal Plants</i>	Kebun Raya Eka Karya Bali	5
12	<i>Komunitas Islam Bonokeling</i>	Pusat Penelitian Sumber Daya Regional	38
13	<i>Dinamika Pelaksanaan Syariah: Perkawinan dalam Kontestasi Agama dan Negara</i>	Pusat Penelitian Kemasyarakatan dan Kebudayaan	76
	Total		879

Before continuing with the examination, the type of error must first be determined. Previous studies had proposed such categorization albeit some differences with each other (Kessler & Ullen, 2005; Brahma & Gall, 2006; Homol, 2014; Stevens, 2016; Kratochvíl, 2017).

However, this paper found that Stevens' method was the most concise and precise in formulating an ideal categorization for determining error in reference list entries. Therefore, Stevens' (2016) method and type of error were adopted for this paper, with minor adjustments (Table 2).

Table 2. Categorization of Error and the Examples

Type of error	Examples
Punctuation	Misuse/absence of comma; period; semicolon; bracket; parentheses.
Capitalization	Capital first letter only in the first word of a title while in CMS style; capital first letter in all title words while in APA style.
Italicization	Not italicizing book title or journal name; italicizing article title.
Spelling and word choice	Using '&' in CMS style; using 'and' in APA style.
Syntax	Not inverting author names in APA style; putting publisher first before location in fact of publication.
Spacing	Double spacing; putting space between volume and number of a journal in APA style; no space between words in a title.
Extraneous information	Including pages number for book entries.
Missing data	Incomplete author names; no year; missing publisher name or location in fact of publication.

Source: Stevens (2016)

The next step was to determine the reference style used for cross-checking the reference list in each manuscript. LIPI Press gives freedom to the author(s) to choose their own preference of referencing style as long as they compose it clearly and consistently to a particular reference style (LIPI Press, 2018). Reference list plagued by unclear, inconsistent and seemingly random-styled entries was observed for their tendency. Tendency was determined by examining degree of similarity between the elements of the entry and rules from a particular referencing style. Take one example of an entry of a reference list from one of the manuscripts.

Buttenheim, Allison M. & Jenna Nobles, (2009). *Ethnic Diversity, Traditional Norms, and Marriage Behaviour in Indonesia*, dalam *Population Studies*, Vol. 63, No. 3.

The example had three elements conforming to CMS style: 1) the second author name was not inverted; 2) article title was in all capital first letters; 3) author first names are spelled out rather than abbreviated. Two elements conformed to an APA style (bracketed year and the use of '&'), while the page number was missing and italicized article title was inconsistent with both style. In this case, the entry was cross-checked against CMS style, as it had the most similarity with. This was done in case per case basis, so other styles might be added as well to ensure fair and valid cross-check. By scanning that tendency in an entire reference list of a manuscript thoroughly, a pattern could be detected, showing the most dominant referencing style in a manuscript. That information was used as the core to determine the reference style for that manuscript

4. RESULTS & DISCUSSION

4.1 Results

From 13 manuscripts that were examined, there were 8 manuscripts which inclined to follow CMS style, 3 leaned towards APA style, and 1 manuscript each that had tendency to Vancouver and MLA style. Result of the cross-checking is presented in Table 3.

Table 3. Cross-checking Reference Lists' Entries Results

No.	Title	Tendency	Error-free entries*	Error-ridden entries**	Total errors	Average error per-entry
1	<i>Science, Technology & Society (STS): Bidang Agrikultural di Indonesia, Sejarah Problematika dan Prospek</i>	CMS	0	175	716	4.09
2	<i>Sistem Pengukur Intersepsi Curah Hujan untuk Pemantauan Komponen Siklus Hidrologi yang Terlalaikan</i>	CMS	0	205	1471	7.18
3	<i>Sistem Pemantauan Pengelolaan dan Distribusi Air Minum untuk PDAM Bangka Barat</i>	MLA	0	34	212	6.24
4	<i>Status Keanekaragaman Hayati Flora Indonesia</i>	CMS	12	42	167	3.09
5	<i>Rekayasa Benefisiasi Pemrosesan Mineral untuk Peningkatan Nilai Tambah Sumber Daya Marginal</i>	Vancouver	0	48	255	5.31
6	<i>Pengelolaan Kebun Raya Daerah: Antara Harapan dan Kenyataan</i>	APA	0	14	40	2.86
7	<i>Keanekaragaman Tumbuhan Pulau Sempu dan Ekosistemnya</i>	APA	49	72	257	2.12
8	<i>Jenis-Jenis Karang di Perairan Teluk Ambon</i>	CMS	0	17	53	3.12
9	<i>Strategi dan Rencana Aksi Konservasi 12 Jenis Pohon Langka Indonesia 2018-2028</i>	APA	0	26	64	2.46
10	<i>Teknik Penyusunan Prosedur Kerja dengan Metode Integrasi Business Process Improvement dan Risk Based Thinking</i>	CMS	16	50	66	1.00
11	<i>Usada: A Book About Traditional Balinese Medicinal Plants</i>	CMS	0	5	20	4.00
12	<i>Komunitas Islam Bonokeling</i>	CMS	23	15	35	0.92
13	<i>Dinamika Pelaksanaan Syariah: Perkawinan dalam Kontestasi Agama dan Negara</i>	CMS	0	76	295	3.88
	Total		100	779	3651	4.15

Noted: *Entries with zero error; ** Entries with one or more errors

From 779 entries that contained error(s), this paper found 1,576 punctuation errors; 396 capitalization errors; 338 italicization errors, 206 spelling and word choice errors; 126 syntax errors; 642 spacing errors; 46 extraneous information errors; and 321 missing data errors. Nine manuscripts had zero error-free entry, meaning their referencing style were completely undetectable before a thorough observation using the tendency method was done. Only four had distinguishable style since they contained entries with perfect accuracy and consistencies to a reference style, although throughout its respective reference list, errors could still be found.

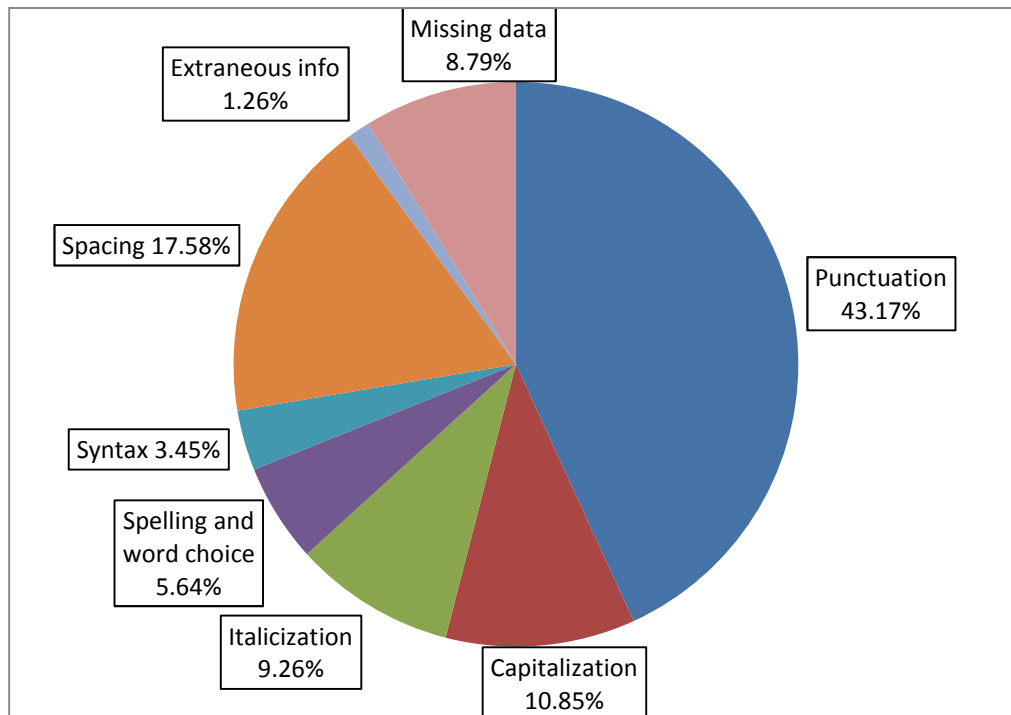


Figure 1. Percentage of error types

4.2 DISCUSSION

The results show that there is a general negligence of the writers in composing reference. It can be seen from the fact that most errors are minor types, such as punctuation, spacing, capitalization, and italicization; combined they make up more than three quarters of total error (80.85%). This is further backed up by the low error on syntax (3.45%). Low percentage of syntax error shows that the writers actually have the general concept of references style and are aware of which style they want to use because each style has slight variation in syntax. Therefore, the overall high error rate is not completely due to the writers' lack of knowledge of reference and styles, but rather comes down to their negligence in doing self-editing and expanding style mastery.

Even more concerning is the fact that major error in the form of missing data is still recorded (8.79%). This paper categorizes this error when writers failed to provide the complete set of reference element. It comprises of the data on author, date, title, and source – publisher name and its location, DOI, or URL (Lee, 2012). Each broken or non-existent element will count toward error on missing data. Example from this type is error on incomplete publisher's information and reference with URL only, without providing author or title information. This will lead to untraceable references that later might expose to the risk of plagiarism. Joob & Wiwanitkit (2018) argue that no amount of plagiarism is tolerable in scientific works to prevent any chance for the writer to cheat. That includes unintentional plagiarism stemming from writers' negligence to recheck and self-edit their citations and references, as this paper has suggested. Therefore, 321 errors on this type are considered still too many.

Meanwhile, comparison with past studies shows some important points. *First*, this studies 4.15 error rate is worse compared to previous result by Stevens (2016); Ullen & Kesler (2012), with 3.53 and 3.4 error rate, respectively. To make it even worse, Stevens' study actually

experimented with undergraduate students, thus implying that LIPI scientists perform worse in composing reference lists. *Second*, the disparity in the results of this paper is similar to Stevens' finding (2016). The lowest and highest error rates recorded in this paper are 1.00 and 7.18, respectively, while 0.67 and 9.00 of lowest-highest error rates are found in Stevens' study. Huge disparity range followed by high total error rate in this paper suggests that the number of writers who are capable and careful in composing reference lists is very rare, confirming the general negligence of the population. *Third*, only one manuscript in this paper uses RMS. All previous studies mentioned as comparison above are RMS-based. This shows that utilization of RMS among LIPI scientist is still low. Higher error rate than those shown by studies using RMS further suggests that RMS utilization does offer benefit for writers in composing reference lists with less error.

However, previous studies have unanimously agreed that no RMS has zero error percentage in generating references (Kessler & Ullen, 2005; Brahma & Gall, 2006; Homol, 2014; Stevens 2016; Kratochvíl, 2017; Kessler & Ullen, 2006), while the accuracy of pre-formatted citations and citations from on-demand tools is only a fraction better (Ullen & Kessler, 2012). Therefore, in generating the ideal references, RMS usage must be followed by manual care and self-edit.

To further discuss about the negligence shown by LIPI scientists that lead to error-ridden references, the paper identified several negative impacts as a result. Upon submission to a publisher, manuscript with broken reference becomes a great burden to reviewers and copy editors. Reviewers will have a hard time trying to browse and check the validity of each entry when it is riddled with error. For copy editors, manuscript with so many mistakes will take huge effort and time in copy editing phase, thus forcing the publisher to spend more resources on that manuscript. Those are the kind of energy as well as time consuming manuscripts for copyeditor. Coincidentally, reference error is one of the most aggravating problems for copy editing workload. In study by Wates & Campbell (2007), it was found that reference errors has contributed to 42.7% of all unanswered copy editing queries that was raised by editors to writers.

In other scenario, bad references will pose problems for the readers. Publishers have different policies and standards in running their editorial process (Gardner, 2011; Cochran & Wulf, 2019). Publisher with strict and highly-controlled copy editing might eliminate most errors in the process, resulting in final product with only small occurrences of minor typos. However, there are publishers with more 'lenient' approach to editorial process, and they tend to leave these errors, including reference error, intact in the published work. This phenomenon is not uncommon, and this will severely impact the readers. In scientific publication, readers often look to reference list and browse through entries in order to get the complete information of a particular in-text citation, usually for further reading or to obtain more data from the original source. In order to do so, they will search that information in the corresponding entry of the reference list. Reference lists that have proper and accurately composed entries will be helpful for readers, as it is easier and faster for them to browse and locate the entry, while error-ridden references will lead to more time and energy spent by readers doing the same thing.

Ultimately, reference error will always be related to plagiarism and ethical problem in all stages of publication, whether it is still a manuscript or published material. Negligence in composing well-written references and not providing them with all the required elements will make them untraceable to their original sources. Untraceable reference hinders the

acknowledgment for the original authors of the cited articles, and improper referencing generally may lead to negative effects for them (Teixeira, et al., 2013), while threatening writers with accusation of plagiarism.

5. CONCLUSION

From those analyses, this paper claims that being neglectful in composing reference list is the main cause of the problem. It can be seen from the fact that writers make numerous minor errors and are forgetful in providing important data of the citation, while they do not seem to have problem in grasping the general syntax of their preferred reference style. This negligence impacts parties involved in the whole chain of publication; from reviewers' difficulties in looking up to the references, copy editor needing more time to edit the reference, writers risking themselves being accused of unintentional plagiarism, and readers having hard time tracing original source of a particular reference for further reading. In an effort to find for a solution, wider application of RMS seems easy to recommend. However, it will be useless demanding wider usage of RMS from the writers while their negligent attitude toward composing references still exists, which is exactly what this paper has found. Before broader utilization of RMS can be applied to reduce error rate, further studies tackling the problem in general negligence of the writers toward composing reference are firstly needed.

BIBLIOGRAPHY

- Brahmi, F.A. & Gall, C. 2006. EndNote® and Reference Manager® Citation Formats Compared to 'Instructions to Authors' in Top Medical Journals. *Medical Reference Services Quarterly*, 25(2), 49–57. https://doi.org/10.1300/J115v25n02_04.
- Cochran, A. & Wulf, K.. 2019. *Editing Is at the Heart of Scholarly Publishing*. di <https://scholarlykitchen.sspnet.org/2019/04/24/editing-is-at-the-heart-of-scholarly-publishing/>.
- Gardner, R. 2011. *Do Publishers Edit Books Anymore?* di <https://rachellegardner.com/do-publishers-edit-books/>.
- Homol, L. 2014. Web-Based Citation Management Tools: Comparing the Accuracy of Their Electronic Journal Citations. *Journal of Academic Librarianship*, 40(6), 552–57. <https://doi.org/10.1016/j.acalib.2014.09.011>.
- Joob, B. & Wiwanitkit, V. 2018. Plagiarism: Either Intentional or Unintentional, It Is Still Plagiarism! *Perspectives in Clinical Research*, 9(3): 151. https://doi.org/10.4103/picr.picr_17_18.
- Karabulut, N. 2017. Inaccurate Citations in Biomedical Journalism: Effect on the Impact Factor of the American Journal of Roentgenology. *American Journal of Roentgenology*, 208(3), 472–74. <https://doi.org/10.2214/AJR.16.16984>.
- Kessler, J. & Ullen, M.K.V 2006. Citation Help in Databases: Helpful or Harmful? *University Libraries Faculty Scholarship*. https://doi.org/10.1300/j295v02n01_03.
- Kessler, J. & Ullen, M.K.V. 2005. Citation Generators: Generating Bibliographies for the Next Generation. *The Journal of Academic Librarianship*, 31(4), 310–16.
- Kratochvíl, J. 2017. Comparison of the Accuracy of Bibliographical References Generated for Medical Citation Styles by EndNote, Mendeley, RefWorks and Zotero. *Journal of Academic Librarianship*, 43(1), 57–66. <https://doi.org/10.1016/j.acalib.2016.09.001>.
- Lee, C. 2012. *Missing Pieces: How to Write an APA Style Reference Even without All the Information*. di <https://blog.apastyle.org/apastyle/2012/05/missing-pieces.html>.
- LIPI Press. 2018. *Pedoman Penerbitan Buku Ilmiah LIPI Press*. Jakarta: LIPI Press.
- Melles, A. & Unsworth, K. 2015. Examining the Reference Management Practices of Humanities and Social Science Postgraduate Students and Academics. *Australian Academic and Research Libraries*, 46(4), 250–76. <https://doi.org/10.1080/00048623.2015.1104790>.
- Santini, A. 2018. The Importance of Referencing. *Journal of Critical Care Medicine (Universitatea de Medicina Si Farmacie Din Targu-Mures)*, 4(1): 3–4. <https://doi.org/10.1515/jccm-2018-0002>.
- Stevens, C.R. 2016. Citation Generators, OWL, and the Persistence of Error-Ridden References: An Assessment for Learning Approach to Citation Errors. *Journal of Academic Librarianship*, 42(6), 712–18. <https://doi.org/10.1016/j.acalib.2016.07.003>.
- Taylor, D.M. 2002. The Appropriate Use of References in a Scientific Research Paper. *Emergency Medicine*, 14(2), 166–70.
- Teixeira, M.C., Thomaz, S.M., Michelan, T.S., Mormul, R.P., Meurer, T., Fasolli, J.V.B., & Silveira, M.J. 2013. Incorrect Citations Give Unfair Credit to Review Authors in Ecology

Journals. *PLoS ONE*, 8(12), 8–11. <https://doi.org/10.1371/journal.pone.0081871>.

Ullen, M.K.V. & Kessler, J. 2012. Citation Help in Databases: The More Things Change, the More They Stay the Same. *Public Services Quarterly*, 8(1), 40–56. <https://doi.org/10.1080/15228959.2011.620403>.

Wates, E. & Campbell, R. 2007. Author's Version vs. Publisher's Version: An Analysis of the Copy-Editing Function. *Learned Publishing*, 20(2), 121–29. <https://doi.org/10.1087/174148507X185090>.