

Opinion
Editing, Writing & Publishing



Similarity and Plagiarism in Scholarly Journal Submissions: Bringing Clarity to the Concept for Authors, Reviewers and Editors

Aamir Raof Memon

Institute of Physiotherapy & Rehabilitation Sciences, Peoples University of Medical & Health Sciences for Women, Nawabshah (Shaheed Benazirabad), Sindh, Pakistan



Received: Apr 2, 2020

Accepted: May 7, 2020

Address for Correspondence:

Aamir Raof Memon, DPT, MPhil, PGD
Institute of Physiotherapy & Rehabilitation Sciences, Peoples University of Medical & Health Sciences for Women, Hospital Road, Nawabshah (Shaheed Benazirabad), 67450, Sindh, Pakistan.
E-mail: memon.aamir.raoof@gmail.com

© 2020 The Korean Academy of Medical Sciences.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iD

Aamir Raof Memon
<https://orcid.org/0000-0002-3203-418X>

Disclosure

The author has no potential conflicts of interest to disclose.

INTRODUCTION

What constitutes plagiarism? What are the methods to detect plagiarism? How do “plagiarism detection tools” assist in detecting plagiarism? What is the difference between plagiarism and similarity index? These are probably the most common questions regarding plagiarism that many research experts in scientific writing are usually faced with, but a definitive answer to them is less known to many. According to a report published in 2018, papers retracted for plagiarism have sharply increased over the last two decades, with higher rates in developing and non-English speaking countries.¹ Several studies have reported similar findings with Iran, China, India, Japan, Korea, Italy, Romania, Turkey, and France amongst the countries with highest number of retractions due to plagiarism.¹⁻⁴ A study reported that duplication of text, figures or tables without appropriate referencing accounted for 41.3% of post-2009 retractions of papers published from India.⁵ In Pakistan, *Journal of Pakistan Medical Association* started a special section titled “Learning Research” and published a couple of papers on research writing skills, research integrity and scientific misconduct.^{6,7} However, the problem has not been adequately addressed and specific issues about it remain unresolved and unclear. According to an unpublished data based on 1,679 students from four universities of Pakistan, 85.5% did not have a clear understanding of the difference between similarity index and plagiarism (unpublished data). Smart et al.⁸ in their global survey of editors reported that around 63% experienced some plagiarized submissions, with Asian editors experiencing the highest levels of plagiarized/duplicated content. In some papers, journals from non-English speaking countries have specifically discussed the cases of plagiarized submissions to them and have highlighted the drawbacks in relying on similarity checking programs.⁹⁻¹¹ The cases of plagiarism in non-English speaking countries have a strong message for honest researchers that they should improve their English writing skills and credit used sources by properly citing and referencing them.¹²

Despite aggregating literature on plagiarism from non-Anglophonic countries, the answers to the aforementioned questions remain unclear. In order to answer these questions, it is important to have a thorough understanding of plagiarism and bring clarity to the less known issues about it. Therefore, this paper aims to 1) define plagiarism and growth in its prevalence as well as literature on it; 2) explain the difference between similarity and plagiarism; 3) discuss the role of similarity checking tools in detecting plagiarism and the flaws on completely relying on them; and 4) discuss the phenomenon called Trojan citation. At the end, suggestions are provided for authors and editors from developing countries so that this issue maybe collectively addressed.

PLAGIARISM

Defining plagiarism and its prevalence in manuscripts

To begin with, plagiarism may be defined as “*when somebody presents the published or unpublished work of others, including ideas, scholarly text, images, research design and data, as new and original rather than crediting the existing source of it.*”¹³ The common types of plagiarism, including direct, mosaic, paraphrasing, intentional (covert) or unintentional (accidental) plagiarism, and self-plagiarism have been discussed in previous reviews.¹⁴⁻¹⁶

Evidence suggests that the first paper accused for plagiarism was published in 1979 and there has been a substantial growth in the cases of plagiarism over time.^{1-5,8,17} Previous studies have pointed that plagiarism is prevalent in developing and non-English speaking countries but the occurrence of plagiarism in developed countries suggests that it is rather a global problem.^{1-4,18-20} As of today (1 April 2020), the search conducted in Retraction Database (<http://retractiondatabase.org/RetractionSearch.aspx?>) for papers retracted for plagiarism found 2,280 documents. Similarly, Scopus search for plagiarism in title of journal articles found 2,159 results. This suggests that the papers retracted for plagiarism are in fact higher than the papers published on this issue. However, what we see now may not necessarily be true i.e., the cases of plagiarism might be higher than we know. Certainly, database search for papers tagged for plagiarism is limited to indexed journals only, which keeps non-indexed journals (both low-quality and deceptive journals) out of focus.^{5,21} Moreover, journal coverage may vary from one database to the other as reported in a recent paper on research dissemination in South Asia.²² Therefore, both the prevalence of plagiarism and literature published on it as reported by database search are most likely “*understated as of today.*”⁵

Reasons for plagiarism: lack of understanding and poor citing practices

Although reasons for plagiarism are complex, previous papers have suggested possible causes for plagiarism by authors.^{16,23-26} One of the major but less known reason for this might be that the students, naïve researchers, and even some faculty members either lack clarity about what constitutes plagiarism or are unable to differentiate similarity index versus plagiarism.^{24,26,27} For example, a recent online survey conducted on the participants in the AuthorAID MOOC on Research Writing found that 84.4% of the survey participants were unaware of the difference between similarity index and plagiarism, though almost all of them had reported having an understanding of plagiarism.²⁴ The same paper reported that one in three participants admitted that they had plagiarized at some point during their academic career.²⁴ Therefore, it is important to have clarity about what constitutes plagiarism and the difference between similarity index and plagiarism so that the increasing rates of plagiarism could be deterred.

The ‘existing source’ or ‘original source’ in the definition of plagiarism refers to the main (primary) source and not the source (secondary) from where the author extracts the information. For example, someone cites a paper for a passage on mechanism of how exercise affects sleep but the cited paper aims to determine the prevalence of sleep disorders and exercise level rather than the mechanistic association. A thorough evaluation finds that the cited paper had used the text from another review paper that talked about the mechanisms relating sleep with exercise behavior. This phenomenon of improper secondary (or indirect) citations may be common among students and novice researchers, particularly from developing countries, and should be discouraged.²⁷

SIMILARITY INDEX

Plagiarism vs. similarity index and the role of similarity checking tools

Plagiarism as defined above refers to the intentional (covert) or unintentional (accidental) theft of published or unpublished intellectual property (i.e., words or ideas), whereas similarity index refers to “*the extent of overlap or match between an author’s work compared to other existing sources (books, websites, student thesis, and research articles) in the databases of similarity checking tools.*”^{9,24} The advancements in information technology has helped researchers get help from various freely available (i.e., Viper, eTBLAST/HelioBLAST, PlagScan, PlagiarismDetect, Antiplagiat, Plagiarisma, DupliChecker) and subscription-based (i.e., iThenticate, Turnitin, Similarity Check) similarity checking tools.^{8,24} Many journal editors use iThenticate and/or Similarity Check (Crossref) for screening submitted manuscripts for similarity detection whereas Turnitin is commonly used by universities and faculty to assess text similarity in students’ work; however, there is a fairness issue that not every journal or university, particularly those from developing countries, can afford to pay for using these subscription-based services.²⁸ For instance, an online survey found that only about 18% participants could use Turnitin through their university subscription.²⁴ Another problem is the way these tools are commonly referred to as i.e., plagiarism detection tools, plagiarism checking software, or plagiarism detection programs. However, based on the function they perform, it would be appropriate to call them differently, such as similarity checking tools, similarity checkers, text-matching tools, or simply text-duplicity detection tools.^{5,8,23} This means that these tools help locate matching or overlapping text (similarity) in submitted work, without directly flagging up plagiarism.²⁴

Taking Turnitin as an example, these tools reflect the text similarity through color codes, each linked to an online source of it; details for this have been described elsewhere.^{23,28} Journal editors, universities and some organizations consider text above specific cutoff values for the percentage of similarity as problematic. According to a paper, 5% or less text similarity (overlap of the text in the manuscript with text in the online literature) is acceptable to some journal editors, while others might want to put the manuscript under scrutiny if the text similarity is over 20%.^{29,30} Another paper observed that journal editors tend to reject a manuscript if text similarity is above 10%.³¹ The study on participants completing the AuthorAID MOOC on Research Writing also found that some participants reported that their institutions consider text similarity of less than 20% as acceptable.²⁴ As an example, the guidelines of the University Grants Commission of India allow for similarity up to 10% as acceptable or minor (Level 0), but anything above is categorized into different levels (based on the percentages), each with separate list of repercussions for students and researchers.³² This approach might miss the cases where the acceptable similarity of 10% comes from a single source, especially if the editors relied on the numbers only. In addition, this approach has the potential for punishing authors who have not committed plagiarism at all. To illustrate this, the randomly written text presented in **Fig. 1** would be considered plagiarism based on the rule of cutoff values. Some authors opine that text with over four consecutive words or a number of word strings should be treated as plagiarized.^{28,33} This again is not a good idea as the text “the International Physical Activity Questionnaire was used to measure ...” would be same in several papers, but this is definitely not plagiarism because the methodology of different papers on the same topic could be similar; so, the decision should not be based on the numbers reflected by similarity detection tools.²⁸ Therefore, it would be prudent not to set any cutoff values for text similarity as it will lead to a *slippery slope* (“*a course of action that seems to lead inevitably from one action or result to another with unintended consequences*”—defined by Merriam-Webster Dictionary) and give “a sense of impunity to the perpetrators.”³²

The number of documents and total citations were used to determine most active journals and institutions. The Journal of Pakistan Medical Association, Pakistan Journal of Medical Sciences and Journal of College of Physicians and Surgeons Pakistan were the most contributing journals. Aga Khan University Hospital (Karachi), Dow University of Health Sciences (Karachi) and Peoples University of Medical & Health Sciences for Women (Nawabshah) were the most prolific institutions.

Similarity Check Text



Fig. 1. Turnitin report for text similarity based on a randomly written text (on April 2, 2020). The author of this paper has access to Turnitin through the University and not to iThenticate. Therefore, Turnitin was used as an example in Fig. 1.

Drawbacks of similarity checking tools

There are a few drawbacks on completely relying on the similarity checking tools. First, these tools are not foolproof and might miss the incidents of translational plagiarism and figure plagiarism.²⁴ Translational plagiarism is the most invisible type of copying in non-Anglophone countries where an article published in languages other than English is copied (with or without minor modifications) and published in an English journal or vice versa.¹⁰ This is indeed extremely difficult type of plagiarism to detect, and different approaches (e.g., use of Google translator) to address it have been recently reported.^{34,35} Nevertheless, there might be some cases where this practice maybe acceptable, such as publishing policy papers (see “*Identifying predatory or pseudo-journals*” – this paper was published in *International Journal of Occupational and Environmental Medicine*, *National Medical Journal of India*, and *Biochemia Medica* in 2017 by authors affiliated with World Association of Medical Editors (WAME) – or “*The revised guidelines of the Medical Council of India for academic promotions: Need for a rethink*” – this paper was published in over ten journals during 2016 by four journal editors and endorsed by members (not all) of the Indian Association of Medical Journal Editors, for example). Second, text similarity in some parts of manuscript (i.e., methods and results) should be weighed differently from other sections (i.e., introduction and discussions) and its conclusions.³¹ In addition, based on the personal experience of the author of this paper, some individuals might use a sophisticated technique to avoid detection of high similarity through the use of inappropriate synonyms, jargon, and deliberate grammatical and structural errors in the text of the manuscript. Third, plagiarism of ideas may be missed by these tools as they can only detect plagiarism of words.^{23,32} Therefore, similarity checking tools tend to underestimate plagiarized text or sometimes overestimate non-plagiarized material as problematic (Fig. 1).^{24,36} It should be noted that these tools serve as only an aid to determine suspected instances of plagiarism and the text of the manuscript should always be evaluated by experts, so “a careful human cannot be replaced.”^{31,37} A few papers published in the *Journal of Korean Medical Science* have presented the examples where plagiarized content was missed by similarity checking tools and later noticed after a careful examination of the text.^{9,10} Finally, plagiarism of unpublished work cannot be detected by these tools as they are limited to online sources only.²³ This is particularly important in the context of developing countries

where research theses/dissertations of students are not deposited in research repositories, and where commercial, predatory editing and brokering services exist.^{10,38} For example, the research repository of the Higher Education Commission of Pakistan allows deposition of doctoral theses only, and less than five universities (out of over 150) across the country have a research repository allowing for deposition of scholarly content.³⁸ Recently some strange trend of predatory editing and brokering services has emerged that offer clones of previously published papers or unpublished work to non-Anglophone or some lazy authors demanding quick and easy route to publications for promotion and career advancement.¹⁰ Although plagiarism of unpublished work would not be easy for experts to detect, this may be possible through their previous experience and scholarly networks.

TROJAN CITATION: PERSONAL EXPERIENCE

A recent experience worth discussion in context to plagiarism comes in the shape of the Trojan citation where someone “*makes reference to a source one time to in order to evade detection (by editors and readers) of bad intentions and provide cover for a deeper, more pervasive plagiarism.*”³⁹ This practice is particularly common in those with an intent of deceiving the readers and playing with the system. A few months ago, the author of this paper was invited to review a manuscript on predatory publishing by a journal. The content of the manuscript appeared suspicious but was not labelled “plagiarized” during the first round of the review. However, during the second round, it was noticed that this was a case of Trojan citation where the author(s) cited the main source for a minor point and copied the major part of the manuscript from a paper published in *Biochemia Medica* (a Croatian journal) with slight modification in the content.⁴⁰ The editor of the journal was informed about this and the manuscript was rejected further processing. This example suggests that careful human intervention by experts is required to highlight the cases of plagiarism.

CONCLUSION

In conclusion, what we know about the growth in the prevalence of plagiarism may be ‘just the tip of the iceberg’. Therefore, collective contribution from authors, reviewers, and editors, particularly from Asia-Pacific region, is required. Authors from the Asia-Pacific region and developing countries, with an expertise on this topic, should play their role by supporting journal editors and through their mentorship skills. Furthermore, senior researchers should encourage and help their honors and master students to publish their unpublished work before it gets stolen by commercial, brokering agencies. They should also work in close collaboration with universities and organizations related with higher education in countries where this issue is not properly addressed, and should facilitate education and training sessions on plagiarism as previous evidence suggests that workshops and online training sessions may be helpful.⁵ On the other hand, journal editors from Asia-Pacific region and developing countries should not judge the manuscripts solely on the basis of percentage of similarity as reflected by similarity checking services. They should have a database of their own where manuscripts about plagiarism in scientific writing, for example, should be sent for review to the experts on this subject. As journal editors may not be experts in all fields, networking and seeking help from experts would be helpful in avoiding the cases of plagiarism in the future. It would be appropriate that the journal editors and the trainee editors, particularly from the resource-limited countries, are educated about the concept

of scientific misconduct and the advancement in knowledge around this area. Moreover, journal editors should publish and publically discuss the cases of plagiarism as a learning experience for others. The *Journal of Korean Medical Science* has used this approach regarding cases of plagiarism, which other journals from the region are encouraged to adopt.^{9,10} Likewise, a paper discussing case scenarios of salami publication (i.e., “a distinct form of redundant publication which is usually characterized by similarity of hypothesis, methodology or results but not text similarity”) serves as a good example of how journal editors may facilitate authors to utilize their mentorship skills and support journals in educating researchers.⁴¹ There should be strict penalties on cases of plagiarism, and safety measures for security of whistleblowers should be in place and be ensured. By doing so, evil and lazy authors who bypass the system would be punished and honest authors would be served. Thus, the take-home message for editors from Asia-Pacific region is that a collective effort and commitment from authors, reviewers, editors and policy-makers is required to address the problem of plagiarism, especially in the developing and non-English speaking countries.

REFERENCES

- Brainard J, You J. What a massive database of retracted papers reveals about science publishing's 'death penalty'. *Science* 2018;25(1):1-5.
[CROSSREF](#)
- Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci U S A* 2012;109(42):17028-33.
[PUBMED](#) | [CROSSREF](#)
- Stretton S, Bramich NJ, Keys JR, Monk JA, Ely JA, Haley C, et al. Publication misconduct and plagiarism retractions: a systematic, retrospective study. *Curr Med Res Opin* 2012;28(10):1575-83.
[PUBMED](#) | [CROSSREF](#)
- Amos KA. The ethics of scholarly publishing: exploring differences in plagiarism and duplicate publication across nations. *J Med Libr Assoc* 2014;102(2):87-91.
[PUBMED](#) | [CROSSREF](#)
- Misra DP, Ravindran V, Wakhlu A, Sharma A, Agarwal V, Negi VS. Plagiarism: a viewpoint from India. *J Korean Med Sci* 2017;32(11):1734-5.
[PUBMED](#) | [CROSSREF](#)
- Jawad F. Plagiarism and integrity in research. *J Pak Med Assoc* 2013;63(11):1446-7.
[PUBMED](#)
- Rathore FA, Farooq F. Plagiarism detection softwares: useful tools for medical writers and editors. *J Pak Med Assoc* 2014;64(11):1329-30.
[PUBMED](#)
- Smart P, Gaston T. How prevalent are plagiarized submissions? Global survey of editors. *Learn Publ* 2019;32(1):47-56.
[CROSSREF](#)
- Baydik OD, Gasparyan AY. How to act when research misconduct is not detected by software but revealed by the author of the plagiarized article. *J Korean Med Sci* 2016;31(10):1508-10.
[PUBMED](#) | [CROSSREF](#)
- Hong ST. Plagiarism continues to affect scholarly journals. *J Korean Med Sci* 2017;32(2):183-5.
[PUBMED](#) | [CROSSREF](#)
- Park S, Yang SH, Jung E, Kim YM, Baek HS, Koo YM. Similarity analysis of Korean medical literature and its association with efforts to improve research and publication ethics. *J Korean Med Sci* 2017;32(6):887-92.
[PUBMED](#) | [CROSSREF](#)
- Yessirkepov M, Nurmashev B, Anartayeva M. A scopus-based analysis of publication activity in Kazakhstan from 2010 to 2015: positive trends, concerns, and possible solutions. *J Korean Med Sci* 2015;30(12):1915-9.
[PUBMED](#) | [CROSSREF](#)
- Roig M. Encouraging editorial flexibility in cases of textual reuse. *J Korean Med Sci* 2017;32(4):557-60.
[PUBMED](#) | [CROSSREF](#)

14. Das N, Panjabi M. Plagiarism: why is it such a big issue for medical writers? *Perspect Clin Res* 2011;2(2):67-71.
[PUBMED](#) | [CROSSREF](#)
15. Das N. Intentional or unintentional, it is never alright to plagiarize: a note on how Indian universities are advised to handle plagiarism. *Perspect Clin Res* 2018;9(1):56-7.
[PUBMED](#) | [CROSSREF](#)
16. Mohammed RA, Shaaban OM, Mahran DG, Attellawy HN, Makhlof A, Albasri A. Plagiarism in medical scientific research. *J Taibah Univ Med Sci* 2015;10(1):6-11.
17. Steen RG, Casadevall A, Fang FC. Why has the number of scientific retractions increased? *PLoS One* 2013;8(7):e68397.
[PUBMED](#) | [CROSSREF](#)
18. Halupa C, Bolliger DU. Faculty perceptions of student self plagiarism: an exploratory multi-university study. *J Acad Ethics* 2013;11(4):297-310.
[CROSSREF](#)
19. Higgins JR, Lin FC, Evans JP. Plagiarism in submitted manuscripts: incidence, characteristics and optimization of screening-case study in a major specialty medical journal. *Res Integr Peer Rev* 2016;1(1):13.
[PUBMED](#) | [CROSSREF](#)
20. Almeida RM, de Albuquerque Rocha K, Catelani F, Fontes-Pereira AJ, Vasconcelos SM. Plagiarism allegations account for most retractions in major Latin American/Caribbean databases. *Sci Eng Ethics* 2016;22(5):1447-56.
[PUBMED](#) | [CROSSREF](#)
21. Memon AR. Revisiting the term predatory open access publishing. *J Korean Med Sci* 2019;34(13):e99.
[PUBMED](#) | [CROSSREF](#)
22. Memon AR. Scholarly publishing and research dissemination in South Asia: some exemplary initiatives and the way forward. *J Pak Med Assoc* 2019;69(9):1349-54.
[PUBMED](#)
23. Meo SA, Talha M. Turnitin: is it a text matching or plagiarism detection tool? *Saudi J Anaesth* 2019;13(5 Suppl 1):S48-51.
[PUBMED](#) | [CROSSREF](#)
24. Memon AR, Mavrinac M. Knowledge, attitudes, and practices of plagiarism as reported by participants completing the AuthorAID MOOC on research writing. *Sci Eng Ethics* 2020;26(2):1067-88.
[PUBMED](#) | [CROSSREF](#)
25. Shashok K. Authors, editors, and the signs, symptoms and causes of plagiarism. *Saudi J Anaesth* 2011;5(3):303-307.
[PUBMED](#) | [CROSSREF](#)
26. Heitman E, Litewka S. International perspectives on plagiarism and considerations for teaching international trainees. *Urol Oncol* 2011;29(1):104-8.
[PUBMED](#) | [CROSSREF](#)
27. Mogull SA. Accuracy of cited "facts" in medical research articles: a review of study methodology and recalculation of quotation error rate. *PLoS One* 2017;12(9):e0184727.
[PUBMED](#) | [CROSSREF](#)
28. Pastor JC. Plagiarism in publications. *Arch Soc Esp Ophthalmol* 2018;93(12):571-2.
[PUBMED](#) | [CROSSREF](#)
29. Peh WC, Arokiasamy J. Plagiarism: a joint statement from the Singapore medical journal and the medical journal of Malaysia. *Singapore Med J* 2008;49(12):965-6.
[PUBMED](#)
30. Swaan PW. Publication ethics--a guide for submitting manuscripts to pharmaceutical research. *Pharm Res* 2010;27(9):1757-8.
[PUBMED](#) | [CROSSREF](#)
31. Mahian O, Treutwein M, Estellé P, Wongwises S, Wen D, Lorenzini G, et al. Measurement of similarity in academic contexts. *Publications* 2017;5(3):18.
[CROSSREF](#)
32. Kadam D. Academic integrity and plagiarism: the new regulations in India. *Indian J Plast Surg* 2018;51(2):109-10.
[PUBMED](#) | [CROSSREF](#)
33. Bouville M. Plagiarism: words and ideas. *Sci Eng Ethics* 2008;14(3):311-22.
[PUBMED](#) | [CROSSREF](#)
34. Spiroski M. How to verify plagiarism of the paper written in Macedonian and translated in foreign language? *Open Access Maced J Med Sci* 2016;4(1):1-4.
[PUBMED](#) | [CROSSREF](#)

35. Masic I. What to do when you have suspect translational plagiarism?-editor's view. *Med Arch* 2018;72(6):466-7.
[PUBMED](#) | [CROSSREF](#)
36. Foltýnek T, Dlabolová D, Anohina-Naumeca A, Razi S, Kravjar J, Kamzola L, et al. Testing of support tools for plagiarism detection. <https://arxiv.org/abs/2002.04279>. Updated 2020.
37. Glänzel W, Braun T, Schubert A, Zosimo-Landolfo G. Coping with copying. *Scientometrics* 2015;102(1):1-3.
[CROSSREF](#)
38. Memon AR, Rathore FA. Moodle and online learning in Pakistani Medical Universities: An opportunity worth exploring in higher education and research. *J Pak Med Assoc* 2018;68(7):1076-8.
[PUBMED](#)
39. Shaw D. The Trojan citation and the “accidental” plagiarist. *J Bioeth Inq* 2016;13(1):7-9.
[PUBMED](#) | [CROSSREF](#)
40. Ferris LE, Winker MA. Ethical issues in publishing in predatory journals. *Biochem Med* 2017;27(2):279-84.
[PUBMED](#) | [CROSSREF](#)
41. Supak Smolčić V. Salami publication: definitions and examples. *Biochem Med* 2013;23(3):237-41.
[PUBMED](#) | [CROSSREF](#)