

Textual Similarity Patterns and Citation Style Errors in Literature Review Chapters of LIS Theses of South Indian Universities:A Study Chikku Balachandran¹; N. S. Harinarayana²

UGC-Senior Research Fellow, Department of Studies in Library and Information Science, University of Mysore, Mysuru¹; Professor, Department of Studies in Library and Information Science, University of Mysore, Mysuru, India²

chikkulisc@gmail.com, harinarayana@lisc.uni-mysore.ac.in

ABSTRACT

This study investigated the PhD students' writing styles of the Reviews of Literature chapter in the theses submitted at southern Indian universities. These days, doctorate students frequently plagiarize original authors' words without giving them the required credit. This practice is common since students frequently copy entire or portions of manuscripts from original authors without giving due credit. The study used corpora constructed from the review of the literature chapter of randomly selected 20 theses accessible from Shodhganga. The results showed that students displayed an array of improper borrowing of texts. While copying text as it is in the original source and failed to paraphrase the content intentionally or unintentionally. Similarly, concerning improper paraphrasing, 'change of reporting verb' was the highest form of error in terms of plagiarism. Among the in-text citation inaccuracy both short quotation without quotation marks and block quotations without separate intended paragraphs were frequent in review of literature chapter.

KEYWORDS: Textual similarity, Plagiarism, Citation error, Content borrowing, academic dishonesty.

1. INTRODUCTION

Copying text from other sources without acknowledging the author properly, which is commonly known as plagiarism, regarded as one of the deadliest sins in academia (Siaputra, 2013)[1]. Plagiarism is deeply rooted in some of the academic environments and students are fully aware of it as a dishonest practice but do it especially when there is a deadline or too much work to accomplish (Pupovac, Bilic-Zulle, & Petroveck, 2008)[2]. This is true as the researchers rely on past studies as a preliminary step or contextualize their latest findings and conclusions to support their claims. They do so through research, which aims at creating new knowledge by employing specific methodologies to accomplish a systematic investigation. This implies that it is inevitable in any research to draw logical conclusions without referring to an already existing treasury of knowledge. Plagiarism resonates in the mind

of every researcher interacting with information resources to synthesize ideas or form other documents for getting marks, grades, promotion, or fulfilling the requirements of a particular program.

Among the notable causes of plagiarism are subjective and objective factors. For the former, they include attitudinal, individual i.e. circumstances, ambitions, competitive academic drive, or, even simple ignorance. While for the latter encompass pressure directed at individuals by society and family, and demand by workplace, lack of rules to guide the maintain individual behavior, among others (Starovoytova, 2017) [3].

For long, organizations have recognized the importance of articulating their position against plagiarism to the extent of inculcating good practices that strengthen the application of ethical principles such as honesty and distancing themselves from bad practices such as stealing in their work environments (Gotterbarn, Miller, & Impagliazzo, 2006). However, this is not a guarantee for implementing a strong policy in these organizations (Gotterbarn, Miller, & Impagliazzo, 2006). By implication, many organizations assume that putting strong policies in place can reduce or eradicate plagiarism, which is not the case often. This difficulty resulted from the heated debates geared toward understanding the causes of plagiarism via deontological and utilitarian perspectives, which look at plagiarism as intentional/unintentional or detectable/undetectable, respectively (Gotterbarn, Miller, & Impagliazzo, 2006)[4].

2. BACKGROUND

Plagiarism increases in a similar proportion as online resources increase varying from 5-80% (Marshall, Taylor, Hothersall, & Pe'rez-Marti'n, 2011) [5]. In other words, plagiarism will continue to increase as long as the internet continues to permeate every nook and crannie of human endeavors, which results in the corresponding growth of studies on plagiarism (Ison, 2012) [6]. Internet-user friendliness often linked with plagiarism as a growing threat to the academia is available in the literature. As captured by Walker, (2010) citing Warn, (2006) [7], using the TOAST plagiarism tool, 74(10.8%) of the sample were found to commit plagiarism and there was an increase of 3.2% to 15.6% verbatim plagiarism in the text. This is to the extent that the mean similarity index of dissertations was 15.1 ($\sigma=13.02$), resulting in categorizing plagiarism into lower level (46%), medium level (11%), and higher level (3%) (Ison, 2012). Previous researchers have failed to link the prevalence of such heinous acts with international students who regard English as a second language coupled with students on distance learning portraying a large of number of male students inclined to practice plagiarism than their counterparts are also available in the literature. This raises concern as students self-reporting reasons for indulging in plagiarism is insufficient rather requires a different perspective dependent upon empirical measurement of plagiarism (Walker, 2010) [8].

While plagiarism has diffused and fragmented, it is not only prevalent among students, but rather academics also get involved in especially self-plagiarism with a varying perception of it among academics and administrators (Bretag & Carapiet, 2007)[9]. For instance, Baz'daric', Bilic'-Zulle, Brumini, and Petrovec'ki, (2012) noted that, in Croatian Medical Journal submission out of 754 manuscripts submitted, 105(14%) were suspicious of plagiarism, 85(11%) conformed as plagiarized manually, 63(8%) were true plagiarism, and 22(3%) were identified as self-plagiarism. The manuscripts mostly came from China (21%), Croatia (14%), and Turkey, (19%) (Baz'daric', Bilic'-Zulle, Brumini, & Petrovec'ki, 2012)[10]. To concur with these findings, plagiarism is even prevalent in speech (Mariani, Francopoulo, & Paroubek, 2017) [11]. To be precise, plagiarism is prevalent in the doctoral theses submitted for the award of higher degrees in different institutions of higher learning (Kumar, 2019) [12]. This necessitated research of this kind to investigate doctoral theses to understand the pattern of plagiarism to aid in quality control of research in

Textual Similarity Patterns and Citation Style Errors in Literature Review Chapters of LIS Theses of South Indian Universities: A Study

universities, and formulating of viable laws, regulations, policies, etc. concerning dissertations, theses, or any other document writing. Hence this study analysis the text similarity pattern and writing style in literature review chapter Ph. D. Theses submitted by LIS doctoral Students.

3. METHODOLOGY

The doctoral theses submitted to different South Indian universities in the subject area of Library and Information Science were selected for the study. A purposive sample of Literature Review chapters from 20 theses have been used for sample selection. The literature review chapter has been selected as the study corpus based on the assumption that among the mandatory chapters, a review of literature is prepared by the synthesis of previous studies to set the background or to establish the research gap. Hence, it is inevitable to acknowledge the previous authors. The chapters were run in the anti-plagiarism software Turnitin to identify the similarity check. The similarity reports were scrutinized manually, and the errors were quantified in excel for further analysis.

4. ANALYSIS AND DISCUSSION

4.1 Overview of sample

Table 1: Overview of sample

Chapter code	Year of submission	of Length chapter	of Number of words	The type of Style manual adopted
STLR61	2013	33	9629	APA
STLR73	2014	64	20301	APA
STLR83	2014	41	13106	APA
STLR58	2015	28	8218	APA
STLR71	2015	54	15388	APA
STLR82	2015	43	11527	APA
STLR56	2016	35	6584	APA
STLR57	2016	31	5899	APA
STLR80	2016	48	11637	APA
STLR64	2017	28	9219	APA
STLR86	2017	33	9015	APA
STLR53	2018	34	12222	APA
STLR68	2018	25	6363	APA
STLR69	2018	46	14218	APA
STLR76	2018	38	11207	APA
STLR88	2018	48	10946	APA
STLR89	2018	28	6683	APA
STLR62	2019	21	9959	APA
STLR67	2019	45	18755	APA
STLR51	2019	24	9254	APA
Average		37.35	11006.5	

Table 1 gives the overview of the study sample in terms of year of submission, number of words, number of references, and the citation manual adopted for rendering the in-text and end-text citations. It is noted that the average length of review of the literature chapter is 37.35 pages and the average number of words is 110006.5. All the study samples selected adopted the American Psychological Association reference manual for formatting the thesis. The year of submission ranges from 2013 to 2019 as the study considered 10 years' time span in which more number of submissions occurred in 2019 in the south Indian universities.

4.2 Pattern of text copying from other sources

Table 2: Pattern of text copying of LIS researchers (N=20)

Category	Occurrence	%
Citation inaccuracy	868	46.072
Improperly paraphrased	686	36.412
Word-to-word copy	330	17.516
Total	1884	100

It is observed from Table 2 that the mostly found mistakes of LIS researchers while borrowing the text from other sources is citation error (46.072%). In this study, the citation inaccuracies refer to the intent citation in which the author is supposed to adhere to one of the reference-style manuals while reporting the research in the form of a thesis. As per The American Psychological Association (APA) style manual, direct quotes up to 40 words (short quotations) should be enclosed in double inverted comma along with author name, year of publication and page number and direct quotes more than 40 words (block quotations) should be in separate indent paragraph with page number (APA, 2020)[13]. The second highest occurrence of text borrowing inaccuracy is the improperly paraphrase with 686 occurrences and word to word copy of content comprises 330 instances in the over all 20 Literature review chapters analyzed.

4.3 String length of word-to-word copying of text

Table 3: String length of word-to-word copying of text

String length	Occurrence	%
15 to 30	237	71.818
31 to 40	93	28.182
Total	330	100

Table 3 shows the number of consecutive words copied by the LIS researchers from the original source without giving proper acknowledgment or following any style manual rules of copying text from other sources. To determine how much material was copied, the researchers manually counted the highlighted content provided Turnitin software. This is clear from the fact that words with an average length of 15 to 30 appear the most frequently in cases of text copying. This suggests that academicians and supervisors should get involved and instruct the researchers on how to properly quote content for citing. While creating their research implications and reporting it as theses or even research articles, the researchers should use suitable content borrowing practises.

4.4 Strategy used in improper paraphrasing

One method for avoiding the straight copying of text from other sources is to use paraphrasing. While paraphrasing, it is necessary to give credit to the original author and make adjust the sentence structure and lexicon without compromising the essence of the text. Failure to do so results in unethical text copying, which is done by introducing or removing certain words from the paragraph or by replacing them with synonyms.

Table 4: Strategy used in improper paraphrasing.

Improperly paraphrased	Occurrence	%
Change of verb	360	52.478
Replacement with synonyms	97	14.140
Addition of new words	112	16.327
Deletion of existing words	66	9.621
Spelling changes	51	7.434
Total	686	100

Table 4 depicts the comprehensive list of strategies used by the LIS researchers for paraphrasing others work. A total of 686 attempts of improper paraphrasing strategies has been identified in which the major categories were changing the reporting verb and addition of new words. The change of verb (52.478%) was the major category employed by the researchers while writing the literature review chapter. It has been noticed that most of the reviews are arranged in chronological order in which the abstract has been copied using the various paraphrasing strategies listed in table 3. Although the researchers attempted to paraphrase the abstract of various published articles, the anti-plagiarism software shows the matches because of the poor paraphrasing techniques employed while reviewing. It can be assumed that these types of dishonesty are employed mainly because of the researchers' poor command over the English language and the unawareness of the proper way of paraphrasing a particular concept or an article altogether.

4.5 Types of citation error

The author of a thesis follows a particular referencing style for the writing and formatting of the thesis, and the errors in both in-text and end text citation cannot be made. As per the definitions of plagiarism, text copies from other sources without putting quotations while including in the text is considered plagiarism. This study considered the Review of Literature chapter as the study corpora to analyse the pattern an practice of LIS researchers while writing the thesis.

Table 5: Types of citation error

Citation error	Occurrence	Average string length
without indent paragraph	639	53.15
without quotation	229	16.5
Total	868	100

Table 5 describes the type of citation errors identified in the review of literature chapters of various 20 theses submitted to different south Indian universities by LIS researchers. Block paragraphs with more than 40 words of direct copying comprise the highest number of errors committed by the LIS researchers. As already discussed in Table 1, Manual of the American Psychological Association mandates to present the direct quotes more than 40 words in special indent paragraphs. Failing to which is considered plagiarism. It is difficult to infer that this type of plagiarism is intentional or unintentional.

CONCLUSION

The results of this study make it very evident that many LIS researchers copy abstracts or even review from other researchers without synthesizing the concept very effectively. Without making any attempt to rephrase the material that was handed to them, students used to copy it exactly, and the literature review suffered the most improper text-borrowing in this regard. To reduce this practice to the bare minimum, text borrowing-related issues should be covered in the curriculum or in the coursework given to doctorate candidates. It is unexpected that PhD students are unaware of text-stealing. To prevent students from copying text from other sources, extra care should be taken to improve their writing and comprehension aspects of language.

REFERENCES

- [1] I. B. Siaputra, (2013). "The 4PA of plagiarism: A psycho-academic profile of plagiarists, *International Journal for Educational Integrity*", *International Journal for Educational Integrity*, vol. 9, 2013.
- [2] V. Pupovac, L. Bilic-Zulle, and M. Petrovecki, "On academic plagiarism in Europe. An analytic approach based on four studies" In: R. Comas and J. Sureda (coords.). *Academic cyberplagiarism*. [online dossier]. *Digithum*, vol.10, 2008.
- [3] D. Starovoytova, "Plagiarism under a Magnifying-Glass", *Journal of Education and Practice*, vol. 8, pp.109–129, 2017.
- [4] D. Gotterbarn, K. Miller and J. Impagliazzo, "Plagiarism and scholarly publications: An ethical analysis," 36th ASEE/IEEE Frontiers in Education Conference M1H-22, San Diego, 2006.
- [5] T. Marshall, B. Taylor, E. Hothersall and L. Pe´rez-Marti´n, "Plagiarism: A case study of quality improvement in a taught postgraduate programme," *Medical Teacher*, vol. 33 pp.375–381, (2011).
- [6] D.C. Ison, "Plagiarism among dissertations: Prevalence at online institutions," *J Acad. Ethics*, vol.10, pp.227–236, (2012).

Textual Similarity Patterns and Citation Style Errors in Literature Review Chapters of LIS Theses of South Indian Universities: A Study

- [7] J. Warn, "Plagiarism software: No magic bullet," Higher Education Research and Development, vol. 25, pp. 195–208, (2006).
- [8] J. Walker, "Measuring plagiarism: researching what students do, not what they say they do," Studies in Higher Education vol.35, pp.41-59, 2010.
- [9] T., Bretag and S.Carapiet, "A preliminary study to identify the extent of self- plagiarism in Australian academic research," Plagiary: Cross-Disciplinary Studies in Plagiarism, Fabrication, and Falsification, vol. pp. 92-103, 2007.
- [10] K. Bazđaric', L. Bilic' - G. Zulle, Brumini and M. Petrovec'ki, "Prevalence of plagiarism in recent submissions to the Croatian Medical Journal," Sci Eng Ethics ,vol. 18, pp. 223–239. 2012.
- [11] J. Mariani, G. Francopoulo and P. Paroubek, "Reuse and plagiarism in Speech and Natural Language Processing publications," Int J Digit Libr., 2017.
- [12] P.K.S. Kumar, "Similarity index of doctoral theses submitted to universities in Kerala: An investigation," Library Philosophy and Practice (e-journal), pp.2130, 2019.
- [13] APA, Publication Manual of the American Psychology Association, 7th ed., Washington: American Psychology Association, 2020, pp.163-164.
-