Checking Similarity in EFL Undergraduate Students’ Research Proposal

I Gede Yoga Permana

English Education Department, STKIP Agama Hindu Singaraja, Buleleng, Indonesia.

ABSTRAK
Penelitian saat ini bertujuan untuk mengidentifikasi kesamaan yang ditemukan dalam kertas kerja penelitian mahasiswa. Dalam melakukan pemeriksaan kesamaan, digunakan perangkat lunak bernama Plagiarism X checker. Seratus makalah penelitian siswa dikumpulkan dan kesamaan diperiksa. Hasil penelitian menunjukkan bahwa kesamaan yang teridentifikasi dalam pekerjaan siswa adalah 22.18% yang termasuk dalam kategori sedang. Persentase kesamaan tertinggi terdapat pada latar belakang penelitian (13.86%) diikuti dengan tinjauan teoritis (5.22%) dan metode penelitian (3.08%). Dilihat dari latar belakang penelitian, persentase kemiripan tertinggi adalah 34% yang termasuk dalam kategori sedang. Sedangkan pada tinjauan teoritis dan metode penelitian kemiripannya kurang dari 15% yang menunjukkan tingkat yang kecil. Beberapa makalah diidentifikasi mengandung paragraf curian karena penggunaan kata-kata yang sama persis dengan sumber aslinya tanpa ada proses penyuntingan. Diskusi dan saran disebutkan lebih lanjut.

ABSTRACT
The current study aimed at identifying the similarity found in undergraduate students’ research working paper. In doing similarity check, a software named Plagiarism X checker was used. A hundred students’ research paper were collected and similarity checked. The result showed that the similarity identified in students work is 22.18% which falls into medium category. The highest percentage of similarity is found in the background of study (13.86%) followed with theoretical review (5.22%) and research method (3.08%). In background of study, the highest percentage of similarity is 34% that belongs to medium level. Meanwhile in theoretical review and research method the similarity is less than 15% that indicates a small level. Some papers were identified contain a stolen paragraph because the use of exact same words with the original sources without any process of editing. Discussion and suggestions are further presented.

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

Technological advances provide easy access to a huge source of information on the internet. This source of information certainly helps everyone, especially academics to find research sources (Anney & Mosha, 2015). However, this convenience is sometimes used to carry out unethical actions such as plagiarism (Ison, 2014; Sutherland-Smith, 2005). Plagiarism comes from the Latin plaga that means to kidnap or steal (Gasparyan et al., 2017). Plagiarism is conceptually defined as an act of using the work or ideas of people without giving proper credit or attribution (Grossberg, 2011; Helgesson & Eriksson, 2018) The act such as copy-paste, translation, self-plagiarism are examples of forms of plagiarism (Weber-Wulff, 2014).

There are several reasons for committing plagiarism. The most common reason found is the lack of knowledge about the concept of plagiarism (Al Darwish & Sadeqi, 2016; Permana & Santosa, 2018). Understanding the concept of plagiarism is very important for students. By understanding the concept, they will be able to avoid acts of plagiarism. The lack of instruction and guidance from teachers and institutions (Konstantinidis et al., 2020; Louw, 2017). The presence of teachers and institutions in providing regulations on plagiarism is considered very crucial in helping students to avoid plagiarism (McGee, 2013). The other reason of committing plagiarism is the absence of plagiarism...
Plagiarism detection applied in university or institution (Bakhtiyari et al., 2014; Batane, 2010; Stapleton, 2012). Plagiarism detection programs strongly provides an overview of plagiarism found in student work.

One of similarity checking tools available is Turnitin. Turnitin is considered the best similarity checking tools because it provides complete information of similarity content and had larger database (Batane, 2010; Shahabi, 2012). However, due to its cost, not much university is able to afford it. Many universities are turning to alternative similarity testing tools such as Plagiarism X Checker. The Plagiarism X checker also gave the link where the similar content can be found. In terms of effectiveness of plagiarism checker, there were studies reported that plagiarism checker had positive benefit both for university and students as well (Paul & Jamal, 2015; Srivastava & Govilkar, 2017). Their studies revealed that when a plagiarism checker was applied in classroom, the students’ academic writing skill significantly improved and they tended to avoid plagiarism because they were afraid of being caught.

Despite of the availability of plagiarism detection, many studies have found cases of plagiarism. In Indonesian context, Sariffuddin et al., (2017) found some students’ papers containing 25% - 75% of plagiarism content through Turnitin check. Sulaiman (2018) also identified that from Dupechecker test, 97.54% from 44 papers contains plagiarism. Manual check was also conducted and found that students mostly did not give proper credit to the sources mentioned in the paper. From the results of previous studies, it can be concluded that plagiarism is indeed happening in the academic world. A more recent study conducted by Arias-chávez and Ramos-quispe (2020) found 28.6% of students work is indicated in level III (25-49% content of similarity), 50% indicated in level IV (50%-74% content of similarity) and 21.4% indicated in level V (above 74%).

In the process of completing study, students are required to write a thesis. Thesis is a part of academic writing which is very closely related to plagiarism (Nadelson, 2007; Pecorari, 2003). In the process of writing a thesis, students are initially asked to write a research proposal. The proposal contains (1) research background, (2) literature review, and (3) research method. Despite testing the similarity of the students’ final thesis, this study focuses on identifying the similarity on students’ proposal as an early detection. In doing the similarity test, the Plagiarism X checker was used to identify the similarity found in students writing. The results of the similarity test on the proposal can provide an overview of the indications of plagiarism in students’ research proposal and can be used as a guidance for students to be more careful in taking or citing sources. Besides that, the result of similarity check in this study can be used as reference for universities that have not yet implemented similarity check in taking further steps to counter plagiarism.

2. METHOD

The current study is a descriptive study employing document study to identify the similarity found in EFL students’ research proposal by utilizing a tool namely Plagiarism X Checker. Plagiarism X checker gives a brief analysis on text by giving the level of similarity in percentage and giving direction to the source in which taken by the writer.

The participants of the current study were EFL students in 7th semester. There were 124 students invited to submit their research proposal to be similarity checked. However, only 100 students submitted.

A descriptive statistics analysis is also used to identify the general percentage of similarity found in students’ research proposal. The percentage of students’ similarity contents were classified based on the report generated by Plagiarism X Checker where <19% is categorized low, 20%-39% is categorized medium, and above 40% is categorized high.
3. FINDINGS AND DISCUSSIONS

The summary of descriptive statistics is presented in Table 1.

| N | Mean   | Range | Small Similarity (≤ 19%) | Medium Similarity (20% - 39%) | High Similarity (≥ 40%) |
|---|--------|-------|--------------------------|--------------------------------|--------------------------|
| 10| 22.18% | 37%   | 40                       | 58                             | 2                        |

In Table 1, the average similarity found in student proposals is 22.18%. In terms of range, the lowest percentage of similarity was 3% and the highest was 40%. According to the similarity level category in the Plagiarism X checker, there are 40 student research proposals that are in the small category, 58 proposals in the medium category and 2 proposals in the high category. The next analysis conducted is the analysis of the level of similarity in the parts of the students’ proposal (background of study, theoretical review, and research method). Table 2 presents the result of the analysis.

Table 2 shows that the highest percentage of similarity is found in the background of study (13.86%) followed with theoretical review (5.22%) and research method (3.08%). In background of study, the highest percentage of similarity is 34% that belongs to medium level. Meanwhile in theoretical review and research method the similarity is less than 15% that indicates a small level.

In general, the level of similarity found in student proposals is at the medium level according to the criteria in Plagiarism X Checker. Deeper looking from the results of the analysis, there are a number of students taking actions that can be categorized as plagiarism by using the exact same sentence as the source used as shown in Table 3.
In Table 3, students directly copy and paste text from the source without paraphrase or proper editing. Copy-paste as an act of plagiarism occurs for number of reasons. The ease and fast of retrieving information from the internet is a factor for students to copy-paste (Anney & Mosha, 2015; Gasparyan et al., 2017). In the context of EFL, students’ inability to paraphrase sentences characterizes low academic writing skills which can trigger plagiarism (Al Darwish & Sadeqi, 2016; Hermansyah & Aridah, 2021; Hu & Sun, 2016). The results of this study also found that similarity is frequently was found in the background of study content. Developing a background of study is indeed not something easy for EFL students especially for those who have poor academic writing skills. Difficulties that are usually faced by students in the writing process are organizing ideas and limited knowledge about grammar and vocabulary (Doró, 2015). In addition, the factor of students doing copy paste is the limited ability to develop a research idea (Alsied & Ibrahim, 2017).

In the thesis writing process, the role of the supervisor is certainly very crucial in relation to the ethics of conducting research. As an act of prevention, supervisor can provide understanding to students about the concepts and consequences of doing (Romanowski, 2021). The absence of a supervisor in the research process. Students tend to choose a shortcut to do plagiarism because of the limited time in completing studies (Permana & Santosa, 2018). In addition, the role of institutions is also needed in this context as the policy maker. Policy is very important in the process of prevention, detection, sanctions and education for all students and teachers (Carroll & Zetterling, 2013). Using similarity checking tools is a must as a step of early detection. However, due to limitation of similarity checking tools, institutions or universities need to develop policy in determining the level of similarity that is allowed or tolerated.

To cope with this issue, the use of plagiarism checker on students’ writing are currently used widely by universities. There are numbers of plagiarism checker software available such as Turnitin, Ithenticate and Plagscan. One and the most well-known software is Turnitin but it is not free. There are several softwares that offer plagiarism checking and the software are free such as Plagiarism X, Grammarly and Quetext. Even though the result given by those free software is not accurate as Turnitin. However, there are several limitations on using plagiarism checker. As found by Atkinson and Yeoh (2008), plagiarism checker cannot detect source or material which were not available in Internet. Related with the finding of the study, the plagiarism checker was also limited on intentional plagiarism. The accidental similarities would be identified as plagiarism by plagiarism detection machine (Bakhtiyari et al., 2014). The phrases which were tagged as plagiarism were actually a common knowledge. This weakness of plagiarism checker was also identified by (Brown et al., 2010). They found Plagiarism checker can produce a false report by identifying common phrases as a plagiarism.

The results of this study imply that the similarity content test process is important to be carried out as early as possible and continuously. The similarity content test will be able to help students map the level of similarity that exists in their research proposals and immediately revise it before it becomes a final product, namely a thesis for bachelor or master's level, and a dissertation at the doctoral level.

4. CONCLUSION

The results of this study identify that there are similarities content in students’ work. The average similarity found in student proposals is 22.18%. In terms of range, the lowest percentage of similarity was 3% and the highest was 40%. According to the similarity level category in the Plagiarism X checker, there are 40 student research proposals that are in the small category, 58 proposals in the medium category and 2 proposals in the high category. The highest percentage of similarity is found in the background of study (13.86%) followed with theoretical review (5.22%) and research method (3.08%). In background of study, the highest percentage of similarity is 34% that belongs to medium level. Meanwhile in theoretical review and research method the similarity is less than 15% that indicates a small level. Some of the similarities in student work are indicated as acts of plagiarism because the sentences used are exactly the same as the sentence in the original source. This research is limited to description of similarity results using one application.
5. REFERENCES

Al Darwish, S., & Sadeqi, A. A. (2016). Reasons for College Students to Plagiarize in EFL Writing: Students’ Motivation to Pass. *International Education Studies*, 9(9), 99. https://doi.org/10.5539/ies.v9n9p99

Alsied, S. M., & Ibrahim, N. W. (2017). Exploring Challenges Encountered by EFL Libyan Learners in Research Teaching and Writing. *IAFOR Journal of Language Learning*, 3(2), 143–158.

Anney, V. N., & Mosha, M. A. (2015). Student’s Plagiarisms in Higher Learning Institutions in the Era of Improved Internet Access: Case Study of Developing Countries. *Journal of Education and Practice*, 6(13), 203–216. https://eric.ed.gov/?id=EJ1080502 on 10th August 2015.

Arias-chávez, D., & Ramos-quispe, T. (2020). Evaluation of the Index of Similarity Detected by Turnitin in Research Projects of a Master’s Degree in Higher Education. *Proceedings of the Second Workshop on Second-Order Quantifier Elimination and Related Topics*, 1–11.

Atkinson, D., & Yeoh, S. (2008). Student and staff perceptions of the effectiveness of plagiarism detection software. *Australasian Journal of Educational Technology*, 24(2), 222–240.

Bakhtiyari, K., Salehi, H., Embi, M. A., Shakiba, M., Zavvari, A., Shahbazi-Moghadam, M., Ebrahim, N. A., & Mohammadjafari, M. (2014). Ethical and unethical methods of plagiarism prevention in academic writing. *International Education Studies*, 7(7), 52–62. https://doi.org/10.5539/ies.v7n7p52

Batane, T. (2010). Turning to Turnitin to Fight Plagiarism among University Students. *Educational Technology and Society*, 13(2), 1–12.

Brown, V., Jordan, R., Rubin, N., & Arome, G. (2010). Strengths and Weaknesses of Plagiarism Detection Software. *Journal of Literacy and Technology*, 11(1), 110–131.

Carroll, J., & Zetterling, C. (2013). Guiding students away from plagiarism. In *Writing integrity workshop, Stockholm*. http://people.kth.se/~ambe/KTH/Guidingstudents.pdf

Doró, K. (2015). Why Do Students Plagiarize? Efl Undergraduates’ Views on the Reasons Behind Plagiarism. *Romanian Journal of English Studies*, 11(1), 255–263. https://doi.org/10.2478/rjes-2014-0029

Gasparyan, A. Y., Nurmashev, B., Seksenbayev, B., Trukhachev, V. I., Kostyukova, E. I., & Kitas, G. D. (2017). Plagiarism in the context of education and evolving detection strategies. *Journal of Korean Medical Science*, 32(8), 1220–1227. https://doi.org/10.3346/jkms.2017.32.8.1220

Grossberg, M. (2011). History and the Disciplining of Plagiarism. In C. Eisner & M. Vicinus (Eds.), *Originality, Imitation, and Plagiarism: Teaching Writing in the Digital Age* (4th ed., pp. 159–172). University of Michigan Press.

Helgesson, G., & Eriksson, S. (2018). Plagiarism in research. *Getting to Good: Research Integrity in the Biomedical Sciences*, 262–273. https://doi.org/10.1007/s11019-014-9583-8

Hermansyah, H., & Aridah, A. (2021). Teachers’ Perception toward the Challenges in Online English Teaching during Covid-19 Pandemic. *Indonesian Journal of EFL and Linguistics*, 6(1), 63. https://doi.org/10.21462/ijefl.v6i1.342

Hu, G., & Sun, X. (2016). Chinese university EFL teachers’ knowledge of and stance on plagiarism. *Comunicar*, 24(48), 29–37. https://doi.org/10.3916/C48-2016-03

Ison, D. (2014). Does the Online Environment Promote Plagiarism? A Comparative Study of Dissertations from Brick-and-Mortar versus Online Institutions. *Journal of Online Learning and Teaching*, 10(2), 272.

Konstantinidis, A., Theodosiadou, D., & Pappos, C. (2020). Plagiarism: Examination of Conceptual Issues and Evaluation of Research Findings on Using Detection Services. *Contemporary Educational Technology*, 4(3), 212–221. https://doi.org/10.30935/cedtech/6104

Louw, H. (2017). Defining plagiarism: Student and staff perceptions of a grey concept. *South African Journal of Higher Education*, 31(5), 116–135. https://doi.org/10.20853/31-5-580

McGee, P. (2013). Supporting academic honesty in online courses. *Journal of Educators Online*, 10(1), 1–31. https://doi.org/10.9743/JEO.2013.1.6

Nadelson, S. (2007). Academic Misconduct by University Students: Faculty Perceptions and Responses. *Plagiarism, Cross-Disciplinary Studies in Plagiarism, Fabrication, and Falsification*, 67–76.

Paul, M., & Jamal, S. (2015). An improved SRL based plagiarism detection technique using Sentence...
ranking. Procedia Computer Science, 46(Icict 2014), 223–230. 
https://doi.org/10.1016/j.procs.2015.02.015

Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language 
writing. Journal of Second Language Writing, 12(4), 317–345. 
https://doi.org/10.1016/j.jslw.2003.08.004

Permana, I. G. Y., & Santos, M. H. (2018). EFL Students’ Perception on Plagiarism. Language in the 
Online & Offline World 6: The Fortitude, May, 128–133.

Romanowski, M. H. (2021). Preservice Teachers’ Perception of Plagiarism: A Case from a College of 
Education. Journal of Academic Ethics. https://doi.org/10.1007/s10805-021-09395-4

Sariffuddin, S., Astuti, K. D., & Arthur, R. (2017). Investigating Plagiarism: The Form and The 
Motivation in Performing Plagiarism in High Education. Journal of Education and Learning 
(EduLearn), 11(2), 172–178. https://doi.org/10.11591/edulearn.v11i2.5994

Shahabi, M. (2012). Comparing Three Plagiarism Tools (Ferret, Sherlock, and Turnitin). Internal 
Journal of Computational Linguistics (IJCL), 3(1), 53–66.

Srivastava, S., & Govilkar, S. (2017). A Survey on Paraphrase Detection Techniques for Indian Regional 
Languages. International Journal of Computer Applications, 163(9), 42–47. 
https://doi.org/10.5120/ijca2017913757

Stapleton, P. (2012). Gauging the effectiveness of anti-plagiarism software: An empirical study of 
second language graduate writers. Journal of English for Academic Purposes, 11(2), 125–133. 
https://doi.org/10.1016/j.jeap.2011.10.003

Sulaiman, R. (2018). Types and Factors Causing Plagiarism in Papers of English Education Students. 
Inspiring: English Education Journal, 1(1), 95–104. https://doi.org/10.35905/inspiring.v1i1.836

Sutherland-Smith, W. (2005). Pandora’s box: Academic perceptions of student plagiarism in writing. 
Journal of English for Academic Purposes, 4(1), 83–95. 
https://doi.org/10.1016/j.jeap.2004.07.007

Weber-Wulff, D. (2014). False Feathers: A Perspective on Academic Plagiarism. Springer.