

EXPLORING THE RELATIONSHIP BETWEEN VOCABULARY MASTERY AND PLAGIARISM AMONG UNIVERSITY STUDENTS

Dang Arif Hartono¹, Stefanus Angga Badara Prima², Jonathan Gultom³, Wahyu Indriyo⁴

Universitas Agung Podomoro

dang.arif@podomorouniversity.ac.id

Abstract

Despite the growing body of literature on plagiarism and academic dishonesty, few studies have explored how vocabulary mastery contributes to plagiaristic behavior. This is unfortunate, as vocabulary is the building block of language proficiency, and without it, one's ability to produce spoken or written texts may be hampered. To address this gap, this quantitative study investigates the relationship between vocabulary mastery and plagiarism among university students. Conducted at a private university in West Jakarta, Indonesia, this study involved 71 students from various majors enrolled in a rhetoric and composition course. Using iThenticate to measure plagiarism and the Vocabulary Levels Test (VLT) to assess vocabulary proficiency, the findings indicate a low average plagiarism rate of 11.56% in the students' essays and no significant correlation between vocabulary mastery and the extent of plagiarism. Further analysis revealed no substantial difference in the amount of plagiarized content between students with lower and higher vocabulary proficiency. These findings suggest that vocabulary mastery alone may not be a significant factor in determining plagiarism behaviors, confirming the multifaceted nature of plagiarism. The study underscores the need for further research into other potential contributors to plagiarism, such as psychological factors and cultural influences, and emphasizes the importance of teaching proper citation practices to combat academic dishonesty.

Keywords: *Plagiarism, Vocabulary Mastery, University Students, Academic Dishonesty, Language Proficiency*

Introduction

Research has highlighted a significant rise in plagiarism within higher education institutions. [1] discovered that in Vietnam, where one university does not use Turnitin software, 91.7% of graduation reports were flagged for plagiarism, compared to 61.7% at an institution utilizing such software. In Austria, [2] reported a substantial incidence of plagiarism among college students in an Austrian university. Similarly, [3] underscored the escalating issue of plagiarism across higher education institutions in sub-Saharan Africa. [4] state that plagiarism is rampant in Indonesian higher education institutions. For instance, [5] found that 44 out of 50 students at a private Indonesian university produced essays containing varying extents of verbatim plagiarism. Furthermore, [6] identified mosaic and verbatim plagiarism as the most prevalent types in another private Indonesian university, despite students having received training on understanding and avoiding plagiarism. These studies underscore the increasing prevalence of plagiarism in universities and colleges globally.

The literature on plagiarism has provided insights into its causes. [7] suggest that the main reason for the prevalent plagiarism among university students is the increased ease of accessing information and technology. The availability of information and documents on the internet and in online libraries has, to some extent, led to an increase in plagiarism. [8] conducted a study on plagiarism in an academic writing course for postgraduate students in Malaysia, and they found that internet sources are among the primary sources of plagiarism identified in the students' writing.

According to other researchers, including [9] and [10], cultural norms and values can also influence plagiarism. [11] contend that in some Asian countries, where education systems heavily focus on rote learning,

students may not perceive plagiarism as ethically problematic. For example, learners from Confucian societies may not have developed as strong notions of text ownership as Western societies, seeing textual borrowing as a sign of respect toward the original work and blurring the line between common knowledge and original ideas (Chen, 2010). Similarly, (Chen, 2010) found that cultural factors, such as an emphasis on social relationships and reciprocity in writing, influence student plagiarism in the Taiwanese context.

Some researchers attribute plagiarism to limited linguistic competence.

(Chen, 2010) highlight that a lack of language skills and difficulty expressing thoughts or concepts can result in students resorting to plagiarism. Similarly, (Chen, 2010) argued that low English proficiency may make students less aware that heavy reliance on direct copying from a source is unacceptable in academic settings. Additionally, (Chen, 2010) found that poor vocabulary command is a significant cause of plagiarism among second-language writers.

Despite the growing body of literature on plagiarism, studies investigating how linguistic competence contributes to plagiarism are still limited. Many published studies focus on other aspects, such as learners' or teachers' perceptions and attitudes toward plagiarism (e.g. Erguvan, 2022; Khatayut et al., 2022; Nketsiah et al., 2023, Tran et al., 2022) institutional policies on plagiarism (e.g., (Chen, 2010); (Chen, 2010); (Chen, 2010)), software or techniques to detect plagiarism (e.g., (Chen, 2010); (Chen, 2010); (Chen, 2010)), and the types or severity of plagiarism committed by students (Chen, 2010; (Chen, 2010); (Chen, 2010)). More recently, research has also examined the role of artificial intelligence in facilitating plagiarism (Chen, 2010; (Chen, 2010)).

The lack of research on how linguistic competence, specifically vocabulary mastery, contributes to plagiarism is unfortunate. Vocabulary is an essential building block of language (Chen, 2010), without which one might find comprehension and production in a foreign language challenging. (Chen, 2010) emphasized the importance of vocabulary by stating that while grammar allows for limited communication, effective communication is impossible without vocabulary. The only published study that looked into vocabulary's contribution to plagiarism was conducted by (Chen, 2010) at an EMI university in China, which found a strong negative correlation between vocabulary size and the plagiarism rate among two hundred students. Given the limited number of studies on how vocabulary mastery may contribute to plagiarism, it is evident that there is a significant gap in the literature that warrants further investigation.

Therefore, this study seeks to address the existing gap in the literature by investigating the relationship between vocabulary proficiency and instances of plagiarism. Specifically, the authors seek to investigate (1) the extent to which learners' vocabulary mastery correlates with the percentage of plagiarized content in their writings and (2) whether there is a difference in the amount of plagiarized content in the writings produced by learners with lower levels of vocabulary mastery compared to those with higher levels of vocabulary mastery.

Method

The study was conducted between February and April 2024 at a private university in West Jakarta, Indonesia. Seventy-one students from two intact classes, comprising 41 females and 30 males, participated in this study. The students' ages ranged from 16 to 22 years, with an average of 18.99 years. They were from various majors: 24 in product design, 16 in accounting, 12 in architecture, 8 in management and construction engineering, 6 in business law, and 5 in urban and regional planning.

This study employed a quantitative research design to explore the relationship between vocabulary mastery and the incidence of plagiarism in students' writing. The main instruments used were the Vocabulary Levels Test (VLT) and the iThenticate plagiarism detection software.

The research was conducted in a mandatory rhetoric and composition course in which students were taught to write argumentative essays. Over several weeks in the first half of the course, the students concentrated on writing an argumentative essay for a mid-semester assessment. In the fourth week, the class covered notable instances of plagiarism, various forms of plagiarism, and its repercussions. Additionally, they were instructed on correct paraphrasing and citation techniques. In the fifth and sixth weeks, the students researched their selected topics, created outlines, and started writing their essays. Peer review sessions took place in the seventh week, providing an opportunity for feedback and revisions. Ultimately, the students submitted their final argumentative essays for mid-semester assessment in the eighth week

The submitted essays were analyzed using iThenticate, a subscription-based plagiarism checker. iThenticate presents the analysis results as a similarity index, indicating the percentage of plagiarized content in the submitted text. A similarity index of 23%, for example, means that 23% of the text matches content in the iThenticate database, suggesting a corresponding percentage of plagiarized content.

To assess the participants' vocabulary mastery, the Vocabulary Levels Test (VLT) developed and validated by Webb et al. (2017) was administered during the first session of the course. This test evaluates learners' English vocabulary proficiency across five levels, ranging from 1,000 to 5,000-word families, with each level scored from 0 to 30. Thus, the total possible scores range from 0 to 150. A score of 26 or higher indicates mastery at a particular level. Vocabulary within the 1,000 to 3,000-word families is considered beginner or high-frequency vocabulary, whereas vocabulary within the 4,000 to 5,000-word families is regarded as intermediate-level vocabulary. Vocabulary from 6,000 to 9,000-word families is classified as advanced.

The collected data were tabulated and analyzed using IBM® SPSS version 26. Following recommendations from experts in second language acquisition research (Larson-Hall, 2015; Plonsky et al., 2014; Plonsky, 2015), we employed a robust statistical technique called bootstrapping to generate descriptive and inferential statistics. Bootstrapping leverages parametric techniques but uses computer-intensive methods to eliminate the requirement for normally distributed data. It is resilient to assumption violations (Larson-Hall, 2015), and more powerful than both parametric and non-parametric statistics (Plonsky et al., 2014). Following LaFlair et al., (2015) suggestions', the specifications for bootstrapping in this study included a simple sampling method, 10,000 resampling iterations, and the calculation of bias-corrected and accelerated (BCa) 95% confidence intervals (CI).

Findings and Discussions

Findings

Table 1 presents the descriptive statistics of the study's findings. As shown in the table, the participants generally performed well on the VLT, achieving average scores of 26 or higher, except at the 3,000-word level, where the average score was slightly below 26.

Table 1. Descriptive Statistics (with Bootstrapping)

| Variables | Min. | Max. | M | BCa 95% CI for M | | SD |
|---------------------|------|------|--------|------------------|--------|--------|
| | | | | LL | UL | |
| 1,000-word families | 22 | 30 | 29.21 | 28.21 | 29.54 | 1.48 |
| 2,000-word families | 11 | 30 | 27.84 | 27.01 | 28.54 | 3.16 |
| 3,000-word families | 7 | 30 | 25.73 | 24.55 | 26.80 | 4.59 |
| 4,000-word families | 10 | 30 | 26.22 | 25.22 | 27.12 | 3.84 |
| 5,000-word families | 2 | 30 | 26.19 | 24.81 | 27.38 | 5.17 |
| Overall Score | 59 | 150 | 135.21 | 130.96 | 138.90 | 16.05 |
| Number of Words | 254 | 1991 | 608.78 | 556.43 | 668.91 | 262.73 |

| | | | | | | |
|--------------------------|---|----|-------|------|-------|-------|
| Percentage of Plagiarism | 0 | 85 | 11.56 | 7.47 | 16.04 | 19.81 |
|--------------------------|---|----|-------|------|-------|-------|

Note: $N = 71$

In our analysis, we also measured the length of the essays produced by the participants. As shown in Table 1, the participants wrote an average of slightly more than 600 words (excluding the titles of their essays, personal information, and references), equivalent to one and a half to two and a half pages. We found this result reasonable for first- and second-year students. It is also interesting to note that the average percentage of plagiarized content in the participants' essays was only 11.56%. This is relatively low and below the university's maximum threshold of 20%.

To explore the relationship between variables, we calculated Pearson's product-moment correlation with bootstrapping. The result is displayed in Table 2.

Table 2. Correlation Matrix between Variables (with Bootstrapping)

| | Percentage of Plagiarism | | BCa 95% CI for r | | Effect Size (r^2) |
|---------------------|--------------------------|-----|--------------------|-----|-----------------------|
| | r | p | LL | UL | |
| 1,000-word families | .12 | .30 | -.00 | .23 | .015 |
| 2,000-word families | .10 | .37 | -.06 | .24 | .011 |
| 3,000-word families | .15 | .20 | -.02 | .29 | .023 |
| 4,000-word families | .06 | .58 | -.11 | .21 | .004 |
| 5,000-word families | .11 | .33 | -.07 | .22 | .013 |
| Overall Score | .13 | .28 | -.03 | .26 | .016 |
| Number of Words | -.14 | .21 | -.28 | .01 | .022 |

Note: $N = 71$

Table 2 shows the correlation between the participants' vocabulary mastery at various levels, the length of their writings, and the percentage of plagiarism in their essays. As shown in the table, there are small correlations between these variables, but none are statistically significant. The effect sizes are also very small, ranging from 0.4% to 2.3%. These results suggest that vocabulary mastery has a negligible relationship with plagiarism and might not influence it.

We divided the participants into two groups to assess the performance differences between less proficient and more proficient students. Students who scored below 26 on the 4,000-word and 5,000-word levels and had an overall score below 130 were classified as beginners, as their vocabulary size did not meet the typical intermediate learner criteria. Based on these criteria, 24 participants were classified as beginners, while the remaining 47 were categorized as intermediate learners. Table 3 compares the performance of these two groups.

Table 3. Comparison of Groups' Performances across Variables

| Variables | Beginner Group ^a | | Intermediate Group ^b | |
|--------------------------|-----------------------------|--------|---------------------------------|--------|
| | M | SD | M | SD |
| 1,000-word families | 28.08 | 2.06 | 29.79 | .46 |
| 2,000-word families | 25.33 | 4.35 | 29.13 | .92 |
| 3,000-word families | 21.50 | 5.33 | 27.89 | 1.95 |
| 4,000-word families | 22.62 | 4.33 | 28.06 | 1.71 |
| 5,000-word families | 21.41 | 6.30 | 28.63 | 1.70 |
| Overall Score | 118.96 | 18.10 | 143.51 | 4.51 |
| Number of Words | 587.33 | 348.79 | 619.74 | 209.41 |
| Percentage of Plagiarism | 6.87 | 15.70 | 13.95 | 21.37 |

Note: a) $n = 24$; b) $n = 47$

Table 3 shows that students in the beginner group had lower scores across all VLT levels than those in the intermediate group. The disparity was particularly evident in the higher levels of VLT (3,000 to 5,000-word families), where the intermediate group consistently outperformed the beginner group. Interestingly, despite the differences in vocabulary mastery, the amount of plagiarized content did not show a statistically significant difference between the two groups. The average plagiarized content was higher in the intermediate group ($M = 13.95$, $SD = 21.37$) compared to the beginner group ($M = 6.87$, $SD = 15.70$). However, this difference was not significant ($t(69) = -1.44$, $p = .156$, 95% CI [-16.92, 2.76]), meaning no significant correlation between vocabulary mastery at any level and the percentage of plagiarism. This result suggests that vocabulary mastery, while crucial for overall language proficiency, may not be a key factor in determining whether a student plagiarizes.

Discussion

Our findings indicate a weak and insignificant correlation between the participants' vocabulary mastery, ranging from 1,000-word to 5,000-word families, and the plagiarized content in their argumentative essays. In other words, vocabulary proficiency at the beginner and intermediate levels may not significantly influence plagiaristic behavior. This result contrasts with [Gardner and Lambert \(1972\)](#), who reported a strong, negative, and significant correlation between vocabulary size and plagiarism.

There are two possible reasons for these discrepancies. First, in our study, the participants received instruction on the types of plagiarism and learned how to paraphrase and cite sources properly. Given the low average of plagiarism found in their essays, they likely effectively applied these techniques. [Gardner and Lambert \(1972\)](#) also found that when students are given proper instructions on how to avoid plagiarism, the average percentage of plagiarized content in their essays becomes low. In our study, this relatively low percentage of plagiarism might have influenced the results, leading to a weak and insignificant correlation between vocabulary size and the percentage of plagiarism in the students' texts.

The second reason is that the students submitted their essays in the eighth week after receiving peer feedback in the seventh week. This additional time allowed them to improve their essays and address any potential plagiarism issues identified by themselves or their peers, thus enhancing the quality of their submissions and minimizing plagiarism. While these explanations are plausible, we could not verify their accuracy as we did not interview the students in this study.

In addition to exploring the relationship between vocabulary mastery and plagiarism, we aimed to determine whether vocabulary proficiency among participants would result in varying levels of plagiarism in their essays. As previously shown in Table 3, our findings reveal clear differences in vocabulary mastery between the beginner and intermediate groups. However, regarding plagiarism, no significant difference was observed between the two groups.

This lack of significant difference suggests that vocabulary proficiency alone may not be a decisive factor in determining the extent of plagiarism in students' essays. In fact, our findings indirectly point to the multifaceted nature of plagiarism, as proposed by [Gardner and Lambert \(1972\)](#).

Plagiarism cannot be attributed solely to linguistic incompetence; many factors may contribute to producing plagiarized content. It also is possible that both beginner and intermediate students faced similar challenges and pressures, leading them to engage in plagiarism regardless of their vocabulary proficiency.

[Gardner and Lambert \(1972\)](#) argue that psychological factors such as competition, lack of confidence, impulsive tendencies, and biased cognitions play a significant role in plagiarism. Unfortunately, we could not explore this possibility further as we did not interview our participants. Future research should consider these psychological dimensions, alongside linguistic proficiency, to better understand the complex motivations behind plagiarism. By

doing so, we can develop more comprehensive strategies to combat academic dishonesty and promote integrity in educational settings.

Conclusion

In conclusion, our findings indicated that while there are differences in vocabulary mastery between beginner and intermediate groups, these differences do not significantly affect the amount of plagiarized content in their essays. The lack of significant correlation suggests that factors other than vocabulary mastery may be more influential in determining plagiarism behaviors.

These results highlight the need for further research into the multifaceted nature of plagiarism, considering factors such as cultural influences, academic pressure, and access to information. Additionally, educational institutions should continue to focus on teaching proper citation practices and raising awareness about academic integrity to mitigate plagiarism.

Future studies could explore other linguistic components, such as syntactic complexity and writing fluency, to better understand their relationship with plagiarism. Additionally, longitudinal studies could provide insights into how students' attitudes towards plagiarism evolve over time and the effectiveness of educational interventions in reducing plagiarism.

References

- Adel, G. M. A., & Wang, Y. (2019). Effectiveness Level of Online Plagiarism Detection Tools in Arabic. *Internet of Things and Cloud Computing*, 7(1), 19. <https://doi.org/10.11648/j.iotcc.20190701.13>
- Ahn, S. (2020). Borrowing Words from a Source in Second-Language Writing. *Studies in Modern Grammar*, 105, 141–165. <https://doi.org/10.14342/smog.2020.105.141>
- Akbar, A., & Picard, M. (2019). Understanding plagiarism in Indonesia from the lens of plagiarism policy: lessons for universities. *International Journal for Educational Integrity*, 15(1), 7. <https://doi.org/10.1007/s40979-019-0044-2>
- Alvi, F., Stevenson, M., & Clough, P. (2021). Paraphrase type identification for plagiarism detection using contexts and word embeddings. *International Journal of Educational Technology in Higher Education*, 18(1), 42. <https://doi.org/10.1186/s41239-021-00277-8>
- Anil, A., Saravanan, A., Singh, S., Shamim, M. A., Tiwari, K., Lal, H., Seshatri, S., Gomaz, S. B., Karat, T. P., Dwivedi, P., Varthya, S. B., Kaur, R. J., Satapathy, P., Padhi, B. K., Gaidhane, S., Patil, M., Khatib, M. N., Barboza, J. J., & Sah, R. (2023). Are paid tools worth the cost? A prospective cross-over study to find the right tool for plagiarism detection. *Heliyon*, 9(9), e19194. <https://doi.org/10.1016/j.heliyon.2023.e19194>
- Ayton, D., Hillman, C., Hatzikiriakidis, K., Tsindos, T., Sadasivan, S., Maloney, S., Bragge, P., Diug, B., & Illic, D. (2022). Why do students plagiarise? Informing higher education teaching and learning policy and practice. *Studies in Higher Education*, 47(9), 1921–1934. <https://doi.org/10.1080/03075079.2021.1985103>
- Chien, S.-C. (2014). Cultural Constructions of Plagiarism in Student Writing: Teachers' Perceptions and Responses. *Research in the Teaching of English*, 49(2), 120–140. <https://doi.org/10.58680/rte201426160>
- Cong-Lem, N., Tran, T. N., & Nguyen, T. T. (2024). Academic Integrity In The Age Of Generative AI: Perceptions And Responses Of Vietnamese EFL Teachers. *Teaching English With Technology*, 2024(1), 28–47. <https://doi.org/10.56297/FSYB3031/MXNB7567>
- Dodigovic, M., Mlynarski, J., & Wei, R. (2016). Vocabulary Size Assessment as a Predictor of Plagiarism. In C. Gitsaki & C. Coombe (Eds.), *Current Issues in Language Evaluation, Assessment and Testing: Research and Practice* (pp. 222–235). Cambridge Scholars Publishing. <https://www.researchgate.net/publication/322617919>

- Doss, D. A., Henley, R., Gokaraju, B., Mcelreath, D., Lackey, H., Hong, Q., & Miller, L. (2016). Assessing Domestic vs. International Student Perceptions and Attitudes of Plagiarism. *Journal of International Students*, 6(2), 542–565. <http://jistudents.org/>
- Ehrich, J., Howard, S. J., Mu, C., & Bokosmaty, S. (2016). A comparison of Chinese and Australian university students' attitudes towards plagiarism. *Studies in Higher Education*, 41(2), 231–246. <https://doi.org/10.1080/03075079.2014.927850>
- Erguvan, I. D. (2022). An attempt to understand plagiarism in Kuwait through a psychometrically sound instrument. *International Journal for Educational Integrity*, 18(1), 26. <https://doi.org/10.1007/s40979-022-00120-1>
- Hartono, D. A., & Prima, S. A. B. (2023). Plagiarism in the Indonesian Higher Education: A Study of Non-English Major Undergraduates. *ELTIN Journal: Journal of English Language Teaching in Indonesia*, 11(1), 39–46. <https://doi.org/https://doi.org/10.22460/eltin.v11i1.p39-46>
- Hopp, C., & Speil, A. (2021). How prevalent is plagiarism among college students? Anonymity preserving evidence from Austrian undergraduates. *Accountability in Research*, 28(3), 133–148. <https://doi.org/10.1080/08989621.2020.1804880>
- Ibrahim, K. (2023). Using AI-based detectors to control AI-assisted plagiarism in ESL writing: “The Terminator Versus the Machines.” *Language Testing in Asia*, 13(1), 46. <https://doi.org/10.1186/s40468-023-00260-2>
- Jereb, E., Perc, M., Lämmlein, B., Jerebic, J., Urh, M., Podbregar, I., & Šprajc, P. (2018). Factors influencing plagiarism in higher education: A comparison of German and Slovene students. *PLOS ONE*, 13(8), e0202252. <https://doi.org/10.1371/journal.pone.0202252>
- Kayaoğlu, M. N., Erbay, Ş., Flitner, C., & Saltaş, D. (2016). Examining students' perceptions of plagiarism: A cross-cultural study at tertiary level. *Journal of Further and Higher Education*, 40(5), 682–705. <https://doi.org/10.1080/0309877X.2015.1014320>
- Khathayut, P., Walker-Gleaves, C., & Humble, S. (2022). Using the theory of planned behaviour to understand Thai students' conceptions of plagiarism within their undergraduate programmes in higher education. *Studies in Higher Education*, 47(2), 394–411. <https://doi.org/10.1080/03075079.2020.1750584>
- LaFlair, G. T., Egbert, J., & Plonsky, L. (2015). A Practical Guide to Bootstrapping Descriptive Statistics, Correlations, T Tests, and ANOVAs. In L. Plonsky (Ed.), *Advancing Quantitative Methods in Second Language Research* (1st ed.). Routledge.
- Larson-Hall, J. (2015). *A Guide to Doing Statistics in Second Language Research Using SPSS and R*. Routledge. <https://doi.org/10.4324/9781315775661>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learning Environments*, 11(1), 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Mireku, D. O., Dzamesi, P. D., & Bervell, B. (2024). Plagiarism in Higher Education (PLAGiHE) within Sub-Saharan Africa: A systematic review of a decade (2012–2022) literature. *Research Ethics*, 20(2), 156–186. <https://doi.org/10.1177/17470161231189646>
- Mohd Habali, A. H., & Fong, L. L. (2016). Plagiarism in Academic Writing Among TESL Postgraduate Students: A Case Study. In *7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings* (pp. 729–740). Springer Singapore. https://doi.org/10.1007/978-981-287-664-5_57
- Moss, S. A., White, B., & Lee, J. (2018). A Systematic Review Into the Psychological Causes and Correlates of Plagiarism. *Ethics & Behavior*, 28(4), 261–283. <https://doi.org/10.1080/10508422.2017.1341837>
- Muluk, S., Yanis, F. R., Dahliana, S., & Amiruddin, A. (2021). Scrutinizing EFL students' plagiarism practice. *Englisia: Journal of Language, Education, and Humanities*, 9(1), 145. <https://doi.org/10.22373/ej.v9i1.10492>
- Nation, P., & Meara, P. (2010). Vocabulary. In N. Schmitt (Ed.), *An Introduction to Applied Linguistics* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203783733>

- Nketsiah, I., Imoro, O., Barfi, K. A., Amoah, E., & Amenorvi, C. R. (2023). Academic integrity: Do socio-demographic differences in perception and awareness of plagiarism matter? *Cogent Education*, 10(2). <https://doi.org/10.1080/2331186X.2023.2273175>
- Nushi, M., & Firoozkahi, A. H. (2017). Plagiarism policies in Iranian university TEFL teachers' syllabuses: an exploratory study. *International Journal for Educational Integrity*, 13(1), 12. <https://doi.org/10.1007/s40979-017-0023-4>
- Patak, A. A., Wirawan, H., Abduh, A., Hidayat, R., Iskandar, I., & Dirawan, G. D. (2021). Teaching English as a Foreign Language in Indonesia: University Lecturers' Views on Plagiarism. *Journal of Academic Ethics*, 19(4), 571–587. <https://doi.org/10.1007/s10805-020-09385-y>
- Pecorari, D., & Petrić, B. (2014). Plagiarism in second-language writing. *Language Teaching*, 47(3), 269–302. <https://doi.org/10.1017/S0261444814000056>
- Plonsky, L. (2015). *Advancing Quantitative Methods in Second Language Research* (L. Plonsky, Ed.). Routledge. <https://doi.org/10.4324/9781315870908>
- Plonsky, L., Egbert, J., & Laflair, G. T. (2014). Bootstrapping in Applied Linguistics: Assessing its Potential Using Shared Data. *Applied Linguistics*, 36(5), 1–21. <https://doi.org/10.1093/applin/amu001>
- Ratna, A. A. P., Purnamasari, P. D., Adhi, B. A., Ekadiyanto, F. A., Salman, M., Mardiyah, M., & Winata, D. J. (2017). Cross-Language Plagiarism Detection System Using Latent Semantic Analysis and Learning Vector Quantization. *Algorithms*, 10(2), 69. <https://doi.org/10.3390/a10020069>
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/S0261444812000018>
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88. <https://doi.org/10.1177/026553220101800103>
- Shang, H.-F. (2019). An investigation of plagiarism software use and awareness training on English as a foreign language (EFL) students. *Journal of Computing in Higher Education*, 31(1), 105–120. <https://doi.org/10.1007/s12528-018-9193-1>
- Shi, L. (2006). Cultural Backgrounds and Textual Appropriation. *Language Awareness*, 15(4), 264–282. <https://doi.org/10.2167/la406.0>
- Sulaiman, R. (2018). TYPES AND FACTORS CAUSING PLAGIARISM IN PAPERS OF ENGLISH EDUCATION STUDENTS. *Journal of English Education*, 3(1), 17–22. <https://doi.org/10.31327/jee.v3i1.471>
- Tran, M. N., Hogg, L., & Marshall, S. (2022). Understanding postgraduate students' perceptions of plagiarism: a case study of Vietnamese and local students in New Zealand. *International Journal for Educational Integrity*, 18(1), 3. <https://doi.org/10.1007/s40979-021-00098-2>
- Tran, U. T., Huynh, T., & Nguyen, H. T. T. (2018). Academic Integrity in Higher Education: The Case of Plagiarism of Graduation Reports by Undergraduate Seniors in Vietnam. *Journal of Academic Ethics*, 16(1), 61–69. <https://doi.org/10.1007/s10805-017-9279-9>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL - International Journal of Applied Linguistics*, 168(1), 33–69. <https://doi.org/10.1075/itl.168.1.02web>
- Wilkins, D. A. (1972). *Linguistics In Language Teaching*. MIT Press.