Journal of Language and Pragmatics Studies, Volume 4 Number 1 (Apr 2025), p. 38-45
e-issn 2984-6051
DOI: https://doi.org/10.58881/jlps.v3i2
https://jurnal.ympn2.or.id/index.php/JLPS

Designing a computer-based assessment system for evaluating writing proficiency in the Indonesian language

Sri Kusuma Winahyu[1], Endry Boeriswati[2], Fathiaty Murtadho[3], Fairul Zabadi[4]

Universitas Negeri Jakarta, Indonesia[1,2,3]
Badan Riset dan Inovasi Nasional, Indonesia[4]
[1]Email:
SriKusumaWinahyu_9906917033@mhs.unj.ac.id

Abstract - Writing is a fundamental language skill that students must master in school-based learning. This study aims to develop a computer-based assessment system for evaluating Indonesian opinion articles, addressing the challenges teachers face in manual writing assessments. Utilizing the initial phases of research and development (R&D) with the ADDIE model, this study conducted a needs analysis through literature reviews, interviews, and questionnaires. Findings indicate the necessity of a computer-based writing assessment system incorporating both machine and human raters. The machine rater evaluates mechanical aspects and vocabulary (word count) using pre-processing techniques in Natural Language Processing (NLP), supported by an Indonesian vocabulary database and punctuation programming. Meanwhile, the human rater, an Indonesian language teacher, conducts assessments via an interactive interface. Testing by four teachers on 40 students revealed that 97.6% of teachers responded positively to the system's assessment process. This research is particularly relevant in the post-COVID-19 era, highlighting the positive role of technology in advancing language education and information technology. The system can be adapted for assessing other language skills, such as reading, and for large-scale applications like the Indonesian language proficiency test.

Keywords: computer-based test; human rater; machine rater; natural language processing; opinion article

1. Introduction

Writing is an essential language skill that students must master in educational settings, alongside listening, reading, and speaking. Within the Indonesian education curriculum, writing is emphasized across all grade levels. For instance, students in grade X learn to produce various texts such as anecdotal, exposition, and complex procedural texts, while grade XI students focus on short stories, rhymes, and reviews of films or dramas. At the grade XII level, students are introduced to writing historical stories, news, advertisements, and opinion articles, which are considered particularly challenging due to their argumentative structure (Wiratno & Santosa, 2019). Opinion articles require students to construct well-structured arguments presenting

Journal of Language and Pragmatics Studies, Volume 4 Number 1 (Apr 2025), p. 38-45
e-issn 2984-6051
DOI: https://doi.org/10.58881/jlps.v3i2
https://jurnal.ympn2.or.id/index.php/JLPS

multiple perspectives on controversial issues, posing a significant challenge in both writing and assessment.

Assessment of opinion articles is particularly complex due to the need to evaluate multiple components, including text structure, argument quality, and linguistic accuracy. This complexity, combined with the time-consuming nature of manual scoring, highlights the need for a more efficient and objective system. The COVID-19 pandemic further accentuated this need, as educators and researchers were compelled to adopt technology-based learning and assessment tools to facilitate distance education. Thus, a computer-based writing assessment system tailored to the Indonesian context is not only timely but also critical for enhancing the quality and efficiency of language education.

The challenges of assessing opinion articles manually underscore the importance of developing an automated system. Manual assessment often leads to inconsistencies and delays, particularly when teachers are tasked with evaluating large volumes of student work. Furthermore, traditional assessment methods lack the scalability and adaptability required to meet the demands of modern education, particularly during crises like the COVID-19 pandemic. Automated assessment systems can provide consistent, objective evaluations while significantly reducing the workload for educators.

In addition to addressing practical challenges, a computer-based assessment system can support the broader goals of language education by promoting self-directed learning. With an automated system, students can receive immediate feedback, enabling them to identify and address their weaknesses independently. Such systems can also be adapted for assessing other language skills, including reading and listening, as well as for conducting large-scale standardized tests such as the Indonesian language proficiency test. Therefore, this study is crucial for advancing both the pedagogical and technological dimensions of language education.

Automated writing assessment has a long history, beginning with the development of essay grading machines in the 1960s. Early systems focused primarily on mechanical aspects of writing, such as grammar and vocabulary, but lacked the capacity to evaluate more nuanced features like rhetorical structure and coherence (Shermis et al., 2010). Subsequent advancements in computational linguistics led to the development of systems capable of assessing text semantics and organization. For example, Latent Semantic Analysis (LSA) has been used to evaluate the semantic similarity between student essays and model answers, achieving accuracy rates exceeding 86% (Ratna et al., 2015).

In the Indonesian context, two notable systems have been developed for automated writing assessment. The first system evaluates the mechanical and lexical aspects of sentences produced by Indonesian language learners, while the second, SIMPLE-O, uses LSA to assess student essays by comparing them to model answers provided by instructors (Ratna et al., 2015). However, both systems have significant limitations. They primarily focus on surface-level features and fail to account for rhetorical and linguistic nuances, making them unsuitable for assessing opinion articles in a high school setting. This study seeks to address these gaps by developing a comprehensive computer-based writing assessment system tailored to the unique requirements of Indonesian language education.

What constraints are faced by Indonesian language teachers when manually assessing opinion articles written by grade XII students? How can a computer-based writing assessment system be designed to address these constraints and meet the needs of Indonesian language teachers?

This study aims to explore the challenges faced by teachers in manually assessing opinion articles and to develop a computer-based assessment system tailored to the Indonesian language curriculum. The system will integrate both machine and human ratters to ensure accurate and comprehensive evaluations of student writing. By addressing the identified constraints, the proposed system seeks to enhance the efficiency and objectivity of writing assessments while supporting the broader goals of language education in Indonesia.

Journal of Language and Pragmatics Studies, Volume 4 Number 1 (Apr 2025), p. 38-45
e-issn 2984-6051
DOI: https://doi.org/10.58881/jlps.v3i2
https://jurnal.ympn2.or.id/index.php/JLPS

2. Method

As mentioned above, in this study, researchers used an R&D method by applying ADDIE model (Nada, 2015). Development research in the scope of education is an attempt to develop an effective product for school use, and not to test theory. The goal of development research is to develop innovative prototypes. In this case, the researcher tried to explore the needs of Indonesian language teachers for a computer-based writing assessment system and its components based on their experience when assessing student opinion articles manually. That experience is as real as it appears to the teacher. The results of extracting information from the teacher were manifested in the development of an assessment prototype for writing computer-based Indonesian opinion articles.

To answer the problems or research questions and achieve the above objectives, the researcher designed the study and assigned the sample. According to Sugiyono (2019), way to select samples in research is to do it purposively, that is, to be selected based on certain considerations and goals. In this case, the selection of a sample of grade XII Indonesian language teachers was carried out with the consideration that opinion articles were produced by grade XII students and the teachers acted as human rater of the articles produced by these students. On that basis, the sample set was four Indonesian teachers of grade XII, namely three from Jakarta and its surroundings and one from North Sulawesi. Later during the trial, each teacher involved his student with a total of 40 students.

In collecting data through a literature study, the researcher conducted a literature study related to the basic competencies of grade XII of high school students. In addition to going through the literature study, to explore the teacher's need for the importance of developing computer-based writing assessment tools, the researcher conducted interviews and gave questionnaires to four teachers.

In the interviews, the researcher intends to explore the needs of Indonesian language teachers for computer-based writing assessment tools. The data were obtained by asking questions related to teacher's experiences, opinions, feelings, and knowledge regarding the manual assessment of writing opinion articles by students as well as questions about their background. Researchers conducted semi-structured interviews, by listening to and recording what the informants said based on the questions above. Then, the data were analyzed by exploring the answers one by one from the results of the interviews with the teacher and using them as a component to build an assessment system. The results of the interview were studied by researcher so that researcher found the need for an assessment form to write opinion articles. While the data analysis was carried out on questionnaires that had been filled in by the teachers to find out the teacher's opinion about the assessment components in the application to be developed.

3. Results and Discussion

Literature study, which is one of the steps for exploring data in this research, yields information that in the Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 24 of 2016 there are basic competencies for grade XII of high school in the 2013 Curriculum, that students are required to be able to compose opinions in the form of articles such as in magazines. Meanwhile, the obstacles faced by teachers in conducting a manual assessment of student opinion articles based on the teacher's answers in the interview are: 1) student statements or thesis are not well structured if the teacher does not provide trigger questions; 2) teachers do not always use the assessment components completely; 3) teachers often find it difficult to complete assessments quickly; and 4) teachers sometimes find it difficult to be

Journal of Language and Pragmatics Studies, Volume 4 Number 1 (Apr 2025), p. 38-45
e-issn 2984-6051
DOI: https://doi.org/10.58881/jlps.v3i2
https://jurnal.ympn2.or.id/index.php/JLPS

objective. These constrain are then processed in data analysis so that it appears the teacher's needs for computer-based writing assessment.

To further clarify what components of the assessment will be included in the assessment instrument, the researcher gave a questionnaire to the teachers who had been previously been interviewed. The questionnaire instrument provided contains the language assessment components that will be included in the assessment application instrument for writing computer-based Indonesian opinion articles. In addition, the questionnaire instrument also contains an application display plan from the information technology side.

The results of the need analysis (literature study, interviews, and questionnaires) were then transformed into a design for developing a computer-based writing assessment program. Therefore, to design the computer-based opinion article writing assessment system, the researchers added: 1) trigger questions; 2) completeness of the assessment component; and 3) scoring system conducted by human rater (guided by an assessment rubric) and machine rater in CBT (WBT) format. Trigger questions are provided to overcome student difficulties in compiling a statement or thesis; completeness of the assessment component is provided so that students get a complete assessment; and in general, the use of the assessment rubric and machine rater in the CBT system is carried out to reduce the possibility of the subjectivity of the assessment.

The manifestation of the teacher's needs into the system is also accompanied by other features in the system. Referring to the first problem: the unstructured student thesis in the resulting article, related to the completeness of the thesis or argument as the main element in an article. This argument should be structured by a thesis supported by data (Toulmin, 2003). Complete arguments are not only supported by data that are in line with the thesis but can also be followed by data that invalidate the thesis. For this reason, in addition to realizing the trigger questions, the assessment system includes a thesis statement and data as part of the content assessment carried out by human rater.

Then, the second problem: the completeness of assessment components, Brown (2003) mentions written components, such as content, organization, discourse, syntax, vocabulary, and mechanics, are applied in this system. In this case, mechanics and vocabulary (word count) were corrected by a machine rater, while vocabulary (word use and word formation), discourse content, organization, and syntax were corrected by a human rater. The application of the five assessment components also distinguishes this system from the two systems mentioned in the introduction section (Ratna et al., 2015). Assessment by machines will automatically be completed earlier when compared to human raters. However, processing to the final grade is carried out simultaneously after the human rater has finished assessing.

The last problem related to teacher objectivity, overcome by using the CBT system with two raters: machines rater and humans' rater (Brown, 2003). Regarding this CBT, Sugiyono et al. (2019) analyzed its use for the implementation of tests at the high school level and the results were very good, 83.34% for product validation tests, 100% for operator user tests, and 88.94% for user tests. Another familiar computational system for evaluating language learning, both tests, and assessments, is the Computer Adaptive Test (CAT). However, CAT is more suitable for determining the level of the test taker because the questions provided will adjust the level of the test taker. If successful at a certain level, test participants will automatically be faced with questions with an increasing level of difficulty (Rezaie & Golshan, 2015). Therefore, CAT does not fit the teacher's needs in the context of this study

The following is an assessment design for writing computer-based opinion articles in the Indonesian language according to the needs of Indonesian language teachers, which is in the form of CBT (Figure 1). In Figure 1, the way the system works is initiated by students who input writing in the form of opinion articles based on trigger questions that have been previously inputted by the teacher. Students write opinion articles in separate columns. Upon completion, the system will conduct an assessment with two raters: machine rater and human rater (Indonesian language teacher). Human rater assesses content, organization, syntax, and discourse using interface

Journal of Language and Pragmatics Studies, Volume 4 Number 1 (Apr 2025), p. 38-45
e-issn 2984-6051
DOI: https://doi.org/10.58881/jlps.v3i2
https://jurnal.ympn2.or.id/index.php/JLPS

assessments. This interface assessment was adapted from a writing test assessment conducted by the Test of English for International Communication (TOEIC) organized by the Educational Testing Service (ETS) (Everson and Hines 2010). The content consists of the thesis and data, as explained above that the thesis and the data supporting it are the main requirements for the content of the opinion article (Toulmin, 2003). The organizational component is prepared for teachers to assess the opening paragraphs, content paragraphs, and closing paragraphs (Deane, 2011). The discourse component consists of cohesion and topic continuity (Halliday & Christian, 2014), while a syntactic component is available to assess sentence structure (Moeliono et al., 2017). Each component assessed by a human rater is presented in a template that allows the teacher to work fairly and objectively because the assessment components are available in ranges and weights (Brown, 2003).

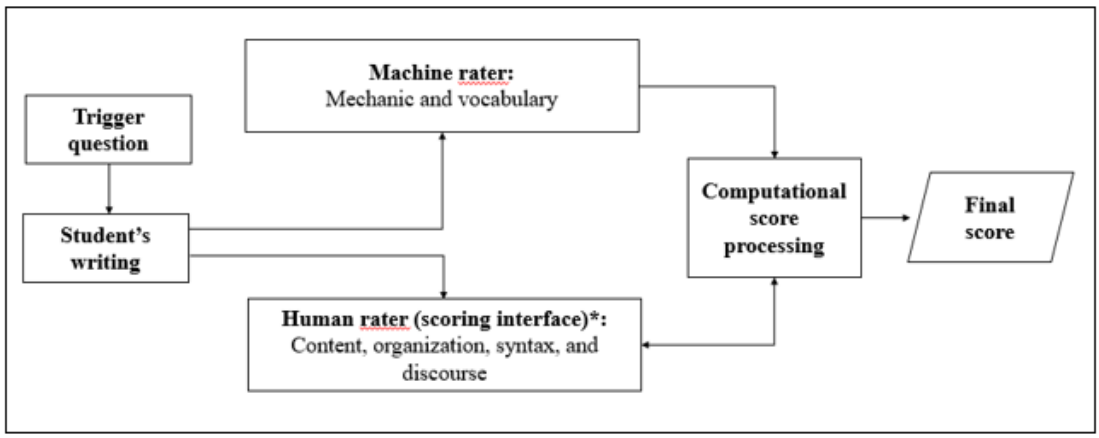

Figure 1. Computer-Based Writing Assessment Design

Machine raters apply to preprocess steps in natural language processing (NLP) to assess mechanics and vocabulary (Gelbukh, 2014; Hyland, 2009). These steps are input text editor, sentence segmentation, case folding, tokenizing, stemming, and stop-words removal (Figure 2). Sentence segmentation is the process of separating sentences and case folding is the process of converting all letters to lowercase, removing irrelevant numbers, punctuation marks, and empty characters. Tokenizing is the process of separating the text into chunks known as tokens. Tokens can be words, numbers, symbols, punctuation marks, and other important entities. Tokenizing the word means to separate the words in a sentence. Tokenizing is done after case-folding so that the sentence does not contain punctuation, capital letters, empty characters, and unnecessary numbers. Stemming is the process of changing affixed words into basic words, filtering or stop-words removal is the stage of taking important words from the token results using a stop-list algorithm (removing less important words) or a wordlist (storing important words) (Nugroho, 2019). In this study, the preprocessing steps in NLP (Figure 2) were used as the basis for calculating the number of words produced by students as well as detecting the accuracy of using punctuation marks and word formation. To support the machine assessment process, the system has a database in the form of a list of Indonesian language vocabulary. In addition, the system also performs programming related to spelling and punctuation.

Journal of Language and Pragmatics Studies, Volume 4 Number 1 (Apr 2025), p. 38-45
e-issn 2984-6051
DOI: https://doi.org/10.58881/jlps.v3i2
https://jurnal.ympn2.or.id/index.php/JLPS

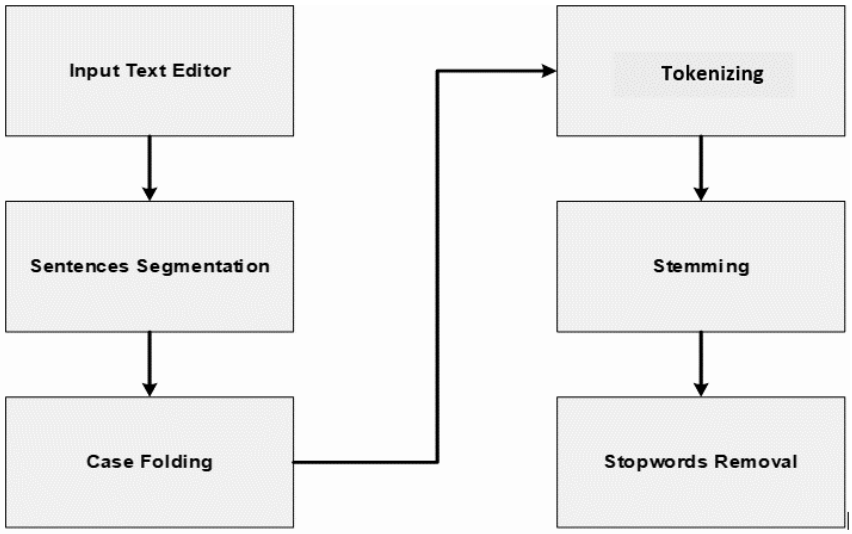

Figure 2. NLP Preprocessing Stage for the Assessment of Mechanics and Vocabulary

After the development stage, prior to the evaluation stage in the form of a trial run, validation is carried out in advance by the material expert validator and information technology expert validator. The instrument used to assess the product being developed is a Likert scale questionnaire with answer categories in the form of choices consisting of: very good, good, enough, less, and very poor with weight 5, 4, 3, 2, 1 respectively. Meanwhile the instrument used by teacher to assess the product after the trial was a Gutmann scale questionnaire with the answer categories being Yes (1) or No (0).

The results of the material expert validator are feasible (75.71%) and very feasible (82.22%). While the result of the analysis of the questionnaire data from the trial results showed that analysis showed that 96.76% of the assessment application for writing Indonesian language opinion article and the assessment components (rubric) in it is approved by the teacher (Figure 3).

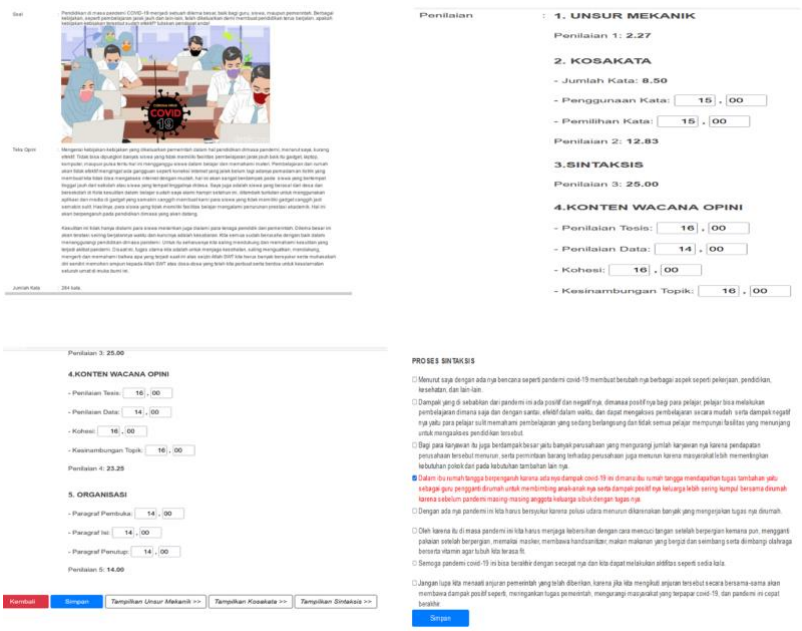

Figure 3. Examples of student writing results and assessment by human rater and machine rater.

Journal of Language and Pragmatics Studies, Volume 4 Number 1 (Apr 2025), p. 38-45
e-issn 2984-6051
DOI: https://doi.org/10.58881/jlps.v3i2
https://jurnal.ympn2.or.id/index.php/JLPS

4. Conclusion

The development of a computer-based writing assessment system in the Indonesian language, specifically designed to evaluate opinion articles written by Grade XII students, represents a significant step forward in addressing the needs of Indonesian language teachers. This system integrates innovative features to overcome the limitations inherent in traditional manual assessment methods, making it a highly relevant and practical tool for both educators and students.

At the core of this system are several key components that enhance its functionality and usability. The first feature is the inclusion of originating questions, which are tailored to prompt critical thinking and coherent expression of ideas in opinion articles. These questions are designed to engage students in a way that stimulates creativity and logical reasoning, which are essential aspects of effective writing. Additionally, the system employs five distinct components of writing assessment—organization, content, language use, mechanics, and style—to provide a comprehensive evaluation of students' writing abilities. This multi-faceted approach ensures a more detailed and accurate assessment of their strengths and areas for improvement.

Another notable feature is the use of Computer-Based Testing (CBT) in the form of Web-Based Testing (WBT). This innovative approach incorporates both machine raters and human raters, blending the efficiency of automated scoring with the nuanced judgment of human evaluators. Machine raters can handle large volumes of submissions quickly, offering preliminary assessments based on predefined criteria. Human raters, on the other hand, add depth and context by addressing elements of writing that require subjective interpretation, such as tone, creativity, and cultural relevance. This hybrid assessment model ensures a balanced and fair evaluation process.

Beyond its immediate use in schools, this computer-based writing assessment system also offers significant benefits for at-home learning. With a stable network connection, students can access the system from their homes, enabling a seamless connection with their teachers. This feature is particularly valuable in contexts where remote learning has become a necessity, such as during the COVID-19 pandemic. By providing continuous access to assessment and feedback, the system supports students in developing their writing skills in a flexible and self-paced manner.

Furthermore, the adaptability of this system extends its potential applications beyond the assessment of writing skills. It can be modified to evaluate other aspects of language learning, such as reading comprehension, listening, and speaking. This flexibility makes it a versatile tool that can address various educational needs within the Indonesian language curriculum. Additionally, the system holds promise for implementation in large-scale testing scenarios, such as the Indonesian language proficiency test. Its ability to handle high volumes of test-takers while maintaining accuracy and consistency positions it as a viable solution for standardized language assessments.

The development of this system also aligns with broader educational goals, such as promoting digital literacy and leveraging technology to enhance learning outcomes. By integrating technology into the assessment process, it prepares students for a future where digital tools play a central role in education and professional life. Teachers, too, benefit from reduced workloads and more efficient assessment processes, allowing them to focus on delivering high-quality instruction and personalized feedback.

In conclusion, the computer-based writing assessment system for the Indonesian language addresses critical challenges faced by teachers and students in evaluating writing skills. Its innovative features, adaptability, and scalability make it a valuable asset for educational institutions. By supporting both classroom and remote learning, and by offering potential applications across various language skills and testing contexts, this system represents a transformative advancement in language education in Indonesia.

Journal of Language and Pragmatics Studies, Volume 4 Number 1 (Apr 2025), p. 38-45
e-issn 2984-6051
DOI: https://doi.org/10.58881/jlps.v3i2
https://jurnal.ympn2.or.id/index.php/JLPS

References

Aldoobie, Nada. (2015). ADDIE Model. *American International Journal of Contemporary Research, 5*(6), December 2015.

Brown, H. D. (2003). *Language Assessment: Principles and Classroom Practices.* California: Longman.

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approach.* California: Sage Publications, Inc.

Deane, P. (2011). Writing Assessment and Cognition. New Jersey: ETS, Princeton.

Everson, P., & Hines, S. (2010). How ETS Scores the TOEIC Speaking and Writing Test Responses. New Jersey: ETS.

Gall, M. D., Borg, W. R., & Gall, J. P. (2003). *Educational Research: An Introduction.* USA: Pearson Education, Inc.

Gelbukh, Alexander. (2009). "Computational linguistics and intelligent text processing." Paper presented at The 10th International Conference, CICLing, Mexico City, Mexico, March 2009.

Halliday, Michael, & Mattiessen, Christian. (2014). *Halliday's Introduction to Functional Grammar.* New York: Routledge.

Hamp-Lyons, L. (2014). Writing assessment in the 21st century: Old, new, and emerging approaches. *Language Testing, 31*(3), 285–296. https://doi.org/10.1177/0265532214530833

Hamp-Lyons, Liz. (2014). Writing Assessment in the Global Context. *Research in The Teaching of English Journal, 48*(3), 353–62.

Hyland, Ken. (2009). *Teaching and Researching Writing* (2nd ed.). New York: Routledge.

Moeliono, A. M., Lapoliwa, Hans, Alwi, Hasan, Sasangka, Sry S. Tjatur, Wisnu, & Sugiyono. (2017). *Tata Bahasa Baku Bahasa Indonesia.* Jakarta: Badan Pengembangan dan Pembinaan Bahasa.

Nugroho, S. K. (2019). "Dasar Teks Preprocessing dengan Phyton." *Medium,* June 18. https://medium.com/@ksnugroho/dasar-text-preprocessing-dengan-python-a4fa52608ffe

Ratna, A. A. P., Purnamasari, P. D., & Adhi, B. A. (2015). SIMPLE-O, the essay grading system for Indonesian language using the LSA method with multi-level keywords. *Conference Proceedings, The Asian Conference on Society, Education & Technology, 2015.*

Ratna, I. W., Santosa, M. H., & Wiratno, T. (2015). SIMPLE-O: An automatic essay grading system for the Indonesian language. *Journal of Computational Linguistics, 21*(4), 412–430.

Rezaie, M., & Golshan, M. (2015). Computer Adaptive Test (CAT): Advantage and limitations. *International Journal of Educational Investigations, 2*(5), 128–37.

Shermis, M. D., Burstein, J., Higgins, D., & Zechner, K. (2010). Automated Essay Scoring: Writing Assessment and Instruction. In *International Encyclopedia of Education* (3rd ed., Vol. 4), edited by Penelope Peterson, Eva Baker, and Barry McGaw, 20–6. Oxford: Elsevier.

Shermis, M. D., Burstein, J. C., & Elliot, S. (2010). Automated essay scoring: A cross-disciplinary perspective. *Assessment in Education, 17*(3), 237–252. https://doi.org/10.1080/0969594X.2010.497523

Sugiyono. (2019). *Metode Penelitian Kuantitatif Kualitatif dan R&D.* Bandung: Alfabeta.

Sugiyono, S., & Rochmadi, T. (2019). Pengembangan Sistem Computer Based Test (CBT) Tingkat Sekolah. *Indonesian Journal of Business Intelligence, 2*(1), 1–8.

Toulmin, S. E. (2003). *The Uses of Argument: Philosophical Studies.* London: Cambridge University Press.

Wiratno, T., & Santosa, R. (2019). *Discourse Analysis and the Teaching of Writing.* Jakarta: Indonesian Language Publishing.

Wiratno, T., & Santosa, R. (2019). *Pengantar Linguistik Umum.* Banten: Universitas Terbuka.

Wiratno, T., & Santosa, R. (2019). Language, Society, and Culture. Surakarta: Universitas Sebelas Maret Press.

Journal of Language and Pragmatics Studies, Volume 4 Number 1 (Apr 2025), p. 25-32
e-issn 2984-6051 DOI: https://doi.org/10.58881/jlps.v3i2
https://jurnal.ympn2.or.id/index.php/JLPS