OPEN ACCESS

European Journal of Teaching and Education
ISSN 2669-0667

Evaluating Student Communication Skills and User Acceptability of an NLP-Based Automated Essay Grading System

Dennis C. Gabon[1*], Albert A. Vinluan[2], and Jennifer T. Carpio[3]

[1] IT Faculty, IT Department, College of Information Technology, Central Bicol State University of Agriculture, Camarines Sur, Philippines

[2] IT Faculty, IT Department, College of Computing Studies, Information and Communication Technology, Isabela State University, Isabela, Philippines

[3] IT Faculty, IT Department, College of Computing Studies, San Beda University, Manila, Philippines

ARTICLE INFO

Keywords:
Automated Essay Grading, Communication Skills, Natural Language Processing, Educational Technology, User Experience

ABSTRACT

This study examined the effectiveness, user experience, and acceptability of a self–developed Natural Language Processing (NLP)-based automated essay grading (AEG) system in evaluating students' communication skills. The system was tested with 35 IT students and 16 IT teachers from Central Bicol State University of Agriculture – Sipocot, College of Information Technology, to assess its reliability and usability. Results showed that the system provided consistent and objective evaluations, with a Cohen's kappa value of 0.797, indicating substantial agreement with human grading. Teachers found the system helpful for reducing their grading workload and appreciated its reliability, while students valued the constructive feedback it provided to improve their writing. Both groups rated the system highly for its accuracy, consistency, and fairness. However, the system showed some limitations in assessing more nuanced aspects of communication, such as coherence and relevance, which should be addressed in future enhancements. Overall, this study highlights the potential of AEG systems to support education by providing reliable, efficient, and actionable assessments of communication skills.

1. Introduction

The use of technology in language learning and assessment has grown significantly, bringing about new opportunities to enhance educational practices. Automated essay grading (AEG) systems, which leverage Natural Language Processing (NLP) and machine learning, are among the tools transforming how writing and communication skills are evaluated. These systems provide immediate, objective feedback on various aspects of essays, such as grammar, organization, and content, while alleviating the heavy workload typically associated with manual grading (Madnani & Cahill, 2018; Hicke, Tian, Jha, & Kim, 2023).

Despite their advantages, AEG systems still face significant challenges. While they perform well in assessing basic language skills, they often need help with more complex tasks, such as

* Corresponding author's E-mail address: denzgabon85@gmail.com, https://orcid.org/0009-0008-0839-3909

Cite this article as:
Gabon, D. C., Vinluan, A. A., & Carpio, J. T. (2025). Evaluating Student Communication Skills and User Acceptability of an NLP-Based Automated Essay Grading System. *European Journal of Teaching and Education, 7*(*3*): 30-40. https://doi.org/10.33422/ejte.v7i3.1531

evaluating the coherence of ideas, the strength of arguments, and the overall persuasiveness of an essay (Ramesh & Sanampudi, 2022). These limitations highlight the need for further refinement to make AEG systems more effective and user-friendly.

At the College of Information Technology at Central Bicol State University of Agriculture (CBSUA), essay-type assessments remain an essential part of the curriculum. These assessments, guided by rubrics, are often paper-based and require significant time and effort from instructors, particularly for open-ended responses. As more educational institutions adopt digital methods, such as computer-based testing, there is a growing need for efficient and reliable systems to support essay grading. This need is underscored by the Commission on Higher Education (CHED) Memorandum Order No. 25, Series of 2015, which emphasizes the importance of essay-type assessments in information technology and computer science programs.

This study aims to evaluate the effectiveness of an NLP-based AEG system in assessing communication skills. It also examines the experiences of students and teachers with the system and explores its acceptability and reliability as a tool for educational assessment. By addressing these areas, the research seeks to contribute to ongoing efforts to enhance the use of technology in education and improve assessment practices in academic institutions.

2. Literature Review

2.1 Effectiveness of Automated Essay Grading in Assessing Communication Skills through Text Coherence

Automated essay grading (AEG) systems have increasingly demonstrated effectiveness in assessing communication skills by leveraging advanced natural language processing (NLP) and machine learning methods. These systems aim to provide consistent, objective, and efficient feedback, a benefit widely recognized in educational contexts where reliable evaluation is critical. Studies reveal that AEG systems utilizing hybrid models, such as those incorporating BERT along with feature engineering, capture both syntactical and semantic dimensions, enhancing the accuracy of communication skill assessments by accounting for complex linguistic interactions within student essays (Prabhu, Kara, & Sanriya, 2022). Transformer-based models contribute to effective grading by generating semantic embeddings enriched with discourse features, leading to more accurate and nuanced scoring (Ait Khayi & Rus, 2024).

Neural network models like Bidirectional Long Short-Term Memory (BI-LSTM) networks have also been instrumental in AEG systems, as they allow for the analysis of logical relationships and semantic coherence, which are essential for evaluating communication proficiency (Ibrahim, Elfakharany, Ramzy, & Hamed, 2022). Additional features, such as Automated Writing Evaluation (AWE) and Grammatical Error Correction (GEC), extend AEG functionality by offering corrective feedback, which has proven beneficial for non-native English speakers aiming to improve both language and communication skills (Wang et al., 2024). AEG systems also support detailed, sentence-level evaluations highlighting specific writing traits, including grammar and coherence, thus contributing to a more comprehensive communication assessment (Hossain & Mustafa, 2023). Recognizing the diversity of student backgrounds, some AEG systems are designed to account for language proficiency and cultural factors, ensuring fair and accurate evaluation across linguistic differences (Poonpon, Manorom, & Chansanam, 2023).

Despite these advancements, AEG technology still faces challenges, particularly in adapting to varied writing styles and maintaining relevance over time, as these systems require

frequent updates and training to sustain accuracy. Nonetheless, AEG continues to emerge as a valuable tool for enhancing assessment practices, offering scalable, equitable opportunities for evaluating communication skills in diverse educational settings worldwide.

2.2 User Experience of Students and Teachers with Automated Grading Systems

The user experience of students and teachers with automated essay grading systems (AEG) is multifaceted, encompassing both positive aspects and challenges. Students generally appreciate the immediate feedback and scoring provided by AEG, which can enhance their writing skills by allowing them to reflect on and improve their work based on the input received (Das et al., 2022). However, there are concerns regarding the transparency and explainability of these systems, as users often need help understanding how scores are derived, which can affect trust and acceptance of the technology (Hall, Seyam, & Dunlap, 2023).

Conversely, teachers find AEG beneficial in reducing the grading workload, especially in large classes, and appreciate the system's ability to provide consistent and objective evaluations (Nguyen, 2017; Lewis, 2013). Despite these advantages, teachers express concerns about the potential loss of control over the grading process and the need for the technology to be advanced enough to replace the nuanced judgment of human graders (Lewis, 2013). Usability issues, such as unfamiliar interfaces and inconsistent design elements, can hinder the overall user experience. Still, user-centered design improvements have significantly improved students' and teachers' effectiveness and satisfaction (Nouriska et al., 2023).

Additionally, while AEG systems like PaperRater have demonstrated acceptable reliability, integrating such tools into educational settings requires careful consideration of pedagogical implications and ongoing research to address usability challenges and improve system transparency (Nguyen, 2017). Overall, while AEG systems offer valuable support in educational contexts, their successful implementation depends on effectively addressing these usability and transparency issues to meet the needs of students and teachers.

2.3 Acceptability and Perceived Reliability of Coherence-Focused Automated Essay Grading Systems

Automated essay grading (AEG) systems have gained significant attention due to their potential to enhance grading efficiency, objectivity, and feedback quality. Advanced machine learning models, such as bi-directional LSTM networks and attention mechanisms, have significantly improved AEG systems' semantic understanding, enabling them to assess essays more effectively (Pradeep & Kowsalya, 2022). These innovations have made AEG systems increasingly comparable to human graders in terms of accuracy and reliability, addressing previous concerns about consistency and fairness (Sharma & Goyal, 2020).

Another significant advantage of AEG systems is their ability to provide immediate feedback, particularly for students aiming to refine their writing skills. These systems foster student engagement and encourage self-assessment by highlighting strengths and pinpointing areas for improvement (Vanga et al., 2023). This aligns with their perceived usability and educational impact, which enhances their acceptance among students.

Despite these advancements, challenges remain. Limitations in evaluating content coherence and relevance, critical for comprehensive grading, have been identified as areas needing improvement (Ramesh & Sanampudi, 2022). Addressing these issues is essential to ensure the broader acceptance of AEG systems in educational settings. These techniques help

mitigate the subjectivity of manual grading while maintaining alignment with educational standards.

3. Materials and Methods

This study utilized a descriptive-evaluative research design. A descriptive method was used to describe the level of communication skills of the students assessed using an automated essay grading system and their experience using it. On the other hand, the evaluative method was used to evaluate the system's effectiveness by identifying its reliability and acceptability.

3.1 Data Collection

This study involved 35 Bachelor of Science in Information Technology students and 16 Information Technology teachers, who were randomly selected to participate in the evaluation of the Automated Essay Grading (AEG) system. The data collection process utilized a survey questionnaire divided into two sections: the first focused on the respondents' experience using the AEG system, while the second gathered feedback on the system's perceived acceptability. The survey aimed to capture both the user experience and the overall acceptance of the system among students and teachers.

To ensure the reliability of the gathered data, the questionnaire was subjected to a reliability test using Cronbach's Alpha, which yielded good to excellent internal consistency. The data analysis was conducted using the Statistical Package for Social Science (SPSS) version 27.0 software, applying statistical tools such as the mean and standard deviation to determine the experience and acceptability of the system. Additionally, Weighted Cohen's Kappa was used to assess the reliability of the system and evaluate the level of agreement between the responses from teachers and students.

While the sample size provided valuable insights, it is important to acknowledge that the relatively small number of participants may limit the generalizability of the findings. Expanding the sample size in future studies to include participants from different institutions would provide a broader perspective on the system's effectiveness across various educational contexts.

Additionally, as the participants were all from CBSUA as an institution, the results may be influenced by specific teaching styles or student backgrounds. A more diverse group in future research would help assess the system's applicability and reliability across various educational settings.

The AEG system, used to evaluate students' communication skills, focused on key aspects such as organization, coherence, and thematic relevance. The system begins by preprocessing the student-submitted essays, which involves tokenizing and cleaning the text to ensure consistency. It then extracts features essential for evaluating communication skills, including the structural organization of ideas, logical progression, and relevance to the assigned topic.

The AEG system was developed using a combination of modern technologies. The front end was built with the Angular framework, creating an interactive user interface, while the backend API was developed using Laravel, a PHP framework. Python was utilized for natural language processing tasks, employing libraries such as NLTK, spaCy, sci-kit-learn, and language-tool-python. These technologies enabled the system to efficiently process, analyze, and evaluate the essays in an automated manner, providing consistent and objective assessments.

While the study yielded promising results, the training data set used to develop the system was based on specific essay types from CBSUA students. This may limit the system's ability

to generalize across different essay types or writing styles, which could be addressed in future studies. Expanding the range of essay prompts and participant diversity would help enhance the system's versatility and improve its ability to assess a variety of writing styles and complexities.

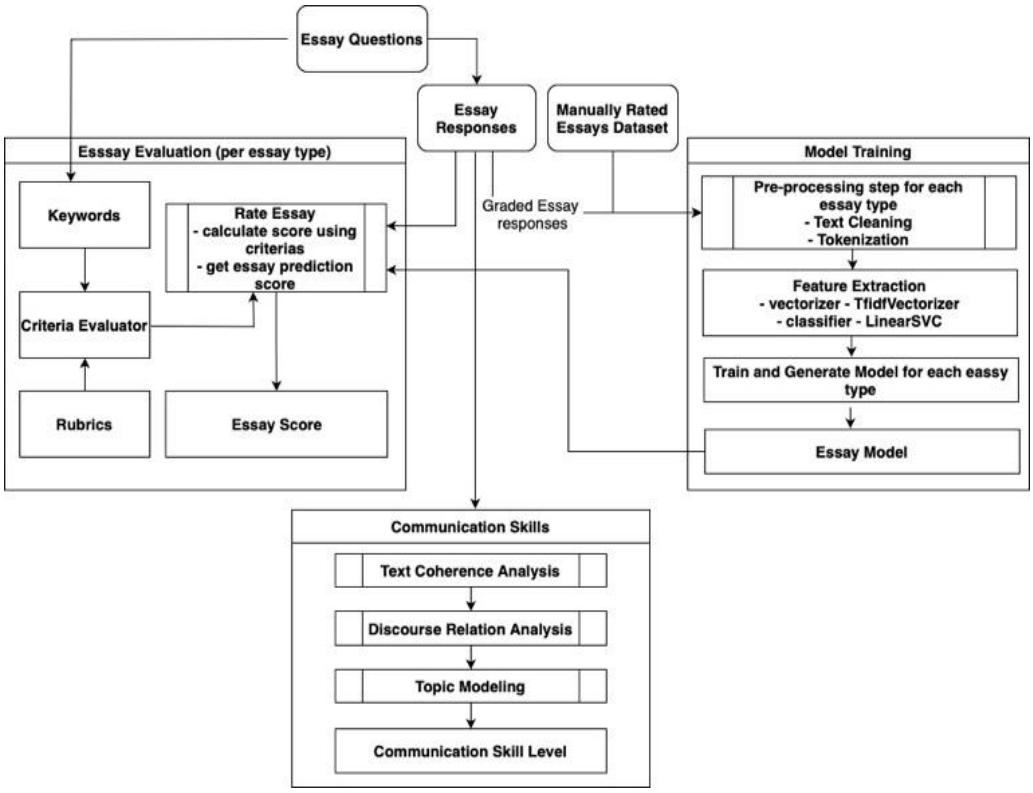

Figure 1. System Architecture

The scoring process, depicted in Figure 1, utilizes machine learning models trained on manually graded essays. These models evaluate the coherence of text, the connections between different parts of the essay, and the thematic consistency of the content. Based on these analyses, the system generates scores reflecting student communication skills. Additionally, the system provides feedback highlighting a prompt that the student did not achieve 50% of the grade or rating in each submitted essay. This workflow ensures that assessments are reliable and actionable, aligning with academic standards and supporting the educational objectives of both students and instructors.

3.2 Model Training and Evaluation

Custom datasets were collected for each essay type, with 175 responses per type. The responses were divided into training (70%) and testing (30%) sets to evaluate the model's accuracy. This approach ensured the system could generalize to new essays while maintaining its evaluation capabilities.

However, the dataset, which was based on specific prompts, limits the model's ability to generalize to other essay types. Future studies should include various essay types to test the system's versatility. Additionally, while the model showed substantial agreement with human grading (Cohen's kappa of 0.797), assessing more complex aspects of communication, such as coherence and argument strength, remains challenging. Future improvements should enhance the system's ability to evaluate these nuanced writing elements.

The AEG system was developed using Angular for the front end, Laravel for the backend API, and Python for natural language processing tasks, utilizing libraries like NLTK, spaCy,

and sci-kit-learn. These technologies enable the system to process and evaluate essays efficiently and consistently.

4. Results and Discussion

4.1 Effectiveness of the Automated Essay Grading System in Assessing Communication Skills

Table 1 presents the respondents' assessed communication skills and the reliability of the automated essay grading (AEG) system. The data reveal that most respondents' communication skills ranged from beginning (34%) to developing (20%), with 43% at the proficient level and only 3% classified as advanced. Regarding reliability, Cohen's kappa (κ) value of 0.797 indicates substantial agreement in the system's communication skills assessment. This level of agreement suggests that the AEG system provides reliable evaluations, with a p-value of less than 0.001 confirming the statistical significance of this reliability.

These findings align with previous studies showing that AEG systems perform reliably in grading complex language skills. For example, Ibrahim, Elfakharany, Ramzy, and Hamed (2022) demonstrated the effectiveness of a deep learning-based AEG system using NLP methods, reporting high reliability in assessing linguistic complexity and coherence. Similarly, Prabhu, Kara, and Sanriya (2022) explored a hybrid model combining BERT with feature engineering, achieving high accuracy in evaluating nuanced writing skills, which aligns with the substantial agreement observed in this study. Furthermore, Poonpon, Manorom, and Chansanam (2023) investigated AEG reliability for non-native speakers, finding that advanced models, such as those incorporating transformer-based embeddings, could achieve consistency comparable to human raters. These studies collectively support the current findings by indicating that well-designed AEG systems can reliably assess communication skills across various proficiency levels, thereby validating the substantial agreement observed with Cohen's kappa value in this study.

Table 1. Assessed the communication skills of the respondents and the reliability of the system

Communication skills of the respondents			**Reliability of the System**	
			Cohen's kappa (κ)	*p-value*
Level	Frequency (N=35)	Percentage	0.797	0.000
Advanced	1	3		
Proficient	15	43		
Developing	7	20		
Beginning	12	34		

Legend:

Cohen's Kappa (k)	Interpretation
$k < 0.20$	*Slight*
$0.21 \geq k \leq 0.40$	*Fair*
$0.41 \geq k \leq 0.60$	*Moderate*
$0.61 \geq k \leq 0.80$	*Substantial*
$k > 0.80$	*Almost Perfect*

Theoretically, this study supports the growing body of research on automated assessment and its role in improving educational systems. The high reliability score aligns with educational

theories that advocate using objective, standardized assessment tools, reducing human bias, and increasing grading consistency (Nicol & Macfarlane-Dick, 2006). The findings also affirm the evolving role of Artificial Intelligence (AI) in reshaping educational assessment models, contributing to a broader understanding of AI's integration into academic environments. This study's results suggest that AEG systems can enhance grading efficiency and support more scalable, objective assessments that align with modern educational theories of fairness and consistency.

On the practical side, the reliability of the AEG system has significant implications for educational institutions, particularly those that face high student-to-teacher ratios. AEG systems can assist in managing grading workloads, providing consistent evaluations across many students. This is particularly beneficial in courses with frequent written assignments, such as writing or language classes. However, practical challenges may arise regarding system integration, with schools needing to invest in training educators and providing technical support to ensure smooth adoption. Additionally, concerns about the system's inability to assess more nuanced aspects of writing, such as argumentative structure and content relevance, must be addressed for broader acceptance and use in educational settings. Future research and development should focus on improving these areas to create a more comprehensive and effective tool for assessing all facets of communication skills.

4.2 Experience of Students and Teachers, its Acceptability and Perceived Reliability When Using the Automated Essay Grading System

Table 2 summarizes students' and teachers' experiences, acceptability, and perceived reliability of the Automated Essay Grading (AEG) system. The results indicate high satisfaction across all evaluated dimensions, with mean scores consistently falling within the "Strongly Agree" range (4.20–5.00). Both students and teachers rated the system highly regarding accuracy, consistency, and fairness. Teachers provided slightly higher ratings for reliability (M = 4.88, SD = 0.34) and ease of use (M = 4.75, SD = 0.45), suggesting their confidence in the system's ability to streamline grading processes while maintaining fair evaluations (Hall, Seyam, & Dunlap, 2023; Nouriska et al., 2023; Nguyen, 2017; Pradeep & Kowsalya, 2022).

Table 2. Combined User Experience, Acceptability, and Perceived Reliability of the AEG System

Area	**Students**		**Teachers**	
	Mean	*Sd*	*Mean*	*Sd*
Accuracy	4.40	0.73	4.69	0.48
Consistency	4.48	0.55	4.56	0.51
Fairness	4.61	0.49	4.88	0.34
Feedback Quality	4.58	0.50	4.50	0.52
Ease of Use	4.69	0.47	4.75	0.45
Efficiency	4.54	0.55	4.56	0.63
Reliability	4.56	0.50	4.88	0.34
Overall Mean	4.55	0.54	4.73	0.45

Legend:

4.20 – 5.00	*Strongly Agree*
3.40 – 4.19	*Agree*
2.60 – 3.39	*Neutral*
1.80 – 2.59	*Disagree*
1.00 – 1.79	*Strongly Disagree*

Students valued the feedback quality (M = 4.58, SD = 0.50) and efficiency (M = 4.54, SD = 0.55), emphasizing the system's role in supporting learning through actionable insights (Das et al., 2022; Vanga et al., 2023). The high ratings for reliability and accuracy align with previous research on AEG system advancements in machine learning (Sharma & Goyal, 2020), while concerns regarding coherence and content relevance remain areas for further improvement (Ramesh & Sanampudi, 2022).

These findings support prior research highlighting AEG systems' benefits and challenges. Transparency and explainability are key factors influencing trust in automated grading (Hall, Seyam, & Dunlap, 2023). As emphasized in earlier studies, user-centered design plays a critical role in usability and effectiveness (Nouriska et al., 2023). The ability of AEG systems to provide quick and constructive feedback aligns with research demonstrating their positive impact on student learning (Das et al., 2022). Systems like PaperRater have also shown how automation can enhance grading reliability (Nguyen, 2017).

Educators have found AEG systems helpful in managing workloads while ensuring fairness in grading (Lewis, 2013). Advances in machine learning, particularly ensemble learning and bi-directional LSTM networks, have improved grading systems' accuracy and consistency (Sharma & Goyal, 2020).

The ability to deliver immediate feedback has also been recognized as an essential aspect of student success (Vanga et al., 2023). However, despite their advantages, AEG systems still require improvements in evaluating coherence and content relevance, which remain challenges identified in prior studies (Ramesh & Sanampudi, 2022). This study contributes to ongoing efforts to refine AEG systems and enhance their role in educational assessment.

Theoretically, these results reinforce the argument that AEG systems can play a transformative role in education by enhancing the speed, fairness, and consistency of grading. The findings confirm that AI systems can significantly contribute to educational equity by offering consistent evaluations free from human bias, thus improving the objectivity of assessments (Chinta et al., 2024). Furthermore, the positive feedback on the system's usability and accuracy supports the theoretical view that technology, when designed with user needs in mind, can enhance the educational experience.

In practical terms, the findings suggest that both students and teachers benefit from the introduction of AEG systems. The system relieves the time-consuming grading process for educators, enabling more time for instructional activities. The high ratings for reliability and feedback quality suggest that students find value in the immediate, actionable insights provided by the system, which helps them improve their writing skills. However, practical barriers to widespread adoption may include the need for technical infrastructure, faculty training, and initial resistance to AI-driven grading systems. To overcome these hurdles, institutions should provide adequate resources, such as training sessions, to ensure that faculty are confident in interpreting and using the system's feedback. Expanding the system's capabilities to assess higher-order communication skills like argumentation and critical thinking will ensure its long-term success and acceptance in educational contexts.

While this study offers valuable insights into the effectiveness of the NLP-based Automated Essay Grading (AEG) system, the small and homogeneous sample size is a limitation. The study was conducted with participants from a single institution, Central Bicol State University of Agriculture (CBSUA), which may not fully represent the experiences and perspectives of students and teachers from other academic environments. As such, the findings might not be directly transferable to other institutions, especially those with different curricular structures, student populations, or institutional settings.

The homogeneity of the sample, consisting primarily of Information Technology students and teachers, could also influence the study's results. Future research should aim to include a more diverse group of participants from various educational contexts to assess whether the system's effectiveness holds across different subjects, teaching methodologies, and student backgrounds. Additionally, expanding the sample size would help ensure the results are more robust and better reflect the diverse ways AEG systems might be perceived and utilized across various educational systems.

5. Conclusion

This study assessed the effectiveness, user experience, and acceptability of a self-developed Natural Language Processing (NLP)-based Automated Essay Grading (AEG) system in evaluating students' communication skills. The findings demonstrate that the system provides reliable and consistent evaluations, as reflected by a Cohen's kappa value of 0.797, indicating substantial agreement with human grading.

The merged user experience and acceptability analysis confirmed high satisfaction among students and teachers. Teachers rated the system highly for reliability ($M = 4.88$, $SD = 0.34$) and ease of use ($M = 4.75$, $SD = 0.45$), suggesting confidence in its ability to streamline grading while maintaining fairness. Meanwhile, students valued the feedback quality ($M = 4.58$, $SD = 0.50$) and efficiency ($M = 4.54$, $SD = 0.55$), emphasizing its role in providing actionable insights for writing improvement.

Despite these strengths, the study also identified areas for enhancement, particularly in assessing coherence and content relevance, which remain challenges in automated grading. Future improvements should focus on refining NLP models to improve the evaluation of logical flow, argument structure, and contextual understanding.

Overall, the findings confirm that the NLP-based AEG system is an effective and reliable tool for assessing communication skills. With further refinements, it has the potential to enhance learning outcomes, streamline grading processes, and support educators in delivering fair and objective assessments in academic settings.

In conclusion, the NLP-based AEG system demonstrated substantial reliability and user satisfaction, but future research should address the current study's limitations, particularly the sample size and homogeneity. Expanding the sample to include a broader, more diverse group of participants will help ensure the findings apply to a wider range of educational settings. Future studies should also explore the system's ability to assess more complex aspects of communication, such as coherence and argument strength, to refine the tool further.

References

Ait Khayi, N., & Rus, V. (2024). Automated essay scoring using discourse external knowledge. *Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI)*. https://doi.org/10.24963/ijcai.2024/791

Chinta, A. S., Sharma, R., & Patel, P. (2024). Navigating fairness, bias, and ethics in educational AI applications. *Journal of Educational Technology, 62*(4), 124-137. https://doi.org/10.48550/arXiv.2407.18745

Das, L. B., Raghu, C. V., Jagadanand, G., George, R. A. R., Yashasawi, P., & Patnaik, V. K. (2022). FACToGRADE: Automated essay scoring system. *Proceedings of the IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 42–48. https://doi.org/10.1109/IAICT.2022.10007407

Hall, E., Seyam, M., & Dunlap, D. (2023). Identifying usability challenges in AI-based essay grading tools. In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial Intelligence in Education. AIED 2023* (Vol. 1831, pp. 1–10). Springer. https://doi.org/10.1007/978-3-031-36336-8_104

Hicke, Y., Tian, T., Jha, K., & Kim, C. H. (2023). Automated essay scoring in argumentative writing: DeBERTeachingAssistant. *arXiv Preprint*. https://doi.org/10.48550/arxiv.2307.04276

Hossain, M., & Mustafa, A. (2023). Automated writing evaluation using sentence-by-sentence scoring model. *Proceedings of the IEEE National Conference on Innovative Management (NCIM)*. https://doi.org/10.1109/ncim59001.2023.10212752

Ibrahim, S. S., Elfakharany, E. F., Ramzy, E. M., & Hamed, H. (2022). Improved automated essay grading system via natural language processing and deep learning. *Proceedings of the IEEE International Conference on Emerging Technologies (ICEET)*, 100–107. https://doi.org/10.1109/ICEET56468.2022.10007407

Lewis, J. (2013). Ethical implementation of an automated essay scoring (AES) system: A case study of student and instructor use, satisfaction, and perceptions of AES in a business law course. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2684803

Madnani, N., & Cahill, A. (2018). Automated scoring: Beyond natural language processing. *Proceedings of the 27th International Conference on Computational Linguistics (COLING)*, 1099–1109. https://doi.org/10.55041/IJSREM29523

Nguyen, V. T. (2017). Automated essay assessment: An evaluation on PaperRater's reliability from practice. *Journal of Creative Practices in Language Learning and Teaching (CPLT), 5*(1), 1–18.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. https://doi.org/10.1080/03075070600572090

Nouriska, S., Untoro, M. C., Afriansyah, A., Praseptiawan, M., Yulita, W., & Ashari, I. F. (2023). User experience answer system automatically with user-centered design and user experience questionnaire-short. Jurnal Ilmu Pengetahuan dan Teknologi Komputer (JITK), 9(1), 81–88. https://doi.org/10.33480/jitk.v9i1.4152

Poonpon, K., Manorom, P., & Chansanam, W. (2023). Exploring effective methods for automated essay scoring of non-native speakers. *Contemporary Educational Technology, 14*(1). https://doi.org/10.30935/cedtech/13740

Prabhu, S., Kara, A., & Sanriya, S. (2022). A hybrid approach towards automated essay evaluation based on BERT and feature engineering. *Proceedings of the IEEE International Conference on Innovative Computing Technologies (I2CT)*, 982–987. https://doi.org/10.1109/i2ct54291.2022.9824999

Pradeep, R., & Kowsalya, M. (2022). An investigation of several models for machine learning-based automated essay grading system. *International Journal of Open Access*. https://doi.org/10.21203/rs.3.rs-1278674/v1

Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: A systematic literature review. *Artificial Intelligence Review, 55*(3), 2495–2527.

Sharma, S., & Goyal, A. (2020). Automated essay grading: An empirical analysis of ensemble learning techniques. In *Computational Methods and Data Engineering: Proceedings of ICMDE 2020* (Vol. 2, pp. 343–362). Springer. https://doi.org/10.1007/978-981-15-7907-3_26

Vanga, R. R., Bharath, M. S., Sindhu, C., Vadivu, G., & Hsu, H. C. (2023). Grade it: A quantitative essay grading system. In *Proceedings of the International Conference on Innovative Computing and Communication* (pp. 317–327). Springer. https://doi.org/10.1007/978-981-99-3010-4_28

Xiaoxiao Wang, I., Wu, X., Coates, E., Zeng, M., Kuang, J., Liu, S., Qiu, M., & Park, J. (2024). Neural automated writing evaluation with corrective feedback. *arXiv Preprint*. https://doi.org/10.48550/arxiv.2402.17613