

Enhancing e-Assessment in Higher Education: An Exploration of an Automated Essay Feedback System

Sze Yie M. Chye

Centre for Innovation in Learning, National Institute of Education, Nanyang Technological University, Singapore
Email: mary.chye@nie.edu.sg

Abstract—This paper details the evaluation of an Automated Essay Scoring (AES) software for the purposes of providing feedback on students' writing. The study aimed to ascertain if faculty and students found the writing draft support provided by the automated feedback of the software accurate, of good quality and useful. AES software subscription for institutional use may then be endeavoured. Using a mixed methods design, consisting of user survey, user interview and analysis of the AI-generated feedback, 33 students (Age 19–25, male and female) enrolled in an academic writing course and 2 faculty members teaching the course participated in this pilot. Generic feedback was given in five categories, namely, Focus & Meaning, Content & Development, Organisation, Language Use, Voice & Style, Mechanics & Conventions. Overall, 60% of the users indicated positivity towards using AES if made available to them although they thought that the AES-generated feedback provided minimal understanding of their strengths and weaknesses in writing due to its generic nature. For the 30% of users who rated highly on the quality and usefulness of the AES feedback, they were mainly focused on the grammar check functionality of the AES. Faculty feedback mirrored students' feedback. The study revealed 2 roadblocks for the adoption of AES.

Keywords—computer aided assessments, e-assessment, automated essay feedback

I. INTRODUCTION

In academia, the task of providing timely and meaningful feedback on student writing poses a significant practical challenge, particularly in the face of increasing time constraints. This challenge is compounded by the need to support low achievers in developing their disciplinary mastery, critical writing and thinking skills. To address these concerns, there is a growing emphasis on prioritizing Assessment for Learning (AfL) strategies, which aim to provide students with personalized and timely feedback to enhance their performance. Findings to a review of courses across initial teacher preparation and graduate programmes from various academic groups also revealed that the essay type assessment task was the most common assessment type

used in the university and often held the largest weightage for course assessments.

Given the above context, one approach that holds promise in meeting the objectives is the use of Automated Essay Scoring (AES) software, which automates the process of providing feedback on student writing. AES software has the potential to streamline the feedback process, alleviate faculty workload, and provide targeted support to students.

Despite the potential benefits of AES software, there is a need for empirical evidence to evaluate its effectiveness in providing feedback on student writing. This study seeks to fill this gap by conducting an evaluation of an AES software to determine if faculty and students perceived the automated feedback generated by the software to be accurate, of good quality and useful. If the evaluation demonstrates positive outcomes, there is potential for institutional adoption of the software which could lead to broader implementation of AES for feedback provision on writing across the university.

This study aims to contribute valuable insights into the role of AES software in providing feedback on student writing and its potential to address pressing concerns in higher education. By evaluating the effectiveness of the AES software, the study seeks to inform evidence-based decision-making and contribute to the ongoing discourse on innovative approaches to writing instruction and assessment.

II. LITERATURE REVIEW

Feedback is a central aspect in writing instruction [1]. Burgeoning research has confirmed the overall advantages of written Corrective Feedback (CF) in enhancing students' writing accuracy [2, 3].

Written CF has been categorized into three types: i) direct, ii) indirect or iii) metalinguistic [4]. However, these categories are focused on feedback on linguistic elements in an essay and do not encompass the range of the feedback provided by AES, which can give feedback not just on linguistic features, such as language use, mechanics, and conventions but also on semantic elements (e.g., focus and meaning, content, coherence and development) and rhetorical features (e.g., persuasion, overall presentation or organization and voice and style) in an essay [5]. Such feedback goes beyond teachers'

underlines or cursors in students' text or their indications on the paper's margins by employing acronyms or shortened forms of error codes. Instead, they tend to be presented in the form of declarative or interrogative sentences [6] and tend to be extended as compared to the more concise written CF given by teachers [7].

This is understandable as faculty do not have time. In a research-intensive institution, faculty are required to conduct research, secure substantial external grant funding while fulfilling their teaching responsibilities [8]. With these competing demands, providing extensive written feedback to large number of classes with huge class sizes would be a challenge. This underscores the need for the present study, which aims to explore AES software as an alternative method or tool to address the issue of providing formative feedback to students efficiently.

AES systems offer numerous benefits, including improving consistency in scoring, providing variety in high-stakes assessments, and reducing time taken to grade and process scripts as well as upholding standardisation by applying a uniform criterion to all essays. However, these systems also come with many drawbacks such as its propensity to extract irrelevant variables during evaluation, the lack of personal relationship between assessors and students and the requirement for a substantial corpus of sample text to train the AES model [9]. Other challenges include how AES are unable to assess novelty in ideas and can mistakenly ascribe lower or higher scores [10]. This can occur when students use longer, less frequent words to trick the system into improving scores [11]. Comparative studies have shown that feedback provided by instructors are of better quality than AES feedback [12]. Another oft-cited shortcoming of AES systems is the black box nature of its scoring algorithms [13], which give rise to a lack of trust and acceptance of AES systems [14].

The limitations discussed above, coupled with the imperative to understand user attitudes, preferences, and experiences within the institution for the purpose of assessing the effectiveness and usability of AES in the unique, practical setting necessitates empirical evidence.

III. MATERIALS AND METHODS

This study employed a mixed methods approach to evaluate the effectiveness of AES software for providing feedback on students' writing. This design integrated both quantitative and qualitative data collection methods to comprehensively assess the perceptions of faculty and students regarding the software's utility and impact.

A. Participants

The participants in this study consisted of a representative 33 students, aged between 19 and 25, who were enrolled in a non-discipline specific skills-based academic writing course with a large cohort of approximately 200 students. This course is taken by students across different academic disciplines with a common assessment task scored using a common rubric. Two faculty members responsible for teaching the course

were also participants. This sample was selected to represent a diverse range of student demographics and academic disciplines.

B. De-identification and Data Security

To ensure participant privacy and data security, measures were implemented to de-identify student submissions and maintain anonymity in the feedback process. Unique login credentials were provided to students to access the AES software, separate from their regular login details associated with their matriculation numbers. Additionally, personal identifiable information was removed from student scripts before submission to the AES system.

C. Training the AI Scoring Engine

To train the AI scoring and feedback engine to learn and replicate the way the rubric has been interpreted in the past by human scorers, 300 past de-identified student scripts to a specific prompt, its assessment rubrics, and previously assigned grades were required. However, only 200 past student scripts from the August 2021 and August 2022 semesters were available for download in the LMS servers and provided to the vendor to be used as the training set for the AI-scoring engine. Scores to the 200 past student scripts were not available as they were purged from the mark entry system.

Since this was lower than the vendor-specified requirement of 300 scripts including scores, the option of using a customised scoring engine was not available. In view of the likelihood of a similar condition of insufficient scripts and unavailability of scores across all courses in the institute, the pilot proceeded with a generic, pre-scored scoring engine to test its utility and applicability. The generic, pre-scored scoring engine is pre-trained on college level essays based on genre and discipline.

D. Method

Students submitted the drafts of their final assignment within the Learning Management System (LMS), where the AES feedback software tool will be integrated under a Staging environment. Once the essays are submitted, feedback on categories related to those in the assignment grading rubric will be automatically generated for user review.

To prevent feedback from affecting assignment grades, access to the AES system for student submissions for AES feedback was provided one day after the final assignment submission deadline. This is to ensure that no significant edits to essays may be made, which may have an adverse effect on students' final grade, since the AI engine is under validation on its accuracy, quality, and usefulness in this current pilot.

E. Data Collection Instruments

A user survey was developed to gather quantitative feedback from students regarding their experiences with the AES software. The survey included questions pertaining to the accuracy, quality, and usefulness of the automated feedback generated by the software.

Additionally, faculty opinions were sought via open-ended email questionnaire and face to face discussion which allowed them to provide qualitative insights and suggestions for improvement.

F. Analysis of AI-Generated Feedback

The feedback generated by the AES software was subjected to qualitative analysis to assess its quality and relevance. A criterion for evaluation was adopted to compare the AES-generated feedback with human-generated feedback. The table of criterion is attached in Appendix B. However, due to the generic nature of the AI-generated feedback, the planned comparative analysis could not be conducted as initially intended.

G. Data Collection and Analysis

Quantitative data collected through the user surveys were analyzed using descriptive statistics to summarize participant responses. Qualitative data from faculty interviews and analysis of AI-generated feedback were thematically analyzed to identify key themes and patterns in participant responses.

H. Ethical Considerations

Ethical approval for the study was obtained from the Institutional Review Board. Informed consent was obtained from all participants prior to their involvement in the study. Throughout the research process, steps were taken to ensure the confidentiality and anonymity of participant responses.

IV. RESULT AND DISCUSSION

A. AI Generated Feedback

Feedback was given in 5 categories, namely, Focus & Meaning, Content & Development, Organisation, Language Use, Voice & Style, Mechanics & Conventions. They were stock feedback that were generic with minor variations generated among scripts. Table I below gives an example of feedback in the first category with the paraphrased sentences indicated in italics, repeated feedback has been indicated in bold and new feedback in normal font.

TABLE I. COMPARISON OF AI-GENERATED FEEDBACK BETWEEN 2 SCRIPTS

Script Number	1st Category: Focus & Meaning
1	Revision Goal 1: Write for your audience.
	1. Look at the beginning of your essay. Highlight, in yellow, the one sentence you wrote that tells your readers what your essay will be about. This is called your thesis statement. If you did not include a thesis statement, add it now.
	2. Your introduction should also include interesting details about your topic to grab your readers' attention. <i>Add a quotation, an interesting fact, an example, a question or an exclamation about your topic to interest your reader.</i>
2	Revision Goal 2: Include information about the central/controlling idea of your essay.
	1. Read your essay and highlight the important details about your central/controlling idea in green. <i>Details include information such as facts, examples, definitions, explanations, or quotations.</i>
	2. Are all the details you highlighted in green about your central/controlling idea? If not, remove them now. <i>Add more facts, examples, definitions, or explanations about your topic.</i>
	Revision Goal 1: Understand and write for your audience.
	1. Your introduction should include interesting details about your topic to grab your readers' attention. <i>Can you think of a clever way to grab your readers' interest? You could add a quote, an interesting fact, an example, a question or an exclamation about your topic to hook your reader.</i>
	Revision Goal 2: Include information about the central/controlling idea of your essay.
	1. Read your essay. Highlight the details about your central/controlling idea in green. <i>Now, add more information about the details you already included in your essay. You might add a sentence to explain an example, a statement, or a quote you included. Perhaps you could give a definition of an important word you used. Or you may choose to give an example of a statement you wrote.</i>

Essentially, the analysis reveals that stock feedback is repeated for all scripts but with the sentence order rearranged and interspersed with some paraphrased versions, which is the essence of the minor variations.

B. User Survey Feedback

Feedback from two groups of users – students and faculty were gathered and is presented below.

The survey instrument used in this study was developed based on a comprehensive review of relevant literature. The items were designed to assess accuracy, quality and usefulness and were refined through an iterative process involving input from experts in the field of Applied Linguistics. To ensure the validity of the survey instrument, content validity was assessed through review with the participating faculty who evaluated the

relevance and comprehensiveness of the survey items. Ethical approval for the survey study was obtained from the Institutional Review Board. Informed consent was obtained from all participants prior to their participation, and measures were taken to ensure the confidentiality and anonymity of their responses.

1) Student feedback

Familiarity and previous experience

Overall, the respondents had never used an AES feedback system and were not at all familiar with AES feedback systems before, despite very frequent engagement in writing assignments for their courses.

Quality and Usefulness of Feedback

The respondents rated differently on the overall quality of the feedback provided by the Automated Essay Feedback System, with 33.3% rating it Excellent, 33.3%

rating it Very Poor and 33.3% rating it Average. The possible reason for the Excellent rating is that the student found the grammar check quite useful (an aspect of the AES students found most helpful and valuable) while the Very Poor rating can be accounted for by the general and basic report generated and irrelevant details picked up, (an unhelpful or confusing feature of the AES which students listed). The Average rating could be because although students found the frameworks and guiding questions provided in the automated feedback good, they also thought the prompts were rather generic. On the confusing aspect of the AES, respondents found the copying and pasting of his essay in the textbox confusing.

Overall, 60% of the respondents did not think that the feedback from the AES helped them understand their strengths and weaknesses in writing very much. Respondents who rated the AES feedback helped Very Much could possibly have rated so, because they mainly considered the grammar checking features.

Potential Future Usage

Regarding potential usage of the AES tool for draft revision and essay improvement, if the AES tool was made accessible prior to students' final submission, 60% of the respondents gave a positive reply, indicating interest in using it. The same respondents felt that the feedback could influence their overall writing process and approach by allowing them to re-read and revise their writing with the help of generic prompts and to make it more concise and grammatically accurate. Respondents who rated negatively on the potential usage felt that the feedback would not influence their writing process nor approach as there were limited gains from the feedback given in this AES pilot, given the generic feedback.

Key Area of Improvement

Overall, the recurrent feedback gathered from the user survey was the generic AES feedback and the limited consideration of the writer's content, context, and intention of the essay. Suggested improvements also focused on the need for specific prompts and attention to parts of the paper where the assignment question was referring to, with one suggestion requesting the checking of the reference list.

User Satisfaction Rating

The overall satisfaction rating was neutral, i.e., neither satisfied nor dissatisfied.

2) Faculty feedback

Open-ended feedback was sought from the 2-participating faculty through a combination of email communication and face to face discussions. This approach aimed to capture faculty perspectives comprehensively and allowed for both written responses and in-depth conversations.

The feedback received from faculty members was mixed. Positive feedback included sentiments that the introduction of AES software for writing feedback was a valuable initiative. Faculty members expressed appreciation for the innovative approach and alignment with broader trends, particularly those advocated by the Ministry of Education (MOE). Faculty members also expressed a keen interest in exploring opportunities for

sustained collaboration in future cohorts. Suggestions included the possibility of transforming the course into a non-graded format, thereby shifting the focus from evaluation to learning enhancement and to leverage on AES software to provide draft support to facilitate iterative writing processes and promote continuous improvement.

Conversely, negative feedback highlighted concerns that the AI-generated feedback was overly generic with similar advice repeated across multiple scripts. Faculty members felt that it would be more beneficial for the feedback to reference specific words, phrases, or sentences from the text to allow for more targeted and contextually relevant feedback. More nuanced and personalized guidance was desired.

V. CONCLUSION

The evaluation of the non-customised, pre-scored AI scoring engine revealed limitations in the effectiveness of the generic feedback provided to students. This was primarily due to two technical roadblocks encountered in this pilot. These roadblocks included the insufficient availability of past student scripts and the absence of recorded scores in the institutional learning management and examination mark entry system. Only 200 out of 300 required past student scripts were available as most of the past scripts from previous cohorts have been purged from the LMS. Similarly, records of scores were unavailable for the 200 past student scripts as they have been purged from the examination mark entry system. As a result, the potential of the AES software to provide personalized and actionable feedback was hindered.

A. Future Directions

To overcome these limitations and enhance the utility of AES software to support student writing, future directions should focus on several key areas. Firstly, efforts should be made to explore solutions for retaining a sufficient number of student scripts and corresponding scores within the institutional learning management and mark entry systems to ensure an adequate dataset for training a customised scoring engine. Furthermore, the implementation of a custom-built AI scoring engine could improve the relevance and specificity of feedback provided to students.

Additionally, future research endeavours should consider involving student participants in the review and utilization of AI-generated feedback to fully harness the intended purpose of the AES as a draft support tool. Allowing student engagement and action on the feedback can facilitate their understanding of the strengths and weaknesses in their writing and promote active learning and improvement.

It may also be valuable to conduct comparative studies to evaluate the capabilities of other existing on-the-shelf AES systems and explore the feasibility of custom-building an AES system tailored to the specific needs and objectives of the institution.

In summary, addressing the identified challenges and exploring the future directions outlined above will further

maximise the potential of AES software in providing effective feedback on student writing and promote academic success.

CONFLICT OF INTEREST

The author declares no conflict of interest.

ACKNOWLEDGMENT

The author wishes to thank Dr. Anitha Devi Pillai, Dr. Chang Qizhong and Dr. Guo Libo who kindly agreed to participate in the pilot. The author also wishes to thank Ang Kheng Yong and Lum Hon Peng who supported in this pilot study.

REFERENCES

- [1] K. Hyland and F. Hyland, "Feedback on second language students' writing," *Language Teaching*, vol. 39, no. 2, pp. 83–101, 2006. <https://doi.org/10.1017/S0261444806003399>
- [2] J. Bitchener and U. Knoch, "The value of written corrective feedback for migrant and international students," *Language Teaching Research*, vol. 12, no. 3, pp. 409–431, 2008. <https://doi.org/10.1177/1362168808089924>
- [3] I. Lee, "Research into practice: Written corrective feedback," *Language Teaching*, vol. 46, no. 1, pp. 108–119, 2013. <https://doi.org/10.1017/S0261444812000390>
- [4] R. Ellis, "A typology of written corrective feedback types," *ELT Journal*, vol. 63, no. 2, pp. 97–107, 2009. <https://doi.org/10.1093/elt/ccn023>
- [5] D. S. McNamara, S. A. Crossley, R. D. Roscoe, L. K. Allen, and J. Dai, "A hierarchical classification approach to automated essay scoring," *Assessing Writing*, vol. 23, no. 1, pp. 35–59, 2015. <https://doi.org/10.1016/j.asw.2014.09.002>
- [6] Y. J. Jong, Y. J. Kim, and O. C. Ri, "Review of feedback in automated essay scoring," arXiv.org, 2023. <https://doi.org/10.48550/arxiv.2307.05553>
- [7] S. Dikli, "The nature of automated essay scoring feedback," *CALICO Journal*, vol. 28, no. 1, pp. 99–134, 2010. <https://doi.org/10.11139/cj.28.1.99-134>
- [8] S. E. Brownell and K. D. Tanner, "Barriers to faculty pedagogical change: Lack of training, time, incentives, and tensions with professional identity?" *CBE Life Sciences Education*, vol. 11, no. 4, pp. 339–346, 2012. <https://doi.org/10.1187/cbe.12-09-0163>
- [9] S. Toranj and D. N. Ansari, "Automated versus human essay scoring: A comparative study," *Theory and Practice in Language Studies*, vol. 2, no. 4, pp. 719–725, 2012. <https://doi.org/10.4304/tpls.2.4.719-725>
- [10] M. A. Hussein, H. A. Hassan, and M. Nassef, "Automated language essay scoring systems: A literature review," PeerJ preprints, 2019. <https://doi.org/10.7287/peerj.preprints.27715v1>
- [11] I. I. Bejar, M. Flor, Y. Futagi, and C. Ramineni, "On the vulnerability of automated scoring to Construct-Irrelevant Response Strategies (CIRS): An illustration," *Assessing Writing*, vol. 22, no. 1, pp. 48–59, 2014. <https://doi.org/10.1016/j.asw.2014.06.001>
- [12] S. Dikli and S. Bleyle, "Automated essay scoring feedback for second language writers: How does it compare to instructor feedback?" *Assessing Writing*, vol. 22, no. 1, pp. 1–17, 2014. <https://doi.org/10.1016/j.asw.2014.03.006>
- [13] Y. K. Singla, S. Parekh, S. Singh, J. J. Li, R. R. Shah, and C. Chen, "AES systems are both overstable and oversensitive: Explaining why and proposing defenses," arXiv.org, 2021. <https://doi.org/10.48550/arxiv.2109.11728>
- [14] V. S. Kumar and D. Boulanger, "Automated essay scoring and the deep learning black box: How are rubric scores determined?" *International Journal of Artificial Intelligence in Education*, vol. 31, no. 3, pp. 538–584, 2021. <https://doi.org/10.1007/s40593-020-00211-5>

Copyright © 2025 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).