

Students' Acceptance of Automated Essay Scorer: Reflection on an Initiative to Mitigate Disruption to Writing Lessons during Pandemic

Mohammad Radzi Manap¹, Norkhairi Ahmad² and Zulkarnin Zakaria³

¹Universiti Teknologi MARA, Malaysia, {moham830@uitm.edu.my}

²Universiti Kuala Lumpur Malaysia France Institute, Malaysia, {norkhairi@unikl.edu.my}

³Kolej GENIUS Insan, Universiti Sains Islam Malaysia {zulkarnin@usim.edu.my}

ABSTRACT

Automated Writing Evaluation (AWE) is an innovation in the field of language teaching and learning with features like portfolio and writing assistant resources has become a useful alternative to support language assessment processes during the pandemic. Like many artificial intelligence-based tools, there is always concern on scoring accuracy, reliability, and acceptance by users. This paper aims to explore language learners' experience in using an AWE called PaperRater (PR) available from the internet. Data was elicited via a questionnaire designed based on the Technology Acceptance Model (TAM) and it focuses on six variables of acceptance namely perceived usefulness, perceived ease of use, user satisfaction, usability, user behaviour and user profiles. Rasch model and descriptive statistical analysis were used in analysing responses from 62 undergraduates. The respondents are found to have a positive level of acceptance towards the use of AWE as depicted by the -1.21 to 2.07 Rasch logit unit. This tool is also perceived to be beneficial for formative learning purposes via students' self-assessment, in the absence of educators in physical classes and limited online access to educators during this pandemic.

Keywords: AWE, TAM, writing assessment.

I INTRODUCTION

Writing tasks are varied in both complexity and purpose, with many elements that encompass the writing process (Roscoe & McNamara, 2013). Being the end user of the technology, most language instructors' objective of incorporating automated rating or scoring tools for writing (henceforth AES) in the classroom is none other than to ease and expedite the assessment of their learners' essay apart from taking advantage of the AES which could also become a tool to increase learners' motivation in writing. Reilly et al. (2014) in their study of comparing instructors' grading and AES-holistic scores in MOOC courses have found that AES is still found to be useful in the teaching and learning process of writing. Chapelle and Douglas (2006) believe that computerized teaching tools and technologies should be effective aids in language classrooms. This could be seen from the literature and

the results of the statistical analysis of user perception on an AES tool in this study.

II STATEMENT OF THE PROBLEM

Whenever an automated scoring tool is utilized, its reliability should be taken into careful consideration (Warschauer and Grimes, 2008). Shin (2012) stresses that even though web-based language testing may "enhance test authenticity and reliability by making possible a rich contextualized input, various response formats, and automated scoring", there has still been very little study conducted to investigate whether online testing can actually work in writing classroom (p.277).

III RELEVANT LITERATURE

Many researchers studied the use of AES in language lessons and identified their potential as well as room for improvement. Peterson (2017), in his study on students at two suburban high schools elicited data on student perceptions of feedback pre- and post-intervention and computer-generated percentile rankings of students' writing skills. The findings indicated that computer-generated feedback increased students' writing efficacy and inclination to revise writing. A study by Nguyen (2017) indicates that the reliability of PaperRater is acceptable and that writing teachers can somehow rely on the functions of this tool as a reference in grading papers.

An appropriate combination of traditional and computerized grading methods can generate effectiveness, especially in large classrooms or with a great number of papers. The tandem integration of language instructor roles and the use of AES for future writing teaching practice has also been propagated by a few researchers with a few recommendations and highlights on pedagogical impacts. Due to this teaching innovation still being at an early evolving stage there are limited studies being carried out. Yinghui and Dan (2015), it is recommended that language instructors opt for conventional teaching of writing along with AES to evaluate students' writing tasks. They found that such a strategy is applicable for classes with a high number of students or for teachers handling a large number of learners and especially in the context of countries like China. This strategy will enable two means of feedback for students' writings and could be based on students' learning levels and teachers' actual needs.

As such AES or AWEs have both merits and drawbacks. This is discovered by Pei-Ling (2015) where approximately two-thirds of her study participants perceived that the integration of machine scoring with language instructors' input became the recommended implementation method for writing classes.

IV METHODOLOGY

This study applied purposive sampling techniques where the samples are undergraduate respondents enrolled in a compulsory academic writing course at a public university. The students were briefed on PaperRater on how to use it as well as the benefits that they can gain. Besides, the respondents will also be briefed on how to monitor their own improvement in writing skills based on the report generated by PaperRater. Respondents will be experimenting with PaperRater at their own will in a duration of two weeks. It is important to stress here that the use of PaperRater is encouraged among the respondent but was not made compulsory. However, both lecturers did inform the students that they are welcome to discuss matters regarding their essay and issues regarding the use of PaperRater if they want to during their meetings with the instructor.

For data elicitation, a questionnaire was designed based on TAM with 6 sections representing aspects such as Demographic Information, Perceived Usefulness, Perceived Ease of Use, User Satisfaction, Usability and General Question. A four-point Likert Scale comprising Strongly Disagree, Disagree, Agree and Strongly Agree was employed. The reliability of this tool is validated through Rasch Model Analysis using Winstep 3.72.3 for internal consistency reliability. According to Abdul Aziz (2010), reliability of a questionnaire is verified via Cronbach alpha (α) value, person and item reliability value, person measure and valid responses. Generally Cronbach alpha value for reliability should range from 0.67 to 1.00 which indicates poor up to excellent value (Fisher, 2007). Higher Cronbach alpha value indicates stronger relationship between the questionnaire items.

In Rasch Model, this value is explained by Kuder-Richardson (KR-20) and coefficient alpha value (Cronbach, 1984). Through the analysis, the instrument in this study indicates the Cronbach alpha value of 0.93 putting it to be in the 'Very Good' category as determined by Fisher (2007). Other reliability indicators include Person Reliability = 0.87, Person separation index = 2.64, Item reliability = 0.88 and Item separation index = 2.68.

V FINDINGS

Basically, the 62 undergraduate respondents made up of 25 males and 37 females from four faculties enrolled in two groups for an academic writing course at a public university. The data was tabulated by describing the respondents' response pattern in all the sections of the questionnaire (Section A to D).

The items per section are listed based on their difficulty level (the most difficult to the easiest) as per Rasch Model Analysis where the higher the position of the item in the Person Item Map (PIM) the more difficult the item is to be endorsed. The results of the additional questions attached to the questionnaire (Section E) were also discussed.

A. Perceived Usefulness

This section is to define the degree to which students believe that PaperRater is able to improve their writing skills. Perceived usefulness of PaperRater in the questionnaire is represented by 5 items.

Table 1. Perceived Usefulness.

SECTION A: Perceived Usefulness		Positive Responses	Logit Value (Rasch)
A3	PaperRater is useful for rapid retrieval of assessment of my essay.	95.1%	-0.21
A4	PaperRater will save the time of lecturers and writers.	88.7%	-0.21
A5	Using PaperRater would improve my writing performance.	91.9%	-0.28
A1	PaperRater enables me to get the assessment/ score of my writing quickly.	95.1%	-0.61
A2	PaperRater allows me to follow up with the errors and weaknesses of my essay anytime and anywhere.	95.1%	-1.21

B. Perceived Ease of Use

Section B of the questionnaire focuses on the students' perceived ease of using PaperRater in improving their writing skills. This section is represented by 5 items.

Table 2. Perceived Ease of Use.

SECTION B : Perceived Ease of Use PaperRater(PR)		Positive Response	Logit Value (Rasch)
B5	It is easy for me to become skillful at using PR to benefit my writing.	90.4%	+0.48
B4	I find PaperRater flexible to interact with.	91.9%	+0.20
B1	Learning to operate PaperRater is easy for me.	96.8%	-0.28
B2	I find it easy to get the benefit of using PR.	96.7%	-0.28
B3	I can easily understand the report generated by PaperRater to improve my writing.	98.4%	-0.34

C. User Satisfaction

User satisfaction is another construct based on TAM included in the questionnaire. This section is used to measure students' satisfaction in using PaperRater in assisting and improving their writing skills.

Table 3. User Satisfaction.

SECTION C : User Satisfaction		Positive Responses	Logit Value (Rasch)
C5	I believe that using PaperRater will increase the quality of my writing.	93.6%	-0.15
C4	I can accomplish writing tasks quickly using PaperRater.	95.1%	-0.21
C1	I am completely satisfied using PaperRater.	93.6%	-0.28
C2	I feel very confident in using PaperRater.	95.2%	-0.34
C3	I found it easy to share information about my writing assessment using PaperRater.	98.4%	-0.54

D. Usability

Usability is a construct to which students' responses to the items are gathered to reflect how they are able to utilize the elements of PaperRater in effort to improve

their writing skills. This construct is represented by 4 items.

Table 4. Perceived Usability.

SECTION D : Usability		Positive Response	Logit Value (Rasch)
D1	It is easy to follow the suggestions /report of PaperRater.	96.8%	-0.28
D3	I found the report in PaperRater to be comprehensive and able to guide me in improving my writing.	93.6%	-0.34
D4	I will use PaperRater in my next writing assignment.	95.2%	-0.34
D2	By using PaperRater, I can easily identify the areas of my writing that need to be improved.	95.1%	-0.61

E. General Question

A number of general questions were also included. Here, in terms of frequency of submission to PaperRater, it was found that 17.7% submitted once, 41.9% submitted twice, 12.9% submitted three times and 27.4% submitted more than 3 times. Other findings reveal that 66.2% are reluctant to share their scores with friends (logit +1.39). Furthermore, the students generally did not agree if PaperRater is used to score essays in final examination (logit +2.07) and they also found it difficult to use PaperRater through their smartphone.

F. Profiling of Respondents

Profiling of the respondents was made based on their responses in the questionnaire. In the Rasch Model, there is a specific formula to calculate the strata value based on the person separation index (2.64). Therefore, the calculation is as $(2.64 \times 4 + 1 \div 3 = 3.85)$. This means the respondents can be divided into 3.85 or 4 groups based on their patterns of response in the questionnaire. These 4 groups are labelled as 'Extremely Receptive' (16 or 25.8%), 'Very Receptive' (18 or 29.03%), 'Receptive' (26 or 41.9%) and 'Resistant' (2 or 3.2%). Overall, majority of the respondents or 96.7% showed that they are receptive of the idea of using PaperRater as a tool in improving their writing skills. However, their acceptance level varied

based on their endorsement of items in the questionnaire. These differences have categorized them into 3 different categories as described earlier. Additional information on the profiling is tabulated in the table below:

Table 5. User Profiles.

Group	N	Gender		Education Level
		Male	Female	
Extremely Receptive	16	5	11	Tertiary Level at Public University
Very Receptive	18	7	11	
Receptive	26	12	14	
Resistant	2	1	1	

G. Instructors' feedback

Generally, the instructors are quite satisfied with the acceptance of the use of PaperRater among their students. Those who used PaperRater were found to be more aware of their mistakes and sections of their essay that need to be upgraded, improved or even corrected based on the report generated from PaperRater. They are seen to have a better control of their own work, more independent and motivated. The instant feedback received by the students from PaperRater, its consistent report and grading were able to provide the students with a sense of direction in their learning and on the instructors' side, an "extra pair of eyes" assisting them with instant feedback on numbers of drafts submitted even though the final grading of the assessment is still done by the instructors themselves.

VI DISCUSSION

Data from this study suggests that PaperRater is reliable, easy and practical to use by both teachers and students. It has the potential to develop a sense of autonomy within language learners and is generally beneficial to students. As such, PaperRater is a good alternative to support lessons in academic writing courses during pandemic. Such automated tool is able to mimic the classroom assistance offered by language instructors where sometimes students might need to ask for feedback to ensure that they are making good progress and confidently proceed with more writing tasks (Likkell, 2012). In fact, the latest version comes together with a plagiarism checker. This makes PaperRater really suitable for tertiary students as good writing processes and originality are aspects

emphasized in academic writing courses at universities. The reliability and consistency in grading by the PaperRater compared to human rater were proven through the work by Nguyen (2017) where it was recommended to be used by writing instructors especially those who are dealing with large classes. Similar claims were made by Manap, et al. (2019) where they found a moderate positive linear relationship between PaperRater and a group of language instructors in grading of written works of the students.

The use of such automated tools like PaperRater becomes more handy during Covid 19 pandemic due to closure of campuses, movement controls and lockdowns. Apart from relying on online interaction with lecturers, tertiary students can now develop learner autonomy by using tools like PaperRater. Sing et al. (2016) believe that AES is also good to be used in tandem by lecturers as physical raters of writing and by students such as for self-study, formative assessment and assessment for learning purposes.

VII CONCLUSION

The use of automated rater tools in language teaching and learning is still not that widespread in Malaysia based on the dearth of literature on such innovative practice. This form of digital technology integration has become timely and highly necessary as a measure to mitigate the disruptions to previous traditional routines of teaching and learning brought about by Covid-19 pandemic. This study has shed light on the reliability and benefits of using PaperRater as vouched by the undergraduates and the language instructors for the academic writing courses. It is able to effectively fulfill functions as an artificial intelligence-based tool that facilitates rating processes, fosters independence learning and guides learners to improve as writers.

REFERENCES

- Abdul Aziz, A. (2010). Rasch Model Fundamentals: Scale Construct and Measurement Structure. Kuala Lumpur: Integrated Advanced Planning Sdn. Bhd.
- Chapelle, C. A., and Douglas, D. (2006). Assessing Language through Computer Technology. Cambridge: Cambridge university Press. 138.
- Cronbach, L. J. (1984). Essentials of Psychological Testing (4th ed.). New York, NY Harper & Row.
- Fisher Jr., W. P. (2007). Rating Scale Instrument Quality Criteria. Rasch Measurement Transaction, 21, 1095. <http://www.rasch.org/rmt/rmt211a.htm>.
- Huaqing, H. (2016). A Survey of EFL College Learners' Perceptions of an On-Line Writing Program. International Journal of Emerging Technologies in Learning. 11(4). 11-15.
- Likkell, L. (2012). Calibrated Peer Review Essays Increase Student Confidence in Assessing Their Own Writing. Journal of College Science Teaching, 41(3), 42-47
- Manap, M. R., Ramli, N. F., Kassim, A. A. M. (2019). WEB 2.0 Automated Essay Scoring Application and Human Essay Assessment: A Comparison Study. European Journal of English Language Teaching. 5(1), 146-162. <http://dx.doi.org/10.5281/zenodo.3461784>
- 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), 2017.
- Peterson, E. K. (2017). The Impact of Computer-Generated Feedback on Student Perceptions of Revision Process. Retrieved from Sophia, the St. Catherine University repository website: <https://sophia.stkate.edu/maed/247>
- Pei-ling, W. (2015). Effects of an Automated Writing Evaluation Program: Student Experiences and Perceptions *Electronic Journal of Foreign Language Teaching*. 12 (1),79–100.
- Reilly, E. & Stafford, R. & Williams, K. & Corliss, S. (2014). Evaluating the Validity and Applicability of Automated Essay Scoring in Two Massive Open Online Courses. *International Review of Research in Open and Distributed Learning*. 15. 10.19173/irrodl.v15i5.1857.
- Roscoe, R. D., & McNamara, D. S. (2013). Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, 105(4), 1010–1025. <https://doi.org/10.1037/a0032340>.
- Sing, Y. N., Chih H. B., Nung, K. L. and Kian, S. H. (2016). Automated Essay Scoring Feedback (AESF): An Innovative Writing Solution to the Malaysian University English Test (MUET). *International Journal on E-Learning and Higher Education*, 4. 130-143.
- Shin, S. Y. (2012). Web-Based language testing. In Coombe, C., et al. (Eds.). *Second Language Assessment* (pp. 274-279). Cambridge University Press.
- Warschauer, M., & Grimes, D. (2008). Automated Writing Assessment in the Classroom. *Pedagogies: An International Journal*, 3, 22–36
- Warschauer, M. & Grimes. (2008). Automated Writing Assessment in the Classroom. *Pedagogies*. 3. 22-36. 10.1080/15544800701771580.
- Yinghui, H. & Dan, Z. (2015). A Comparative Study of Teacher Feedback and Automated Essay Scoring in College English Writing. *International Journal of Linguistics and Communication*. 3 (2), 82-97 ISSN: 2372-479X (Print) 2372-4803. DOI: 10.15640/ijlc.v3n2a0