Perceptions of automated writing evaluation (AWE) feedback on L2 writing

Yang, Hsiao-hui

楊小慧

Assistant Professor

Applied English Department

Abstract

This study is aimed to explore Taiwanese students' perceptions of automated written evaluation (AWE) feedback in a L2 writing class. The research tool applied in the study is *Pigai,* the automated writing evaluation system (AWE) in the Chinese interface. There are a total of 32 students enrolled in a writing class with the same number of students (16) in the experimental and control group. The participants were asked to take a pretest and posttest, which applied to the intermediate level of the GEPT (General English Proficiency Test) writing exam. Another source of data analysis is the questionnaire responses after the treatment. The results of the questionnaire's analyses demonstrate that students generally had positive opinions of *Pigai's* suggestions for their writing. However, some students reported that the incorporation of automated feedback did not always lead to visibly better-amended versions from the students. They needed to have teachers' feedback or guidance about the writing recommendations from the tool. The use of AWE technology in EFL writing classrooms and its implications for writing teaching are also discussed.

Keywords: Automated Writing Evaluation (AWE), L2 writing, Feedback, Perception

I.   Introduction

Automated writing feedback (AWE) is an area that has received growing attention in L2 teaching and learning. It was referred to as automated essay scoring (AES) in earlier times and has received a lot of close attention in L2 teaching writing (Li, 2021). This was because of the rapid development of Artificial Intelligence (AI) and Natural Language Processing (NLP) which not only alters how people gain access to and obtain information but also accelerates the accumulation of knowledge resources.

With the rapid development of technology, numerous L2 writing tools have been developed. Included among these are sophisticated automated writing evaluation (AWE) systems that aim to provide more precise and adaptable options for writing recommendations. The impacts of AWE feedback are typically measured by looking at how pupils' writing improves (e.g., Li, Link, and Hegelheimer, 2015). How students interpret the information they receive from the computer has received much less consideration. Therefore, understanding L2 learners' perceptions towards AWE feedback is essential given that learner perceptions are "crucial determinants in their performance as writers" (Zamel, 1987, p. 699) and that students may ignore the feedback when their expectations about the feedback are not met (Swain, 2006).   It has thus become a hotly debated topic how to teach students to search for and use information efficiently on the AWE systems in order to improve their writing problems. To fill up these crucial gaps, the current study used a mixed-methods research approach to examine the effects of the incorporation of *Pigai*, the AWE system with Chinese interface and instructions, with a focus on (1) how students view the usefulness of automated written feedback and (2) how the incorporation of *Pigai* affects the revision quality of students' writing.

II.   Literature Review

A. Automated Writing Evaluation

Automated Writing Evaluation (AWE) is a computer program that evaluates written texts automatically, awarding a general grade and/or offering comments on areas like grammar, mechanics, content, organization, vocabulary use, or style (Warschauer & Ware, 2006). It was initially created with the intention of awarding written texts summative scores. The tool has developed over the past few decades and can now offer thorough automatic feedback (Nunes et al., 2021). As a result of giving

students more chances to organize, draft, and revise written materials with the aid of AWE feedback, the use of AWE has grown in popularity in educational environments such as schools and universities (Grimes & Warschauer, 2010). The tool can also assist in lessening the effort of teachers in EFL/ESL writing classrooms where big class sizes are frequently the norm. It was recognized to be able to provide individual feedback in numerous draft writings (Warschauer, 2006).

The quality of AWE feedback might be related to students' different proficiency levels. For example, Yang and Meng (2013) discovered, in a CALL peer review exercise, that less proficient students improved more during text revision than more competent students did after their online feedback instruction on error repair. The weaker students were more adept at detecting and correcting both local (i.e., grammatical) and global (i.e., text development, organization, and style) problems in their own and peers' texts. Teachers may need to apply different methods in applying AWE tool in a writing classroom. To combat these challenges and make automated feedback meaningful, Hoang and Kunnan (2016) propose a mixed pedagogy in which the writing instructor acts as a liaison between the students' writing and the automated feedback. Research indicates that this type of feedback is the most successful (Choi & Lee, 2010) because teachers may provide instruction on how to utilize the software most effectively, provide clarification of problems (Hoang & Kunnan, 2016), and reduce errors generated by the program.

B. Challenges of AWE

Though AWE systems have been applied widely in the L2 writing classroom, there remain unsolved challenges and problems. Depending on their corpora and processing capacities, existing AWE systems may generate more or fewer linguistic errors or false alarms, and their judgment of student text quality may vary, at least with variances in a topic, genre, or rhetorical mode (Lai, 2010; Ranalli, 2018).

Another study by Liao (2016) also pointed out that teachers' guidance is crucial in leading to effective use of AWE tools. Liao demonstrated that without a teacher's guidance, the autonomous usage of AWE might frustrate pupils and consequently hinder their writing education. With these inherent constraints in mind, language teachers must be aware that the adoption and implementation of AWE in their own classrooms requires strong pedagogical designs and considerations for writing instruction. (Chen & Cheng, 2008) Moreover, several researchers have also brought forth a number of concerns with the use of AWE. One of these concerns is the feedback's ambiguity, as AWE makes no specific recommendations for students to enhance their capacity to provide coherent, relevant messages in their writing (Lai,

2010). Additionally, AWE feedback is predefined by computer programming, which restricts its capacity to provide deep meaning negotiation and has a negative impact on the writing's ability to grow its content (Chen & Cheng, 2008). Another argument against the adoption of AWE is that it disadvantages students who have less technological experience (Hoang & Kunnan, 2016). These objections to the use of AWE must also be considered when teaching writing.

C. *Pigai*

Apart from corrective feedback, it offers holistic scoring, ranking, highest and lowest scores, and end comments. Its scoring model is calibrated with a large corpus incorporating standard English, students' English essays, another English textbooks. *Pigai* generates an overall score for an essay by calculating its quantitative differences from those texts in its corpus regarding four areas: vocabulary, sentence, structure and organization, and content relevance. The default formula for scoring preset as vocabulary 43%, sentence 28%, structure organization 22 %, and content relevance 7 %, but users can adjust the weighting of each area according to assessment purposes. Similarly, the default length of an essay is between 100 to 500 words, which can also be adjusted. The accuracy rates of identifying mechanics errors and grammar errors reported by Bai and Hu (2017) are 98 % and 59 %, in comparison to 50 % and 63 % documented by Dikli and Bleyle (2014) for *Criterion.* The major functions of *Pigai* are listed in Table 1.

Table 1. The major functions of *Pigai* automated writing evaluation system

Functions	Descriptions
Identifying spelling errors	Errors in spelling and capitalizing
Identifying mechanical errors	Errors in the use of punctuations
Identifying content words-related errors	Errors in using the morphological form of nouns, verbs, and pronouns; errors in ranking the order of different adjectives and adverbs; misusing adjectives as adverbs and vice versa.
Identifying function words-related errors	Misusing or lacking articles, prepositions, and conjunctions

Identifying collocation errors	Grammatically incorrect collocations (e.g., too much things); Non-native like expressions
Identifying syntactic errors	Errors leading to an incomplete sentence structure
Giving suggestions on lexical words	Collocation suggestions, synonyms recommendations
Giving tips on syntactic use	Sentence structure suggestions

III. Method

A. Participants

The present study was conducted in the context of an 8-week English writing class in a university of technology in Taiwan. English writing is a mandatory course required for all English-major students in the two-year program. Since a writing course requires a lot of teachers' efforts in correcting students' compositions and providing feedback, the writing course in the program was divided into two classes taught by two instructors at the same course schedule. The participants in this study were 32 students enrolled in an English writing course taught by two teachers. All of the individuals were approximately 20-21 years old in the two-year program. Students' proficiency levels are diverse from A1 to B2 level.

The subjects were thus separated into experimental and control writing groups according to their student identification number with each group of 16 students. During the 2020 semester, one class was instructed on how to use *Pigai* to fulfill their writing assignments. The other class, which served as the control group, received traditional teacher-led writing instruction, with teachers serving as the primary source of control throughout the writing process.

B. Instrument

The AWE tool used for the present study is *Pigai*. This system provides an overall score and general comments on students' writing, as shown in Figure 1.

Figure 1 A screenshot of feedback and overall score by *Pigai*

Additionally, the system provides diagnostic recommendations in relation to a range of formal characteristics of writing in written feedback for each sentence in the composition as shown in Figure 2. It was shown that the AWE systems pointed out the possible vocabulary misuses (*course* and *coarse*) and synonyms recommendations (pros and cons). For EFL learners, it is helpful for the AWE tool to provide synonyms and antonyms from certain words so that they can diversify word usages and choices in the composition as demonstrated in Figure 2.
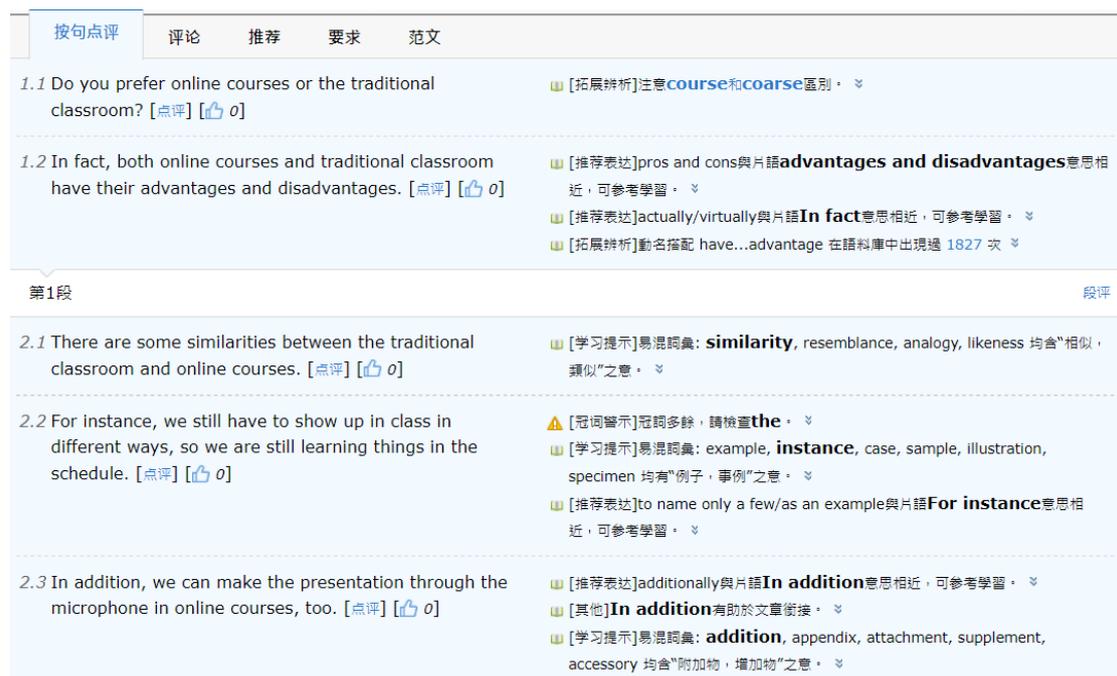
Figure 2 A screenshot of lexical choices by *Pigai*

In addition to pointing out lexical and punctuation errors, *Pigai* also provides descriptions of students' grammatical mistakes. (Figure 3)

Figure 3 A screenshot of verb errors suggested by *Pigai*

C. Procedure

As the purpose of the present study is to explore the effects and students' perceptions of using *Pigai*, an AWE tool, between an experimental group and a control group, a writing composition pretest was conducted at the beginning of the spring semester 2020 before the students in the experimental group were trained to use the AWE tool. The writing topic of the pretest was adopted from a GEPT writing

test at an intermediate level, the General English Proficiency Test, which was created and is given by the LTTC. The test was targeted to assess Taiwanese English learners of all levels and intended audience for the General English Proficiency Test, which was created and is given by the LTTC. This exam complies with Taiwan's English education system, satisfies the unique self-assessment requirements of English learners in Taiwan, and offers organizations or schools a benchmark for assessing the English proficiency levels of their job candidates, employees, or students. The GEPT has received wide recognition since it was launched in 2000. As of 2021, over 8.5 million English learners, from students of all school levels

With the funding of the department, each student in the treatment group was assigned a *Pigai* account. All students in both groups took the pre-test in writing in the second week and the posttest in the ninth week of class. To ensure that the pre- and post-tests were comparable, they were administered under identical writing conditions. Both tests were identical in terms of text type (opinion essay), setting (classroom), length (120 to 150 words), and duration (within 40 minutes) without the use of reference materials such as online dictionaries or the internet. Prior to using *Pigai*, they were required to disable the autocorrect and spelling and grammar features of Microsoft Word to prevent the software from pre-screening the essays for errors. In addition, students were instructed not to seek assistance from their peers and teachers when completing their essays and grammar self-study. After the post writing test, students in the treatment group were asked to fill out a perception survey in using *Pigai* via Google Forms. A 5-point Likert-scaled questionnaire (from 1=strongly disagree to 5=strongly agree) with twelve items was made to gather information regarding the participants' perceptions toward the use of *Pigai* (Appendix A). Those items were written based on the previous studies related to AWE uses (Hyland & Hyland, 2006).

IV．**Results and Discussion**

The purpose of the current study was to examine students' perceptions to using *Pigai* and whether there was a significant difference between the treatment and control groups.

The comments from the treatment group support the use of Pigai for L2 writing. The findings indicate that Pigai editing tools helped students understand grammar rules, exercise self-directed learning, pay attention to forms, and identify discrepancies between their usage and the correct form. The qualitative data also reveals that learners perceive affordances differently and are influenced differently by these affordances in terms of writing correctness.

Table 2. Results of participants' general perceptions toward the use of *Pigai*

Item	Mean	SD
General perceptions		
1.The feedback can help me correct grammar errors in writing.	4.32	.82
2. The feedback can help me correct vocabulary misuses in writing.	3.88	.86
3 The feedback can help me correct punctuation errors in writing.	3.82	.78
4. The feedback can help me correct article errors in writing.	4.24	.86
5. The suggestions of synonyms by Pigai are helpful to me.	3.68	.72
6. I can understand feedback by *Pigai*	4.22	.78
7. I need teachers' help to understand the feedback by *Pigai.*	4.36	.92
8. I need classmates' help to understand the feedback by *Pigai*.	3.66	.78
9. Overall, I think *Pigai* can help me improve the quality of my writing.	3.85	.92
Overall Experience		
10. Overall *Pigai* can help me improve my grammar.	4.34	.68
11. Overall *Pigai* can help me enlarge my vocabulary.	4.54	.90
12. Overall *Pigai* is a useful writing tool.	4.18	.88

As shown in Table 2, the participants' impressions of the use of *Pigai* were generally favorable, and many of these positive perceptions are comparable with the results of prior studies. Participants in both our study and that of Dikli and Bleyle (2014), for example, reported that they felt comfortable using AWE tools and that they could use the tool to improve their vocabulary and grammar. Our participants believed that *Pigai* could enlarge their overall grammar and lexical knowledge, which partially echoes the empirical findings of Liao (2016) and Ranalli (2018) that AWE corrective feedback can enhance L2 learners' writing performances.

V. Conclusion

The purpose of the current study was to determine how a group of Taiwanese students studying English as a foreign language would respond to the addition of automated feedback in the writing classroom. The study focused on how students perceived *Pigai* system and how integrating this AWE tool affected the quality of their revisions. The results demonstrate that despite the observed technological limitations, students were positive about the feedback provided by *Pigai*, and that the integration of AWE feedback did not always produce observable improvements in the students' revised drafts. Moreover, some students reported that they need teacher's explanations on the feedback suggested by *Pigai*. A more detailed training session in teaching the tool prior the writing class is thus needed.

Several practical implications can be drawn from the present study for the use of AWE technology and teacher intervention in EFL writing instruction. First, the research supports the use of AWE tools by EFL students who are still learning the language. Teachers may need to monitor students' interpretations of the machine-generated comments so that students may acquire the accurate writing skills in their writing assignments. The study also serves as a reminder that teaching writing is not always made simpler through technology use in language classes. In order to have effective corrective feedback, teachers may need to make more efforts in designing pedagogical activities in order to maximize the benefits and minimize the issues caused by the AWE technology, rather than assuming that the additional feedback source will help students to a greater extent. Finally, the results of this study suggest that in exam-driven, accuracy-oriented EFL contexts, language-related writing issues may be overemphasized. As a result, writing instructors in these contexts may want to think about guiding students toward a more comprehensive understanding of what constitutes a good writing.

References

Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: effects on L2 writing accuracy. *Computer Assisted Language Learning.* https://doi.org/10.1080/09588221.2021.1936071

Chen, C. F. E., & Cheng, W. Y. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness if EFL writing classes. *Language learning & Technology*, *12*(2), 94-112.

Dikli, S., & S. Bleyle. (2014). Automated essay scoring feedback for second language writers: how does it compare to instructor feedback? *Assisting Writing*, *22*, 1-17.

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning, and Assessment*, *8*(6), 4-43.

Guo, Q., Feng, R. L., & Hua, Y. F. (2021). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning.* https://doi.org/10.1080/09588221.2021.1879161

Hoang, G. T., & Kunnan, A. J. (2016). Automated Essay Evaluation for English Language Learners: A Case Study of *MY Access. Language Assessment Quarterly*, https://doi.org/10.1080/15434303.2016.1230121.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83-87.

Lai, Y. (2010). Which do students prefer to evaluate their essays: Peers of computer program. *British Journal of Educational Technology*, *41*(3), 432-454.

Liao, H. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach, *System*, *62*, 77-92.

Li, J., S. Link, and Hegelheimer. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, *27*, 1-18.

Li, J., & Li, M. (2022). Assessing L2 writing in the digital age: Opportunities and challenges. *Journal of Second Language Writing*, *57*, https://digitalcommons.georgiasouthern.edu/writing-linguistics-facpubs/251

Li, M. (2021). *Researching and Teaching Second Language Writing in the Digital Age*, Palgrave Macmillan Publishing Co.

Nunes, A., Cordeiro, C., Limpo, T., & Castro S. L. (2021). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning*, *38*(2), 599-620.

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning, 31*(7), 653-674.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, *10*(2), 157-180.

Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing*, *27*(3), 291-300.

Zamel, V. (1987). Recent research on writing pedagogy. *TESOL Quarterly*, *21*(4), 697-715.

Appendix A Student questionnaire: perceptions of using *Pigai*

1.The feedback can help me correct grammar errors in writing.
2. The feedback can help me correct vocabulary misuses in writing.
3 The feedback can help me correct punctuation errors in writing.
4. The feedback can help me correct article errors in writing.
5. The suggestions of synonyms by Pigai are helpful to me.
6. I can understand feedback by *Pigai*
7. I need teachers' help to understand the feedback by *Pigai.*
8. I need classmates' help to understand the feedback by *Pigai*.
9. Overall, I think *Pigai* can help me improve the quality of my writing.
10. Overall *Pigai* can help me improve my grammar.
11. Overall *Pigai* can help me enlarge my vocabulary.
12. Overall *Pigai* is a useful writing tool.

中文摘要

　　本研究旨在探討台灣學生對第二語寫作課中自動寫作評估（AWE）反饋的感知研究。研究中應用的研究工具是*Pigai*(批改網)，以中文界面為主的自動寫作評估系統（AWE）。研究對象包括寫作課32名學生，實驗組和對照組的學生人數相同（各16人）。資料分析的主要來源以全民英檢中級寫作題目為前後測、並納入自動寫作評估系統問卷回覆。問卷分析結果表明，學生普遍對*Pigai*(批改網)的寫作建議持正向態度。然而，一些學生回覆表達自動反饋的說明不總能提供更好或清楚的修正版本。他們仍需要從該工具中獲得教師對寫作建議的反饋或指導。研究最後也提供自動寫作評估系統技術在英文為外語教學寫作課堂中的使用及其對寫作教學的影響。

關鍵字：自動寫作評估，第二語言寫作，反饋，感知