

# Effect of an Automated Writing Evaluation System on Students' EFL Writing Performance

Limei ZHANG<sup>a\*</sup> & Zheng HUANG<sup>b</sup>

<sup>a</sup>*Nanyang Technological University, Singapore*

<sup>b</sup>*Shanghai Jianqiao University, China*

\*limeizh2008@yahoo.com

**Abstract:** Recently, a growing number of studies were conducted on the use of automated writing evaluation (AWE) systems in writing classrooms. As an online writing evaluation system, Pigai has been used more and more widely in college-level English as a Foreign Language (EFL) writing courses. The current study examines the accuracy of the feedback from Pigai and the effect of the AWE system on Chinese college students' EFL writing performance. 97 students from an academic writing course were invited to participate in the study. This study addresses the following two research questions: 1) How accurate is the feedback from Pigai in terms of the errors it can identify in students' writing? 2) What is the effect of using the AWE system on students' essay performance? Results showed that the average precision rate of Pigai is 58% and the accuracy varies across error types. In addition, Pigai improved students' writing performance significantly in terms of the quality and length of essays. The findings will inform discussion of whether and how to integrate the use of online writing evaluation systems into writing classrooms.

**Keywords:** automated writing evaluation, Pigai, writing performance

## 1. Introduction

Recently, a growing number of studies were conducted on the use of automated writing evaluation (AWE) systems in writing classrooms, which includes Criterion (ETS), Project Essay Grade (PEG) Writing (Measurement Incorporated), My Access! (Vantage Learning) etc. (Attali, 2004; Chodorow, Gamon, & Tetreault, 2010; Chen & Cheng, 2008; Lavolette, Polio, & Kahng, 2015; Ranalli, Link, & Chukharev-Hudilainen, 2016). Researchers found that the use of AWE systems can affect students' writing performance positively. For example, it was reported that the use of Criterion led to a significant decrease in the number of errors in learners' resubmissions (Ranalli, et al., 2016).

Pigai, an AWE system developed in mainland China, has been used more and more widely in college-level English as a Foreign Language (EFL) writing courses. However, few studies have been conducted on the system itself and its effect on Chinese EFL students' writing performance. The current study examines the accuracy of the feedback of Pigai and its effect on EFL students' writing performance.

## 2. Literature Review

Automated writing evaluation (AWE) systems, which combine automated essay scoring with automated feedback (Grimes & Warschauer, 2010), provide students cycles of writing practices and formative feedback, which can reduce demands on teachers. Previous studies have shown that AWE systems can be as reliable as or more reliable than human raters in assigning scores.

An accurate AWE system can provide students with accurate information to target relevant areas of revision, improvement and learning (Attali, 2004). Consequently, they can help improve students' writing performance (Lavolette, et. al, 2015; Ranalli, et. al, 2016). The current study examines the accuracy of the feedback from Pigai and the effect of the AWE system on Chinese college students' EFL writing performance.

### 3. Research Questions

There are two research questions in the study, as follows:

- 1) How accurate is the feedback from Pigai in terms of the errors it can identify in students' writing?
- 2) What is the effect of using the AWE system on students' essay performance?

### 4. Research Methods

#### 4.1 Participants

There were 97 first-year Chinese college students participating in this study. They were from nine different departments in a Chinese university.

#### 4.2 The Writing Task

Participating students were required to write an essay entitled "How will AI affect our life?" The writing prompt was as follow:

Recently, Google's AlphaGo defeated Lee Sedol, the world Go Champion, 4 to 1 in a five-game match. The machine's sweeping victories have once again made AI (artificial intelligence) a hot topic. Some people welcome the progress and expect AI to benefit mankind in more fields. Some others fear that AI will eventually get out of control. What is your view? How will AI affect our life? Write an essay in response to the questions. Give reasons to support your points of view.

#### 4.3 Procedure

The students were first required to complete the essay using Pigai and submit it to get the score and feedback from Pigai. Next, they were required to revise their essays following the feedback from Pigai and get a new score. Meanwhile, human raters corrected the two essays.

#### 4.4 Data analysis

Descriptive statistics were calculated first, which includes mean and standard deviation of the scores of students' essays. Accuracy ratings of the error types were checked. Pair-sample *t*-tests were conducted to compare the scores of different versions of students' essays.

### 5. Results and Discussion

#### 5.1 Descriptive statistics

Table 1 shows the descriptive statistics of the participants' compositions.

Table 1 Descriptive Statistics of Participants' Compositions

	Mean	SD
First draft (Pigai)	86.31	4.27
Second draft (Pigai)	88.96	3.34
Teachers' score-1	10.54	1.38
Teachers' score-2	11.28	1.29
Word Count-1	243.00	59.10
Word Count-2	248.23	60.59

#### 5.2 Accuracy

Results showed that there were 16 types of error identified by Pigai, including four types of grammar mistakes, ten types of usage mistakes and two types of mechanic mistakes. Among the 471 mistakes identified, 272 were correct, indicating the average precision rate is 58 %. For mechanic errors the precision rate is as high as 75%. Correlation between human rater' score and Pigai' is .59,  $p < .05$ . Table 2 shows the accuracy rating for ten most commonly identified error types.

Table 2 Accuracy ratings for ten most commonly identified error types

Error type	No. of errors	Precision	Not precise	% Precision	Category
Spelling mistakes	103	92	11	89.3	M
Wrong articles	17	12	5	70.6	U
Determiner noun agreement	30	21	9	70.0	U
Ill-formed verbs	21	14	7	66.7	G
Capitalization	43	26	17	60.5	M
Subject-verb agreement	102	53	49	52.0	G
Preposition error	8	4	4	50.0	U
Part of speech error	6	3	3	50.0	U
Senses error	10	4	6	40.0	G
Collocation error	32	9	23	28.1	U

Note: G=grammar; M=mechanics; U=usage

### 5.3 Effect of using AWE on students' essay performance

Pair-sample *t*-tests were conducted with students' first draft and second draft which were scored by Pigai. Similarly, pair-sample *t*-tests were conducted with the number of words they wrote for the two rounds of essay-writing as well as the two drafts teacher-scored essays. Results showed that  $t(96)=-9.67, p<.05$ ;  $t(95)=-7.27, p<.05$ ;  $t(96)=-4.15, p<.05$ .

As the results indicated, the feedback students received from the automated writing evaluation system, i.e., the Pigai has improved their performance significantly, in terms of the quality as well as the length of the essay. This has been supported by the results of teachers' evaluation.

## 6. Conclusions and Suggestions for Future Research

The study examines the accuracy of the feedback from Pigai in terms of the errors it can identify in students' writing and its effect on students' writing performance. It was found that the average precision rate of the AWE system was 58%. With the use of Pigai, students' writing performance was significantly improved.

Due to limited time and resources, the current study was tentative and exploratory in nature. It is hoped that future studies can be conducted with longer duration and larger sample sizes through multiple data collection methods.

## References

- Attali, Y. (2004). Exploring the feedback and revision features of Criterion. *Journal of Second Language Writing, 14*, 191–205.
- Chen, C., & Cheng, W. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology, 12*(2), 94–112.
- Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for english language learners: Feedback and assessment. *Language Testing, 27*(3), 419–436.
- Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment, 8*(6), 1–44.
- Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning and Technology, 19*(2), 50–68.
- Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2016). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology, 37*(1), 8–28.