

Does the Automated Essay Scoring System Reduce Motivation to Learn English Writing? A Case Study of Chinese College Students

Zhixin Gui, Duoduo Zhang, Yiming Bai, and Jieqian Ouyang

School of Design, Hunan University, Hunan, 410082, P.R. China

ABSTRACT

Artificial intelligence technology has given rise to the Automated Essay Scoring (AES) system, which has greatly reduced the workload of English teachers. It can give overall scores, error feedback and writing suggestions in seconds, after students have uploaded their English essays. In China, the most widely used AES system is called Pigai, which has been shown to help improve English writing skills, but has caused resistance from some students who believe it reduces motivation and creativity in English writing. Using a case study approach, this study investigated 29 Chinese English learners who had used Pigai. Based on the Cognitive Process Theory of Writing, quantitative analyses were used to assess motivation in three writing stages: planning, translating and reviewing. The results showed that the negative impact on students' motivation occurred in the planning and translating stages for three main reasons: lack of interpretability, unsuitability of recommendation algorithms and inadequate feedback mechanisms. In addition, based on the results of the questionnaire and interviews, we made some design suggestions for the AES system to improve user experience.

Keywords: Automated essay scoring, Writing, Motivation, Pigai, User experience design

INTRODUCTION

Writing composition is a significant factor for measuring test-takers' ability in the English language exam (Hussein, Hassan and Nassef, 2019). There is an extremely large group of English language learners in China who are usually asked to write an essay on a specific topic. Teachers then grade these essays according to specific scoring criteria. However, manual scoring is subjective and the scorers are influenced by factors such as Judge severity (Lunz, Wright and Linacre, 1990) and Context effect (Hughes and Keeling, 1984), so the scoring of the same essay may vary widely. To obtain more accurate scores and improve efficiency, Automated Essay Scoring (AES) systems have been developed over the past decades and have been rapidly developed in China. AES systems typically use natural language processing and machine learning techniques to automatically analyse the quality of essays and assign scores (Dikli, 2006). In addition, it provides immediate and corrective feedback on the quality of second language learners' writing and is considered

an important supportive computational tool in educational settings (Zupanc and Bosnić, 2017). AES systems have been used in large-scale, high-stakes writing assessments such as the TOEFL (iBT), GRE and GMAT as well as in low-stakes settings such as regular classroom practice. They appear to have become a valid and reliable alternative to human scorers. One of the most widely used AES systems in China is Pigai, which is commonly used for low-stakes writing assessment in classrooms.

Most of the existing research focuses on assessing the reliability of AES systems, meaning whether and to what extent it can replace human scoring. There has been much literature demonstrating high agreement between scores generated by AES systems and human (Attali, 2004; Rudner, Garcia and Welch, 2006; Shermis et al., 2010). In China, a number of scholars have also demonstrated the effectiveness of local AES systems. Song (2019) found that Pigai not only improved college students' overall writing proficiency, but also had a positive impact on all three indicators except text structure, with extremely significant performance gains in two indicators, vocabulary and sentence. Cheng (2020) found that Pigai's rich resources led to an increase in students' knowledge and a significant advantage in boosting learning confidence. However, controversy and opposition also exist. Some studies concluded that computer software cannot score students' writing as effectively as humans (Herrington and Moran, 2001; Cheville, 2004; McCurry, 2010). Most AES systems did not take into account the context of writing, such as target audience and genre (Patout and Cordy, 2019). Focusing on vocabulary level and grammatical analysis, it can serve only as a cursory assessment of general writing ability and needs improvement in terms of logic, relevance and fluency (Min, 2019). It has been emphasized that AES systems that support writing strategy development and encourage self-monitoring to improve macro-level text quality (e.g., argumentative structure, rhetorical moves) are sorely lacking (Strobl et al., 2019). Several scholars have studied Chinese AES systems other than Pigai, named iWrite and Bingo. The former was noted as not being a valid tool for assessing writing and predicting human scores, and it should be limited in its use (Qian, Zhao and Cheng, 2020). The latter is considered to be less accurate than manual scoring to reflect individual student differences (Zeng, 2014).

After a critical review of the existing literature, we found that most studies focus on the AES system itself and rarely address the impact of the widespread use of the AES system on English language learners. In fact, its distinguishing feature from manual scoring, online real-time feedback, has largely changed the way English writing is done. In order to get good grades, some students even cheat the AES system by manipulating the scoring rules, such as repeating the same sentences so that illogical compositions can be scored high (Powers et al., 2002). Few scholars have focused on emotions or motivation, examining whether the use of the AES system leads to psychological resistance or contempt among students, thus reducing motivation to learn English writing. It is this point that prompted the present study. We believe that this research is relevant because it will provide useful insights into the application of AES systems in China, elucidate how AES systems have affected English writing learners, and thus provide theoretical support for how to design better

AES systems. The purpose of this paper is to examine the following issues: Does the AES system reduce the motivation to learn English writing? What are the potential effects of the AES system on the different phases of writing? What are the reasons for the negative impact of the AES system?

About Pigai

Pigai (www.pigai.org) was developed by Beijing Ciwang Technology Co., Ltd. It has been selected by over a thousand schools as a platform for their students' English writing. Based on the corpus and cloud technology, the system compares the distance between students' compositions and the standard corpus and maps them into scores. It is claimed to be able to give feedback on English essays in one to two seconds with scores, overall comments and per-sentence reviews for critique. In this process, each essay is measured on 192 indicators, which are weighted and averaged to create an overall score and scores in four dimensions: vocabulary, sentence, text structure and theme. Pigai also points out spelling errors and grammatical mistakes and provides extended knowledge, advanced vocabulary and example sentences. Students can revise their essays based on the feedback from the system and then submit them again until they receive a satisfactory score.

METHODS

Participants

Twenty-nine undergraduate students from Hunan University were recruited to voluntarily participate in this experiment. The average age of the participants was 20 ± 2 years. All had recently used Pigai and could therefore make accurate judgments about the questions associated with it. They were asked about their English proficiency level based on College English Test Band 4 (CET-4) scores prior to the experiment. This is a national test administered by the Chinese Ministry of Education and is very popular in China, with a total score of 710. Three of the participants scored below 425 (failed), 14 scored between 425 and 520, 8 scored between 521 and 600, and 5 scored above 600.

Validation Framework

One of the most widely circulated and influential models of writing is the Cognitive Process Theory of Writing (Hayes and Flower, 1986), which views writing as a complex process of thinking and problem solving. The process consists of three major systems: the task environment, the writer's long-term memory and the writing process. The writing process is divided into three stages: planning, translating and reviewing. We use this model to analyse the process of English writing in the AES system environment. First, in planning, which consists of three sub-processes: generating ideas, organizing ideas and setting goals, the user interface of Pigai provides the topics and requirements for each English writing. After students read, the first thing they do is to understand the meaning of the topic and follow the requirements to generate

ideas in their minds. Next, in translating, students begin to generate sentences from the ideas in their heads to form a draft. This process takes place in the edit box of Pigai's user interface, which allows them to enter and modify text. Finally, in reviewing, students are given a score and suggestions for changes by Pigai. They can choose to make changes to the errors pointed out or, alternatively, they can ignore the suggestions and submit the score directly.

The central question of this study is whether the AES system reduces motivation to learn English writing. In order to collect fine-grained data, we divided motivation into three sub-indicators, namely, planning motivation, translating motivation and reviewing motivation, based on the Cognitive Process Theory of Writing described above. Before the experiment began, we crafted a questionnaire based on a five-point scale, which meant that each question had five options (2= fully positive, 1= positive, 0= uncertain, -1= negative, -2= fully negative). In this way, it is clear to compare the motivation of different writing stages. A score > 0 means positive and a score < 0 means negative. The main body of the questionnaire contains three sets of questions corresponding to the planning, translating and reviewing stages of writing. First, for the planning stage, two questions were set to find out whether Pigai influenced the students' writing ideas and in which dimensions this influence occurred. Since students knew that the overall score given by Pigai depended mainly on the scores of the four dimensions (vocabulary, sentence, text structure and theme), this was likely to lead students to use the most advanced vocabulary, complex sentences and fixed text structures possible, thus limiting their writing ideas. Second, for the translating phase, we set three questions to investigate whether writing on Pigai's interface leads to slackness and resistance, compared to writing traditionally on paper. Finally, for the reviewing phase, we set two questions to examine how students perceive the ratings and suggestions given by Pigai, and to what extent they would follow the suggestions, i.e., how strongly they are motivated to revise. At the end of the questionnaire, we interviewed the students for about 15 minutes to understand their dissatisfaction with their experience of using Pigai.

RESULTS

Based on the valid data collected from the questionnaire, we conducted a quantitative analysis using SPSS and Excel software. The reliability and validity of the questionnaire were first tested. The results showed that our questionnaire was statistically feasible (Cronbach's alpha = 0.870 > 0.8). And the communality values of all research items are higher than 0.4, which indicates that the research item information can be effectively extracted.

The overall results showed that the negative impact on motivation to learn English writing occurred in the planning and translating stages, with scores of -0.15 for both (see Table 1). Precisely, 43.33% of the participants believed that using an AES system like Pigai would negatively affect the planning stage of writing, and 26.67% believed that it would negatively affect the translation stage of writing. In order to get a high score on Pigai, students tend to cater to its evaluation criteria by using as many advanced vocabulary and sentences as possible. The difference is evident from the statistics. Participants placed

Table 1. The average score of motivation in the three phases of writing (2= fully positive, 1= positive, 0= uncertain, -1= negative, -2= fully negative).

Phases	Average score	Related Questions	Average score
Planning	-0.15	Q1. How did you feel when you were told that your English essay would only be viewed and graded by Pigai?	-0.4
		Q2. How did writing on Pigai's user interface influence your conception?	0.1
Translating	-0.15	Q3. How comprehensive was the feedback you received from Pigai during the writing process?	-0.17
		Q4. How motivated were you to write after getting familiar with Pigai?	-0.27
		Q5. How helpful was Pigai in improving your English writing?	0.00
Reviewing	0.40	Q6. How did you feel after seeing the score given by Pigai?	0.17
		Q7. How strong was your willingness to revise your essay after seeing the advice given by Pigai?	0.63

0.37, 0.45, 0.07 and -0.07 emphasis on vocabulary, sentence, structure and theme, respectively (see Table 2). This suggests that using an AES system like Pigai does affect students' motivation to learn English writing, as evidenced by a greater focus on vocabulary and sentences rather than the structure and theme of the essay.

Details of the data are next discussed in order according to the three writing phases. First, in the planning phase, participants reported a slightly negative emotional disposition (mean score of -0.4 in Q1), indicating that it would have a slightly negative psychological impact on students when English essays were only viewed and scored by Pigai. In detail, 23.33% of the participants chose "negative", 20% chose "completely negative", and 36.67% chose "unsure ". The distribution of the data shows that the degree of negativity varies from person to person. In addition, the overall attitude of the participants was close to neutral (mean score of 0.1 in Q2) as to whether writing on Pigai's user interface affects conceptualization. It is worth noting that 36.67% of the participants thought that Pigai would limit their writing ideas. Second, in the translating phase, which is a process that may involve

Table 2. Average scores of participants' effort in four dimensions (vocabulary, sentence, text structure and theme).

Dimensions	Average score
Vocabulary	0.37
Sentence	0.45
Text Structure	0.07
Theme	-0.07

multiple cycles, the user completes the first draft, gets feedback from Pigai, and may revise it again to form a second draft. Therefore, the questionnaire asked about the comprehensiveness of the feedback given by Pigai, the motivation to write with Pigai and the effectiveness of Pigai in improving writing. From the results, the mean scores for these three aspects were -0.17 , -0.27 and 0.00 , respectively. This indicates that the feedback given by Pigai was less than satisfactory, that the participants' motivation to write was mildly negative, and that the effectiveness of Pigai for writing improvement was considered insignificant. It is worth pointing out that 26.67% of the participants felt that their motivation to write decreased after using Pigai for a period of time. Finally, in the reviewing phase, after receiving the score given by Pigai, most of the participants had a positive mood, which means they were able to be convinced of the score given by the AES system and next, reassuringly, developed a strong willingness to revise. This suggests that Pigai can have a significant positive impact in the reviewing phase.

The results of the stepwise regression analysis are reported next. On the one hand, we investigated the relationship between participants' motivation and their efforts on the four dimensions of writing (vocabulary, sentence, text structure and theme). The results showed that "theme" was significantly and positively associated with changes in motivation (see Table 3). After using Pigai for a period of time, those who were less focused on theme were more likely to have negative emotions. On the other hand, we also investigated the correlation between the comprehensiveness of the feedback given by Pigai and motivation in the three stages of writing. The value of the regression coefficient for Q3 was 0.353 (see Table 4), which implies that the comprehensiveness of the feedback given by Pigai had a significant positive relationship

Table 3. Stepwise regression analysis results (a).

	Regression coefficients	VIF
Constant	2.115 (5.345**)	-
Theme	0.356 (3.037**)	1.000
Sample size	29	
R ²	0.255	
Adjust R ²	0.227	
F value	F (1,27) = 9.226, p = 0.005	

Table 4. Stepwise regression analysis results (b).

	Regression coefficients	VIF
Constant	0.305 (0.623)	-
Q3	0.353 (2.952**)	1.32
Q6	0.387 (2.474*)	1.523
Sample size	29	
R ²	0.619	
Adjust R ²	0.573	
F value	F (3,25) = 13.530, p = 0.000	

with the change in motivation to learn English writing. For example, those who felt that Pigai's feedback was not comprehensive enough were more likely to be less motivated to write. In addition, the regression coefficient value for Q6 was 0.387 (see Table 4), which implies that students' emotions after seeing the scores given by Pigai were significantly and positively related to changes in motivation to write. For example, those students who were in a worse mood after seeing the scores were more likely to be less motivated.

DISCUSSION

The user interviews helped us understand why Pigai negatively affected motivation during the writing phase of planning and translating, and what dissatisfaction participants had with the existing AES system. Based on the interview results, we make the following recommendations for the experience design of the AES system. First, users need to be guided to familiarize themselves with the functional modules of the AES system, especially when they first start using it. It is recommended to add some graphic guidance to the interface or provide a resident customer service portal in the sidebar. Second, the vocabulary recommendation algorithm needs to be systematically improved to highlight high-frequency words and important phrase collocations, while removing those that are not commonly used or too simple. Provide a way for users to give feedback, such as a "Like" or "Dislike" button next to the suggestions given by the AES system. This will help the system's algorithm to be continuously improved by eliminating some of the less good suggestions. Third, the AES system should improve the interpretability of scores, clearly tell users what the scores consist of, and appropriately reduce the weight of the two dimensions of text structure and theme. When giving the total score, the scores of the different sub-dimensions should also be provided to help the user identify which dimensions need to be worked on to improve. Last but not least, teachers and peers should be involved in scoring essays rather than relying exclusively on the AES system. The ultimate goal is to create a multifaceted feedback mechanism that combines automated essay scoring, teacher scoring and online peer scoring.

CONCLUSION

This study uses a mixed research approach to assess how Pigai, one of the most widely used AES systems in China, affects students' motivation to learn English writing. The Cognitive Process Theory of Writing was introduced to assess motivation in the three writing phases of planning, translating and reviewing. The results illustrate that the AES system leads to negative effects in the planning and translating phases, mainly due to the lack of interpretability, the inappropriateness of the recommendation algorithm and the inadequacy of the feedback mechanism. Based on the above study, we provide useful insights into the user experience design of AES systems, which can help developers and designers create better products. The limitation of this study is that the sample is not broad and rich enough. The participants were all undergraduate students at Hunan University and should be extended to

English language learners from different regions and grades. In addition, this study only focuses on the user experience of the student side of the existing AES system and does not involve the teacher side. Regarding this part, we plan to continue to explore it in future studies to make the AES system better serve students and stakeholders.

REFERENCES

- Attali, Y. (2004). Exploring the feedback and revision features of Criterion. *Journal of Second Language Writing, 14*(3), 191–205.
- Cheville, J. (2004). Automated scoring technologies and the rising influence of error. *The English Journal, 93*(4), 47–52.
- Cheng, L. (2020). A study on the utility of online automatic assessment system in teaching English writing in college with Pigai as an example. *International Public Relations* (in Chinese).
- Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment, 5*(1).
- Hughes, D. C., & Keeling, B. (1984). The use of model essays to reduce context effects in essay scoring. *Journal of Educational Measurement, 21*(3), 277–281.
- Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *American psychologist, 41*(10), 1106.
- Herrington, A., & Moran, C. (2001). What happens when machines read our students' writing?. *College English, 63*(4), 480–499.
- Hussein, M. A., Hassan, H., & Nassef, M. (2019). Automated language essay scoring systems: A literature review. *PeerJ Computer Science, 5*, e208.
- Lunz, M. E., Wright, B. D., & Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied measurement in education, 3*(4), 331–345.
- McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers?. *Assessing Writing, 15*(2), 118–129.
- Min, Z. (2019). An Empirical Study of the Feasibility of Automated Essay Scoring in Teaching EFL Writing in China. In *Proceedings of the 5th International Conference on e-Society, e-Learning and e-Technologies* (pp. 29–33).
- Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2002). Stumping e-rater: challenging the validity of automated essay scoring. *Computers in Human Behavior, 18*(2), 103–134.
- Patout, P. A., & Cordy, M. (2019). Towards context-aware automated writing evaluation systems. In *Proceedings of the 1st ACM SIGSOFT international workshop on education through advanced software engineering and Artificial Intelligence* (pp. 17–20).
- Qian, L., Zhao, Y., & Cheng, Y. (2020). Evaluating China's automated essay scoring system iWrite. *Journal of Educational Computing Research, 58*(4), 771–790.
- Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of IntelliMetric™ essay scoring system. *The Journal of Technology, Learning and Assessment, 4*(4).
- Shermis, M. D., Burstein, J., Higgins, D., & Zechner, K. (2010). Automated essay scoring: Writing assessment and instruction. *International encyclopedia of education, 4*(1), 20–26.
- Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & education, 131*, 33–48.

-
- Song, Y. (2019). An empirical study on the effect of automatic essay scoring system on college students' English writing ability. *University Education*, 10 (in Chinese).
- Zeng, J. (2014). Exploring the application of Bingo English intelligent essay review system in college English writing teaching. *Journal of Wuzhou College*, 24(1), 104–108 (in Chinese).
- Zupanc, K., & Bosnić, Z. (2017). Automated essay evaluation with semantic analysis. *Knowledge-Based Systems*, 120, 118–132.