

Feedback Analysis of PIGAI Online Scoring System in Business English Writing

Bo Hu

School of Foreign Languages
Hubei University of Education (HUE)
Wuhan, China
robbiewho@126.com.cn

Bo Hu

School of Literature, Law and Foreign Languages
Wuchang University of Technology (WUT)
Wuhan, China
robbiewho@126.com.cn

Abstract—The study explores the effect of evaluation feedback provided by PIGAI—an extensively-used online automated essay scoring system in China. Case study method is employed for the analysis of 2 groups of 10 sample essays of Business English to compare and discuss the online feedback data. The results show that the PIGAI scoring system highlights the appropriate expressions, enlarges knowledge of synonyms and collocations, and distinguishes confusing words. But there is still room for improvement for its inability to accurately identify errors or provide appropriate suggestions associated with textual coherence and organization unity.

Keywords—PIGAI; feedback analysis; business writing; case study

I. INTRODUCTION

The Automated Writing Evaluation (hereinafter referred to as AWE) system is a computer-based scoring system for evaluating and judging essay text [1]. The most important feature of AWE is its instant evaluation and feedback. The immediate feedback of the automated scoring system encourages students to polish their essays and improve their autonomous learning ability, thus better achieving the goal of “assessment for learning”. Timely and objective feedback is beneficial for writing teaching and the improvement of students’ writing skills.

II. RESEARCHES ON AWE AND FEEDBACK

The Automated Writing Evaluation (AWE) system is widely used in grading and evaluating writings in various texts as well as in Computer-Aided Language Learning (CALL). In the United States of America and other English-speaking countries, the integration of network technology and foreign language teaching gave birth to the idea of online automatic scoring system, which has been widely used in TOEFL, GMAT and other tests; Chinese researchers also pay attention to automatic scoring system in writing, and many universities and companies independently develop their own writing scoring platforms, the most representative of which is “PIGAI Online Scoring System (hereinafter referred to as PIGAI). There are more than 5,000 universities, colleges and secondary schools apply PIGAI in writing and translation teaching in China.

A. Researches on AWE

In view of different automatic essay scoring systems, researchers from many countries have carried out relevant review and inquiry, suggesting that the automatic scoring system can not only reduce teachers’ workload, but also provide effective and timely feedback on student essay writing [2]. It’s widely acknowledged that the immediate feedback of the automated scoring system encourages students to polish their essays and improve their autonomous learning ability, thus better achieving the goal of “assessment for learning”. Chinese researchers study empirically the application and role of the PIGAI in college English writing teaching, especially focus on the feedback of AWE in a case study and concluded that the reliability of AWE in scoring lower quality essays is to be improved due to some defects [3].

B. Feedback of AWE System

Feedback is a key factor in stimulating and consolidating learning. Fitzgerald defines feedback as “any form of change at any time in the writing process, including identifying the gap between the actual and ideal states of the essay, determining the parts to be modified in the text, and how to modify”[4]. Research on feedback is usually based on two theories: “Focus-on-form” and “Noticing Hypothesis”. “Focus-on-form” put forward that, in addition to emphasizing the meaning of verbal expression, in the process of language communication learners should pay due attention to the use of language in correct form, and such attention is triggered by the understanding process in verbal communication and the difficulties in expressing themselves [5].

The “Noticing Hypothesis” theory presents two important concepts of cognitive linguistics: Noticing and Awareness [6]. Learners must first pay attention to language phenomena and realize the gap between their own expression and target expression, then carry out cognitive comparisons to acquire related language knowledge to develop interlanguage. The application of these

The research is supported by the “College Teaching Research Program 2015” of Hubei Provincial Department of Education (Grant No. 2015440) and WUT.

theories in writing teaching is mainly manifested in the error correction in writing feedback. The AWE system often draws the learners' attention to inappropriate expressions and encourages them to correct such errors by pointing out the inappropriate expressions and providing advice for correction [7].

III. PIGAI AND ITS OPERATING PRINCIPLE

A. PIGAI

The PIGAI provides a system of three major functions: instant online feedback, interactive learning and process monitoring. It is a corpus and cloud computing technology based online automatic scoring system to provide service of evaluating English essays in China. By calculating the distance between the students' compositions and the standard corpus, it instantly generates a score for student's composition, expository notes upon the work and detailed evaluation. It is applied to effectively improve the teacher's efficiency in grading students' English writing and to promote students writing skills through self-study and practice.

There are three main functions: reviews, sentence reviews, self-training. Students submit their essays into the PIGAI, and then the system automatically analyzes the composition, providing scores for the essays in a second. In addition to an overall score, the system also remarks upon the four related dimensions of vocabulary, sentences, text structure and content in details, and finally gives comprehensive comments. The comments generally point out the weaknesses and highlights of the students' essays. Firstly, the feedback helps the students locate their writing skills and knowledge capacity in general; they would be clear about their assignments to progress and plan for it.

B. Operating principle

If students submit a composition entitled "Profile of the Royal Company", the PIGAI gives the following comments: The connection between expressions of the composition is sound, and the essay is rich in vocabulary of process and procedure; the sentence pattern is varied and the use of clauses is abundant. The author is highly skilled and flexible in choice of vocabulary. It is recommended that the author use more academic vocabulary. "Sentence-by-sentence evaluation" is the main function of PIGAI. Receiving the feedback, students can be very intuitive to see their own mistakes. In evaluating the composition, the system will give specific sentence comments, the content of which can be as specific and concrete as focusing on the spelling, punctuation, grammar, words, text coherence and so on.

To score each essay, the system decomposes each essay into 192 sub-dimensions to measure and form a total score and the scores of 4 dimensions: vocabulary, sentences, text and content after being weighted average, and generate digital statistical information of vocabulary, paragraphs, sentences and spelling errors, grammatical errors, collocation errors, Chinglish expressions and other feedback information. Students can modify and resubmit their essays based on such information and feedback, and the system evaluates the writing process continually. Centering on whether the system provides accurate and effective feedback or not, this case study discusses and analyzes the scoring process and result.

IV. METHOD AND PROCEDURE

Based on the Business English Writing and the curriculum requirements, researchers use the PIGAI scoring system to grade sample compositions of different level in Business English Writing to tentatively explore the validity and effect of feedback provided by the system according to statistics of varied evaluation feedback and analyze the influence of automatic essay scoring systems on Business Writing Teaching.

A. Research method

This case study centers on the effect of AWE in Business writing for students majoring in Business English at Hubei University of Education (HUE hereafter). The subject consists of two parts: five sample business essays are quoted from Shirley Taylor's *Model Business Letters, Emails & Other Business Documents* (7th Edition), and five business compositions composed by the third year students majoring in Business English, are collected at the Provincial English Model Center of HUE.

B. Research procedure

The research procedure includes comparing the two sets of scores, one of which was given by PIGAI and the other by the teachers, given to five sample essays of Business English and five student's essays. Then it analyzes in details the online feedback information of the two groups of essays with high scores and low grades. The study mainly discusses the validity of the feedback proposed by the PIGAI scoring system.

The researchers obtained the scores given by writing teachers, and also recorded scores given by the machine by submitting the 5 sample essays into the PIGAI system, then calculated the percentage of the complete consistency of the five subcategories and the degree of adjacent consistency. Researchers also checked the relevance of the person/machine scoring by using the Pearson relevance test. In the analysis of the feedback proposed by the PIGAI system, researchers selected two well-organized essays with high score deviation between human and machine scoring and two poor-written essays, so as to explore the overall evaluation and evaluation of varied classification. Researchers also collected statistics of sentential evaluation to illustrate the advantages of the PIGAI scoring system and analyze the related deviations.

V. RESEARCH RESULT AND FEEDBACK ANALYSIS

A. Research Result

Zhang and Sheng as in [3], explained the person/machine scoring deviation through a consistency and relevance test of person/machine scoring and found that in the 15-point scoring system, person/machine scoring produced exactly different scores and the full consistency of the person/machine scoring is 60% (≥ 13 points, 8-11 points, 5-8 points), among which a composition of 11 points was rated 8 points by the PIGAI scoring system, and essays of 2 points was given 5 points. In this study, the average scores of the subject graded by human and machine are both 8 points, while the standard deviation of human scoring is greater than that of the machine scoring. Researchers of this study received similar results as Zhang and Sheng did in their CET-4 samples research.

From the overall evaluation, the PIGAI online scoring system grades the essays of 13-15 points higher while grades the essays of 1-3 lower. In order to get a better understanding of the system's evaluation on vocabulary and syntax, we analyze the feedback of different scoring objects by the system one by one. The system also provides feedback of classified evaluation provided automatically by the PIGAI scoring system.

In addition to the table of feedback statistics, the PIGAI scoring system also provides evaluation feedback sentence by sentence, including more than twelve aspects: Highlight, Learning Tips, Correction Tips, Synonym Discrimination, Wrong Sentence, Low Frequency Warning, Collocation Errors, Noun Errors, Verb Errors, Article Error, Wrong Spelling, Wrong Capital Letter or Small Letter and Wrong Punctuation. "Highlight" is positive feedback, suggesting some excellent expression in the essay, such as the effective use of proverbs. "Learning Tips" and "Synonym Discrimination" are supplement to extend students' language knowledge. "Correction Tips", "Wrong Sentence", "Low Frequency Warning", "Collocation Errors", "Noun Error", "Verb errors", "Article Error", "Wrong Spelling", "Wrong Capital Letter or Small Letter" and "Wrong Punctuation" are all negative feedback, indicating that there are certain problems in the writing. In general, essays with high scores have more positive feedback, and those with low scores get more negative feedback.

B. Feedback Analysis

1) Vocabulary

From the perspective of vocabulary, sample essays of 13-15 points contain more types than students' essays of 1-3 points. The use of extra-syllabus vocabulary is also more than essays of 1-3 points. In other words, high-grade essays show a richer vocabulary type and more complex vocabulary. But the number of token and word length did not show any significant differences between high-grade essays and low-grade essays. And even three low-grade essays have more tokens than high-grade ones. This shows that although the number of words is often an important standard in the AWE scoring, in addition to essay length, PIGAI also considers the diversity of words, as well as complexity and intricacy of sentences and paragraphs.

2) Sentence

From the perspective of sentence, the average sentence length of low-grade essays is shorter than that of high-grade essays, and the sentences with one to nine words are more than essays of high scores. On the other hand, sentences with twenty to thirty words are not used in essays of low scores. What's more, wrong expressions are frequent in such essays. Another feature of low-scored essays is the lack of adequate and proper use of adjectives. At present, the PIGAI scoring system is to consider increasing the scoring on different writing styles, such as abstracts, papers, letters, poetry and so on.

3) Paragraph

From the perspective of paragraph, the system takes the paragraph number and word/sentence number in each paragraph of low-grade and high-grade essays into consideration, while it pays no attention to content development, coherence, unity and other aspects. And the system misjudges the expressions in Résumé writing and other forms of Business writing due to its lack of special formats. In addition, the paragraph number and word/sentence number in each paragraph and other information can not affect the results of the scoring, but the content development of different paragraphs, coherence, unity and other aspects are often more important indicators of the quality of an essay, so the system needs to provide related feedback.

C. Evaluation Feedback

In addition to the table of feedback statistics, the PIGAI scoring system also provides evaluation feedback sentence by sentence, including more than twelve aspects: Highlight, Learning Tips, Correction Tips, Synonym Discrimination, Wrong Sentence, Low Frequency Warning, Collocation Errors, Noun Errors, Verb Errors, Article Error, Wrong Spelling, Wrong Capital Letter or Small Letter and Wrong Punctuation. "Highlight" is positive feedback, suggesting some excellent expression in the essay, such as the effective use of proverbs. "Learning Tips" and "Synonym Discrimination" are supplement to extend students' language knowledge. "Correction Tips", "Wrong Sentence", "Low Frequency Warning", "Collocation Errors", "Noun Error", "Verb errors", "Article Error", "Wrong Spelling", "Wrong Capital Letter or Small Letter" and "Wrong Punctuation" are all negative feedback, indicating that there are certain problems in the writing. In general, essays with high scores have more positive feedback, and those with low scores get more negative feedback. Table 3 presents the positive and negative feedback for essays of two extreme grades.

D. Findings

After analyzing the feedback, researchers find that the positive feedback by the system can locate the students' appropriate expressions and compare such expressions with the corpus. While negative feedback can detect errors and draw the attention of learners, which helps learners identify and correct such errors. The system can also expand distinction of synonyms and other confusing words. As Van der Linden concluded, the effectiveness of intelligent feedback reflects the following aspects: (1) to provide personalized feedback to the input compositions; (2) to point out the type of errors; (3) to analyze the errors; (4) to stimulate students' self-correction, as in [3]. It is through analyzing the feedback that the learners pay attention to the form of language use, compare the difference between their interlanguage system and the target language system, confirm the correct use, correct the error, and expand the more important content of grammar and text. Therefore, feedback has positive effect on improving students' language skills and promoting writing teaching.

However, feedback also has obvious problems, such as negative feedback on correct expressions and positive feedback on unqualified expressions. In addition, the system only provides detailed negative feedback on vocabulary, while on content, coherence, textual structure, it make no systematic review or suggestion for improvement, except for overall evaluation. So the system also needs to be improved in many aspects, especially in the evaluation of writing content and thinking ability assessment, which are the major problem confronting SWE systems in the world and a new development direction of modern online essay scoring system.

VI. SUMMARY

AWE requires the "check and balance" of human scorers, so the use of the system requires the role of peers and teachers. In fact, the productivity of human language determines that no matter how gigantic the corpus capacity is, it is difficult and impossible to cover all the forms of human expression. It is therefore believed that the combination of rationalism in the NLP (based on the method of transformational linguistics) and empiricism (based on the analysis of large-scale corpus and statistical linguistic models) is the future of NLP development.

In this case study, two groups of sample essays of five grades as a sample for PIGAI system. The results found that the system provides feedback on the levels of vocabulary, sentences, text and other, but the feedback information is not comprehensive and accurate enough. The system can give effective information for students' knowledge development, positive evaluation on highlight expressions, and suggestions to correct the wrong expressions. But the accuracy of the feedback still has room for improvement, and the field of the evaluation needs to cover textual unity and coherence.

ACKNOWLEDGMENT

Thanks are due to Prof. Yang Xiao at WHUT and Prof Li Daoshun at WUT for assistance with the case study and to Wang Yanhua (HUE) for valuable discussion.

REFERENCES

- [1] Shermis, D. & J. Burstein. *Automated Essay Scoring: A Cross-Disciplinary Perspective*, Mahwah, NJ: Lawrence Erlbaum, pp. 9-21, 2003.
- [2] Attali, Y. Exploring the feedback and revision features of Criterion, Paper presented at the National Council on Measurement in Education (NCME), San Diego, CA, pp. 17-53, 2004.
- [3] Zhang Li & Sheng Yue. Feedback Effect of JUKU Automated Writing Evaluation System—A Case Study, *TEFLE*, vol.163, pp. 38-44, 2015
- [4] Fitzgerald, J. Research on revision in writing, *Review of Educational Research*, vol.57, pp.484, 1987
- [5] Long, H. Focus on form in task-based language teaching, In R. Lambert & E. Shohamy (eds.) *Language Policy and Pedagogy*[C], Amsterdam /Philadelphia: John Benjamins Publishing Company, pp. 121-131, 2000.
- [6] Schmidt, R. The role of consciousness in second language learning, *Applied Linguistics*, vol.11, pp. 123-143, 1990.
- [7] Ellis, R. Principles of instructed language learning, *System*, vol.33, pp. 1701-1721, 2005.