

An Experimental Research on Applying Automated Essay Scoring System to College English Writing Course

Zhang Qiang¹

¹ School of Foreign Languages, Shandong Jiaotong University, China

Correspondence: Zhang Qiang, 5001 Haitang Road, Jinan, Shaodong Province, 250023, China. Tel: 86-186-6017-6618. E-mail: 170983040@qq.com

Received: April 24, 2014 Accepted: May 5, 2014 Online Published: May 28, 2014

doi:10.5430/ijelt.v1n2p35 URL: <http://dx.doi.org/10.5430/ijelt.v1n2p35>

Abstract

A lot of network teaching and learning software and programs related to English teaching have emerged in recent years due to the development of computer technology and the application of the new network technology. Automated essay scoring system provides a feasible platform to the reform of college English writing courses. The author first introduces the development of automated essay scoring system in western countries as well as in China, and then proposes some strategies on the application of AES system in college English writing course. After that the author carries on an experiment and proves that the students' writing ability could be improved significantly after being assisted by the automated essay scoring system. Finally the author proposes some potential problems of the automated scoring system which refers to the deficiency of students' writing strategies and teachers' difficulties in adjusting their new roles in such a new circumstance, and then the author provides some useful suggestions on how to solve these problems.

Keywords: automated essay scoring system, writing courses, writing ability, training

1. Introduction

It is generally acknowledged that the importance of English writing ability should be equal to English reading ability, listening ability and oral expression. But in the current English teaching in China, the phenomenon of "mute English"(it indicates that some Chinese students can only read and write the English language but have difficulty in speaking it) is too obvious, because the experts and teachers have put their main effort to improve the students' oral expression and reading ability, while the improvement of students' English writing skills is neglected. In fact, English writing ability is one of the most important categories in English learning. Through writing, the students' knowledge of English would be constantly consolidated and thus it can pave the way for the comprehensive development of English skills.

In the practical teaching of college English writing, the improvement of the students' writing ability is in a slow speed because most of the teachers' essay correcting burden is too heavy, the period of correction is long, and the students have less opportunities to practice. Therefore, it would be a good solution to apply a practical automated essay scoring system to the college English writing courses.

Automated Essay scoring (for short AES) system research was first carried out in the western countries, in 1999 the system found its way into the practical application stage. The relevant studies of Automated Essay scoring system in western countries are mainly as follows: First of all, Elliot and Mikulas (2004) had undertaken a research on the relationship between automatic evaluation system called *My Access* and writing, the research is focused on whether the application of this system will improve the students' writing scores in the state test. After the experiment researchers found that the application of the system in classroom teaching can positively improve the writing ability of the students. But due to the lack of the control group in the experiment, it is difficult to judge whether the improvement of students' writing scores was affected by online automatic essay scoring system. Shermis (2004) extracted the experimental group and control group from the same class in a random way, and then carried on the comparative study according to the two groups of students. The experimental results show that there are substantial differences between the control group and experimental group of students after the state exam. Vantage Learning

(2007) studied on the validity of the *My Access*, this research proved that online automatic essay scoring system can improve students writing scores in the state test. Rich et al (2008) had also achieved similar research results in the application of automatic essay scoring system. The research on computer automated scoring system in China started later than the western countries, but it has achieved certain achievements in recent years, such as Bingo English intelligence composition review system. And since 2009, more and more schools and colleges began to use the *Bingo system*. The time for the application of AES systems in China is very short. Despite the fact that a lot of experts show positive attitudes towards it, the practical effects of automated essay scoring system in college writing course remains to be examined. This is the fundamental purpose of the experimental research.

2. Strategies on the Application of Automated Essay Scoring System in College English Writing Course

The automated essay scoring system is mainly applied to two aspects: on one hand, it is used for large-scale test, essay examination is an essential and effective assessing method in various kinds of language tests, while the artificial grading is very difficult and the reliability is relatively low; on the other hand, the automated essay scoring system could be used for English essay teaching. If the students want to improve the ability of writing, they must do lots of exercises and get the teachers' targeted feedback, but in the actual writing classes, the teachers could not have enough time and energy to grade each essay carefully and give the students the detailed feedback. Therefore, the automated essay scoring system is of high practicability. The author proposes some strategies on the application of AES system in college English writing course.

2.1 Computer Scoring Should be Combined with Teachers' Instruction and Evaluation

The current automated scoring system could never "appreciate and evaluate" the essays on the technical level, for example, the computer cannot find the humorous expression or the unique style of reasoning. This requires the combination of computer scoring and teachers' evaluation. Teachers can choose the well-written or poor-written essays based on the evaluation of the AES system and comment on them in class; the teachers can also guide the students to design and organize the content of the essays; teachers can also summarize and comment on the problems concerning writing style, grammar, vocabulary, etc, and then help students to solve these problems. In fact, the use of automated scoring system can save the teachers' precious time and energy to help the students to improve their writing ability more efficiently. And if the teachers only rely on the automated scoring system and do not give any instructions or comments, the students will get more nervous and irritable and their improvement on writing ability will be limited.

2.2 Combining Timed Composition and Composition after Class

At present the purpose of training the ability of writing for a lot of non-English major college students is to get high marks in various exams, namely, their writing motivation is examination-oriented. For the examination-oriented training, the teacher had better adopt the timed essay writing mode in the scoring system. Teachers could give the title of the essay to students and limit the deadline of submission for thirty or forty minutes. In this way, the students could practice the ability to arrange the essay's structure and the ability to make sentences so that they can express their meanings clearly. They could also form the concept of arranging time reasonably and complete the essay within the allotted time. The teacher could assign the students some other topics in the form of homework, and the students should also finish the essays within the prescribed time limit.

2.3 Combination of the Product Writing Approach and the Process Writing Approach

The traditional product writing approach regards writing as a kind of result, the students are required to finish an essay in a time limit, and then the teacher gives scores and comments to their essays. Handing in an essay makes the students feel "once and for all"; they only care about their essays' grade and do not take their teacher's comments seriously, thus making the same mistakes repeatedly appear in their essays. While process writing approach requires students to modify the same essay for many times and write whatever they want; it puts more emphasis on ideological content and expression of the essays. It focuses on the feedback of the essays and requires several revisions of them, believing that revision is a very important procedure in the process of writing. The students should modify their essays again and again to improve their cognitive ability, thinking ability and writing ability. There are a lot of researches on process writing approach, but in the practical teaching, the biggest problem is that teachers don't have time and energy to review such a large number of essays. The appearance of automated essay scoring system is really a breakthrough: it not only reduces the teachers' heavy workload but also makes the teachers to focus on broadening the students' horizon and encouraging students to develop their writing ability. The students can use the feedback from the AES system to complete the revisions from the first draft to the final draft, and enjoy the

happiness of writing during the constant revising process.

3. Research Methodology

3.1 Research Object

The author selected two classes in Shandong Jiaotong University as the research object. The students in the two classes had already taken the English writing courses in the same semester. And the two classes were divided into the control group and the experimental group. The experimental group has 55 students, and there are 23 males and 22 females. The control group has 43 students, and there are 18 males and 25 females. The age ranges from 19 to 21. The control group was assigned to take the conventional English writing courses and the teacher was supposed to adopt the conventional teaching and scoring methods. While in the experimental group, the teacher had used the automated essay scoring tool called Bingo system. The author was the teacher of writing course for both of the two groups.

3.2 Research Tool

The automated essay scoring system called Bingo English intelligence composition review system in this research is developed by Bingo Network Science and Technology Ltd. This system is established on the basis of standard of native speakers' corpus and it could analyze the students' essays from the aspect of writing contents, structure, wording, sentence patterns, grammar and writing format and specification. Meanwhile it also provides the writing database such as thesaurus and the editing tools of checking grammar, spelling and genre. It combines the function of scoring, feedback and writing assistance and it is a helpful and efficient tool for college English writing courses. The system has been used in Chinese universities and high schools for more than two years, and every trial had been undergone the artificial detection. It has been confirmed to be able to judge a student's actual writing level with the accuracy which is more than 95%.

3.3 Data Collection

This research has lasted for one semester and the research data came from three sources. (a) The questionnaire for students before and after the experiment. These two questionnaires mainly inspected the English writing motivation of the students in the experimental group before and after the experiment. The questionnaire mainly consisted multiple choice questions and each choice was classified according to *Likert Scale* from "totally agree" to "totally disagree". (b) The interview for students after experiment. Six students from the experimental group were interviewed. The interview adopted the semi-open style and asked the students to explain the influence of the automatic essay scoring system on their writing process from three aspects: the difference between systematical feedback and teacher's feedback, the benefits for their writing ability and the limitations of the system. (c) The writing scores of the students before and after experiment. Before the experiment, the students were assigned to take a standardized writing test, the topic was chosen from the Bingo system. The students were asked to finish their essays in the schedule time and then their scores were stored in the Bingo system. After the experiment, the students in two groups participated in another standardized writing test and their scores were also rated and stored in the Bingo system.

3.4 Data Analysis

There are 55 students in the experimental group and 43 students in the control group. The author used SPSS 20.0 to compare the difference between the two groups in the pretest and the posttest by Independent-samples Test to find out whether there is a difference on the writing ability between the two groups' students. The author also compared the pretest and posttest scores of the two groups by Paired-samples T Test to find which groups' students' writing ability have developed faster.

4. Research Findings

The author used SPSS 20.0 to analyze the scores of the students in experimental and control group in both pretest and posttest by Independent-samples Test method and obtained the following data:

Table 1. Overall Scores in Both Pretest and Posttest of Experimental Group and T Test

Assessment Criteria	Item	N	Mean	Std.Deviation	Sig. (two-tailed)
Overall Scores	Pre-Experiment	55	2.386	.3906	.000
	Post-Experiment	55	3.960	.7486	
Contents	Pre-Experiment	55	2.253	.4385	.000
	Post-Experiment	55	3.871	.8305	
Structure	Pre-Experiment	55	2.500	.4074	.000
	Post-Experiment	55	4.013	.8415	
Genre	Pre-Experiment	55	2.511	.5124	.000
	Post-Experiment	55	4.050	.6471	
Wording	Pre-Experiment	55	2.089	.3667	.000
	Post-Experiment	55	3.608	.7772	
Fluency	Pre-Experiment	55	2.508	.4352	.000
	Post-Experiment	55	3.911	.6575	
Grammar and Spelling	Pre-Experiment	55	2.453	.4746	.000
	Post-Experiment	55	4.121	.7338	

In Table 1, under the premise of 5 points as the full score, the average score of experimental group in pretest was 2.386, but the average score after posttest was 3.950, so we can conclude that the average score of experimental group has been promoted significantly. Meanwhile, the standard deviation of overall scores ranges from 1 to -1 before and after the experiment, this indicated that the students' scores were quite even on this dimension. In order to examine whether the improvement is significant for the students in experimental group, the author had carried on Paired-sample T test and the result indicated that the sig value of the total score in pretest and posttest is $.000 < 0.05$, and this can prove that the total score of experimental group had been promoted significantly.

Table 2. Overall Scores in Both Pretest and Posttest of Control Group and T Test

Assessment Criteria	Item	N	Mean	Std.Deviation	Sig. (two-tailed)
Overall Scores	Pre-Experiment	43	2.448	.4615	.004
	Post-Experiment	43	2.851	.6866	
Contents	Pre-Experiment	43	2.349	.5815	.002
	Post-Experiment	43	2.910	.878	
Structure	Pre-Experiment	43	2.673	.5042	.002
	Post-Experiment	43	3.108	.6184	
Genre	Pre-Experiment	43	2.519	.4988	.000
	Post-Experiment	43	3.049	.6140	
Wording	Pre-Experiment	43	2.119	.4307	.001
	Post-Experiment	43	2.551	.6606	
Fluency	Pre-Experiment	43	2.465	.5579	.001
	Post-Experiment	43	2.986	.5716	
Grammar and Spelling	Pre-Experiment	43	2.565	.5564	.001
	Post-Experiment	43	3.062	.6066	

From Table 2 we could see that the average score of control group in pretest was 2.448, and the average score after posttest was 2.851 which had increased 0.403 points. Compared with the increase of score in experimental group (1.564), the increase of score in control group is not so significant. Through the Paired-sample T test, we could see that the sig value of the total score is $0.004 < 0.05$, this indicates that the students' writing ability in control group was also improved.

Table 3. Overall Scores in Posttest of Control Group and Experimental Group and T Test

Assessment Criteria	Item	N	Mean	Std.Deviation	Sig. (two-tailed)
Overall Scores	Control	43	2.851	.6866	.000
	Experimental	55	3.950	.7486	
Contents	Control	43	2.910	.878	.000
	Experimental	55	3.870	.830	
Structure	Control	43	3.108	.6184	.000
	Experimental	55	4.109	.8415	
Genre	Control	43	3.049	.6140	.000
	Experimental	55	4.050	.6471	
Wording	Control	43	2.551	.6606	.000
	Experimental	55	3.608	.7772	
Fluency	Control	43	2.986	.5716	.000
	Experimental	55	3.911	.6575	
Grammar and Spelling	Control	43	3.062	.6066	.000
	Experimental	55	4.121	.7338	

In Table 3, we could see that after the experiment, the average score of experimental group was 3.950, while the average score of the control group was 2.851, that is, the average score of experimental group had exceeded the control group by 1.1 points. In order to examine the statistical significance of the difference between the experimental and control group after experiment, the author carried on an Independent-samples Test and found out that the sig value of total score of both experimental and control group is $0.000 < 0.05$. So the statistics showed that in terms of writing ability, the students in experimental group had gained a more significant improvement than the students in the control group after one semester.

5. Discussion: Potential Problems of AES System

Automated essay scoring system is a new way of grading the students' essays, but there are still some problems in the practical application:

5.1 The Deficiency of Students' Writing Strategies

Because high school English education mainly focuses on the cultivation of the students reading ability and there is little requirement on English writing ability, the students in high schools are always short of writing practice. Generally speaking, the high school students could not master the writing strategies very well. After entering the college, this phenomenon is particularly prominent. During the process of correcting students' essays, we can often find that the discourse structures in some students' essays are not clear and logical. In the experiment of essay teaching mode based on automated essay scoring system, we can often note that the discourse structure of students' essays are not very clear, and we can also find a phenomenon that the ideas of the students in the essays are disordered. So we can conclude that the students' writing strategies were insufficient. Peng Li thinks that writing strategies include language input strategy, macrostructure strategy, imitation strategy, error correction strategy and writing practice strategy. Combined with the five specific strategies, the author made a random oral interview with the students after the experiment. The author found that the students are not aware of the writing strategy; they do not form the habit of drawing outline or draft, and never consider how to arrange the paragraphs. They seldom revise their essays by themselves, when they meet the problem of expressing their ideas, most of the students don't know how to solve it. They are also not good at imitating and avoidance and the language in their essays are not quite native. While the experimental results confirmed the effectiveness of automated essay scoring system in the essay teaching, but in order to improve the students' comprehensive ability in English writing, we should strengthen the cultivation of students' writing strategies, and let the students learn to write their essays consciously and actively, enjoying the experience of self-improvement and growth.

5.2 The Difficulties in Adjusting the Teachers' Role

The automated essay scoring system can save the teachers a lot of time in correcting students' essays and reduce the pressure of English teachers' work. But because the teachers do not mark the specific errors on the essays of the students like before, some students may think that their teachers' work is not earnest. So, in the process of essay

teaching based on automated essay scoring system, the teachers must adjust their roles, using the extra time for a better guidance of students' essay writing. The teachers should give different writing guidance, supervision and feedback according to the students' writing ability, and often communicate with students. The teacher should also understand the student's difficulty in writing, and strengthen the emotional interaction with the students and eliminate the misunderstanding. For the students who are lagging behind, the teachers should pay more attention to the syntactic structure and word collocation in their essays. For the students who do well in writing, the teacher should give the macro guidance of discourse structure and rhetoric devices and make them be aware of cultivating the good writing habits and improve their writing strategy. The application of automated essay scoring system does not mean that the teachers do not have to score and review the student's essays anymore. The role and function of the teachers should be diverse; they could become the helpers, regulators and responders in the students' writing activities.

6. Conclusion

The English essay teaching mode based on automated essay scoring system focuses on the current situation that the number of English class are limited and the working load of the teachers are very heavy which result in the insufficient writing practices for the students. It makes full use of the advantage of automated essay scoring system and applies it to process writing approach creatively and thus provides the students a network writing platform which integrates the grading and feedbacks, then the teachers could combine the peer review and comments from the AES system to instruct the students' writing and finally promote the students' writing ability. This new kind of writing teaching mode alleviates the teachers' heavy working load of correcting essays and improves the students' independent writing ability. But there are still some limitations of the current automatic essay scoring systems, for example, the current essay scoring systems can only give comments about the imperfection of the language, spelling errors, etc., but these systems cannot suggest how to correct or improve them. Moreover, the related research is relatively weak on the feedback tools of essay writing.

To make the essay scoring more helpful to the daily teaching and feedback system, we should focus on the following three aspects in the future research: 1. The target feedback from the students. From the viewpoint of teaching, the targeted and personalized feedback is more important than accurate composition grading. If a teacher only give a grade to the student's essay, no matter how accurate the grade is, it is not enough. As for the students, he needs to know what kind of problems exist in his essay and what mistakes need to be corrected or avoided; for the teacher, he has to know the language points that the students had acquired and find the language points that still need further explanation. These goals can only be achieved by the detailed feedback through the computer essay evaluation system. 2. Universality. Universality refers to realizing the scoring model's aim, that is, to train it for only one time and use it for a lot of times, and it is applicable to the review of non-specific topic composition. Now that we adopt non-specific topics for a general computer grading system, we should process the language and content respectively. When a general automated scoring model would be constructed, supplemented by the content evaluation method without training, such as the method of text clustering or key words, we can put it into the use of college English essay teaching. 3. The clear division between human and computer interface. With the contemporary technology, it is not realistic that the essays review could be completely done by machines automatically. So the development direction of automated essay scoring system is to combine the scoring approach of both human and machine, and design a best reviewing process that combines human participation and computer work, making a maximum mutual complement between human and machine. Of course, although the advanced automated essay scoring system has been introduced into the traditional classroom teaching, the teachers should not simply treat it as a simple auxiliary tool. They should integrate this new system into the whole teaching system creatively to optimize the teaching effect. This requires every teacher to bring their subjective initiative into play and continuously update their teaching philosophy, cognition and approach.

References

- Anson C. M. (2003). *Responding to and assessing student writing: The uses and limits of technology*. In Takayoshi, & B. Huot (Eds.), *Teaching writing with computers: An introduction*. New York: Houghton Mifflin Company.
- Attali, Y. (2004). Exploring the feedback and revision features of criterion. Paper Presented at the *Annual Meeting of the National Council on Measurement in Education*, San Diego.
- Baron D. *The College Board's new essay reverses decades of progress toward literacy* [EB/OL]. *The Chronicle of Higher Education*, 2005: B14. Retrieved November 20, 2010, from

<http://www.english.illinois.edu/-people-/faculty/debaron/essays/sat.htm>

- Black, P., & D. William. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74. <http://dx.doi.org/10.4135/9781446250808.n2>
- Bruner, J. (1985). *Vygotsky: a historical and conceptual perspective*. Cambridge: Cambridge University Press.
- Chen, E., & E. Cheng. (2008). Beyond the design of automated writing evaluation: Pedagogical practice and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12, 94-112.
- Cheng Y. (2004). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing*, (4), 313-335. <http://dx.doi.org/10.1016/j.jslw.2004.07.001>
- Daly, J. A., & D. A. Wilson. (1983). Writing apprehension, self-esteem, and personality. *Research in the Teaching of English*, (4), 327-341.
- Elliot, S., & C. Mikulas. (2004). The impact of MY Access! Use on student writing performance: A technology overview and four studies. *Annual Meeting of American Educational Association 2004*, San Diego, April.
- Fitzgerald K. R. (1994). Computerized scoring? A question of theory and practice. *Journal of Basic Writing*, 13(2), 3-17.
- Granger S. (2004). Computer learner corpus research: current status and future prospects. In Connor, & T. Upton (Eds.), *Applied Corpus linguistics: A Multidimensional Perspective*. Amsterdam & Atlanta: Rodopi, pp. 123-45.
- Grimes, D. (2008). *Assessing automated assessment: Essay evaluation software in the classroom*. <ftp://ftp.ics.uci.edu/pub/grimesd/AutoAssessment.pdf> 2010-06-203.
- Hoon, T. (2006). *Online automated essay assessment: Potentials for writing development*. Retrieved from <http://ausweb.scu.edu.au/awO6/papers/refereed/tan3/paper.html>
- Keith T. Z. (2003). Validity and automated essay scoring systems. In Shermis M. D., & Burstein J. C. (Eds.), *Automated Essay Scoring: A Cross—disciplinary Perspective*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 147-168.
- Kay, H., & T. (1998). Dudley. Evans. Genre: What teachers think? *ELT Journal*, 52(4), pp. 308-314.
- Lee, I. (2004). Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing*, 13(3), 285-312.
- Rich, C., H. Harrington, J. Kim, & B. West. (2008). Automated essay scoring in state formative and summative writing assessment. Paper Presented at the *Annual Meeting of the American Educational Research Association*, New York.
- Rudner, Lawrence, Phill G. (2001). An overview of three approaches to scoring written essays by computer. *Practical Assessment, Research & Evaluation*, 7(26). Retrieved October 14, 2010 from <http://PARE-online.net/getvn.asp?v=7&n=26>
- Shermis, M. D., & Burstein, J. (2003). Introduction. In Shermis, M. D., & Burstein, J. (Eds.), *Automated Essay Scoring: A Cross—disciplinary Perspective*. Mahwah, NJ: Lawrence Erlbaum Associates, xii.
- Shermis, M., J. Burstein, & L. Bliss. (2004). The impact of automated essay scoring on high stakes writing assessments. Paper Presented at the *Annual Meeting of the National Council on Measurement in Education*, San Diego.
- Warschauer, M., & D. Grimes. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 22-36.
- Wild F, Stahl C, et al. (2005). Factors Influencing Effectiveness in Automated Essay Scoring with LSA. *Frontiers in Artificial Intelligence and Applications*, p. 125.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second Language Development in Writing. Measures of Fluency, Accuracy & Complexity*. Honolulu: Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.
- Yang Y., Buckendahl C. W., Juszkievics P. J., et al. (2002). A review of strategies for validating computer automated scoring. *Applied Measurement in Education*, 15(4), 391-412.