A Comparative Study of Teacher Feedback and Automated Essay Scoring in College English Writing

HU Yinghui[1] & ZHAO Dan[2]

Abstract

Based on some relevant theories of process writing approach, noticing hypothesis, output hypothesis and interaction hypothesis, through literature review of teacher feedback as well as automated essay scoring, this study conducts a comparative study by means of tests, writings tasks and questionnaires, aiming to explore the effectiveness and students' perception of automated essay scoring compared with teacher feedback so as to find out an effective way to improve college students' English writing competence in China and strive for a proper way to give students sufficient and effective feedback with regard to their errors in writing as well.

Keywords: teacher feedback; automated essay scoring; English writing

1. Introduction

As the core of the process-oriented approach, feedback has gradually become more and more popular in foreign language teaching. With the advance of science, automated essay scoring (AES), as a new type of feedback, has become a hot topic in the world. However, the empirical study of AES system is rare in China. Moreover, few such studies have ever been conducted among college students. And only a number of studies, have been carried out to compare AES feedback with other source of feedback but proved to have some problems in reliability and validity. Therefore, based on the theories of process writing approach, noticing hypothesis, output hypothesis, interaction hypothesis, this study aims to explore the effectiveness and students' perception of AES compared with teacher feedback (TF) from perspective of organization, content, grammar, spelling and punctuation in a writing, and try to answer some research questions and finally find out a proper way of giving feedback to college students' English writings in a big class so as to improve students' writing proficiency

2. Literature Review

2.1 Feedback Theories

Originally, feedback was a fundamental concept in cybernetics, which refers to the output of information that is return to the input terminal in the control system, and have an impact on the output information. Some experts define "feedback" as the comments or information that learners heard from teachers or other learners. In the field of writing, feedback is defined as the "input from a reader to a writer" (Ashwell, 2000), with which, the reader provides the modified information for the writer so that the writer could express his point of view more accurately and communicate with others more easily.

Process writing theory introduces the concept of "feedback" to the teaching of writing. It regards writing process as a communicative activity between reader and writer (teacher and student, student and student, student and computer scoring system), which emphasizes that writing is a process of repetitive modification and feedback exists in every fabric of the process.

---

[1] School of Foreign Languages, Shandong Normal University, Jinan 250014, China.
[2] School of Humanities, Shandong Management University, Jinan 250357, China.

Swan's output hypothesis claims that comprehensible input is not enough to the successful acquisition of second language. If you want to use the target language fluently and accurately, the output also played a crucial role (Swain, 1985). According to the theory of noticing hypothesis, in the research of second language acquisition, the noticing and feedback are interconnected and interdependent: feedback can promote the noticing and attract attention; noticing can stimulate feedback and improve learning. Swain (1995) proved in his study that the feedback enables language learners to notice the difference between mother tongue and target language, which makes learners to rebuild their assumption about target language and to reconstruct the language structure. In the theory of interaction hypothesis, feedback also plays an important role because it emphasizes the interaction between professor and learners in the process of foreign language learning.

2.2 Feedback Classifications

Feedback can be classified as TF, peer feedback and AES in terms of the source of feedback. AES regards computer or computer system as the provider and processor of feedback information and give comments to students' composition. Therefore, the types of feedback are varied and each form of feedback has its advantages and disadvantages.

2.2.1 Teacher Feedback (TF)

TF is the exchange of information between teachers and students. Currently, in second language learning, feedback is often reduced to the teacher's response to the student's composition. Generally speaking, there are two kinds of TF: conferencing and written comments. As a direct way, conferencing is a face-to-face discussion by teachers and students. It is also regarded as a face-to-face oral interaction. In this study, TF is narrowed down to teacher's comments on students' writings, especially written comments. Written comments are the most popular and common type adopted by most English teachers. As the name indicates, written comments refer to the written information provided by the teacher in student's writing. It is the process of providing some commentary on student's work in which a teacher reacts to the idea in print, assessing a writing's strengths and weakness, and suggesting directions for improvement. This feedback is typically written on blank spaces on the students' writings for returning to the students. The aim of the written comments is to guide the students so that they can try to self-correct their own compositions.

2.2.2 Automated Essay Scoring (AES)

AES is defined as the computer technology that evaluates and scores the written prose (Shermis & Burstein, 2003). They are mainly used to help overcome time, cost, reliability, and generalizability issues in writing assessment (Chung & O'Neil, 1997). As a relatively young field, AES has only a 40-year history. Ellis Page is generally regarded as the pioneer of AES, for he designed a computer grading program named Project Essay Grader (PEG) in 1966. Utilizing the statistical capabilities of computers, Page looked for the kind of textual features that could be extracted by computers from the texts and then applied multiple linear regression to "determine an optimal combination of weighted features that best predicted the teachers' grades" (Kukich, 2000). Although Page's pioneering work seemed promising, AES tools did not gain popularity for the next three decades. Even though, AES tools will become a trend in writing teaching and correcting.

However, there are fewer such researches on AES in China. Facing so large number of non-English majors including more than 100 students in one single class, it is really heavy for a college English teacher to accomplish the work of composition correcting. In this case, some domestic experts in China proceed to develop some relevant systems, for example, Professor Liang Maocheng presided over the development of "large-scale automatic scoring system", which has been applied for national patent in 2005 and backed by lots of prominent scholars. At the same time, some companies also began to develop this kind of software. Two of the most famous are Bingo and Jukuu English Essay Assessor.

2.2.3 Jukuu

Jukuu is built for Chinese English learners and there is only Chinese version. Jukuu is more localized and easier to use.

Thus, Jukuu is taken as one of the tools for the teaching experiment of this study. Jukuu (located at http://www.pigai.org/) built by Beijing University of Posts and Telecommunications, is partly one of English Writing Intelligent Tutoring Systems based on corpus and cloud computing. By calculating the similarity between students' writing and standard English from corpus, Jukuu offers holistic scores and general comments as well as sentence comments on students' composition immediately, which can help students improve their writing proficiency easily. Figure 1 shows the operating principle of Jukuu

Figure 1: The Operating Principle of Jukuu
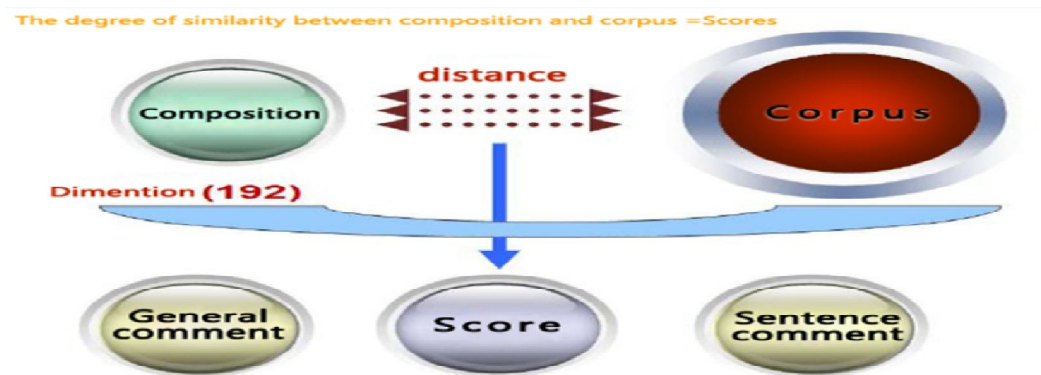
Jukuu provides two different web-pages for teacher and students. The student's web-page and the teacher's web-page are specifically introduced in the following.

Student's Web-page: The system provides a convenient platform for the students to do writing exercise. Students do not need to install any software, as long as network is available. At the same time, it enables students to do writing practice whenever and wherever possible. Once the student completes the assignment and submits it, he or she will get a general assessment and a sentence-by-sentence assessment at once.

Teacher's Web-page: As an intelligent management platform, Jukuu not only provides users with such functions as creating online classes, assigning writings, managing classes, but also adds some personalized options. It scores writing both for content and style and examines various features of writing, including vocabulary, word usage, subordinate clause and sentence structure. As a part of the usual procedure of Jukuu, each writing is compared to every other one in a set. The writings that are extremely similar to others in the corpus are usually examined by Jukuu. Meanwhile, it also allows teachers to trace any student's writing-modifying trail. Jukuu might make all these shinning points possible and in return, the shinning points might well make Jukuu be easily accepted by English teachers and learners in China.

However, Jukuu as a AES also has some drawbacks. To what extent can it be accepted by English teachers and students in China? How well can it assess students' writings? What aspects does it lay emphasis on? All these questions may be shown by the results in this study.

2.3 Previous Studies of TF and AES

2.3.1 Previous Studies of TF

Plenty of surveys inquiring students' perceptions and preferences about TF have been carried out in LI and L2. In order to gain information about ESL students' perceptions towards teacher written comments on content, organization and errors in their writings, Radecki and Swales (1988) conducted a study and surveyed 59 ESL students and interviewed eight of them, which suggested that approximately 87% of the students not only appreciated teacher comments on content and organization of their writing but also expected teachers to correct surface errors, while the minority of the students would rather receive short comments and a grade or even only a grade on their writings. As for the linguistic errors, they considered it was the teachers' responsibility to correct them.

Hedgeock and Lefkowitz (1994) randomly selected 110 ESL students and 137 EFL students and conducted a survey to investigate the differences between attitudes of ESL students towards TF and that of EFL students. On the whole, both groups of students admitted the positive attitudes to TF on their writings.

However, there was a prominent difference between ESL and EFL students in that ESL students have a strong preference on structure, content and writing style while EFL students highly concerned about form-focused feedback. This indicates that students' perceptions about what makes up useful feedback vary according to the orientation and demands of the specific learning context. 320 students from eight schools were asked to fill in a questionnaire regarding TF on their writings in Lee's study (2004) which aimed at exploring students' perceptions towards TF. The results showed that the students were aware of teachers' error corrections, but 76% of the students indexed that they wanted their teachers to give direct correction on all errors in writing and 83% of them said they would like their teachers to mark errors comprehensively in their writing. Besides, it was found that 45% of the students thought that it was their teacher's responsibility to make error correction.

Several studies investigated the positive effect of TF in students' perception. For example, Ferris (1995) pointed out that students took seriously TF on grammar, content and organization, and that 93.5% of the students investigated reported development in writing after they received TF. Feedback could inform the writing process, shaping and molding it, and the nature of writing process is fundamentally social (Arndt, 1993). At the same time, several researches have found out some negative results about the effect of TF in student writing. Knoblauch (1981) pointed out that TF was unsuccessful, no matter how written feedback was delivered in margins or at the end of the writing, in red or black pen, through verbal commentary or correction symbols. Zamel (1985) criticized that the teachers rarely offer specific strategies for revising the writing, impose abstract rules and standards, write contradictory comments, respond to writings as fixed and final products, misread student writings, provide vague prescriptions and make arbitrary corrections. Leki (1990) noted that students felt hostility about teachers' attempts to take over their ideas and writing, so they were not willing to pay much attention to TF and try to understand it. Rollison (2005) summarizes "TF has been criticized as being over-general, inconsistent, idiosyncratic incomprehensible, formulaic, vague, inaccurate, meaningless to the student and abstract."

In China, after reviewing the researches on TF in some well-known Chinese journals, the empirical studies on students' perceptions towards TF are rare and the studies between TF and AES are less. In China, only a few studies have been conducted on how students reacted to TF, and the results are different. Some scholar pointed out that students took TF very seriously and regarded it consistently as the most effective feedback in writing. Some found out that the students preferred TF to peer feedback and attached more importance to TF by a comparative study of TF and peer feedback through students' drafts, questionnaires and interviews. In a secondary school in Hong Kong, Tsui and Ng (2000) showed that ESL learners significantly adopted and perceived more teacher's comments than peer's, but they also thought that peer comments played some roles that could not be given by teachers. In general, when asked to state a preference, the learners always take TF as their first choice because they think that TF focuses more on grammar and tends to be more useful. However, there is no denying that some researches have questioned the effectiveness of TF as a way of improving students' writing. It is discovered that some teachers and students in China are dissatisfied with the current teachers' written feedback with the hope to make it better. Although the researches at home and abroad on students' perceptions of TF vary in results, students' preference for teachers' written comments is conspicuous. In China, English composition pedagogy has always been a major subject for research, but in practice, less attention has been paid to the students' perspective and role in the process of how they react to, and what sense they make of their teacher's comments. That's one of the reasons why the present study is motivated, aiming at comparing students' perception between TF and AES, so as to solicit enlightenment for future English writing pedagogy.

2.3.2 Previous Studies of AES

AES has been under development since the 1960s. In the last decade, the dispute on the effectiveness of AES has never ended. Some scholars think that the application of AES to evaluate students' composition can not only save time and effort, but save cost with high reliability. Of course, there are also some scholars or organizations that take negative attitude towards AES.

The CCCC (the Conference on College Composition and Communication) (2004) questioned the criteria and principles of AES, and claimed that this kind of feedback is "against the nature of the interpersonal communication in writing". A comparative research between TF and AES was conducted by Lee (2007), and the results indicated that the control group and the experimental group has no significant difference in the range of article length, structure and the final score.

Hutchison (2007) reported that AES paid more attention to the structure and organization, and not too much concerned about the abstract intrinsic factors such as the interest of articles, and the relevance to topics and so on. However, Grimes & Warschauer (2008) suggest that the impact of such software depends to a large effect on how it is used. When AES is used to replace writing for a teacher or other audience, students are dissatisfied. However, when AES is used as part of a social writing process—in which, for example, students write earlier drafts for review by the software, and later drafts to be submitted to the teacher or published online for peers—the results are more positive .

 Compared with the study abroad, the empirical study of AES system in China is rare. Many scholars concentrate on the introduction of AES abroad and the development of AES tools in China. Few studies have ever been conducted to explore the college students' perceptions towards AES in China. And so far, a number of studies have been carried out to compare AES with other ones. For example, Liang Maocheng & Wen Qiufang (2007) evaluated and compared three representative AES systems against the major assessment criteria in writing scoring in the field of language testing. Based on the comparison and evaluation, they argued that these systems have reliability problems with the human assigned scores used for training the systems' models, as well as validity problems with the scores the system assigned. They also analyzed the implications that the three systems offered. The study sheds important light on the development of an AES system in China.

Although the results of the researches abroad are positive, it doesn't mean that those AES tools could be applied directly to the correction of Chinese students' compositions. Considering this condition, this study takes advantage of Chinese AES tool -- Jukuu to explore the effectiveness of AES compared with TF and then explores more about the students' perceptions towards AES. By means of investigation, the study might shed some light on teaching of college English writing as well as the correcting of students' writings.

3. Methodology

3.1 Design of the Study

Based on the research aims mentioned above and some related studies at home and abroad, this study aims at getting to know students' perceptions between TF and AES and finding an effective way to improve college students' practical writing competence in China. The three research questions are specifically addressed as follows:

1) What are the respective focuses of TF and AES on the correction of students' writing works?
2) Which feedback will boost college students' English writing proficiency, TF or AES and are there any differences about the effects on different aspects of language skills (such as chapter structure, language expression and so on) produced by these two kinds of feedback ?
3) What are the students' attitudes toward TF and AES in college?

In order to answer the three research questions, two classes are selected, one as the control class (CC), and the other as the experimental class (EC). The experiment lasts for 16 weeks, each student finish 4 pieces of writing during the 16-week English writing teaching. The writings of EC were assessed by Jukuu, while the writings of CC were corrected by teachers. In addition, after students finished the post-test, two questionnaires were carried out in two classes to find out their overall perceptions towards TF and AES. After experiment all the needed data from two English tests (pre-test and post-test), the writing tasks and questionnaires were carefully collected and analyzed by SPSS 19.0 quantitatively.

3.2 Participants

The participants in this study are 80 sophomores from two parallel classes of Electrical and Mechanical Engineering Department from Shandong Xiehe University in Jinan (the size being 40 respectively).

Among the 80 participants, 68% are male and 32% are female, ranging in age from 19-21 by learning the textbook of *New Times College English*. To ensure that all the students have the same level of English, the results of their final test are analyzed in the second semester. It is proved that the two classes have no significant difference. At the same time, the data of standard deviation are also at the same level that proved that the 80 participants have the same level of English. As a result, two classes are randomly classified as follows: Class One is EC which practices AES and Class Two is CC which implements TF. There are four English classes a week. All the participants from the two classes were not informed before the experiment for the purpose of obtaining an objective results.

3.3 Research Instruments

The instruments adopted in this study mainly include pre-test, post-test, the writing tasks, two questionnaires and SPSS Version 19.0, which is the acronym of Statistical Package for Social Sciences Version, to analyze the collected writing scores. The pre-test is used to certify whether the students from EC and CC are at the same level in writing competence. The purpose of the post-test is to see whether there are any significant differences between EC and CC in writing proficiency. The questionnaires were distributed to both participants after they finished the post-test and the writing tasks assigned by writing teacher in class and in Jukuu grading net to find out their overall perceptions towards TF and AES. Finally, all the needed data from pre-test, post-test, the writing tasks and questionnaires were carefully collected and analyzed by SPSS 19.0 quantitatively.

3.4 Procedures

The entire experiment lasted for one academic term from September to December in 2014. In this study, two groups, namely EC and CC, were examined. Compared with the CC group, the EC group had to receive AES training. Figure 2 shows the general experimental procedures being carried out in this empirical study.

Figure 2 Diagram of Research Procedure in the Experiment
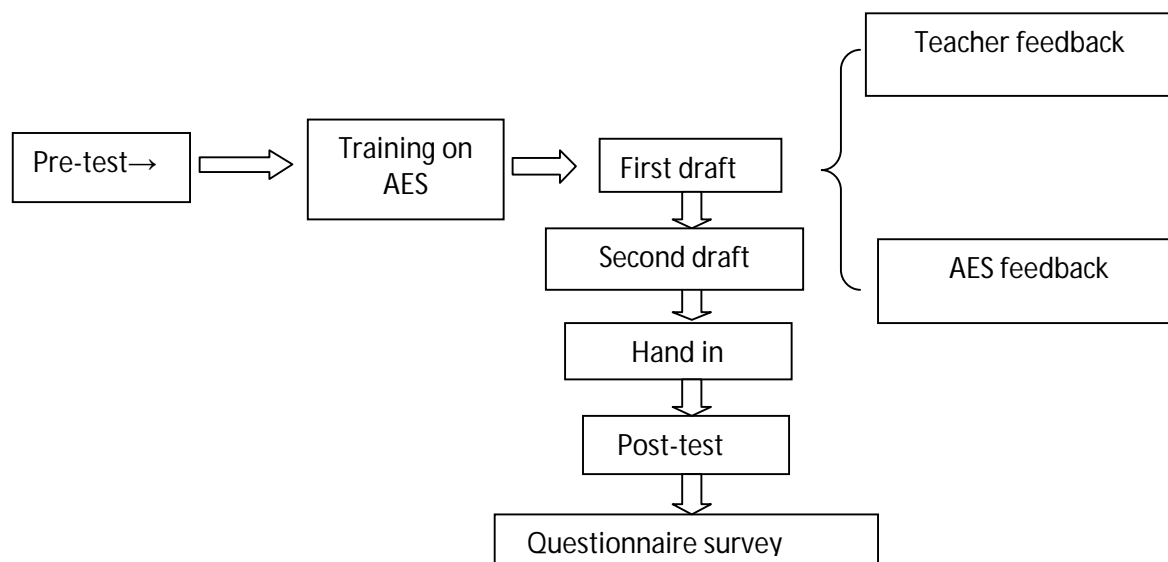
3.4.1 Pre-test

In order to test whether there was significant difference in English writing ability between the two classes, the pre-test was adopted in the study and was carried out in CC and EC separately in the first week of the semester in September, 2014. The test material was taken from PRETCO (Practical English Test for Colleges). Within 45 minutes during the class period, all the participants in both EC and CC were required to write a English writing about 120 words on the same topic (see Appendix I) and then their writings were gathered immediately. All the participants were told that the marks of their pre-test accounted for certain proportion of their class performance so as to make sure that all students would consider this pre-test seriously and ensure the quality of their writings. Three English teachers checked and scored the compositions anonymously so as to guarantee its reliability and justice.

The score of each part of the composition and the total score of their compositions were recorded by the three teachers seriously for further viewing of statistic analysis and discussion.

3.4.2 Training on AES

After pre-test, EC was given some knowledge about Jukuu network with the help of PPT, which not only includes the directions for usage but also several examples of students' self-correction. And then, after class all EC students were required to enroll in a login name and wrote a short essay on the network. After the training, students became more familiar with Jukuu and showed their interest in the network.

3.4.3 Composition Writing

During the experiment, all the students in the two classes were required to finish four compositions (see Appendix II) which are all cited from the PRETCO. As a valid and reliable test paper, PRETCO is not only suitable for testing students' writing competence but helpful to make sure that the degree of difficulty with writing materials is almost the same. The wring procedure was as follows: (1) the first daft; (2) TF or AES feedback; (3) the second daft. Before the first draft, EC students have got AES training. Considering students' familiarity with the step and ways of the two kinds of feedback, they were given the last three times of writing materials for experiment.

The compositions of EC group were assessed by Jukuu, while the written writings of CC group were valued by TF. For the writing that the teacher gave feedback to, the English teacher assigned a writing task in class firstly, and required students to hand in their writing within a week. After the students handed in the first draft, the teacher would grade the writing and give feedback from different aspects, such as the discourse structure, content, language expression, grammar and so on. Then the first draft was returned to the students again. And students could voluntarily decide to revise the writing or not. To the revised draft, the teacher would give comments and re-grade it. As for the writing which would be assessed by Jukuu, the English teacher also assigned the writing task in class, but the students were required to write on Jukuu grading network. After the students submitted their writings online, Jukuu would make comments from many aspects, such as the choice of vocabulary, spelling, structure and so on and give score immediately. The students could revise the compositions based on the system's suggestions, if they are not satisfied with the score. Once the writing was resubmitted, the students could get a new score. Multiple drafts could be submitted in this experiment. In this way, all the students are required to finish 4 writing tasks.

3.4.4 Post-test

Just like the previous pre-test, a post-test (see Appendix I) was taken simultaneously in CC and EC at the end of December. After all the post-test papers were analyzed as before, SPSS 19.0 was used to analyze whether there are significant differences between EC and CC in writing proficiency and each part of their compositions after the whole experiment.

3.4.5 Questionnaire Survey

As a vitally important tool, after all the participants finished the post-test and the four writing tasks assigned by writing teacher in class and in Jukuu grading net, two questionnaires, which includes questionnaire on TF and questionnaire on AES, were delivered to both classes to find out their overall perceptions towards TF and AES. Each questionnaire consists of 20 question items and was employed anonymously. The 20 items can be broadly classified into three sections. The first section, covering Items 1-7, aims to elicit students' emotional changes to the two different kinds of feedback. The second section, involving Items 8-14, inquires into the students' attitude towards the effectiveness of the two types of feedback. The third section, including Items 15-20, mainly focuses on the students' viewpoints on the guiding significance of the two kinds of feedback on the future amendment of writing skills and proficiency. To be easily understood by the students, the words used in these two questionnaires are in Chinese. An interval of the two questionnaire survey lasts for one week so as to get real and objective data. The results of the statements were measured by Likert-type scale of five points ranging from strongly agree to strongly disagree. The response rates of delivery and collection of the twoquestionnaires were both 100%.

3.5 Data Collection

As the essential part of results analysis and discussion, data collection was seriously executed. All the needed data from two English tests (pre-test, and post-test), the four writing tasks and the two questionnaires after experiment were collected. Meanwhile, all the collected writing tasks were sorted carefully according to the category of errors of Chandler (2003) (see Appendix III). Each questionnaire was classified carefully while some incomplete questionnaires were not used. After each test was taken, the data collected from pre-test, post-test and writing tasks were input into computers in the form of excel tables so as to store the data of EC and CC for the following data analysis.

4. Results Analysis and Discussions

4.1 Respective Focus of TF and AES on the Correction of Students' Writing

After finishing the writing tasks, the author collected and classified the last three times' writing works in terms of Chandler's Category of Errors (see Appendix III) and employed the statistic analysis instrument SPSS 19.0 to further analyze these data and the outputs are stated in the following table (see Table 1).

Table 1: Percentage Statistics of the Writing Tasks

Content of feedback	AES feedback	P	TF	P
Discourse structure	20	8.33%	75	33.12%
Content	33	13.75%	65	26.76%
Language expression	90	37.50%	50	24.19%
GSP	80	33.33%	31	13.72%
Other things	17	7.08%	5	2.21%
Total number of feedback	240		226	

Note: P=The proportion of the total number of feedback; GSP= Grammar, spelling, and punctuation.

From the above statistics, we can see that AES has more strips of feedback than TF and the distribution of feedback types is also different. Specifically, the feedback of AES focuses more on language expression, grammar and spelling errors, and the feedback on discourse structure and content is relatively less. The distribution of TF is relatively uniform. It would be specially mentioned that TF lays more stress on discourse structure and content than AES.

In terms of discourse structure, both TF and AES have proposed suggestions on the modification of transition, cohesion and coherence, but TF also puts forward suggestions on the article layout and the unity of a writing. As for content, both groups figure out the irrelevant problems. In addition, TF makes some useful suggestions for the extension and deepening of the article. About grammar, spelling and punctuation errors, there are about 13 types of mistake, and tense, spelling and capital errors are the top three. Meanwhile, AES group also modified the errors of clauses. With respect to vocabulary, there are no significant differences between two groups and they mainly focus on the improper use of vocabulary and matching errors.

From the previous study, we know that Jukuu scores points by calculating the similarity between students' writing and standard English from corpus. As a machine, it inevitably has some weak points in flexibility and comprehensiveness. Therefore, once it pays more attention to grammar and vocabulary, relatively the attention on discourse structure and content would be paid less to. This is the reason why the feedback of AES lays its focus more on language expression, grammar, spelling and punctuation errors, while the focus on discourse structure and content is relatively less. The results show that AES pays more attention to vocabulary and common grammar mistakes such as spelling, lexical collocation and so on, and does not concern too much about structure, content and organization of the writing.

TF is more time-consuming compared with AES, so usually it is impossible to examine a writing carefully. For most Chinese teachers of English, they only have one chance to scan the writing from beginning to end. Therefore, teachers tend to take these aspects more seriously such as cohesion and coherence, the article layout and the extension and deepening of the article, but neglect some errors in grammar and vocabulary. The conclusion shows that TF was inferior to peer and online feedback in polishing language expression and enriching ideological content.

4.2 Results and Discussion of Pre-test and Post-test Writing

4.2.1 Scores of Two Classes in Pre-test and Post-test Writing

All the participants in EC and CC took part in pre-test at the same time, at the beginning of the experiment. Then the relevant statistics were input into SPSS19.0 and the output of Independent-samples T-test is like what Table 2 and Table 3 show in the following.

Table 2: Group Statistics of Pre-test in EC and CC

	class	N	Mean	Std. Deviation	Std. Error Mean
Pre-test	EC	40	11.55	1.93890	.33216
	CC	40	11.26	1.98435	.34312

As Table 2 shows, the mean score of EC is 11.55, which is rather similar to that of CC (11.26). While the mean score in EC is a little bit higher than that in CC (11.55>11.26), the disparity is only 0.29 which is not so significant. Moreover, the standard deviation is 1.93890 in EC which is almost the same as CC (1.98435), too. Therefore, according to the two items of statistics, it is claimed that the average writing proficiency of these two classes is almost at the same level. A further analysis of pre-test would be carried through, for the purposes of examining if there exist any significant differences among the participants in each class. Therefore, the following Independent Samples T-Test (See Table 3) just illustrates this question clearly.

Table 3: Independent Samples T-test Comparison of Pre-test in EC&CC

		Levene's Test for Equality of Variances		T-test for Equality of Means							
										95% CI	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Pre-test	Equal variances assumed	.334	.645	.140	80	.876	.03334	.35747	-.84932	.96547
	Equal variances not assumed			.140	79.955	.876	.03334	.35747	-.84934	.96548

Note：95%CI= 95% Confidence Interval of the Difference

According to Table 3, the significant difference (sig.) of Levene's Test for Equality of Variances is 0.645 (>0.05), which surely indicates that the variances of scores in pre-test of the two classes have no significant difference. Furthermore, the Mean Difference is merely 0.03334 and Sig. (2-tailed) is 0.876 (> 0.05), which also signifies that the mean scores between EC and CC have no obvious difference. In addition, 95% Confidence Interval of the Difference is from -0.84932 to 0.96547 and obviously it includes 0, which also signifies the two classes have no significant difference and nearly own the same writing level before the study, which can insure the validity and reliability of this experiment at the beginning of the experiment.

The post-test is conducted in the same way as the pre-test at the end of the experiment. The results are brought out after inputting the scores into SPSS19.0, and thus the analysis of the relevant statistics could be stated clearly. *Group Statistics of Post-test in EC and CC* shows that the mean score in EC is 14.7479, which is about 0.5164 points higher than that of CC (14.2315). Obviously, the disparity is insignificant. Moreover, the standard deviation is 1.72984 in EC which is very similar to CC (1.73456). It means that the students' writing proficiency has no significant progress in both EC and CC. From *Independent Samples T-test Comparison of Post-test in EC and CC*, we can see that the Sig. (2-tailed) 0.864 is higher than p (0.05), which also means there is no obvious differences in mean scores between EC and CC. Moreover, 95% Confidence Interval of the Difference is from -0.73489 to 0.16743 and it obviously does contain 0, which also can prove students have insignificant change in their writing competence after this experiment.

In a word, the students' writing proficiency has insignificant change in both EC and CC, and that is to say AES and TF almost have the same effect on student's writing proficiency.

In order to inspect if there are some significant differences in students' composition scores in EC and CC respectively between pre-test and post-test after the whole term's experiment, a further Paired Samples T-test in EC and CC would be separately carried through in the following.

4.2.2 Paired Samples T-test of Pre-test and Post-test in EC and CC

Table 4: Group Statistics of Pre-test and Post-test in EC

		Mean	N	Std. Deviation	Std. Error Mean
EC	Pre-test	11.55	40	1.939	.332
	Post-test	14.75	40	1.730	.281

Table 5: Paired Samples T-test Comparison in EC

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
EC	Pre-test-post-test	-3.2	0.209	.051	-3.534	-2.722	2.768	39	.022

As Table 4 illustrates, the mean score of the post-test in EC is 14.75 which is about 3.2 points higher than that of pre-test(11.55). Therefore, the students in EC has made significant progress. Besides, Table 5 shows that the Sig. (2-tailed) is 0.022(<0.05), which obviously demonstrates that there are significant differences in writing scores between pre-test and post-test within EC. That is to say, the mean score of students' writing has been significantly higher improved in post-test than that in pre-test. Therefore, the result just validates that students in EC have made certain progress in their writing proficiency after a whole term's writing practice experiment.

Meanwhile, the data in *Group Statistics of Pre-test and Post-test in CC* show that CC group has gained improvement in writing proficiency. The reason is that the mean score in pre-test is 11.26 and the gap between pre-test and post-test is 2.971. Obviously the disparity is significant. Meanwhile, Sig. (2-tailed) of two tests in *Paired Samples T-test Comparison in CC* is 0.03, which validates there is obvious difference in students' writing proficiency between pre-test and post-test in CC. So it can be summarized that the students' writing proficiency in CC has been enhanced.

4.2.3 Discussion of Pre-test and Post-test Writing

On the strength of statistical results of Paired Samples T-test Comparison within EC and CC, it can be found that the mean scores in both EC and CC have significant difference between the pre-test and post-test. That is to say, students' writing proficiency in both EC and CC has been enhanced and students in the two classes all make great progress in English writing.

However, in the pre-test, the difference of mean scores between EC and CC has no significance, which indicates there is no obvious difference in students' writing proficiency between the two classes at the beginning of the experiment. In the post-test, the difference of mean scores between two classes is also not significant, which means students' writing proficiency has insignificant change in both EC and CC, and the AES and TF almost have the same effect on student's writing proficiency.

4.2.4 Results and Discussion of Writing Skills in Pre-test and Post-test Writing

In order to further study whether AES and TF have different effects on each aspect of students' compositions which includes structure, content, language and grammar etc, the mean scores concerning four aspects for pre-test and post-test of CC and EC are calculated and the results are shown in Table 6 and Table 7.

Table 6: Mean Scores Concerning Four Aspects for Pre-test of CC and EC

	Experimental Class		Control Class		
	Mean	Std. Deviation	Mean	Std. Deviation	Sig.(2-tailed)
Structure	3.956	.712	4.174	.703	.534
Content	4.305	.698	4.325	.654	.154
Language	4.615	.703	4.603	.734	.078
Grammar etc.	4.421	.745	4.230	.765	.106

Table 7: Mean Scores Concerning Four Aspects for the Post-test of CC and EC

	Experimental Class		Control Class		
	Mean	Std. Deviation	Mean	Std. Deviation	Sig.(2-tailed)
Structure	4.445	.734	5.402	.687	.030
Content	5.101	.762	5.382	.710	.015
Language	5.593	.693	5.132	.721	.024
Grammar etc.	5.645	.706	5.065	.745	.031

From Table 6 we can see that there is no significant difference between two classes on the mean scores concerning four aspects in the pre-test. That is to say, there is no obvious difference between the two classes in terms of different language aspects respectively. Totally different from Table 6, Sig (2-tailed) of the four aspects in Table 7 are all lower than 0.05, which means that in the student's post-test composition there is obvious progress in the structure, content, expression and grammar. In order to further explore the effects of TF and AES on the four aspects of writing skills, the mean scores of CC and EC concerning four aspects for the pre-test and post-test are separately conducted and the results are listed in Tables 8 and 9.

Table 8: Mean Scores of CC Concerning Four Aspects for the Pre-test and Post-test

Control Class	Pre-test		Post-test		
	Mean	Std. Deviation	Mean	Std. Deviation	Sig.(2-tailed)
Structure	4.174	.703	5.402	.687	.027
Content	4.325	.654	5.382	.710	.035
Language	4.603	.734	5.132	.721	.062
Grammar etc.	4.230	.765	5.065	.745	.079

According to Table 8, students in CC have made more obvious improvement in discourse structure and content for the reason that the Sig.(2-tailed) are lower than 0.05 on the aspects of structure and content. In other words, TF does contribute to the improvement of students' writing skill in structure and content.

Table 9: Mean Scores of EC Concerning Four Aspects for the Pre-test and Post-test

Experimental Class	Pre-test		Post-test		
	Mean	Std. Deviation	Mean	Std. Deviation	Sig.(2-tailed)
Structure	3.956	.712	4.445	.734	.245
Content	4.305	.698	5.101	.762	.164
Language	4.615	.703	5.593	.693	.032
Grammar etc.	4.421	.745	5.645	.706	.015

From the above Table 9 it can be obviously seen that the Sig.(2-tailed) in the aspects of language and grammar respectively are 0.032 and 0.015 which are separately lower than p (0.05), so the students in EC make more obvious progress in vocabulary and grammar. That is to say, AES focuses more on the aspects of grammar and vocabulary on the process of writing correction.

Based on the above data analysis of the experiment, we can conclude that the two different kinds of feedback have totally different points of focus on the aspects of students' compositions. As for TF, there is significant improvement in discourse structure and content, compared with other two aspects. Different from TF class, the language expression and grammar are significantly enhanced in AES class. It proves once again that both TF and AES respectively have various effects on the aspects of students' writing.

4.3 Students' Attitude towards TF and AES

In order to check what the differences between students' perceptions towards TF and AES, the questionnaire was adopted. The questionnaire includes two pieces (questionnaire on TF and questionnaire on AES). The whole body is composed of 20 items which can be broadly classified into three sections. Items 1-7 constitute the first section, which aims to elicit students' affection to the two different kinds of feedback. The second section, involving Items 8-14, inquires into the students' attitude towards the effectiveness of the two types of feedback. The third section, including Items 15-20, mainly focus on the students' viewpoints on the guiding significance of the two kinds of feedback on the future amendment of writing skills and proficiency.

After carefully collecting the questionnaires, the data are put into SPSS 19.0 and the results show that mean scores of students' affection to the two different kinds of feedback are very close (TF=3.245; AES=3.272), which reveals that students' affection to TF and AES has no significant difference and the Sig. 0.073 in *Paired Samples T-test Comparison in Students' Affection* is higher than significance level 0.05, which further confirms the result. After that, the mean scores of students' attitude towards the effectiveness of the two types of feedback are 4.334 and 3.213 respectively, which indicates that there is significance difference between students' perceptions towards TF and AES in terms of effectiveness. In addition, sig, in *Paired Samples T-test Comparison in Students' Attitude to Effectiveness* is 0.000, much lower than 0.05, which demonstrates that students think TF does work better in the aspects of effectiveness

At last, we have got the *Mean Scores of Students' Attitude to Guiding Significance,* which is 4.253 in TF group, while the mean score of AES is 3.168, much lower than TF. That is to say, there is significance between students' perception towards TF and AES in terms of providing guidance for future writing. In addition, sig. in *Paired Samples T-test Comparison in Students' Attitude to Guiding Significance* is 0.013, much lower than 0.05, which further proves the result.

5. Discussion of the Findings

On the basis of previous data analysis and discussion, the major findings of this study are as follows:

Firstly, AES pays more attention to language expression, grammar, spelling and punctuation errors. On the contrary, TF lays more stress on discourse structure and content. As for different aspects of English writing, the two kinds of feedback have their own emphases. In terms of discourse structure, both TF and AES have proposed suggestions on the modification of transition, cohesion and coherence, but TF also puts forward suggestions on the article layout and the unity of the whole body in a writing. As for content, both groups figure out the problem of digressions and irrelevance. In addition, TF makes some useful suggestions for the extension and deepening of the article. Concerning grammar, spelling and punctuation errors, there are about 13 types of mistake with tense, spelling and capital errors at the top three. Meanwhile, AES group also modifies the errors of clauses.

With respect to vocabulary, there are no significant differences between two groups and they mainly focus on the improper use of vocabulary and matching errors.

Secondly, both TF and AES are effective in enhancing college students' writing proficiency, although the results of Independent Samples T-test indicate that the two kinds of feedback almost have the same effect on students' writing proficiency. On the one hand, the mean scores in both EC and CC have significant difference between pre-test and post-test. That is to say, students' writing proficiency in both EC and CC has been enhanced and students in the two classes all make great progress in English writing. On the other hand, the mean scores in pre-test and post-test have no significant difference, which means AES and TF almost have the same effect on the student's writing proficiency.

It has been found that two different kinds of feedback have totally different focus on the aspects of students' compositions. As for TF group, students have made significant improvement in discourse structure and content rather than language expression and grammar. The reason is that TF is more time-consuming compared with AES, so usually it is impossible to read or correct a writing carefully and most Chinese teachers of English only have one chance to scan the writing quickly. However, different from TF class, the language expression and grammar are significantly enhanced in AES class. Because, as a machine, AES pays more attention to grammar and vocabulary, so it inevitably has some weak points in flexibility and comprehensiveness. Therefore, once it pays more attention to grammar and vocabulary, the attention on discourse structure and content would relatively be paid less to.

Thirdly, as the questionnaires indicate, students prefer TF in terms of its effectiveness and guiding significance in their future writing, but they do not deny the function and necessity of AES. With respect to the effectiveness of two kinds of feedback, students think TF is more accurate and fairer, and TF can provide more useful feedback opinions than AES. Moreover, they believe TF can provide comments in every aspects like discourse structure, content, logic and coherent problems, while AES is very mechanical. In addition, students believe TF has profound guiding significance for their future writing, especially in the discourse structure, content and organization, etc. However, in terms of affection, the students show no distinction as to which one they prefer better. At the same time, the data offer a new angle of integration of TF and AES for the future writing teaching practice.

6. Pedagogical Implications

Firstly, teachers should adopt a variety of feedback modes and adequately take the advantages of them. Although the two types of feedback are different from each other in effectiveness and perception, there is no doubt that both of them play vital role in enhancing students' writing proficiency and language skills. Moreover, if students always receive the same form of feedback from the teacher, they will lack motivation to continue to write because of being bored. Therefore, we should flexibly adopt a variety of feedback modes in our present English writing teaching according to the different levels of students' cognitive competence and language proficiency. Time-consuming is one big drawback of TF, while AES is convenient for its time saving. So AES can effectively deal with the core problems of big-class and excessive workload to English teachers in China. As the study indicates, TF could strengthen students' consciousness of text-structure and effectively broaden ideological content of the article, while AES has positive effect on promoting language competence and grammatical accuracy. In other words, these two modes have their own advantages. Only by combining the two types of feedback effectively and properly, can the reasonable feedback be given to students' writing works. Therefore, employing AES together with TF in a big English class may prove to be more reasonable and practical to help students to improve their writing proficiency.

Secondly, teachers should give full play to the role of guidance, management and supervision and ensure the effect of feedback. Teachers should not only try to be the daunting embodiment of authority raters and supervisors, but try to be mentors, participants, organizers, coordinators and readers who can equally communicate with the students. As a result, in the implementation of TF, teachers should establish equal, friendly and cooperative relations with students and ensure that TF could be understood and accepted by the students through the way of simplifying the general suggestions of modification. In the course of AES, if necessary, teachers should guide students to proactively rewrite their writing tasks and seriously think about the feedback and make active response to it.

7. Suggestions for Future Research

In the first place, future researcher in this area might attempt to include a larger population to compare their previous academic backgrounds. For example, English majors vs. non-English majors, and high-achiever vs. low-achiever student writers. In the second place, if possible, a longer time should be lasted in the further study by future researchers. In order to provide better finding results for the research purpose, the questionnaires should be designed more carefully and data analysis should be more comprehensive at the same time, while an interview should be carried out after the questionnaires. In the third place, future researchers could respectively select one kind of AES software at home and abroad, and focus on one or two aspects to make a comparison. The results may have profound guiding significance for both English teaching and software development.

8. Conclusion

Above all, this comparative study have explored the three research questions and found out the answers to them. In terms of the effectiveness, the data revealed that TF can point out more problems on discourse structure, content, logic and coherent and it is far fairer and more accurate than AES in students' opinion.

As far as guiding significance is concerned, it is believed that TF is conductive to helping teachers to evaluate students' writings in many aspects, such as building a better discourse structure, enriching the writing contents, making up a logical and coherent composition, etc. The data proved that TF is more beneficial compared with AES. Therefore, TF is of vital importance in students' mind and students prefer TF.

However, the study does not deny the function and necessity of AES, which has been proved that students' affection to TF and AES has no significant difference. In a nutshell, the data offers a new angle of integration of TF and AES for future writing teaching practice with some pedagogical implications and suggestions for future research. Considering the fact that there are few such studies being conducted, it is hoped that language teachers may choose TF along with AES to evaluate students' writings tasks and especially it is advisable for English teachers in a big class in China to integrate two means of feedback in writings according to students' learning levels and teachers' actual needs.

References

Arndt, V. (1993). Response to writing: using feedback to inform the writing process. In M. Brock & L. Walters (Eds.), *Teaching Composition around the Pacific Rim:Politics and Pedagogy.* Cleveland, Philadelphia & Adelaide: Multilingual Matters, 90-116.

Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, (9): 227-257.

Chung, K. & O'Neil, H. (1997). *Methodological approaches to online scoring of theses.* ERIC reproduction service .

Ferris, D. R. (1995). Teaching ESL composition students to become independent self-editors. *TESOL Journal*, (4):325-328.

Grimes, D. & Warschauer, M. (2008). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Language, and Assessment*, 4):45-49.

Hedgcock, J. & Lefkowitz, N. (1994) Feedback on feedback: assessing learner receptivity to teacher response in L2 composing. *Journal of Second Language Writing*, (3): 141-163.

Hutchison, D. (2007). An evaluation of computerized thesis marking for national curriculum assessment in the UK for 11-year-olds. *British Journal of Educational Technology*, (38): 977-989.

Knoblauch, C. H., & L. Brannon. (1981). Teacher commentary on student writing: The State of the Art. *Freshman English News*, (2):1-4.

Kukich, K. (2000). Beyond Automated Essay Scoring. *The Journal of Technology, Learning, and Assessment*, (9):56-60.

Lee, I. (2004). Error correction in L2 secondary writing classrooms: the case of Hong Kong. Journal of Second Language Writing, (3): 285-312.

Lee, I. (2007). Feedback in Hong Kong secondary writing classrooms: assessment for learning or assessment of learning? *Assessing Writing*, (12): 180-198.

Leki, I. (1990). Coaching from the margins: Issue in written response. In B. Kroll (ed.) *Second Language Writing: Research Insights for the Classroom*. Cambridge: Cambridge University Press, 43-56.

Liang Maocheng & Wen Qiufang. (2007). A Critical Review and Implications of Some Automated Essay Scoring Systems. *Computer-Assisted Foreign language Education*, (5): 18-24

Radecki, P, & Swales, J. (1988). ESL student reaction to written comments on their written work. *System,* (16), 355-365

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal,* (1): 23-30.

Shermis, M. & Burstein, J. (2003). *Automated Essay Scoring: A Cross Disciplinary Perspective.* Mahwah, NJ: Lawrence Erlbaum Associates.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), Input in Second Language Acquisition. Rowley, MA: Newbury House, 235-254.

Swain, M. (1995) Three functions of output in second-language learning. In G. Cook and B. Seidlhofer (Eds.), Principle and Practice in Applied Linguistics: Studies in Honor of H. G. Widdowson, 125-144.

Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, (2):147-170.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, (19): 79-101.

Appendix I Test Papers

1. Pre-test Paper

You have received the invitation, but you will not be able to attend the party. Write a letter to Mr. Smith, explaining the reason why you cannot go and how you plan to make up for it. You should write approximately 120 words. Do not sign your own name at the end of the letter. Use "Wang Lin" instead. You do not need to write the address.

2. Post-test Paper

Write a letter of complaint. You placed an order for clothes on March 15th, however, the store has delayed the delivery for a week. After you received the pack, there are some problems in quality. So you decided to write a letter of complaint to the storekeeper-Wang Lin. You should write approximately 120 words. Do not sign your own name at the end of your letter. Use "Li Ming" instead. You do not need to write the address.

Appendix II Writing Tasks

Assignment One

Your American friend Peter Jones is coming to your hometown for a one-week visit．Write to him about：

1) the weather conditions of your hometown for the season；
2) the accommodation arrangements for him；
3) the schedule for his sightseeing activities．
You should write approximately 120 words．Do not sign your own name at the end of your letter．Use "Wang Lin" instead．You do not need to write the address.

Assignment Two

A middle school lacks some qualified teachers of English. Write a letter of recommendation of Mr. Zhong Yong, a teacher of English at a middle school. Your letter should refer to the following points:

1) Personal data 2) His work experience 3) His qualifications

You should write approximately 120 words. Do not sign your own name at the end of your letter. Use "Zhang Ping" instead. You do not need to write the address.

Assignment Three

Your friend Li Ming has written to invite you to go to his hometown along with him and you are willing to accept his invitation. Write a reply to Li Ming

1) to express your appreciation and acceptance of his invitation
2) to ask about his schedule for the trip
3) to ask about what necessary preparations you need to make.
You should write approximately 120 words. Do not sign your name at the end of your letter. Use "Wang Lin" instead. You do not need to write the address.

Assignment Four

Suppose you had a bad cold during your last trip to a famous mountain. The tourist guide has helped you a lot. Write a letter of thanks to the travel agency. Your letter should include:

1) a description of your case
2) thanks to the tourist guide and the travel agency
You should write approximately 120 words. Do not sign your own name at the end of the letter. Use "Wang Lin" instead.

Appendix III Category of Errors and Examples

Category of errors	Examples
Sentence structure	I suggest you to go with me.
Meaning unclear	I thought how to divide my cabins in a good way.
Idiom	She asked me to lay the table.
Punctuation	I wanted her to come. because she knew the way.
Capitalization	She said Love was very important.
Delete	She told to me her secret.
Insert	He is listening me.
Wrong word	He is becoming to mature.
Agreement	She have much to do.
Verb tense	If I know, I would tell you.
Verb voice	When it be happens, we will see.
Plural	Two woman come.
Article	He is a important guest.
Pronoun	She is a friend of myself.
Spelling	Chrismas