

The Most Preferred and Effective Reviewer of L2 Writing among Automated Grading System, Peer Reviewer and Teacher

Tsai, Min-Hsiu^{1,*}

¹Department of Modern Languages, National Pingtung University of Science and Technology, Pingtung County, Taiwan, R.O.C.

*Correspondence: Department of Modern Languages, National Pingtung University of Science and Technology, Pingtung County, Taiwan, R.O.C. Tel: 886-0966-996-107. E-mail: minliecool@gmail.com

Received: June 30, 2017

Accepted: July 11, 2017

Online Published: August 14, 2017

doi:10.5430/wje.v7n4p60

URL: <https://doi.org/10.5430/wje.v7n4p60>

Abstract

Who is the most preferred and deemed the most helpful reviewer in improving student writing? This study exercised a blended teaching method which consists of three currently prevailing reviewers: the automated grading system (AGS, a web-based method), the peer review (a process-oriented approach), and the teacher grading technique (the product-oriented approach) in a Writing (IV) class involving 22 technological sophomore students of Modern Languages Department. The questionnaire results indicated the participants preferred the teacher as the reviewer to their peers followed by the automated grading system and considered the teacher the most effective in helping their writing. Three L2 teachers including one native speaker of English reviewed an essay which was the only and the most inconsistent case between a human rater and a machine rater in the study (2.3 vs. 3.6). This case surfaced an essential problem that the automated grading system couldn't detect and correct expressions transferred from L1. Data also revealed that teachers without training, their grammatical error identification rates are respectively 82.9%, 31.4% and 74.3%. After training, student reviewers could detect and correct from 70.2 to 79.3 percent of grammar errors on average.

Keywords: *automated grading system (AGS); peer review; English as foreign language (EFL); second language writing (L2 writing)*

1. Introduction

Second language (L2) teaching and research often lag behind first language at least ten years (Susser, 1994). From the very beginning, L2 research centered on Contrast Analysis. When scholars discovered L2 errors couldn't be all attributed to L1 transfer, L2 research focus turned to error analysis. Analyzing L2 learners' errors which didn't help L2 learning made the error analysis enterprise almost come to the end. Thanks to the advance of technology. The computer-aided tools such as knowledge-based parsers, grammar checkers, discourse processing analyzers, automated grading systems, and L2 corpus brought L2 research back to life.

Automated grading system and peer review are two current ways that might reduce writing teachers' loading in grading students' work and giving corrective feedback. Nevertheless, they are not thoroughly trusted. Consequently, reducing the number of students in a class seems to be the only way to resolve the problem temporarily. However, peerScholar developed by the University of Toronto Scarborough, Canada (Pare & Joordens, 2008) integrated and employed the web 2.0 concept to develop an online peer review tool for assessing critical thinking and written text, and thus claimed this technique could make the writing classes return to a quite large class size possible.

Lai (2010) compared the effectiveness of AGS with peer evaluation (peer review) and found that "EFL learners in Taiwan generally opted for peer review over AGS." The finding inspired the researcher to conduct a research to find out who the most preferred reviewer would be if the teacher joined the force for comparison.

The researcher taught a writing (IV) class of 22 technological sophomore students of Applied Modern languages Department. The instruction blended three assessing methods together: the automated grading system (web-based computer-aided evaluation method), the peer review (a process-oriented approach), and the teacher grading technique (the product-oriented approach). Even though this is a case study of writing instruction and pretty much action-research-like, substantial findings follow.

2. Literature Review

2.1 Peer Review

Coit (2010) depicted the history of peer review coming to play in the writing teaching industry:

Peer review has long been used by academics to give other academics feedback on a paper, text, or other piece of work they have written or are writing. Since the middle of the 20th century, the term peer review has come to designate a method used to select and qualify research which has been submitted for publication.

Starting around the 1970s, the idea of using peer review in L1 writing classes developed as a logical consequence of teachers using the process approach to teaching writing. There were several reasons for this. To begin with, attempting to check and give advice on the many process steps, the interim drafts, as well as final papers could easily become over-demanding and too time-consuming for teachers. On the one hand, teachers were praising the process approach to teaching writing for motivating their students to write more drafts than they had under the product approach, but on the other hand, they often found it hard to keep up with all of the feedback required of them for the different drafts. Consequently, process teachers began to feel the need to turn to peer review for assistance in giving interim feedback.

As a consequence, some process writing teachers began to think of peer review as a valid alternative to teacher-centered feedback on mappings, outlines, and other interim feedback they would normally have provided (adapted from Coit, 2010, p. 56-57).

Tons of research studies reported students hold positive attitude toward peer review activity (Berg, 1999a; Berg, 1999b; Bartles, 2003; Min, 2006; Chen, Yi-Hsuan, 2009; Yang, 2011). Chen, Yi-Hsuan (2009) found peer feedback raised students' awareness of surface-level errors and fostered their critical thinking on writing contents. Peer review activity was viewed as a complement of teacher feedback and an effective way of cooperative learning to EFL student writers (p. ii).

Teaching writing in L1 and L2 is totally different. Young (1978) stated that "the idea that writing in the L2 was mainly a mechanical tool to be acquired through exercises in spelling and grammar." Coit (2010) continued Young's statement: "it would continue long after changes in the theories of writing in the L1 had become well established" and "Research about teaching writing in the L2 generally lagged behind the developments which were taking place in the teaching of writing in the L1" (p. 59). Thus, even though the process approach dominated much of the research on writing in L1 during the 1970s, it did not have the same effect on research carried out on the teaching of writing in L2 classrooms (Coit, 2010, p. 59). However, with the promotion by ESL textbooks and publications, "by the late 1980s [in the United States] process writing pedagogies had reached the mainstream of ESL writing instruction" (Susser, 1994 cited by Coit 2010, p.60).

On the contrary, Applebee (1986) carried out a study to look at actual teaching practice for process oriented writing in L1 classrooms and found few papers went beyond the first draft, and even on the first draft, 60 percent showed no revisions of any kind. Thus, he concluded the process-oriented writing was failing and there had been no widespread movement toward process-oriented assignments in American schools and colleges (Applebee, 1986, cited by Coit, 2010, p. 58).

Researchers such as Mendonca & Johnson (1994) and Yang (2011) pointed out that even though peer review method was deemed helpful in enhancing students' writing, doubt of pupils' ability in evaluating their peers' writing arose. Nelson & Murphy (1993) reported neither the reviewers nor the receivers of peer review took it seriously enough. Besides, Nelson & Carson (1998) indicated that students clearly preferred teacher feedback to interim peer review. Coit (2010) wrote in her dissertation that more than 88% of the students in her experiment agreed with the corrections made by their teachers. In contrast, 75% of the participants who received peer feedback said they didn't always agree with the corrections their peers had made (p.194). However, Coit (2010) proposed the use of Student Empowered Peer Review (SEPR) to develop dialogical academic writing and claimed extra writing practices were as beneficial as teachers' correction feedback; students who exercised more writing practices gained more average scores than those who did not; and empowering students to score the final drafts significantly improved their average grades on the structure part but not on the language mechanics part. Berg (1999b) and Min (2006) emphasized the importance of trained peer response for success in student revision types and writing quality and demonstrated the training process.

2.2 From Error Analysis and Automated Grading System

Error analysis prevailed and reached its height in the 1970s (Dagneaux et al., 1998, p.163), when “many researchers were convinced that behaviorism and Contrast Analysis Hypothesis had been inadequate explanation for second language acquisition” (Lightbrown & Spada, 2006, p. 35) because second language (L2) learners’ errors couldn’t be all attributed to their first language (L1). Quite a variety of studies examined all kinds of errors made by language learners from different aspects, such as tense and aspect errors (Huang, 1994), collocation errors (Wang, 2001; Chen, 2002; Hsueh, 2003), relative clause errors (Chen, 2004), article errors (Shih, 2004), and errors types (Huang, 2001).

Error analysis (EA) suffered from a number of weaknesses as being pointed out by Dagneaux, Dennes & Granger (1998, p.164):

1. Error analysis (EA) is based on heterogeneous learner data;
2. EA categories are fuzzy;
3. EA cannot cater for phenomena such as avoidance;
4. EA is restricted to what the learner cannot do;
5. EA gives a static picture of L2 learning.

Many techniques and tools were borrowed from several natural language processing (NLP) field (Lonsdal & Strong-Krause, 2003, p. 61), including knowledge-based parsers such as ALEK (Assessing Lexical Knowledge) and ICICLE system (Interactive Computer Identification and Correction of Language Errors) (Schneider & McCoy, 1998), grammar and spelling checkers such as *CorrectEnglish*, *White Smoke*, discourse processing analyzers (Mitsakaki & Kukich, 2000), and other hand-crafted knowledge-based sources such as *Word Smith*. More importantly, the automated essay grading systems (AEGS) which integrated some or all of the above techniques and tools were introduced to the error analysis enterprise. The current tools available for automated essay grading include Project Essay Grade (PEG), Intelligent Essay Assessor (IEA), Educational Testing Service (ETS 1), Electronic Essay Rater (E-Rater), Conceptual Rater (C-Rater), etc. (Valenti, Neri, & Cucchiarelli, 2003).

Burstein & Chodorow (1999) and Lonsdale & Strong-Krause (2003) both pointed out that automated grading systems that were originally designed for scoring native English speakers’ written essays significantly differed from grading the L2 learners’ written articles, especially those of lower language proficiency. Furthermore, both Lonsdale & Strong-Krause (2003) and Tsai (2010) concluded special care should be taken to assist the grading procedure while employing AGS for rating the articles that fell into the two extreme ends.

2.3 Writing Theories

Writing theories have advanced from Product Oriented (roots on the learning theory of Behaviorism), Process Oriented (Cognitive Constructivism), Post-Process/Social Constructionism (social Constructionism) to Dialogical Genre Studies/New Rhetoric (dialectical Activity Theory) (Coit, 2010, p.41). And the methods of teaching writing have shifted from outlines, controlled, guided, editing, brainstorming, mappings, free-writing, writing drafts, intervention, revision, rhetorical analysis, discourse analysis, genre analysis, to stabilized-for-now genres analysis of activity systems (Coit, 2010).

Currently prevailing genre writing in L1 writing class bears the idea to prepare students for future workplace. A genre-based approach to academic writing may require learners to write recount texts, instruction texts, one-sided argument texts, two-sided argument texts, explanation texts, classifications texts, and/or blended texts (Johnson & Crombie, 2010). The field of writing teaching was one of the last to have been influenced (Coit, 2010). In view of these teaching theories and methods, writing teachers should be aware of where we went, where we are, where we are going. Besides, L2 writing teachers might want to try it from time to time to keep up with the pace.

3. Methodology

3.1 Participants

The present study involved an English Writing (IV) class of 22 sophomore students from the Modern Languages Department at National Pingtung University of Science and Technology. Among them, three are boys and the others are girls. All of the subjects had taken Writing (I), (II) and (III) courses before the Writing (IV) class because four writing courses in a row are required in the Department. Their English proficiency level averaged about 520 points on the TOEIC. However, one particular case among the participants had reached 920 points on the TOEIC, while the least proficient student gained only 430 points on the same test.

3.2 Instrument

The reviewers employed in the study for assessing students' essays include a web-based automated grading system, the students (peer reviewer), and the teacher. The instrument consists of: a. *Statistica*, the statistic software, was used to analyze the correlation between the human rater's and the e-rater's holistic scores; b. A questionnaire was conducted to discover the participants' preference of certain reviewers; c. three Peer Editing Sheets adopted from Folse et al (2010) for peer review activity in class; and d. three L2 English teachers exercised error tagging technique which is generally practiced in corpus studies.

3.3 Procedure

This study exercised a blended writing instruction that incorporated a computer-aided grading system, peer review activity, and the traditional teacher scoring technique. The textbook used for the class was *Great Writing 4: Great Essays* (Folse et al, 2010, third edition), and the students were required to write six essays which were categorized into five types of essays:

1. Descriptive type: Describe a Movie/TV Program
2. Narrative type: Experiences of Being punished
3. Comparison type: Internet Classroom vs. Traditional Classroom
(This writing practice was graded by an automated grading system.)
4. Cause-Effect type: What Are Three Common Causes of Motorcycle Accidents?
5. Argumentative type: Three options of prompts-
 - a. Do you think students should be penalized for missing classes?
 - b. Is a passing score on an English achievement test necessary for international students to enter a university?
 - c. Is day care beneficial for children under the age of 5?
6. A revised version of the fifth essay.

In the two-period (100 minutes) class meeting time every week, a total of 36 hours a semester, the researcher as the teacher taught the first hour by introducing the textbook content and pinpointing the focuses and requirements of a specific essay to write in the second hour. For the first three essays, the researcher didn't require timed writing. Instead, students were allowed to write in the whole period of time which is 50 minutes in class, and they were permitted to bring in all kinds of dictionaries or references deemed useful for their writing. For the last two essay practices, students were required to write at least 300 words in 30 minutes. The specific number of words was requested to provide at the end of their essay for reference.

Most of the topics or prompts for writing were suggested by the textbook but elected by the students in the class. Once the most students agreed on a topic, all of the students wrote with the same prompt, except the third composition which was assigned by the teacher because the third writing was planned to submit to a web-based automated grading system. Therefore, the prompt must be chosen from a suggested list of the grading system. All of the 5 essays were teacher graded, but the second, the fourth, and the fifth essays were also peer reviewed (Figure 1). For the fifth writing practice, the students asked to grant them more freedom in selecting a topic to write with. Therefore, three prompts were left for them to choose from after they had elected 3 prompts from 5 original ones. The participants actually wrote six essays. Nevertheless, the sixth composition was a revision of the fifth essay.

The students were instructed to type in their third essay which is the comparison essay: *Internet Classroom vs. Traditional Classroom* after the midterm due to time consideration and the account setup reason. Thus, at the time students entered their essays into the computer, they had received the corrective feedback from the teacher and they were allowed to type with the best revision they had done.

Peer review activities were managed right after the students had finished writing their essays each time in class. At the end of the semester, a questionnaire was given to them for responses.

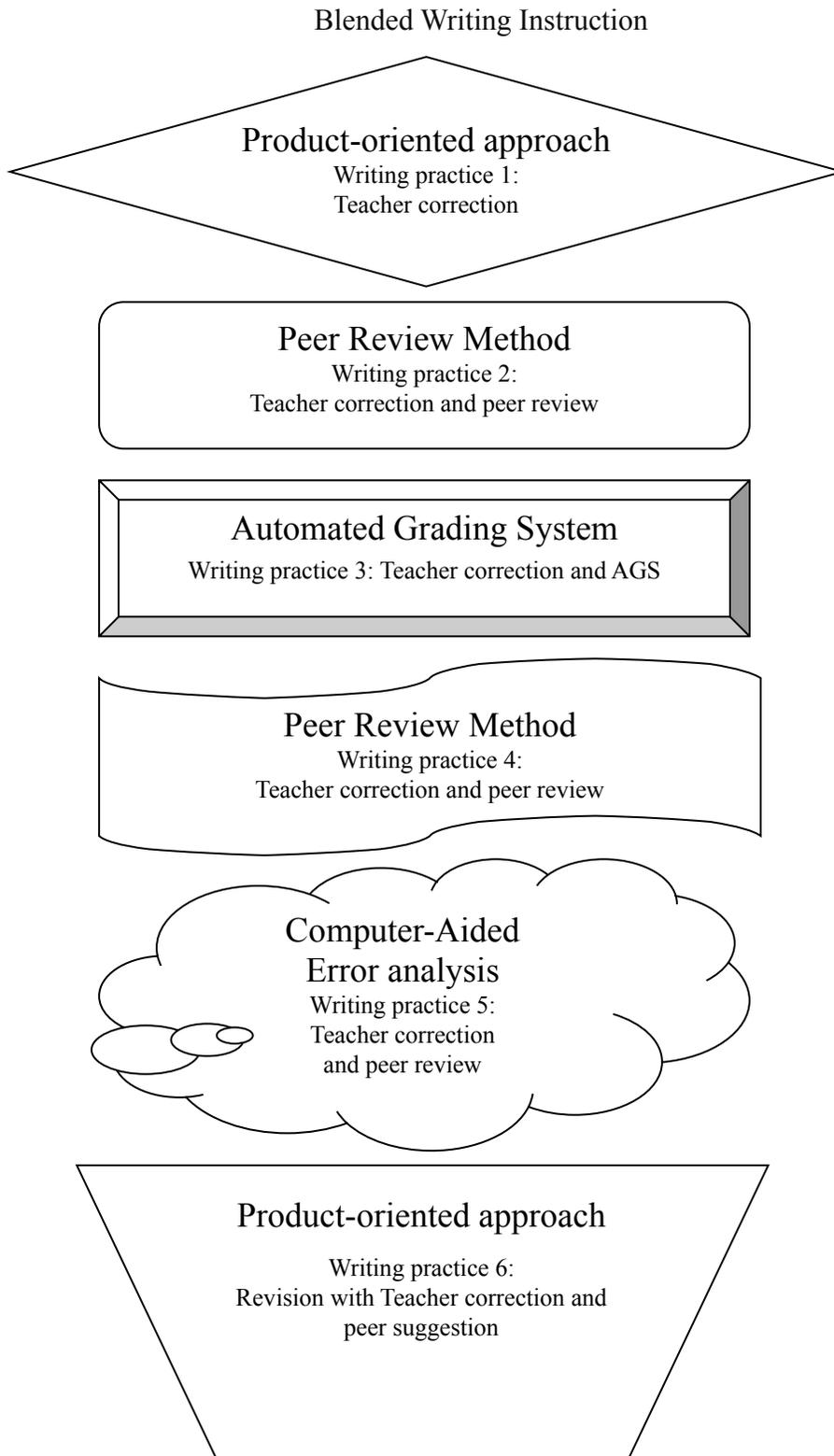


Figure 1. Writing Practices and Different Reviewers

3.4 Research Questions

This study blended three teaching approaches together in writing instruction. Data were collected to explore four research questions.

Question 1: What is the agreement/consistency rate between human rater and e-rater?

Question 2: How competent are students in reviewing their counterpart's writing work?

Question 3: What are the error identification rates of teacher grading?

Question 4: What are students' preferred and deemed effective reviewer(s) among the three?

4. Finding and Discussion

4.1 Research Question 1: What is the Agreement Rate between Human Rater and E-Rater?

In order to answer the first research question, two methods of examination are employed which are correlation and agreement/consistency rate.

4.1.1 Correlation

The students were asked to type their third essays: *Traditional Classrooms vs. Internet Classrooms*, into the grading system, and the teacher used the management function on the system to get the results as follows:

Table 1. Results from Human Rater (Holistic Score Only) and the Machine Rater (Six Measures)

	Holistic score		Focus & Meaning	Content & Development	Organization	Language Use and Style	Mechanics & Convention
	Human rater	E-rater					
1-1		N/A	N/A	N/A	N/A	N/A	N/A
1-2		1.7	1.7	1.5	1.4	1.5	1.4
1	3.4	2.6	2.6	2.3	2.1	2.4	2.2
2	4.2	3.3	3.2	3.0	2.9	3.4	3.3
3	3.2	3.5	3.4	3.1	2.9	3.5	3.3
4	2.6	2.4	2.3	2.2	2.2	2.4	2.2
5	3.4	3.3	3.2	2.9	3.0	3.6	3.4
6	4.1	3.5	3.4	3.1	3.0	3.6	3.5
7	3.0	2.9	2.9	2.6	2.4	2.8	2.7
8	2.9	3.0	2.9	2.6	2.5	2.9	2.9
9	2.3	3.6	3.6	3.3	3.1	3.5	3.5
10	4.3	3.8	3.7	3.4	3.3	3.9	3.7
11	4.0	4.0	3.9	3.4	3.6	4.0	3.8
12	1.8	2.0	1.9	1.8	1.8	2.0	1.9
13	4.6	4.5	4.4	4.0	3.9	4.5	4.2
14	3.2	3.8	3.8	3.4	3.2	3.6	3.8
15	2.8	3.1	3.0	2.6	2.7	3.1	3.0
16	1.8	2.2	2.1	2.0	2.0	2.4	2.3
17	4.6	4.0	3.8	3.5	3.3	4.0	3.8
18	(2.9)	N/A	N/A	N/A	N/A	N/A	N/A
mean	3.30	3.26	3.18	2.89	2.82	3.27	3.15
18-1	2.9	4.0	3.8	3.4	3.5	4.1	4.0
18-2	2.9	3.3	3.2	2.8	2.9	3.3	3.3

Twenty-three articles were typed in the grading system by the participants. Two students entered their essays more than once, thus, only the last copies of the individuals stayed. Four students didn't type their essays into the grading system. Therefore, the total valid essays being retrieved from the system were 18 pieces. The overall average (holistic score) was 3.26 points on a 6-point scale, and the average score of the Language and Style among the five skills measured by the system was the best (3.27), whereas the average score of the Organization, 2.82, was the worst. All in all, the five measured skills were quite evenly developed. The researcher's scores averaged 3.30 which were pretty close to the machine grades.

One essay was first "unscorable" to the system because of a repeated paragraph mistakenly made by the student author but was graded 4.0 by a human rater from the system later. The system correctly detected the problematic essay but generated a false message regarding the unscorable reason just as Chen et al (2009) mentioned that

MyAccess gave a considerable number of false alarms. In order to find out the actual score, the researcher deleted the repeated paragraph and re-submitted to the system, thus obtained the results: holistic score 3.3.

In order to make the machine rater's scores comparable with human rater's, the researcher re-assign the scores based on the new version of essays that the students keyed in the grading system. The results showed that the correlation ranged from 0.59479, 0.786739, 0.780593 to 0.729358, all at significant level ($p < .05$). The correlation values depended on which holistic score (0, ϕ , 3.3, or 4.0 respectively) (Table 2) was assigned for Essay No. 18 and entered for computation. If we considered the AGS was unable to assign the score and gave 0 point or ϕ for the Essay No. 18, then the correlations between the e-rater and the human rater are 0.59479 and 0.786739 respectively. However, if we allowed the human rater of the AGS coming to rescue, then the correlation was 0.780593. Frequently, the system didn't detect the repeated paragraph, thus the correction is 0.729358. This outcome echoes Burstein & Chodorow's contention (1999) that the correlation between e-rater scores and those of a single human reader are about .73; correlations between two human readers are .75 (p.69). Lonsdale & Strong-Krause's (2003) reported their LG parsing system agreed 67% of the time with human raters (p.65).

The consistency rate could be due to the difference between experienced and inexperienced human raters. Pare & Joordens' study (2008) is one of the cases. In Pare & Joordens' study (2008), students got online to assess each other's written abstracts and critical thinking with a 10 point scale and the "agreement level" computed with Pearson correlation coefficient between the expert markers' and the peer markers' average marks was found to be $r(131) = 0.27$ at the significant level ($p < 0.003$). Throughout the study, Pare & Joordens were very upset about the low "agreement level" and blamed the students for inexperience. They conducted the second experiment by asking experienced graduate student TAs to be the expert markers and warned the students who participated in the online peer review activity that their marking behavior might be monitored for inconsistencies and lack of variation. Eventually, their "mark the marker" intervening that requested students to mark on their peers' marks and comments by selecting a label of Not Useful, Useful, or Very Useful from a drop down menu worked to increase the Pearson correlation coefficient by $r(115) = 0.45$ at the significant level ($p < 0.001$).

Table 2. Correlation Values Obtained from Entering Differently Assigned Scores for Essay No. 18

Correlation (The holistic score of Essay No. 18 was entered with 0.) Marked correlations are significant at $p < .05000$ N=18 (Casewise deletion of missing data)				
Variable	Means	Std Dev.	HR	AGS
HR	3.283333	0.874643	1.000000	0.594790
AGS	3.083333	1.018216	0.594790	1.000000
Correlation (The Essay No. 18 was seen as missing case.) Marked correlations are significant at $p < .05000$ N=17 (Casewise deletion of missing data)				
Variable	Means	Std Dev.	HR	AGS
HR	3.305882	0.896152	1.000000	0.786739
AGS	3.264706	0.687333	0.786739	1.000000
Correlation (The holistic score of Essay No. 18 was entered with 3.3.) Marked correlations are significant at $p < .05000$ N=18 (Casewise deletion of missing data)				
Variable	Means	Std Dev.	HR	AGS
HR	3.283333	0.874643	1.000000	0.780593
AGS	3.266667	0.666863	0.780593	1.000000
Correlation (The holistic score of Essay No. 18 was entered with 4.0.) Marked correlations are significant at $p < .05000$ N=18 (Casewise deletion of missing data)				
Variable	Means	Std Dev.	HR	AGS
HR	3.283333	0.874643	1.000000	0.729358
AGS	3.305556	0.688965	0.729358	1.000000

In contrast, take the researcher as an example. The researcher has used *MyAccess* several times and is completely aware that it doesn't give 6 or even 5 points often. As a result, the researcher's rating creates a high agreement rate.

Moreover, a human rater especially in the instructor's position might be more concerned about whether the trained objectives being fulfilled or not, such as the use of the block method or the point-to-point method for comparison essay, the hook, the topic sentence, the supporting information, the conclusion, the connectors, etc. All these could influence the correlation value.

4.1.2 Agreement Rate

Burstein & Chodorow's (1999) study: *Automated Essay Scoring for Nonnative English Speakers* created a new term, agreement percentage, by calculating the percentage of the exact and adjacent scores assigned by e-rater and human rater. Burstein et al. (1998, p.206) defined the agreement rate by stating that "in accordance with human interrater 'agreement' standards, human and e-rater scores also 'agree' if there is an exact match or if the scores differ by no more than one point (adjacent agreement)." In the present study, only Essay No. 9 had the rating difference more than one point (2.3 vs. 3.6). By definition the agreement rate was 94.4 percent (17 cases out of the 18 within one-point difference). Again, this high agreement rate could be possibly due to the researcher's awareness of the "custom" of the grading system that *MyAccess* seldom gives scores 5 or 6 to sophomores. Nevertheless, if the adjacent agreement definition is taken to decimal because *MyAccess* assign scores with decimals in this case, Essay No. 5, 7, 8, 11 and 13 are the cases in the study. Then, the agreement rate is 27.8 percent (5 cases out of the 18 within zero-point-one difference) which is much lower than the 94.4 percent of agreement rate.

4.2 Research Question 2: How Competent Are Students in Reviewing Their Counterpart's Written Work?

Peer review approach has been practiced for quite a few decades in both L1 and L2 teaching. Many report the positive effect on improving student writing and students' welcoming attitude toward the class activity. On the other hand, educators looked into students' capability in detecting and correcting their counterparts' lexical, syntactic, and semantic errors as well as giving proper suggestions for organization, development, and creation.

In determining how competent the students are in reviewing their counterpart's written work, two sets of collected data were looked into: a. Students' midterm and final grades are used to measure their competence of detecting errors and correcting them; b. the content of Peer Editing Sheets was analyzed to assess their reviewing quality.

4.2.1 The Midterm and Final Exam Results

In this teaching practice, the researcher collected all the errors students made on their written work and turned them into the midterm and final exams by asking students to detect the errors and correct them (Appendix 1 & 2). More specifically, errors selected from writing practice (I), (II) and (III) were used for the midterm test, while errors chosen from writing practice (IV) and (V) were tested on the final exam. The total was 50 errors for the midterm and 100 errors for the final test assigned with a 100 point scoring scale. The number attached at the end of each sentence indicates how many errors the sentence has. The results shown on Table 3 accounted for students' capability/competence of correcting writing errors. It can range from 38 percent to 95 percent. The average scores of the midterm and the final exam are 70.2 and 79.3.

Actually, in the beginning the researcher planned to use students' midterm and final exam scores as evidences of students' competence in reviewing their peers' work and didn't mean to release and teach the error items collected from all the essays. However, a student pleaded the teacher to release and teach all the error items in class before the midterm and another requested before the final exam. Therefore, in considering students' right to know and prepare for the tests, the researcher released all the error items online and taught them. This episode shows that grammar training before adopting the peer review approach is really essential.

Table 3. The Midterm and Final Exam Results (N = 22)

	Max	Minimal	Mean	Mode
Midterm test	92	38	70.2	80 (5 cases)
Final exam	95	60	79.3	83 (5 cases)

4.2.2 Peer Editing Sheets

Each participant in the study was supposed to complete three Peer Editing Sheets. However, respectively, 14 valid Peer Editing sheets for the narrative essay (Table 4), 18 sheets for the cause-effect essay (Table 5) and 19 pieces for the argumentative essay (Table 6) were handed in. The student editors' comments are collected and displayed on Table 4, 5, and 6. The researcher also comments on them.

Table 4. Content of the Peer Editing Sheet for the Narrative Essay

Peer Editing sheets for the narrative essay: <i>Describe a movie/TV program</i> (N=14)	
1: Is the hook interesting? If not, how could it be made more interesting?	★12: Yes ★ 1: Not bad 1: Left blank: ▲Do you believe devils?
2: How many paragraphs are going to be in the essay?	▲13: Three ▲ 1: Five
3: What action or event does each topic sentence show?	▲All of the students made comments for the first three paragraphs.
4: Is there a good ending to the action of the story? If not, can you suggest a change to the ending?	9: Yes 2: No ▲Nowadays, due to technology, we can catch the news at the first time by many ways. ▲Writer should write more in the concluding paragraph. 2: Left blank: ▲You can tell readers how you feel in this movie. It seems that the story still goes on without ending. 1: Maybe: ▲This movie is different. It's very exciting and interesting. If you have time, how about trying to watch it.
5: What kind of the ending will the story have a moral prediction or revelation?	▲8: Identified one kind of the ending 6: Left blank
6. Do you think this essay will have enough information? Does the story leave out anything important? Write suggestion here.	8: Yes 1: Yes also: ▲In my opinion, we should cherish our earth, otherwise we don't have any place could [to] live. 3: No: ▲Writer should talk to readers why she thinks that the story has educational meaning and how interesting she thinks. ▲Your essay isn't related to your topic. Maybe you should change your topic or rewrite your essay. ▲Why does meteor come up? 1: Left blank ▲You can describe more movie details. Tell readers how terrible it is. 1: More will fine: ▲How about write more sentences make the whole article become complete.
7: The best part of the outline is: _____	▲[Everyone identified a paragraph or two in response to this question.]

8: Questions I still have about the outline: _____

3: Yes:

- ▲Don't any other things happen in this movie?
- ▲Everyone are [is] afraid [missing *of*] the end of the world. They believe the day will [missing *be*] coming.
- ▲Why does the writer's hook from movies change to TV programs?

2: No

★4: Left blank

1: indicated a specific paragraph without additional comments on

1: pointed out a certain paragraph:

- ▲The content in this paragraph is not so specific.

▲indicates the comment made by the student editor was deemed proper.

★indicates the comment made by the student editor was deemed improper due to not answering the question, not constructive, or not pointing out a way to make a change.

The square brackets [] indicate they are the researcher's notes. The student editors' comments sometimes include grammatical errors. The researcher sometimes tries not to correct them and let them stand real.

The researcher's comment on the Peer Editing Sheet for Narrative essay:

1. The students are all able to identify the number of paragraphs, the action or event written in certain paragraph, and the best part of their peer's essay (Question 2, 3, and 7).
2. Six out of fourteen couldn't identify the ending type because on the day of instruction eight students claimed for an official leave for an important school event (Question 5).
3. The students' comments were generally true, proper and constructive enough to their peers (Question 4, 6, and 8).
4. For Question 1, most of the students seemed not competent enough in judging whether the hook is good or bad. This is also possible that they didn't know how to help their peer construct a better hook in response to the continuous question in Question 1, thus they had to say the hook was good. The textbook exemplified good hooks, and the researcher emphasized them and thought they weren't hard to understand. This is probably one of the very few items on the peer editing sheets that the researcher didn't agree with most of the students' comments. Some examples of hooks done by the student writers are listed in the follows to show the difference between the students' comments and the teachers' expectation. Those are deemed good hooks by the student reviewers, but the researcher didn't agree with them totally:

There are more and more different kinds of dramas playing on TV nowadays.

There are many kinds of movies. Which type of that do you like?

Seeing movies is not only common but also popular entertainment in this world.

Nowadays, there are a lot of different kinds of movies that we can see, such as...

The following are examples of good hooks written by the other students in the class:

If you were the only one person who was alive in this world, how do you feel?

Can you imagine that one day when you wake up and suddenly you find that you can not move your hands, your fingers, your head, your neck, your body, and your feet?

In view of this Peer Editing Sheet, most descriptive comments made by the student editors were proper and constructive, even though a small portion of them was not answering the question, constructive or providing a way to make a change. On the contrary, attention should be given to those who easily check "yes," while they shouldn't. One last note is regarding the design of the Peer Editing Sheet which is advised to ask more open-ended questions. A well-designed Peer Editing Sheet is really thought-provoking and worth of spending 20 to 30 minutes on it.

Table 5. Peer Editing Sheet Content of the Cause-Effect Essay (N=18)

Peer Editing sheets for the cause-effect essay: <i>What are three common causes of motorcycle accidents?</i> (valid N=18; off topic N=1)	
1: What kind of essay will this be—a focus-on-causes essay or a focus-on-effects essay? Can you tell this from the thesis statement? If not, what changes can you suggest to make the purpose of the essay clearer?	▲18 Focus on causes. 18: Yes [2 of the 18 Yes: ▲Write more in the paragraph. ▲Make more paragraphs.]
2: Read the topic sentence for each body paragraph. Is it related to the thesis? If not, mark the topic sentences that need more work.	17: Yes [1 of the 17 yes: ▲Each body paragraph should be a sentence.] 1: Sure
3: Do the supporting details related to the topic sentences? If not, which paragraph(s) need to be developed further?	16: Yes 1: Almost 1: No: ▲You have to add a topic sentence before your reasons.
4: The best part of the outline is _____.	[Everyone identifies a best part of the outline.]
5: Questions I still have about the outline:	8: No [1 of the 8 No: It's perfect. Very clearly [clear].] 1: Left blank Good content. ▲A little disorder and maybe [the writer] can write more details. ▲Some grammar should improve [be improved]. ▲I feel the first paragraph and the third paragraph are similar. ▲How the accidents happened? ▲You need to give examples. Conclusion should stand alone to next paragraph. ★Beautiful handwriting. ★I think time is not enough so the writer cannot talk more details. ★I can't know what she will tell me next.

▲ indicates the comment made by the student editor was deemed proper.

★ indicates the comment made by the student editor was deemed improper due to not answering the question, not constructive, or not pointing out a way to make a change.

By analyzing the Peer Editing Sheets, Question 1, it may be questionable whether the students who wrote “Yes,” had a reason to do so beyond not wanting to explain further. On the other hand, when they say “No,” most of the time they have a good reason, especially those who provided comments. After assessing all of the three Peer Editing Sheets, a conclusion was made that the students’ comments were generally true, proper and constructive enough to their peers. However, they are weak in identifying good hooks as well as counterargument and help their peers to re-write them. Concerning the Peer Editing Sheets, they are provided by a commercial textbook. Instructors might want to design on their own because those sheets as mentioned above do not ask about grammatical items at all which are very concerned by student writers. Besides, critiques questioning about students’ competence in reviewing their peers’ writing always look for evidences from student reviewers’ grammatical error identification rate. Future Peer Editing Sheets are suggested to include a few questions of grammatical errors and of which should be able to finish in 30 minutes on the task.

Table 6. Peer Editing Sheet Content of the Argumentative Essay (N=19)

Peer Editing sheets for the cause-effect essay: (three options of prompt)	
a. Do you think students should be penalized for missing classes?	
b. Is a passing score on an English achievement test necessary for international students to enter a university?	
c. Is day care beneficial for children under the age of five?	
1. Is the hook interesting? In other words, does it catch the reader's attention?	<p>17: Yes ★[Student H's hook states: "At school, many classes must be taken." This can't be considered to be an interesting hook.]</p> <p>2: No [Two hooks being recognized as not interesting were: ▲You must have this kind of experience when you are a little boy or little girl. ▲Nowadays, a great number of parents both have jobs in order to make more money.]</p>
2. Is the writer's opinion clear in the thesis statement?	All of the nineteen student editors indicated their peer's opinion was clear in the thesis statement.
3. Do the topic sentences in the body paragraphs support the thesis?	<p>17: Yes 2: No ★[One of the being accused students did not hand in her written essay at the end of the semester. Thus we don't have a clue to her topic sentences. However, the other wrote in her first paragraph advocating home care. Her second, third, fourth, and fifth topic sentences are listed in the follows:</p> <p>First, home care is more convenient than day care. Second, home care can develop relationship with children than day care. Third, home care can teach children by your own. In contrast, other parents will say day care is much more convenient than home care because they can concentrate on their work, and the baby sitter can take care of their children well. In the end, no matter many advantages or disadvantages about the day care and home care, it depends on you.]</p> <p>[The researcher's note: These topic sentences stay closely and support the writer's thesis. If anything should be blamed about, it is the conclusion that does not take sides. Even though the writer had spent most of the volume saying home care is better than day care, she held back to the middle for retaining objectivity. This happens quite commonly in Chinese students' comparison essays. The researcher had eased them to take sides before writing by saying it wouldn't be judged or graded on which side they took.]</p>
4. In each paragraph, do the supporting details relate to the topic sentence?	<p>16: Yes 3: No ▲[Student H wrote in the first paragraph: "At school, many classes must be taken. If you take the class, and don't go to class, will your teacher be angry?...I think students shouldn't be penalized for missing class..." This creates a discrepancy.] ▲[Student C wrote in the last paragraph: "Can you say day care is not a good thing? Not actually. Some busy parents think day care is their only choice if they don't have enough money to hire a full-time babysitter. But parents' love and company is the most important thing for their child. If you can't afford love and enough company to your child, why do you want a child at first? So, please think twice before your step. Can you offer enough love and company to your child? Or just the money paid for the day care center?" The supporting details in this paragraph do not closely relate to the topic sentence.]</p>

▲[Student W wrote in the third paragraph: “Second, home care can help your babies learn things, like language or walking faster than day care... Also, when your baby is sick or feel uncomfortable, you’ll find out soon.” The supporting details do not relate to the topic sentence.]

5. Are the counterargument and refutation strong?

14: Yes

5: No

▲[Student L doesn’t even have a counterargument paragraph.]

▲[Student H did state the points of the counterpart, but did not refute.]

▲[Student W wrote in the counterargument paragraph: “On the contrast, other parents will say they care is much more convenient than homecare, because they can contract on their work, the babysitter can take care of their children well. And some parents will say that develop relationships is on weekend, not every day or all day. They think children should develop relationships with other children. Or some parents have different opinions about teaching children by themselves. They will think that the profession.” Student W did state the points of the counterpart, but did not refute them.]

▲ [Student LZ wrote in the counterargument paragraph: “Although there are opposed opinions, they still has an advantage that sent children to the daycare center. The advantage is that children can meet other children and learn earlier. Because children get together with other children at five years old, they can know how to communicate and get along with others. Student W did state the points of the counterpart, but did not refute them.”]

[One student didn’t hand in her essay.]

6. Does the writer restate the thesis in the conclusion?

17: Yes

2: No

▲[Student L didn’t restate the thesis in the conclusion]

▲[Student W wrote in the conclusion paragraph: “In the end, no matter their many advantages or do the disadvantages about the daycare and home care, but it is all depends with you. Parents should choice the best way to take care their children.”]

7. The best part of the outline is:

[Everyone recognized the best part of the thesis.]

8. Questions I still have about the outline: _____

5: Yes

▲Body paragraph.

▲Counterargument and refutation.

▲No counterargument.

▲It’s not very specific and many grammars are wrong.

▲How many scores do you think it is a proper score?

8: No

6: Left blank

▲indicates the comment made by the student editor was deemed proper.

★indicates the comment made by the student editor was deemed improper due to not answering the question, not constructive, or not pointing out a way to make a change.

The square brackets [] indicate they are the researcher’s notes. The student editors’ comments sometimes include grammatical errors. The researcher sometimes tries not to correct them and let them stand real.

4.3 Research Question 3: What Are the Error Identification Rates of Teacher Grading?

4.3.1 The Product-Oriented Approach: Teacher Grading

This writing class was primarily organized with a traditional teacher-center method or the product-oriented approach but infused with the peer review (process-oriented method) and the machine rating (also process-oriented method) methods. Take the essay No. 9 on Table 4 as an example because this is the only essay that the researcher's rating doesn't reach an agreement with the automated grading system or differ by more than one point (2.3 vs. 3.6). At the first glance, the researcher thought the assigned score 2.3 couldn't have been wrong because the student writer wrote very little about what it is in the Internet classroom in the third paragraph but stating what's not any more in the traditional classroom. Besides, the grammatical errors were not changed very much based on the instructor's suggestion, and the hook and the topic sentences were not attractive enough as the instructor emphasized several times in the class that they would be the focuses to look at. Also, the instructor look into if the trained objectives such as the hook, the topic sentence, the supporting information, the conclusion, the comparison method, the connectors, and the exercise of first, second, third...etc. are added to his text. In addition, the student author used quite a few Chinese expressions such as "Eight o'clock was the first class in a day" and "Present students are well-being" that a machine rater couldn't possibly detect. All these caused the low points from a human rater, and therefore "inconsistency or disagreement" existed between two reviewers.

Even though teachers traditionally bear the extreme authority for grading student work, Kuo (2008) challenged it by examining the quality of non-native English speaking teachers' error correction and found the teachers participating in the experiment identified and corrected 78 percent of errors on student composition, while the accuracy rate in correcting the de-contextualized short essays was as low as 48%. After an experimental treatment, the accuracy rates increased by 8 percent in both contextualized and de-contextualized content. The causes of teachers' unnecessary corrections included false or partial understanding of English grammar and usage, focusing on style, content and correcting from readers' perspective.

4.3.2 Error Types

In the present study, three English teachers from two private universities of science and technology and one from a public university of science and technology (one native speaker and two non-native speakers) exercised a corpus-study method by tagging error types to the Essay No. 9. The results showed (Table 7, 8, 9 and 10) the total error types identified by the three teachers were 18 and the error counts in the writing example were respectively 29, 11 and 26. The average errors being recognized were 22. Counting the total identified errors but excluding the overlaps, there are 35 errors in the essay according to the reviewers' assessment. The identification rates were respectively 82.9%, 31.4% and 74.3%. The average identification rate (62.9%) is much lower than that in Kuo's study (78%) in 2008 probably because the teachers in the study were not made aware that the experiment was ongoing. The error counts of the top identified frequencies were word choice, tense, L1 transfer, the third person singular, and the plural. The results mostly resembled Tan's report (2008) that the top four commonly made errors in her study of Taiwanese college EFL students were word choice, verb form, missing subject, and verb tense.

In view of the results on Table 7, the count of error types can't be said absolutely exact due to reasons such as a. insufficient definition, b. grammar training, and c. various correction ways. The example of insufficient definition being provided to the teacher is the tag [ing] which refers to the present continuous *-ing* form. Teacher 3 correctly identified the second error in the fourth paragraph but tagged it with [ing], while Teacher 1 sorted it into the [word choice]. An example of grammar training is the "skill cram school" in the third paragraph. A native speaker of English may not know it's a transfer from L1 but feel odd and tag it [word choice]. Sentences such as "Eight o'clock was the first class in a day" and "Present students are well-being" are also difficult for native speakers of English to categorize the sentences into [L1 transfer]. Various correction ways could influence the error types being sorted into, for instance, the sentence "Some went to school by walk." Teacher 1 may want to change the "by walk" to "on foot," therefore, tags it with [expression]; while Teacher 2 may intend to correct the verb form with the gerund *walking*, thus tags it with [word choice]. The third teacher considers this is an error of verb form, as a result, she tags it with [tense -ing]. However, all of them correctly identified the error but tagged it differently. One additional note is the spelling error does not appear in this study because the students in the study have ways and tools to correct them while they are typing. But spelling error is ranked high among the top error types in some researchers' studies. More importantly, this case surfaced a problem that an essay full of Chinese expression could be rated high in the automated grading system but low in human rating. This is also the reason that causes the grading system and the human rater inconsistent in rating.

Table 7. Error Types Identified by Three L2 Teachers

	Teacher 1	Teacher 2	Teacher 3
frequency	5 [word choice]	4 [word choice]	10 [word choice]
	4 [L1 transfer]	3 [tense]	6 [tense]
	3 [III person singular]	1 [plural]	2 [article]
	3 [tense]	1 [L1 transfer]	2 [plural]
	2 [irregular past]	1 [III person singular]	2 [regular past, -ed]
	2 [plural]	1 [omission]	1 [tense, ing]
	2 [regular past -ed]		1 [grammar]
	2 [space error]		1 [III person singular]
	1 [adverb for comparative adjective]		1 [noun]
	1 [article]		
	1 [expression]		
	1 [noun]		
	1 [passive voice]		
	1 [relative pronoun]		
	0 [possessive]		
	0 [ing]		
	0 [auxiliary]		
	0 [copula]		
Total identified	29	11	26

Table 8. Error Type Count for Essay No. 9 (Teacher 1)

Essay No. 9 tagged with error types	Error type count	Overall
As time passed [tense], the way of teaching become [III person singular] more and more developing.	1 [tense]	5 [word choice]
The way of teaching, the teaching place and the teaching time are totally different. What [space error] is the different [noun] between " Internet class" [space error] and "Traditional class" ? Which one is much better for students? Let me compare these two different type [plural] of teaching.	1 [III person singular]	4 [L1 transfer]
	2 [space error]	3 [III person singular]
	1 [noun]	3 [tense]
	1 [plural]	2 [irregular past]
		2 [plural]
		2 [regular past -ed]
		2 [space error]
		1 [adverb for comparative adjective]
In the past, we need [regular past -ed] to go to school for class everyday. Some people needed to get up early because they lived far away from school. Some went to school by walk [expression]. Eight o'clock was the first class in a day [L1 transfer]. Teachers used the blackboard and chalks to start teaching. It was a very rare chance to go to school for former [word choice] people, so everyone who could go to school to study was hard-working. Teachers taught some basic subjects like Chinese and Math. Students finished their final class [word choice] at four o'clock. After class, they can [irregular past] play some interesting games in the field.	1 [regular past -ed]	
	1 [expression]	
	1 [L1 transfer]	1 [article]
	2 [word choice]	1 [expression]
	1 [irregular past]	1 [noun]
		1 [passive voice]
		1 [relative pronoun]
		0 [possessive]
		0 [ing]
		0 [auxiliary]
		0 [copula]
Nowadays, most parents worry about their children losing in the base, so they sent [irregular past] their kids to cram school earlier and earlier [L1 transfer]. Present students [article] are well-being [L1 transfer]. They don't need to get up early to school and go to the class on foot. Their parents will [tense] drive their cars or ride them to school. The time which [relative pronoun] school starts its first class become [III person singular] more [adverb for comparative adjective] later than before. The thing	1 [irregular past]	
	3 [L1 transfer]	
	1 [article]	
	1 [tense]	
	1 [relative pronoun]	
	1 [adverb for comparative adjective]	
	2 [III person singular]	

teachers use to teach also change [III person singular]. Teachers use computers to start their classes. It has become a basic right for people to receive education, so more and more school [plural] were established. Teachers teach more international [word choice], such as English or other languages. Students ended their classes at about six o'clock, but they need [regular past -ed] to go to skill cram school [L1 transfer]. Students have no time to relaxing [word choice].	1 [plural] 2 [word choice] 1 [regular past -ed]
As time passed [tense], do you think whether education becomes better or not? I don't think so. Maybe the government should find a better way to reforming [word choice]. I hope our education system can change [passive voice].	1 [tense] 1 [word choice] 1 [passive voice]

Total identified

29

29

Table 9. Error Type Count for Essay No. 9 (Teacher 2)

Essay No. 9 tagged with error types	Error type count	Overall
As time passed [tense], the way of teaching become [III person singular] more and more developing. [word choice] The way of teaching, the teaching place and the teaching time are totally different. What is the different between "Internet class" and "Traditional class?" Which one is much better for students? Let me compare these two different type of teaching.	1 [tense] 1 [III person singular] 1 [word choice]	4 [word choice] 3 [tense] 1 [plural] 1 [L1 transfer] 1 [III person singular] 1 [omission]
In the past, we need [tense] to go to school for class everyday. Some people needed to get up early because they lived far away from school. Some went to school by walk [word choice]. Eight o'clock was the first class in a day. Teachers used the blackboard and chalks to start teaching. It was a very rare chance to go to school for former people, so everyone who could go to school to study was hard-working. Teachers taught some basic subjects like Chinese and Math. Students finished their final class at four o'clock. After class, they can play some interesting games in the field.	1 [tense] 1 [word choice]	
Nowadays, most parents worry about their children losing in the base, [word choice] so they sent their kids to cram school earlier and earlier. Present students are well-being. They don't need to get up early to school and go to the class on foot. Their parents will drive their cars or ride them to school. The time which school starts its first class become more later than before. [word choice] The thing teachers use to teach also change. Teachers use computers to start their classes. It has become a basic right for people to receive education, so more and more school [plural] were established. Teachers teach more international [omission], such as English or other languages. Students ended their classes at about six o'clock, but	2 [word choice] 1 [plural] 1 [omission] 1 [L1 transfer]	

they need to go to skill cram school [L1 transfer].
Students have no time to relaxing.

As time passed [tense], do you think whether 1 [tense]
education becomes better or not? I don't think so.
Maybe the government should find a better way to
reforming. I hope our education system can change.

Total identified	11	11
------------------	----	----

Table 10. Error Type Count for Essay No. 9 (Teacher 3)

Essay No. 9 tagged with error types	Error type count	Overall frequency
As time passed [tense], the way of teaching become [III person singular] more and more developing. The way of teaching, the teaching place and the teaching time are totally different. What is the different [noun] between [article]" Internet class" and [article] "Traditional class" ? Which one is much better for students? Let me compare these two different type [plural] of teaching.	1 [tense] 1 [III person singular] 1 [noun] 2 [article] 1 [plural]	9 [word choice] 6 [tense] 2 [article] 2 [plural] 2 [tense, -ed] 1 [tense, ing] 1 [grammar] 1 [III person singular]
In the past, we need [tense, -ed] to go to school for class everyday. Some people needed to get up early because they lived far away from school. Some went to school by walk [tense, - ing]. Eight o'clock was the first class in a day. Teachers used the blackboard and chalks (plural) to start teaching. It was a very rare chance to go to school for former people [word choice], so everyone who could go to school to study was hard-working. Teachers taught some basic subjects like Chinese and Math. Students finished their final class at four o'clock. After class, they can play [tense, -ed] some interesting games in the field.	2 [tense, -ed] 1 [tense, ing] 1 [word choice]	1 [noun] 1 [ing]
Nowadays, most parents worry about their children losing in the base [word choice], so they sent [tense] their kids to cram school earlier and earlier. Present [word choice] students are well-being [word choice]. They don't need to get up early to school and go to the class on foot. Their parents will drive their cars or ride them [word choice] to school. The time which school starts its first class become more later [grammar] than before. The thing [word choice] teachers use to teach also change [tense]. Teachers use computers to start their classes. It has become a basic right for people to receive education, so more and more school [plural] were [tense] established. Teachers teach more international [word choice], such as English or other languages. Students ended [tense] their classes at about six o'clock, but they need to go to skill [word choice] cram school. Students have no time to relaxing [tense].	7 [word choice] 5 [tense] 1 [grammar] 1 [plural]	
As time passed, do you think whether [word choice] education becomes better or not? I don't think so. Maybe the government should find a better way to reforming [ing]. I hope our education system can change.	1 [word choice] 1 [ing]	
Total identified	26	26

4.4 Questionnaire Results

A questionnaire was conducted for students to respond a week before the final exam. The results shown on Appendix 3 indicated that the majority of the subjects participating in the study agreed upon that four-semester writing classes for an English major student and six composition practices per writing class were just right. Interestingly, 9 students out of 16 (56.25%) said 6 composition practices in a writing class were just right (Question 5), whereas, when they were reconfirmed with regard to how many were enough, seven students (the mode in a statistic sense) indicated four would be enough (Question 6 on the questionnaire). Thus, this is fair enough to say that six practices are within the range of their toleration. In the results of Question 7, 8 and 9, respectively ten, ten, and nine out of sixteen participants (62.50%, 62.50% and 56.25%) agreed the Department should offer accesses to writing-aid software such as grammar checkers, vocabulary and sentence pattern practices, and automated grading system, *Criterion* or *MyAccess* for instance. Responding to the evaluation of reviewers' usefulness to the respondents' writing, nine considered the automatic grading system's assistance was limited and should be used occasionally (56.25%), while eight of the sixteen participants deemed peer review was innovative and should be used more often (50.00%).

When asked to rank the order of preference and effectiveness among the three reviewers-the automated grading system (*MyAccess* in this case), the peer and the teacher, a high portion of respondents (to be precise, 56.25% and 62.50% for Questionnaire Question 12 and 13 respectively) preferred teacher to peer followed by the automated grading system.

Concluding from an empirical study, Tsao (2006) found the product-oriented method for teaching English writing class was more effective than the process-oriented approach in improving students' grammar and diction, particularly effective in enhancing the grammar skills of less proficient writers. This study therefore attested Tsao's (2006) study mostly.

5. Conclusion and Suggestion

This study is somewhat Action-Research-oriented because this study not only asks questions to seek answers, but also discovers more problems and observes the differences. In answering the first question: What is the consistency rate or agreement rate between a human rater and a machine rater? The study examined the Pearson correlation coefficient and the agreement rate defined by the exact-plus-adjacent ratio, and suggested the correlation rate is 0.780593 and the agreement rate is 94.4% in this case.

Regarding the second question: How competent are students in reviewing their counterpart's writing work, the researcher went through the students' midterm and final exam which were a collection of errors from their essays and looked into the quality of the Peer Editing Sheets and found students generally do a great job on the written comments but have trouble distinguishing the good hooks from the bad ones. While organizing the peer review activity, instructors are advised to design their own Peer Editing Sheets by adding a few items of grammatical questions, if possible.

With the advancing technology, traditional peer review method has launched integrating and employing the web 2.0 concept to develop an online peer review tool for assessing written text. PeerScholar developed by the University of Toronto Scarborough, Canada (Pare & Joordens, 2008) was originally designed to address the need for writing and critical thinking assessment in the Introductory Psychology course which enrolled over 2000 students every year at the university. Classroom peer review activity has expanded its social interaction to the web via the aid of technology, and makes the writing class return to a much larger size possible just as Pare & Joordens claimed in 2008. However, the correlation rate is relatively low as the evidences provided by Pare & Joordens themselves. The approach seems not to be prevailing yet.

From the results of the questionnaire, the preference and effectiveness orders for student writing were found to be the teacher prior to the peer and the automated grading system. Even though teachers and researchers have tried to use the automated grading system and the peer review method for reducing teachers' loading, the methods have been concluded not being able to completely replace the human teachers in terms of preference and effectiveness.

Leki's (1991) pointed out that error analysis focused on grammar errors but didn't help improve learners' writing, while students considered surface errors were important to be corrected by teachers (Leki, 1991). Besides, error analysis retained too often "a static, product-oriented type of research, whereas L2 learning processes required a dynamic approach focusing on the actual course of the process. The static and fixed environment seems to have been claimed to be changed by the Genre writing approach which is described as a moving and dynamic environment (stabilized-for-now) and leads researchers to look closely at the social motives for writing influenced by different

social contexts (Coit, 2010, p.78).

Bernd Susser pointed out in 1994 that “until recently, a gap of at least 10 years could be found in the research describing recommended teaching methods to be used to teach writing in the L1 and the L2. As a consequence, writing instructors in these classrooms were not so eager to adopt the new changes that accompanied the process approach.” In view of the critique, genre writing may not be the only answer for L2 writing instructors, but it is worthy of trying from time to time for keeping up with the L1 teachers’ pace.

References

- Applebee, A. N. (1986). Problems in process approaches: Toward a reconceptualization of process instruction. In A. R. Petrosky & D. Bartholomae (Eds.), *The teaching of writing: Eighty-fifth yearbook of the National Society for the Study of Education, Part II*. Chicago: National Society for the Study of Education, 95-113.
- Bartles, N. (2003). Written peer response in L2 writing. *English Teaching Forum*, 41, 34-36.
- Berg, E. C. (1999a). The effects of trained peer response on ESL students’ revision types and writing quality. *Journal of Second Language Writing*, 8, 215-241. [https://doi.org/10.1016/S1060-3743\(99\)80115-5](https://doi.org/10.1016/S1060-3743(99)80115-5)
- Berg, E.C. (1999b). Preparing ESL students for peer response. *TESOL Journal*, 8, pp. 20–25.
- Burstein, J., & Chodorow, M. (1999). Automated Essay Scoring for Nonnative English Speakers. *Proceedings of a Symposium on Computer Mediated Language Assessment and Evaluation in Natural Language Processing*. <http://portal.acm.org/citation.cfm?id=1598847>.
- Burstein, Jill, Karen Kukich, Susanne Wolff, Chi Lu, Martin Chodorow, Lisa Braden-Harder & Mary Dee Harris (1998). Automated Scoring Using A Hybrid Feature Identification Technique. In the *Proceedings of the Annual Meeting of the Association of Computational Linguistics*, August, 1998. Montreal, Canada.
- Burstein, Jill, Karen Kukich, Susanne Wolff, Chi Lu & Martin Chodorow (1998). Enriching Automated Scoring Using Discourse Marking. In the *Proceedings of the Workshop on Discourse Relations & Discourse Marking, Annual Meeting of the Association of Computational Linguistics*, August, 1998. Montreal, Canada.
- Burstein, Jill C., Lisa Braden-Harder, Martin Chodorow, Shuyi Hua, Bruce Kaplan, Karen Kukich, Chi Lu, James Nolan, Don Rock & Susanne Wolff (1998). *Computer Analysis of Essay Content for Automated Score Prediction*. ETS RR 98-15.
- Chen, C. S. (2004). *A study on errors in English relative clauses made by senior high school students in Taiwan*. Unpublished thesis, National Kaohsiung Normal University.
- Chen, Hao-Jan. (2009a). Developing statistic-based and rule-based grammar checkers for Chinese ESL learners. *Proceedings of the 2009 LTTTC International Conference on English Language Teaching and Testing*. Taipei: Taiwan.
- Chen, Hao-Jan. (2009b). Evaluating two web-based grammar checkers-Microsoft ESL Assistant and NTNU Statistical Grammar Checker. *Computational Linguistics and Chinese Language Processing*, 14(2), 161-180.
- Chen, Howard Hao-Jan, Chiu, Stan Tsuo-Lin; & Liao, Pearl (2009). Analyzing the grammar feedback of two automated writing evaluation systems: My Access and Criterion. *English Teaching and Learning*, 33(2), 1-43.
- Chen, P. (2002). *A corpus-based study of the collocation errors in the writings of the EFL learners in Taiwan*. Unpublished thesis, National Taiwan Normal University.
- Chen, Yi-Hsuan. (2009). *The effects of peer and teacher feedback on college EFL student writing*. Unpublished thesis, National Pingtung Institute of Commerce.
- Coit, Caroline. (2010). *Developing dialogical academic writing through the use of student empowered peer review*. Unpublished dissertation, The University of Cologne, Germany.
- Dagneaux, E., Dennes, S., & Granger, S. (1998). Computer-aided error analysis. *System*, 26, 163-174. [https://doi.org/10.1016/S0346-251X\(98\)00001-3](https://doi.org/10.1016/S0346-251X(98)00001-3)
- Folse, K. S., Muchmore-Vokoun, A., & Solomon, E. V. (2010). *Great Writing 4: Great Essays* (3rd ed.). US: Heinle, Cengage Learning.
- Hsueh, S. C. (2003). *An Analysis of lexical collocational errors in the English compositions of senior high school EFL students*. Unpublished thesis, National Kaohsiung Normal University.

- Huang, H.S. (1994). *An Analytic study on tense errors committed by Chinese learners of English in southern Taiwan*. Unpublished thesis, National Kaohsiung Normal University.
- Huang, S. L. (2001). *Error analysis and teaching composition*. Unpublished thesis, National Tsing Hua University.
- Johnson, D., & Crombie, W. (2010). *A Genre-based Approach to academic writing*. Taipei: Taiwan TungHwa Bookstore. LTD.
- Kuo, Fang-Ying. (2008). *Non-native English Speaking Teachers' Written Error Correction: Quality and Causes of False Corrections*. Unpublished theses, Tamkang University.
- Lai, Y. H. (2010). Which do students prefer to evaluate their essays: Peers or computer program? *British Journal of Educational Technology*, 41(3), 432-454. <https://doi.org/10.1111/j.1467-8535.2009.00959.x>
- Leki, I. (1991). The Preferences of ESL Students for Error Correction in College-Level Writing Classes. *Foreign Language Annals*, 24, 203–218. <https://doi.org/10.1111/j.1944-9720.1991.tb00464.x>
- Lightbrown, P. M., & Spada, N. (2006). *How languages are learned*. China: Oxford University Press.
- Lonsdale, D., & Strong-Krause, D. (2003). Automated rating of ESL essays. *Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing 2*, 61-67. <https://doi.org/10.3115/1118894.1118903>
- Mendonca, C., & Johnson, K. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, 28, 745-769. <https://doi.org/10.2307/3587558>
- Miltsakaki, E., & Kukich, K. (2000). The role of Centering Theory's Rough-Shift in the teaching and evaluation of writing skills. *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*. <https://doi.org/10.3115/1075218.1075270>
- Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, 15, 118-141. <https://doi.org/10.1016/j.jslw.2006.01.003>
- Nelson, G. L., & Carson, J. G. (1998). ESL students' perceptions of effectiveness of peer response groups. *Journal of Second Language Writing*, 7, 113-131. [https://doi.org/10.1016/S1060-3743\(98\)90010-8](https://doi.org/10.1016/S1060-3743(98)90010-8)
- Nelson, G. L., & Murphy, J. M. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts? *Tesol Quarterly*, 27(1), 135-141. <https://doi.org/10.2307/3586965>
- Pare, D. E., & Joordens, S. (2008). Peering into large lectures: examining peer and expert mark agreement using peerScholar, an online peer assessment tool. *Journal of Computer Assisted Learning*, 24(6), 526-540. <https://doi.org/10.1111/j.1365-2729.2008.00290.x>
- Schneider, D., & McCoy, K. F. (1998). Recognizing syntactic errors in the writing of second language learners. *Proceedings of the 17th international conference on Computational linguistics*, 2, 1198-1204. <https://doi.org/10.3115/980432.980765>
- Shih, Y. J. (2004). *A study of Taiwanese learners' use of the English article system*. Unpublished thesis, National Taiwan Normal University.
- Susser, B. (1994) Process approaches in ESL/EFL writing instruction. *Journal of Second Language Writing*, 3(1), 31-47. [https://doi.org/10.1016/1060-3743\(94\)90004-3](https://doi.org/10.1016/1060-3743(94)90004-3)
- Tan, Hui-mien (2008). A study of EFL learners' writing errors and instructional strategies. *Journal of Kun Shan University*, 4, 113-122. <https://doi.org/10.1080/01626620.2012.717033>
- Tsai, M.H. (2010). The Consistency Between Human Raters and an Automated Essay Scoring System in Grading High School Students' English Writing. *Action in Teacher Research*, 34(4), 328-335.
- Tsao, Chia-Hsiu. (2006). On the Relative Effectiveness of the Process-and Product-Oriented Approaches to Teaching English Composition. *Bulletin of National Taiwan Normal University*, 51(1), 23-39.
- Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education*, 2, 319-330.
- Wang, C. S. (2001). *A study of the English collocational competence of English majors in Taiwan*. Unpublished thesis, Fu Jen Catholic University.
- Yang, Shih-hsien. (2011). Exploring the effectiveness of using peer evaluation and teacher feedback in college students' writing. *The Asia-Pacific Education Researcher*, 20(1), 144-150.

Appendix 1

Midterm for Writing (IV) Class, Dept. of Applied Foreign languages, NPUST**Error Correction:**

The number at the end of each sentence indicates the number of errors in the sentence needs to be corrected.

1. There were many bottle of milk and juice on the short desk. (1)
2. Unfortunately, it was found by a classmate and the classmate told to the teacher the truth. (1)
3. I always stick to my own opinion, and never listen to other people's advise. (1)
4. Once I decided to do something, no one can stop me. (1)
5. Sometimes, maybe is because of my age, I often said what I want to say to my parents. (2)
6. Have you ever punished by your parents or teachers? (1)
7. Chinese parents not only want their children learn a lesson, also hope them don't do the same things again. (2)
8. I don't know what should I say. (1)
9. But I don't think that is a good way, because many children don't learn a lesson from punishment. (1)
10. It's means everyone has their own childhood. (2)
11. When it comes to my funny troubles are too many to be cited. (1)
12. After, she cleaned it, she punished me again. (2)
13. I always looked forward to go to adventure with my team. (1)
14. I disagree about my parents' thoughts. (1)
15. Have you ever been punished by telling a lie? (1)
16. He still mad at us. (1)
17. Childhood is filled with fun, naughty, stubborn and laughter. (1)
18. Comparing with traditional class, this way is more attractive to students. (1)
19. You will miss those important information. (1)
20. Students have two choices to attend class. One is the Internet class. Another is the traditional class. (1)
21. Although study at home through Internet is very convenient, it is much more boring than goes to school. (2)
22. There are more and more technological tools been invented. (1)
23. That's very trouble. (1)
24. Students in Internet class are more freer than the students in traditional class. (1)
25. Going to school is a old-fashion thing. (2)
26. The traditional class is not convenience as the Internet class. (2)
27. Do you know that what are the reasons for this situation? (1)
28. We keep discuss until the bed time. (1)
29. Why Fierce Wife is more popular than the other dramas? (2)
30. The story happened to a happy couple who name was Jack and Marry. (3)
31. We don't know that everything we do are cruel. (1)
32. I can feel how much does Holly miss Mike. (1)
33. It told me some things that I've never known. (1)
34. The movie told us how did the earth formed. (2)
35. I think the movie, "World Invasion," it's a exciting alien movie. (2)
36. I found this movie by accidentally. (1)
37. There are still something that is not so good in the movie, such as a lack of a good plot and too many fight scenes make it looks like a war movie. (2)

Appendix 2

Final Exam for Writing (IV) Class, Dept. of Applied Foreign languages, NPUST**Error Correction:**

The number at the end of each sentence indicates the number of errors in the sentence needs to be corrected.

1. It's better to have a test score than don't have one. (1)
2. Opponents who think that there should not be a real qualification of English proficiency to enter a college. (1)
3. Why it causes many different opinions? (1)
4. They must never think how hard do their parents make money to pay for their cost of living. (1)
5. Under the various condition, everyone has himself/herself solutions and viewpoint. (2)
6. Some parents are not have enough time. (1)
7. In my opinion, absent from class is not a bad thing. (1)
8. We'd rather staying at home to sleep than going to a useless class. (2)
9. There are more and more women go outside to earn money. (1)
10. They think that parent should accompany with their children because it is much more safety and also benefit for children. (2)
11. Did you ever missed class? (1)
12. In addition, how students learn things without going to classes? (1)
13. Also, if we indulge them, what a awful students we will have? (2)
14. We must should do efforts to change it. (1)
15. The problem not only affect students and teachers. (1)
16. For some busy parents, day care center is their best choose. (1)
17. They should responsible for their absence of classes. (2)
18. In addition, they can consider to have not too many children. (1)
19. Money is the second essential necessary which behind children. (2)
20. The children under five are too small to sent the day care center. (1)
21. It is not easy to find a person or a day care center that we reliable. (1)
22. Although there are opposed opinions, there still has a advantage that sent children to the day care center. (2)
23. Therefore, "lonely" may be your only friend. (1)
24. Children only can stay in other places where teachers can teach them. (1)
25. One is that the development of mind. (1)
26. Making money is difficult and tired. (1)
27. The teacher shouldn't punish on he/she. (1)
28. They are all can be accepted. (1)
29. In the other hand, some said that the teacher had rights to penalize students for missing classes. (1)
30. When it comes to college, almost everyone would think of stay up late, play in pub and the most common one, skip classes. (3)
31. Skip classes have became really common. (2)
32. When you listen to this kind of issues, what your reaction and answer is? (1)
33. We may stay at home company with them. (1)
34. They want to touch or eat anything they can see. (1)
35. They don't know what is dangerous items. (2)
36. The quality of day care is good or not is a question. (1)
37. Because there are too many children in one day-care class, teachers and baby-sitters are too less. (1)
38. They don't have enough to care every children in the same time. (3)

39. Children may get hurt very easily without being attention. (1)
40. There are a lot of reasons for don't agree on it. (1)
41. They should responsible for their absence of classes. (2)
42. For some busy parents, day care center is their best choose. (1)
43. Should students skip the classes is always noticed in this society. (1)
44. This school policies are not fair to students. (1)
45. They had better approve students of skip classes sometimes. (1)
46. Students will buy textbooks which has many information. (2)
47. On the other hand, speak of internet class, it can save both teacher and students own time. (2)
48. The goal of these two ways are same. (1)
49. The students of Internet class can control their time and also can decide how many times they want to take the class. (1)
50. The students take Internet class aren't have a real teacher because they can't to be with the teacher. (3)
51. Although, the technology makes our way of learning more variety than before, there are still some good and bad point of the new choice. (2)
52. Recently, there are many traffic accidents happened in our life. (1)
53. As we know, there are more and more teenagers they drive a car or ride a motorcycle with no license. (1)
54. Some of them just for fun, because they can show to their peers and make his/her status higher than others. (1)
55. This behavior is very dangerous because we might didn't remember what we did. (1)
56. The amount of motorcycles has increased in two recently decades. (2)
57. Run through the red light is the most common bad habits. (2)
58. This is usually lead to terrible result. (1)
59. If you are riding your motorcycle in a dark place without any light, may happen some unexpectedly accidents. (2)
60. There are many reasons cause accidents. (1)
61. There are still a lot of drivers don't obey the rule. (1)
62. Driver doesn't have good hobby when they are driving cars. (1)
63. The reasons of accidents are variety. (1)
64. Slow down, paying attention and follow the rules are the best way to avoid the accident. (2)
65. If they pay not enough attention on other cars, they may hit them. (1)
66. Have we ever think about why the car accidents happened so often? (1)
67. Maybe they think wait for the light to turn green is wasting their time. (1)
68. People usually have colorful night lifes. (1)
69. Most of car accidents can be avoid if we pay more attention on ourselves. (2)
70. Let us to work hardly to create a safer society. (2)
71. The number of car accidents definitely go down. (1)
72. I think the causes of motorcycle accidents can be prevented. (1)
73. First, everyone has their own driving style. (1)
74. Due to convenient, many junior or senior high school students use motorcycle but not bicycle as their transportation. (1)
75. It will be surprised to everyone because the rate of accidents is becoming higher and higher year by year. (1)

Appendix 3: Questionnaire results of the English Writing (IV) Class

Question	Items to respond	count	percentage
1. Major in _____ Department		16	100%
2. Gender	a. Male	3	18.75%
	b. Female	13	81.25%
3. I think four-semester writing classes are:	a. too many	2	12.50%
	b. too few	5	31.25%
	c. just right	7	43.75%
	d. no comment	2	12.50%
4. I think there should be _____ semester(s) of writing classes.	a. one	0	0%
	b. two	2	12.50%
	c. three	2	12.50%
	d. four	9	56.25%
	e. six	2	12.50%
	f. eight	1	6.25%
	g. more than 8 (please specify the number)	0	
5. I think 6 composition practices in a writing class are:	a. too many	2	12.50%
	b. too few	2	12.50%
	c. just right	9	56.25%
	d. no comment	3	18.75%
6. I think _____ composition practice(s) would be enough for a semester.	a. one	0	0
	b. two	0	0
	c. three	3	18.75%
	d. four	7	43.75%
	e. five	1	6.25%
	f. six	3	18.75%
	g. more than 6 (please specify the number)	2	12.50%
7. I think the Department should offer some writing software for student use because it can point out grammatical errors and provide example essays and writing style.	a. strongly agree	3	18.75%
	b. agree	10	62.50%
	c. disagree	2	12.50%
	d. strongly disagree	1	6.25%
8. I think the Department should offer some writing software for student use because it can help student practice of vocabulary and sentence patterns.	a. strongly agree	3	18.75%
	b. agree	10	62.50%
	c. disagree	3	18.75%
	d. strongly disagree	0	0
9. The Department should offer access to the automated grading systems such as MyAccess or Criterion.	a. strongly agree	3	18.75%
	b. agree	9	56.25%
	c. disagree	2	12.50%
	d. strongly disagree	2	12.50%
10. I feel the automated grading system to my writing is _____.	a. Substantially helpful. Every composition practice should include it.	2	12.50%
	b. Innovative. It can be used more often.	3	18.75%
	c. Limited. It should be used occasionally.	9	56.25%

	d. Not helpful. It's not proper for using.	2	12.50%
11. I feel the peer review to my writing is _____.	a. Substantially helpful. Every composition practice should include it.	2	12.50%
	b. Innovative. It can be used more often.	8	50.00%
	c. Limited. It should be used occasionally.	6	37.50%
	d. Not helpful. It's not proper for using.	0	0
12. The order of the reviewing methods I prefer is _____.	a. teacher > peer review > automated grading system (AGS)	9	56.25%
	b. teacher > AGS > peer review	4	25.00%
	c. peer review > AGS > teacher	0	0
	d. peer review > teacher > AGS	2	12.50%
	e. AGS > teacher > peer review	1	6.25%
	f. AGS > peer review > teacher	0	0
13. The order of the reviewing methods I deem most useful to my writing is _____.	a. teacher > peer review > automated grading system (AGS)	10	62.50%
	b. teacher > AGS > peer review	4	25.00%
	c. peer review > AGS > teacher	0	0
	d. peer review > teacher > AGS	1	6.25%
	e. AGS > teacher > peer review	1	6.25%
	f. AGS > peer review > teacher	0	0