Educational Technology International
2011, Vol. 12, No. 2, 71-94

Integration of Computerized Feedback to Improve Interactive Use of Written Feedback in English Writing Class

Jaeho CHOI*
Handong Global University
Korea

How can an automated essay scoring (AES) program, which provides feedback for essays, be a formative tool for improving ESL writing? In spite of the increasing demands for English writing proficiency, English writing instruction has not been effective for teaching and learning because of a lack of timely and accurate feedback. In this context, AES as a possible solution has been gaining the attention of educators and scholars in ESL/EFL writing education because it can provide consistent and prompt feedback for student writers. This experimental study examined the impact of different types of feedback for a college ESL writing program using the *Criterion* AES system. The results reveal the positive impact of AES in a college-level ESL course and differences between the teacher's feedback and the AES feedback. The findings suggest that AES can be effectively integrated into ESL writing instruction as a formative assessment tool.

Keywords : Automated essay scoring, Written feedback, ESL/EFL, English writing, Technology integration

* Senior Researcher, Handong Global University
jc2st@virginia.edu

Introduction

English as a Second Language (ESL) or English as a Foreign Language (EFL) instruction has become a major educational concern for both English speaking and non-English speaking countries due to a significant increase of English in global communication (Crystal, 2003). For example, the instantaneous transmission of information through the Internet, which is disproportionately in English, is a hallmark of modern society. Therefore, educators need to address the growing need for students to be proficient in English (Crystal, 2003). Timely and adequate feedback has been generally recognized as crucial for learning (Bransford, Brown, Cocking, & National Academy of Sciences, 2000), and written feedback is perceived as the important feedback in English writing instruction. Error correction is most commonly adopted in writing classes. This helps students recognize their errors, correct grammatical forms and expressions and revise their drafts (Cowie, 1995; Fathman & Whalley, 1990; Ferris, 1995).

However, giving timely and adequate written feedback is challenging to teachers in ESL and EFL writing classes in that effective feedback requires intensive, high-quality instructional efforts of teachers. This is challenging for teachers who cannot give high quality feedback for every piece of student writing (Schulz, 2001). It is also demanding for ESL students who must edit and revise their drafts in response to teacher feedback (Silva, 1990). For instance, research on ESL teachers' feedback has revealed that it is often both inconsistent and inaccurate (Cowie, 1995; Truscott, 1996; Zamel, 1985). Although this problem is partly caused by insufficient time, teachers' lack of linguistic and pedagogical knowledge is possibly a significant factor hindering quality feedback. This situation is exacerbated when students are learning English in a non-English speaking country, lacking qualified English teachers and instructional resources. Truscott (1996) raised doubts about the effectiveness of written feedback in ESL writing and insisted that error corrections, which are time- and energy-consuming, did little to improve writing. Therefore, in spite of the increasing world-wide demand for English writing proficiency, English writing

instruction has not been effective; getting the proper feedback in a reasonable time still remains a problem.

In terms of timely and accurately feedback, automated essay scoring (AES) is gaining attention as an alternative instructional tool providing consistent and timely feedback. AES is an adaptation of computer technology that diagnoses essays and gives holistic scores and feedback over several writing domains, such as grammar, usage, style, mechanics, and organization. This is done by combining artificial intelligence and language analysis tools (Shermis & Burstein, 2003). Since 1966, when Ellis Page introduced the Project Essay Grader as an automated essay grading system, on-going development of AES along with technological innovation in artificial intelligence has increased the reliability and validity of AES as an essay rating system (Dikli, 2006). Most research on AES has concentrated on its use in summative assessment as evidenced by its reliability and validity (Attali, 2004; Attali & Burstein, 2006; Burstein, 2003; Elliot, 2003; Keith, 2003; Page, 2003; Warschauer & Ware, 2006). Although research on AES feedback as a formative assessment tool has explored its impact to improve writing proficiency, but the results are not confirmative to conclude its benefits due to lack of rigorous research design and objectivity; most studies were conducted by the program developers (Warschauer & Ware, 2006). Thus, in spite of the popularity of AES in English writing classrooms, the effects and instructional application of AES have not been explored enough. This study examined how AES feedback could be integrated in English writing class as a formative assessment tool contributing to improving writing quality.

Literature review

Writing as an Interactive Process

According to the *process writing* approach, writing is an interactive process

between learners, peers, and teachers that includes planning, drafting, receiving feedback, and revising. The theoretical perspective of this approach is well represented by Michael Long's second language acquisition (SLA) theory, *Interactive Hypothesis*. His interaction hypothesis posits that interactions between learners, peers, and teachers can be designed or modified to make language input more meaningful to learners (Brown, 2000). The hypothesis is supported by Krashen's theory of meaningful input, which suggests that learning occurs when learners perceive inputs that are meaningful to them. For instance, native speakers or teachers use comprehension checking, clarifying, or paraphrasing to make conversation meaningful to non-native speakers. Although Long's hypothesis is concerned, first, with oral communication, it can be easily applied to written communication (Blake, 2000). Feedback on students' writing, for instance, involves corrective interaction between students, peers, and teachers. Because such interaction occurs to negotiate meaning within a social context, the input is meaningful to learners (Brown, 2000). By extending Krashen's focus on learners to interactive relationships between learners, peers, and teachers, Long's interaction theory connects Vygotsky's (1978) theory of the Zone of Proximal Development (ZPD) with second language learning. ZPD is the distance between the learner's current cognitive level and the higher level which can be reached with the help of others. Therefore, with supportive and corrective interactions with teachers and/or native speakers, language learners develop their second language proficiency beyond their initial competency levels. In other words, students in ESL/EFL writing classes can interact with teachers and peers through written feedback and improve their writing competency throughout the writing process. These theoretical perspectives suggest that effective feedback works as a formative assessment tool, which enhances interactive learning.

Written Feedback

In writing class, written feedback is perceived as the critical interaction in

classrooms. By extending the theories discussed in the previous section, the point is made that students improve their writing by interacting with teachers and peers in writing. While SLA theories support the positive effects of written feedback, the results of other research on written feedback contrast them. Further empirical studies need to be made to ascertain whether such feedback effectively improves English writing.

The most common type of written feedback is corrective feedback, which focuses on grammatical accuracy. Truscott (1996) reviewed studies made by Kepner (1991), Semke (1984), and Sheppard (1992), and thereby triggered debates concerning the efficacy of corrective feedback for ESL writing. He suggested that corrective feedback is not only ineffective to improve writing but actually impedes a student's improvement because it cannot accommodate the complex process of second language acquisition. Schulz (2001) argued that effective corrective feedback for students' writing cannot be provided when a teacher in a typical ESL or EFL program has over 100 students. Such teachers are reluctant to give writing assignments due to the burden of work they entail (Kepner, 1991). Another problem occurs when EFL teachers give inaccurate and inconsistent feedback (Cowie, 1995). According to Truscott (1996), such corrective feedback does not improve ESL student writing.

However, other researchers have attributed the results of those studies to the lack of rigorous experimental research. They have concluded that previous research has not addressed the relative effectiveness of various types of feedback and the wide range of teachers' competency in providing it. Ferris (1999; 2004) questioned Truscott's analysis, commenting the lack of robust comparisons between a treatment group which received corrective feedback and a control group that did not. On the contrary, Ferris presented results of several studies demonstrating how error correction improves students' writing (Ferris & Roberts, 2001; Lee, 1997). Others conducted studies on detailed features of feedback that revealed a complex dynamic between the type of error and the type of feedback. For instance, Ferris,

Chaney, Komura, Roberts, and McKee (2000, cited by Bitchener, Young, & Cameron, 2005) investigated the impact of direct feedback, which overtly indicates and explains errors, and indirect feedback, which simply alludes to errors. They found that specific types of feedback were more or less effective in improving grammatical accuracy of a student's writing. For instance, in the short term, the direct feedback group improved grammatical accuracy of writing more than the indirect group. Whereas, in the long term, the indirect group performed better in reducing grammatical errors than the direct group. This research not only proved the positive impact of feedback on writing, but also revealed that some types of feedback were more effective than others (Ferris, 1999; Ferris & Roberts, 2001).

Although results of previous studies are not congruent to conclude the positive effects of written feedback, they agree that timely and appropriate written feedback is effective to improve writing. After intensive review of literature, Ferris(2006) suggests that written feedback needs to be presented timely and frequently during the interim of writing process, containing whole aspects of writing including grammar, usage, organization, and idea development. Thus, the effects of written feedback are dependent on when and how the feedback is given to students.

Automated Essay Scoring (AES)

With increasing need on effective feedback practice in writing class, the attention to assistive feedback tools, such as AES, is growing, which could provide timely and adequate written feedback. AES is computer technology that analyzes and evaluates an essay, utilizing artificial intelligence and linguistic analysis approach. Automated essay scoring (AES) systems typically offer holistic scores and corrective feedback on specific linguistic features of essays (Shermis & Burstein, 2003). Since Page (2003) developed the Project Essay Grader, the continuing development of computing technology and linguistic knowledge has improved the validity and reliability of AES scoring. In addition, corrective feedback has been

included for instructive use of AES. However, research on AES has primarily concentrated on the validation of holistic scores (Warschauer & Ware, 2006). The reliability and validity of AES have been investigated by comparing rating grades of AES and human essay raters (Page, 2003; Warschauer & Ware, 2006). Accumulated studies on the relationships between AES scores and human raters suggest that AES is highly accurate and reliable (Attali, 2004; Attali & Burstein, 2006; Burstein, 2003; Elliot, 2003; Keith, 2003), and this has increased interest in its instructional applications impact (Shermis, Burstein, & Bliss, 2004; Shermis, Garvan, & Diao, 2008; Warschauer & Ware, 2006).

ESL scholars and educators are expecting AES to address such problems in current ESL feedback practices. As a formative assessment tool, AES can provide prompt, accurate, and consistent feedback on essays and, thus, compensate for insufficient feedback (Shermis et al., 2008). While the effectiveness of AES feedback as a formative assessment has not been fully examined (Warschauer & Ware, 2006), a few research results have shown positive correlations between AES feedback and students' improvement (Elliot & Mikulas, 2004; Shermis et al., 2004; Shermis et al., 2008). For instance, Shermis and his colleagues (2008) examined whether getting holistic scores improve writing performance including fluency, accuracy, and holistic quality. The results indicating positive relationship between writing improvement and use of AES suggested that getting formative evaluation, holistic scores, encouraged students to revise more frequently and seriously. However, the effects of other AES feedback, such as feedback on grammar, usage, mechanics, and style, are not clearly examined in that how AES effectively assist students to improve writing. Dikli's(2007) study presented that AES feedback is different to human instructors' feedback in terms of length, frequency, and students' responses. Her study suggests that the feedback of AES and a human instructor would differently contribute to improvement of writing.

Research Questions

This study investigates the impact of written feedback on the quality of writing, differing types of integration of an AES program in ESL writing class. The experimental comparison between different types of feedback is expected to provide empirical evidence for the effective use of AES feedback as well as its integration model with ESL writing instruction. Thus, this study examined three different types of feedback and measured the resulting effects on the quality of essays in answer to the following research questions.

1) Can written feedback effectively improve the quality of ESL writing?

2) Are there any differences between the effectiveness of each of the three feedback types on the quality of ESL writing?

3) What are the differences between teacher and AES feedback?

Materials and methods

Participants and setting

The participants of this study were three instructors and twenty eight students from intensive summer ESL courses offered by the English Language Center of a university, located in a large metropolitan area in the eastern United States. The participants were international students who enrolled in one of the three ESL writing courses: Beginner, Intermediate, and Advanced. The courses were eight weeks long and focused on fostering English writing proficiency appropriate to each level. For instance, the Beginner course aimed to build English writing proficiency at the sentence level, emphasizing grammar and various sentence structures, and the Intermediate course focused on paragraph level proficiency. The Advanced course focused on developing essay level writing.

Instrument

Criterion, an AES program, has been used as an instructional and evaluation tool for this study. *Criterion* is an online writing service developed by Educational Testing Services (ETS), which provides evaluation and diagnostic feedback on students' writing by providing holistic scores(see Figure 1) and targeted feedback in five writing domains(see Figure 2): *Grammar, Usage, Mechanics, Style,* and *Organization and Development*. It also provides additional tools for writing, such as a planning tool, further instructions for revising, and interactive communication between the students and the teacher. However, all other *Criterion* tools, except diagnostic feedback, were blocked in order to eliminate any possible impact they might have on the quality of students' writing and to focus on the impact of different types of feedback. Once the instructor creates assignments and registers students, students can access *Criterion* and submit their papers (see Figure 3). If students submit an essay, *Criterion* provides prompt feedback(see Figure 1 & 2). Students can then review the feedback and revise their drafts. In addition, teachers can review students' *Criterion* feedback given by and add line-by-line or general comments through *Criterion*.

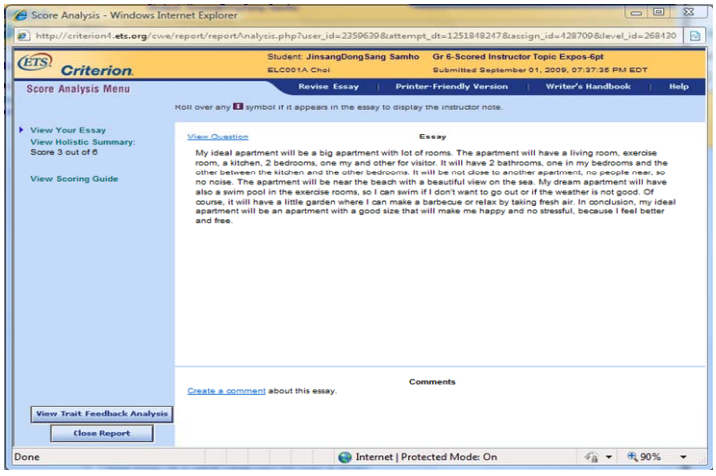

Figure 1. *Criterion* screen capture – AES feedback of holistic score

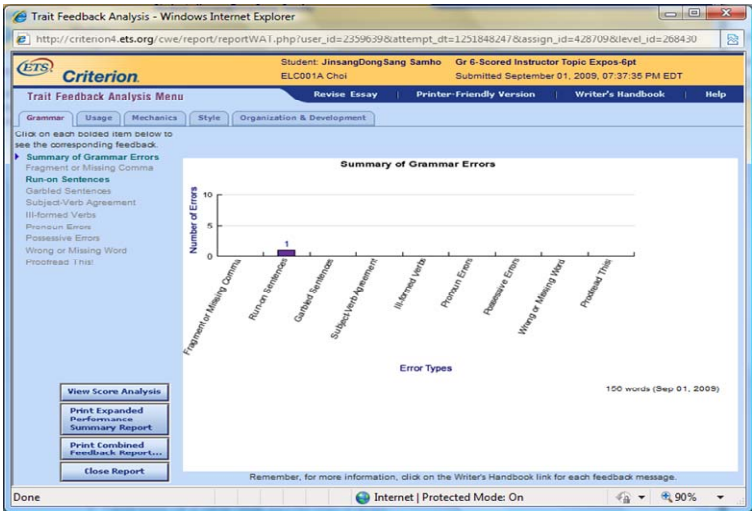

Figure 2. *Criterion* screen capture – AES feedback across five linguistic domains

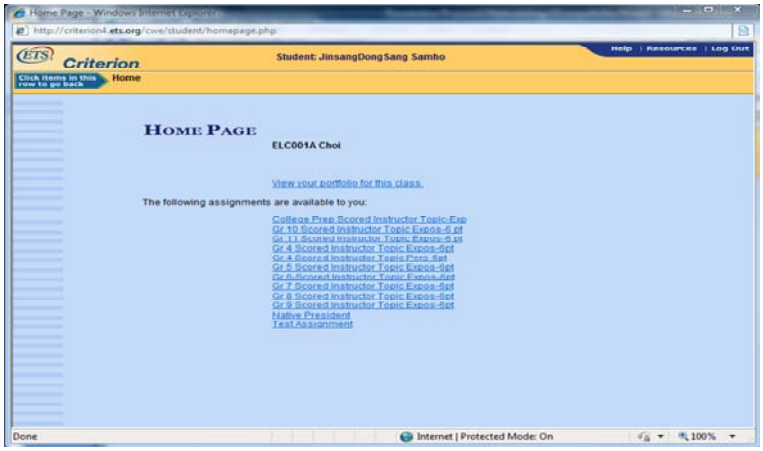

Figure 3. *Criterion* screen capture – student assignment page

Design and Data Collection

This study employed a randomized experimental design with three types of feedback. Before implementing research, instructors underwent a *Criterion* user training and had meetings with the researcher and an administrative staff member to discuss research design and procedures. To examine the effects of different types

of feedback, students of each course were randomly assigned to one of the three types of feedback groups: AES group, Teacher group, or Combined group. The AES group received written feedback from AES only, the Teacher group, from a teacher only, and the Combined group, from both teacher and AES (see Table 1). Students in each course received identical writing lessons and assignments; it was only the type of feedback they received which was different. AES written feedback is indirect feedback indicating the types of errors rather than providing correct forms. AES feedback includes a holistic score, general comments, corrective comments and suggestions across five domains of writing: grammar, usage, mechanics, style, and organization. It was given promptly after submission of an essay. On the contrary, teacher feedback was given one or two days later after submission. Teacher written feedback also covers linguistic features and idea development as well. The types of teacher feedback were mixture of direct and indirect feedback.

Table 1. The number of students and sets of first-revised drafts

	The number of students				The number of writing prompt			
Group	Advanced	Inter-mediate	Beginner	Total	Advanced	Inter-mediate	Beginner	Total
AES	2	3	5	10	3	4	14	21
Teacher	2	3	4	9	4	6	12	22
Combined	1	3	5	9	2	7	10	19
Total	5	9	14	28	9	17	36	62

Students were given one writing task each week during the three-week long intervention period. With each writing prompt, students wrote first drafts, received feedback on these drafts only once, and revised them, using *Criterion*. Each feedback group received one of three types of feedback between their first and revised drafts. Participants accessed the *Criterion* online service during their computer-lab sessions and after class. Students' writing performance data were

collected from instructors and *Criterion*. Records of 62 sets of writing (first and revised drafts) and written feedback from AES and teachers were collected across courses and feedback groups (see Table 1). To measure the improved quality of writing, five kinds of assessment data were collected as follows.

- *Improvement of AES Holistic Scores:* Holistic scores of the first drafts and the revised drafts, graded holistically from 1 to 6 by *Criterion,* were collected. Improvement was measured by subtracting the first draft score from the revised draft score. The improvement of holistic scores measured the improvement of the overall quality of writing between first drafts and their revised versions.
- *Number of Reduced Errors:* The number of errors in, and comments on, the first and the revised drafts were collected from *Criterion* across four traits including *Grammar, Usage, Mechanics*, and *Style*. The data regarding *Organization and Development* were excluded for analysis because they were difficult to quantify and usually inappropriate to sentence- and paragraph-level writing. To assess the reduction in errors, the number of errors per 100 words in the first and the revised drafts were compared. The purpose of measuring the number of reduced errors was to evaluate any improvement of writing accuracy.
- *Total Amount of Feedback:* The number of feedback comments for each trait was considered in determining the total amount of feedback provided for each draft. The total amounts of AES and teacher feedback were compared in order to locate any differences between AES and teacher feedback.
- *Student Response Rate:* Student response rates to both AES and teacher feedback were measured by checking whether each error was corrected or not in the revised drafts. Student response rate indicated the students' attitude towards feedback given by AES and teachers.
- *Proportion of Feedback across Traits:* The distribution of AES and teacher feedback for each trait and their relative proportions were measured and compared to find any differences between AES and teacher feedback.

Results

Collected data were examined to answer the research questions of this study. For the statistical analysis, this study used SPSS 17.0. The results of analysis were presented for each research question as follows.

Table 2. One-sample *t* test of dependent variables

Dependent Variables	Type	*N*	Mean	*SD*	*t*	*df*	*Sig.*
Improvement of holistic score	AES	21	.19	.40	2.17	20	.04*
	Teacher	22	.27	.46	2.81	21	.01*
	Combined	19	.53	.61	3.75	18	.00*
	All groups	62	.32	.50	5.03	61	.00*
Number of reduced grammar errors every 100 words	AES	21	.19	.94	.91	20	.37
	Teacher	22	.50	.74	3.15	21	.00*
	Combined	19	1.00	2.37	1.84	18	.08
	All groups	62	.55	1.50	2.87	61	.01*
Number of reduced usage errors every 100 words	AES	21	.69	1.15	2.74	20	.01*
	Teacher	22	.09	1.30	.32	21	.75
	Combined	19	.90	1.26	3.13	18	.01*
	All groups	62	.54	1.26	3.37	61	.00*
Number of reduced mechanics errors every 100 words	AES	21	.08	.47	.80	20	.43
	Teacher	22	-.42	.69	-2.83	21	.01*
	Combined	19	.18	2.15	.36	18	.72
	All groups	62	-.07	1.29	-.40	61	.69
Number of reduced style errors every 100 words	AES	21	.01	.30	.22	20	.83
	Teacher	22	.04	1.61	.10	21	.92
	Combined	19	.41	1.62	1.10	18	.28
	All groups	62	.14	1.32	.86	61	.39

*$p<.05$

Research Question 1: Can written feedback effectively improve the quality of ESL writing?

One-sample *t* tests were conducted to examine the improvement of writing quality in each feedback group regardless of types of written feedback. The results revealed that all feedback groups showed improved holistic writing quality in their revised drafts in terms of AES holistic scores. While the number of style errors was not significantly reduced in any group, the numbers of grammar, usage, and mechanics errors were significantly reduced in some of the three feedback groups (see Table 2).

Research Question 2: Are there any differences between the effectiveness of each of the three feedback types on the quality of ESL writing?

To examine this research question, the impact of different types of feedback was examined in terms of the five dependent variables: improved *Criterion* holistic score and the number of reduced errors in grammar, usage, mechanics, and style. For the analysis of data, one-factor fixed effect ANOVA and follow-up post hoc test, Bonferroni method of multiple comparisons, were adopted. The results of one-factor fixed effect ANOVA test revealed that the types of feedback significantly influenced the improvement of holistic scores (see Table 3 & 4); $F(2, 59) = 3.35$, $p = .04 < .05$. Bonferroni pair-wise comparisons among the three feedback groups indicated that the Combined group ($M = .55$, $SD = .60$) performed significantly better than the AES group ($M = .17$, $SD = .37$), $p = .04 < .05$ (see Figure 4). The difference between AES and Teacher ($M = .30$, $SD = .45$) groups was not statistically significant, $p = 1.00 > .05$ (see Table 5). The types of feedback did not have significant effects on reduction of errors in any writing domains: grammar, usage, mechanics, and style (see Table 4). Although there were no statistically significant differences across the types of feedback, the Combined feedback group showed the best performance of error reduction in all writing domains (see Figure 4).

Table 3. Descriptive statistics of writing performance data

Dependent variables	Type	*Mean*	*SD*	*N*
Improved holistic score	AES	.17	.37	21
	Teacher	.30	.45	22
	Combined	.55	.60	19
Number of reduced grammar errors	AES	.19	.94	21
	Teacher	.49	.74	22
	Combined	1.00	2.37	19
Number of reduced usage errors	AES	.69	1.15	21
	Teacher	.09	1.30	22
	Combined	.90	1.26	19
Number of reduced mechanics errors	AES	.08	.47	21
	Teacher	-.42	.69	22
	Combined	.18	2.15	19
Number of reduced style errors	AES	.01	.30	21
	Teacher	.04	1.61	22
	Combined	.41	1.62	19

Table 4. One-factor fixed effect ANOVA tests of writing performance

Dependent variables	*Source*	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Improved holistic score	Feedback	1.53	2	.76	3.35	.04*
	Error	13.44	59	.23		
	Total	14.97	61			
Number of reduced grammar errors	Feedback	6.67	2	3.34	1.51	.23
	Error	130.59	59	2.21		
	Total	137.26	61			
Number of reduced usage errors	Feedback	7.44	2	3.72	2.44	.10
	Error	90.07	59	1.53		
	Total	97.51	61			
Number of reduced mechanics errors	Feedback	4.28	2	2.14	1.30	.28
	Error	97.28	59	1.65		
	Total	101.56	61			
Number of reduced style errors	Feedback	1.96	2	.98	.59	.58
	Error	103.76	59	1.76		
	Total	105.72	61			

*$p<.05$

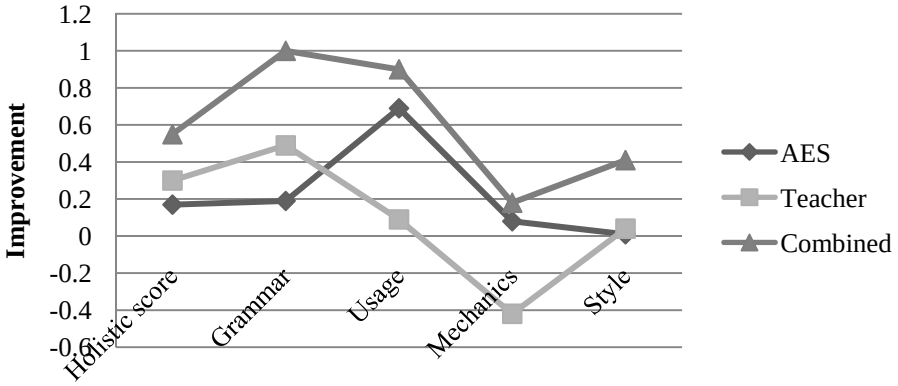

Figure 4. Writing performance across three different feedback groups

Table 5. Bonferroni multiple comparisons for the improvement of holistic scores

(I) Types of feedback	(J) Types of feedback	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
AES	Teacher	-.13	.15	1.00	-.49	.23
	Combined	-.39	.15	.04*	-.76	-.01
Teacher	AES	.13	.15	1.00	-.23	.49
	Combined	-.26	.15	.27	-.63	.11
Combined	AES	.39	.15	.04*	.01	.76
	Teacher	.26	.15	.27	-.11	.63

*p<.05

Research Question 3: What are the differences between teacher and AES feedback?

The analysis results revealed several differences between the teacher and AES feedback. In the analysis, teacher and AES feedback were compared in terms of the number of comments, the student response rate, and the proportion of feedback in each trait; differences between them then were identified. A Linear Mixed model analysis indicated that the type of feedback had significant effects on the total

amount of feedback; F (1, 33.94) = 23.37, p < .001. The test of Estimated Marginal Mean indicated that the amount of AES feedback (M = 6.99) is much less than the amount of teacher feedback (M = 13.68) per each writing draft (see Table 6). In addition, the type of feedback did not significantly influence the student response rate to feedback; F (1, 23.21) = 1.33, p = .26 > .05. Although the differences were not statistically significant, the test of Estimated Marginal Mean indicated that the mean rate in the AES group (M = .62) is slightly lower than the mean in the Teacher feedback (M = .74) (see Table 6). The distribution of AES and teacher feedback for each writing category indicated the differences between these two types of feedback as shown in Figure 5. For grammar and style, the teacher and AES showed similar proportions of feedback. In both types of feedback, grammar feedback was almost 40% of the total feedback. There were, however, obvious differences in the usage and mechanics traits. The teacher's feedback for mechanics was about 21% of total amount of feedback, more than three times of the amount of mechanics feedback (6%) in AES. The proportion of usage feedback in teacher feedback was 20%, which was less than half of the proportion of usage feedback in AES (46.2%). Last, the content feedback was given by the teacher alone. The examination of total amount, student response, and proportion of feedback indicates that teacher feedback and AES feedback are different. The differences suggest the potential of a complementary combination of both types of feedback. For instance, teacher's feedback on content could supplement AES feedback which cannot provide comments on meaning.

Table 6. Differences between AES and teacher feedback

Variables	Type III Tests of Fixed Effects					Estimated Marginal Means			
	Source	Numerator df	Denominator df	*F*	*Sig.*	Type	Mean	Std. Error	df
Total number of feedback	feedback	1	33.94	23.37	.00*	AES	6.99	.69	35.3
						teacher	13.68	1.20	33.5
Response rate	feedback	1	23.21	1.33	.26	AES	.62	.06	23.9
						teacher	.74	.09	22.9

*p<.05

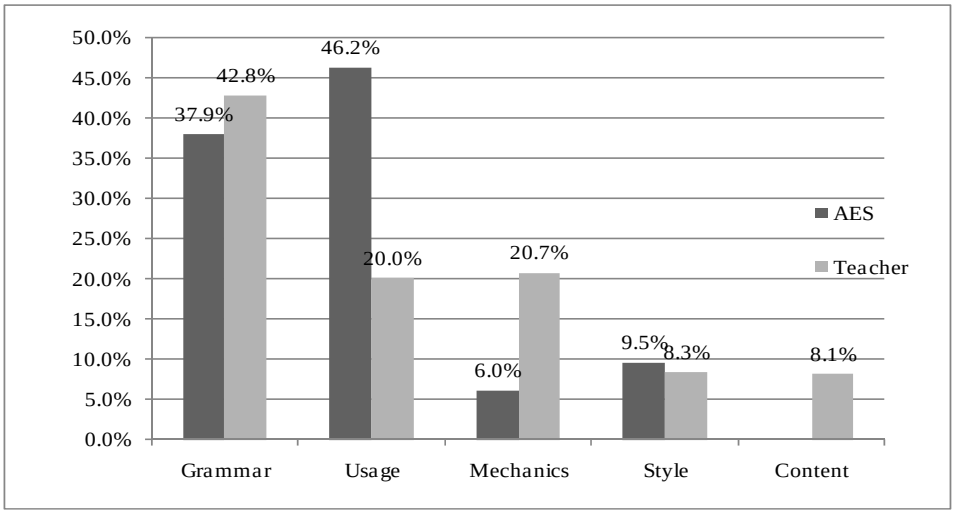

Figure 5. Distribution of feedback across five categories of feedback

Conclusions

In this study, revision practices based on given feedback were observed to improve the quality of writing regardless of feedback types. Specifically, the Combined feedback group usually showed the greatest improvement in measurements of holistic writing quality, while the AES feedback group achieved results equivalent to those of the Teacher feedback group. The insignificant effects of the types of feedback in other writing domains seems to be caused by dispersion of holistic effects across domains, which supported by the relative excellence of the Combined feedback group in those domains. In addition, the differences and similarities between AES feedback and teacher's feedback were observed. These findings suggest several implications for educators.

First, regardless of types of written feedback, the writing process with feedback can be useful in improving the quality of writing. Although the intervention period was short and feedback was given only once for each writing task, the improvement of the writing was statistically significant. This indicates that the writing process,

including receiving feedback and revising drafts, is effective for ESL writing education, and written feedback facilitates and encourages the revision process

As a result of this study, it was found that AES feedback can help ESL students improve the quality of their writing through the revision process in spite of limited use of AES. Although the teacher feedback group achieved slightly better performance, the AES group showed almost equivalent improvement. The results seem to be caused by the effect of feedback to encourage students to revise, which was observed by Shermis and his colleagues (2008), in spite of differences between AES and teacher feedback. In addition, it revealed that written feedback focused on grammar accuracy, presented by AES, could contribute to improvement of writing. These results indicate the potential of AES as an instructional tool for ESL writing, which provides formative assessment in the interim of writing process.

In addition, this study suggested that AES and teacher feedback can be combined effectively. The outstanding performance of the Combined group, in improving the quality of writing, could be attributed to the complementary features of both types of feedback. This result is congruent with Ferris' (2007) conclusion of effective feedback: feedback is effective when it is presented with whole aspects of writing. Considering the time and efforts for such feedback, AES could be effectively combined with teachers' feedback. In other words, a human instructor can focus on the aspects not covered in AES feedback, such as ideas, structure and development, and content while AES provides consistent, timely, and detailed feedback on grammar accuracy. With the combined feedback, students in the Combined group were assumed to have had more feedback (input), work with a higher sense of community (learning community), and have a better understanding of feedback (meaningful input) through corrective communications with a teacher.

Thus, the results of this study propose how effectively AES can be integrated with teacher feedback. Understanding the differences and similarities between AES and teacher feedback is necessary for integration of AES in writing class.

Limitations and suggestions for future studies

The limitations of this study, evident in research design and application, need to be addressed in future research. The most significant limitation of this study was the low sample size (student $N = 28$, set of writing- first/revised drafts $N = 62$) and the specific population from a summer ESL program at a university. As a result, the findings are hard to be generalized for comprehensive ESL program. Weak intervention was another critical limitation. The intervention was three weeks long with three writing tasks; feedback was given only once for each writing task. In terms of developing proficiency, a three-week period is insufficient time to develop effective writing skills. Although the results of this study reveal statistically significant improvement between the first drafts and the revised drafts, they do not indicate any improvement of writing proficiency but evidence the effectiveness of feedback and revision. Another limitation of this study is the fidelity of research implementation. For instance, some teachers could not follow the research procedures related to teacher feedback; thus, their data were discarded due to possible contamination by other factors.

These limitations need to be addressed in designing future research. First, a longitudinal study with large population is recommended to examine the development of writing proficiency and to generalize the findings. At least, a semester-long intervention period is required because the minimum unit of evaluation of instruction and curriculum is usually a semester. In addition, future studies need to be conducted in different programs with a diverse ESL population to determine whether ESL writing needs to be generalized across grades, students' English proficiency levels, and course curricula. Last, although this study limited the number of feedback and other functions of AES to accommodate participants' situations and to focus on the effects of feedback, future study should examine the effects of AES with full functions and investigate each function in terms of correlation.

References

Attali, Y. (2004). *Exploring the feedback and revision features of Criterion.* Paper presented at the National Council on Measurement in Education (NCME) Held between April, 12

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater v. 2. *Journal of Technology, Learning, and Assessment, 4*(3), 1-31.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*(3), 191-205.

Blake, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology, 4*(1), 120-136.

Bransford, J. D., Brown, A. L., Cocking, R. R., & National Academy of Sciences - National Research Council, Washington,DC.Commission on Behavioral and Social Sciences and Education. (2000). *How People Learn: Brain, Mind, Experience, and School. Expanded Edition.* National Academies Press.

Brown, H. D. (2000). *Principles of language teaching and learning.* NY: Longman.

Burstein, J. (2003). The e-rater scoring engine: Automated essay scoring with natural language processing. In M. D. Shermis, & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 113-121). Mahwah, NJ: Lawrence Erlbaum Associates.

Cowie, N. (1995). Students of Process Writing Need Appropriate and Timely Feedback on Their Work, and In Addition, Training in Dealing with That Feedback. *Saitama University Review, 31*(1), 181-194.

Crystal, D. (2003). *English as a global language.* Cambridge: Cambridge University Press.

Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment, 5*(1), 2006-2012.

Dikli, S. (2007). Automated essay scoring in an English as a second language setting.

Unpublished doctoral dissertation, The Florida State University.

Elliot, S. (2003). IntelliMetric: From here to validity. In M. D. Shermis, & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 71-86). Mahwah, NJ: Lawrence Erlbaum Associates.

Elliot, S., & Mikulas, C. (2004). *The impact of MY Access!™ use on student writing performance: A technology overview and four studies.* Paper presented at Annual Meeting of the American Educational Research Association, San Diego, CA.

Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178-190). New York: Cambridge University Press.

Ferris, D. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29*(1), 33-53.

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*(1), 1-11.

Ferris, D. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here?(and what do we do in the meantime…?). *Journal of Second Language Writing, 13*(1), 49-62.

Ferris, D. R., Chaney, S. J., Komura, K., Roberts, B. J., & McKee, S. (2000). *Perspectives, problems, and practices in treating written error.* In Colloquium presented at International TESOL Convention, Vancouver, B.C., March 14–18, 2000.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes How explicit does it need to be? *Journal of Second Language Writing, 10*(3), 161-184.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short-and between-pre-and-post-test effects of written error correction. In K. Hyland, & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104)

Keith, T. Z. (2003). Validity of automated essay scoring systems. In M. D. Shermis, & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 147–167). Mahwah, NJ: Lawrence Erlbaum Associates.

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal, 75*(3), 305-313.

Lee, I. (1997). ESL Learners' performance in error correction in writing: Some implications for teaching. *System, 25*(4), 465-477.

Page, E. B. (2003). Project essay grade: PEG. In M. D. Shermis, & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 43-54) Mahwah, NJ: Lawrence Erlbaum Associates.

Schulz, R. A. (2001). Cultural Differences in Student and Teacher Perceptions concerning the Role of Grammar Instruction and Corrective Feedback: USA-Colombia. *The Modern Language Journal, 85*(2), 244-258.

Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals, 17*(3), 195-202.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23*(1), 103-110.

Shermis, M. D., & Burstein, J. (2003). *Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shermis, M. D., Burstein, J. C., & Bliss, L. (2004). The impact of automated essay scoring on high stakes writing assessments. *Annual Meeting of the National Council on Measurement in Education,*

Shermis, M. D., Garvan, C. W., & Diao, Y. (2008). *The Impact of Automated Essay Scoring on Writing Outcomes*. Paper presented at the Annual Meetings of the National Council on Measurement in Education, New York, NY, March 25-27, 2008.

Silva, T. (1990). Second language composition instruction: Developments, issues, and directions in ESL. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 11-23). NY: Cambridge University Press.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*, 327-369.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research, 10*(2), 1-24.

Zamel, V. (1985). Responding to Student Writing. *Tesol Quarterly, 19*(1), 79-101.

Jaeho CHOI

Ph.D., Senior Researcher, Handong Global University. Interests: Written Feedback, Peer Feedback, Automated Essay Scoring (AES), Online Learning, Competency-based Learning, Instructional Design, English Writing

E-mail: jc2st@virginia.edu

Received: August 30, 2011 / Peer review completed: October 16, 2011 / Accepted: October 20, 2011