English Language & Literature Teaching, Vol. 21, No. 1 Spring 2015

Korean EFL College Writers' Perceptions of *Criterion*[*]

Eun-Hee Lee
(Kyung Hee Cyber University)

Lee, Eun-Hee (2015). Korean EFL college writers' perceptions of *Criterion*. *English Language & Literature Teaching, 21*(1), 199-216 .

Previous research has reported positive results on the effectiveness of automated feedback programs, and some of the studies stated that Korean EFL writers preferred automated feedback to human feedback. This study explored Korean EFL writers' perceptions of the automated feedback program *Criterion*. Seventy-two Korean university students whose proficiency levels were intermediate and low participated in the current study and answered a questionnaire at the end of the semester. The analysis of the questionnaire indicated that the participants preferred receiving teacher's feedback during the writing conference to receiving *Criterion* feedback. The main reason was that they could negotiate meaning during the writing conference if they had questions. However, participants sometimes did not know how to revise their essays with the given feedback from the automated tool because they could not ask for clarification from the automated program. The results of bivariate regression analysis indicated that only one factor, organization and development, contributed to the participants' satisfaction levels of *Criterion* feedback. The results of this study can provide language instructors and researchers who plan to use an automated tool with information about language learners' perceptions and the strengths and weakness of *Criterion*.

[automated writing evaluation/writing feedback]

I. INTRODUCTION

As many studies (Egbert & Hanson-Smith, 1999; Ihm, 2005; Nam & Kim, 2009; Warschauer & Meskill, 2000; Warschauer, Turbee, & Roberts, 1996) have already proved the effectiveness of computer-assisted language learning (CALL), the current interest is on investigating the characteristics that make CALL effective rather than

[*] Parts of this study were presented at the Second International Conference of English Scholars Beyond Borders in Bangkok, Thailand (February, 2015).

evaluating its effectiveness. Results of previous research have shown that CALL can motivate learners and help them to be self-regulated language learners (Fang, 2010; Simmering, Posey, & Piccoli, 2009). On this basis, automated writing evaluation programs have been considered vital for meeting the language instructor's challenge of providing prompt feedback on a large volume of writing. In writing instruction, providing learners with appropriate and timely feedback has been considered to be a critical element in determining learning efficacy (Cha, 2007; Elis, 1997). However, providing feedback on written assignments is time-consuming, so instructors have difficulty in providing the quantity and frequency of feedback that learners want (Shim, 2013). Due to these difficulties, automated writing evaluation has been developed over the decades, and some programs have already been commercialized.

Previous researchers (Kim, 2014; Shim, 2013) have claimed that using an automated writing tool in writing classrooms was effective in revising EFL writers' essays. However, they compared the first and second drafts in order to track changes, and the changes were considered as a symbol of proficiency improvement if there were changes in the second draft. Truscott (1996) claimed that mere changes of the forms could not be considered evidence of learners' proficiency improvement, and providing feedback on grammar could be even harmful. Korean learners tend to think that revision of writing means correcting grammatical mistakes and checking for typos. They often have difficulty in revising paragraph development and organization areas. Therefore, the current study attempted to explore Korean EFL university writers' perceptions about the effectiveness of the automated program at their university.

Hypothesis 1: Low- and intermediate-level participants will show different levels of satisfaction on *Criterion* feedback and the writing conference.

Hypothesis 2: The participants will prefer *Criterion* feedback to human feedback.

Hypothesis 3: Grammar, organization & development, and mechanical factors will be positively related to the participants' satisfaction levels with *Criterion* feedback.

II. LITERATURE REVIEW

Truscott (1996) claimed that grammar correction in L2 writing was harmful, but Ferris (1999) argued that error correction was necessary. Ellis, Sheen, Murakami, and Takashima (2008) agreed that error correction was helpful in EFL writing classes. Previous research (Ferris, 1997, 2003; Hyland & Hyland, 2006; Kang, 2008; Kim, 2007; Kim, 2014; Long, 1996; Shin & Seong, 2013; Swain, 1995; Schmidt, 1990) proved the

importance of feedback in order to improve language learners' writing abilities. They claimed that appropriately provided feedback on writing assignments could facilitate and encourage writers' improvement.

On the basis of the effectiveness of feedback, research interests have moved to investigating the effectiveness depending on feedback type. Kang (2008) investigated the effects of three different types of feedback on EFL writers' proficiency improvement: teachers' feedback, peer feedback, and self-revision. The results indicated that the participants' proficiency levels played a central role in deciding their preferred feedback type. Most participants preferred the instructor's feedback to peer feedback and self-revision. However, advanced-level participants received more benefits from peer feedback and self-revision, but less proficient participants made the most of the instructor's feedback.

Kim and Ha (2014) investigated university writers' reactions to their instructor's feedback and peer feedback. Their participants positively adopted the instructor's feedback and peer feedback. However, they adopted peer feedback more because their peers provided detailed feedback with some related examples while the instructor gave them some coded feedback and correction symbols without any examples. Kim and Ha interpreted these results as participants having a preference for comprehensible feedback; they adopted peer feedback more positively because they did not understand the teacher's correction symbols. Sim and Heo (2014) also claimed that interaction with instructors increased the participants' level of understanding about feedback, and writing conferences with teachers enhanced Korean university writers' motivation, engagement, and interest.

Even though previous research has noted the importance of prompt and appropriate feedback in writing, providing prompt feedback has been difficult because it takes a lot of time and effort, especially when teachers have to deal with a large volume of essays. Automated writing evaluation (AWE) systems can provide immediate feedback on writing, and they have been significantly developed over the last two decades due to their convenience and promptness.

Warschauer and Ware (2006) explained and compared the automated evaluation programs of three companies that had been widely used: *Criterion*, *My Access*, and *Intelligent Essay Assessor*. They claimed that non-native writers could receive more frequent and individualized feedback from these automated programs, but the instant evaluation function needs to be carefully used in specific teaching and learning contexts.

Vantage Learning (2007) claims that nonnative writers need to practice writing frequently and need to receive prompt feedback in order to increase writing proficiency. They believe that their product, *My Access*, can help a large number of writers to receive appropriate feedback in a timely manner. Even without the program developers' claim,

many researchers (Attali, 2004; Burstein, Chodorow, & Leacock, 2004; Fang, 2010; Kim, 2010, 2011; Kim, 2014; Mon & Pae, 2011; Otoshi, 2005; Shim, 2013; Warschauer & Grimes, 2008) have reported positive effects of automated writing evaluation programs. Fang (2010) investigated Taiwanese writers' perceptions of *My Access* and found that the participants thought that the tool was helpful in revising their essays and in developing writing skills, especially in correcting forms rather than content. However, they doubted its ability as an essay-grading tool. Lee (2008) also investigated the benefits of *My Access* in EFL classrooms along with the reliability of the program. She found that the program facilitated the participants' writing and motivated them to revise their essays. She concluded that the use of the automated tool could be beneficial in Korean EFL writing classrooms.

Among three programs, *Criterion* was developed in the mid-1990s and has been administered by Educational Testing Service (ETS). It has become popular as it provides writers with prompt feedback for a large volume of essays. *Criterion* provides feedback on five areas: grammar, usage, mechanics, style, and organization/development. Shim (2013) investigated the most frequently appearing error types and the influence of *Criterion* on Korean female writers' proficiency improvement. Her intermediate-level participants were satisfied with using *Criterion* as they could use the program to find their errors at any time.

Attali (2004) investigated whether or not *Criterion* was helpful for nonnative writers when revising their essays and found positive outcomes. The participants' subsequent revisions after using *Criterion* showed improvement on revising essays. Kim's (2014) participants also showed overall improvement of writing abilities after using *Criterion*. Kim compared Korean female university writers' first and second drafts in order to explore the effectiveness of *Criterion*. After her participants wrote their first drafts, they revised their first draft with the feedback from *Criterion* and wrote the second draft. The researcher tracked changes between the first and econd drafts and concluded that her participants showed overall improvement after using *Criterion* regardless of their proficiency levels—advanced or low level. The researcher claimed that the tool helped her participants in the self-editing process. Follow-up interviews with the participants showed that they were satisfied with the convenient and prompt feedback of *Criterion* and its function of error detection. However, they stated that its explanations about their errors were not sufficient. Despite its shortcomings, *Criterion* enabled writers to self-edit their own essays and helped them to be more autonomous learners. Participants also had a positive attitude toward the instructor's oral and written feedback, and one of her participants responded that she preferred the instructor's oral feedback, as it was easier to understand.

Moon and Pae (2011) and Otoshi (2005) explored what *Criterion* could and could not

provide. Moon and Pae (2011) found that *Criterion* did not provide appropriate feedback on "basic sentence structures, subject-verb relationship, awkward meaning construction, and formulaic expressions" (Moon & Pae, 2011, p. 130). Moon and Pae also investigated the short-term effects of the *Criterion* feedback and found that about 44% of the feedback was not reflected in students' revisions. Moreover, 18% of the *Criterion* feedback was incorrect, and this incorrectness led to failures in the participants' revisions; they failed to rectify 38% of their errors. The *Criterion* feedback did not precisely pinpoint errors, so participants wanted to receive instructors' feedback or to have a writing conference.

Otoshi (2005) also found that *Criterion* could not detect all the errors and provide appropriate feedback on five categories: verbs, word choice, nouns, articles, and sentence structure. The researcher concluded that the tool should not be used by itself in evaluating or providing feedback due to its shortcomings. Kim (2010) also claimed that *Criterion* could not provide appropriate and direct feedback at the sentence level. Her participants were more responsive to feedback about grammar, vocabulary, content, and organization than feedback regarding sentence structure. They were satisfied with the mechanical or functional strengths of *Criterion* such as promptness, easiness, and convenience, but the absence of direct feedback made them dissatisfied.

Kim (2014) and Shim (2013) compared their participants' first and second drafts in order to investigate the effectiveness of *Criterion*. After tracking changes in drafts and comparing them, they concluded that the automated tool was effective for revising writing. However, Truscott (1996) and Kepner (1991) claimed that giving feedback on grammar was not helpful for improving learners' writing abilities. Also, tracking form changes in writing on the basis of automated feedback is limited in determining the role of the automated tool in learners' proficiency improvement; the participants might mechanically change the forms without any actual learning. The current study focuses on the participants' opinions on using *Criterion* in order to find strengths and weaknesses of the program. Kang (2008) and Kim (2014) claimed that writers' proficiency levels influenced their preferred feedback type and the influence of feedback. Thus, this study hypothesizes that writers' proficiency levels will influence their preferred feedback type, either automated feedback or human feedback. Also, there might be a relationship between non-native writers' preferred areas of revision and their satisfaction level of using *Criterion*.

III. METHOD

1. *Criterion*

Educational Testing Services (ETS) developed *Criterion* in the mid-1990s, and the program provides feedback on five areas: grammar, usage, mechanics, style, and organization & development. Users go to the *Criterion* website (http://*Criterion*.ets.org) and create user accounts. After filling in the blanks with their personal information and entering codes, they are allowed to access the class that their instructor created. Users can see a summary page and individual error page (ETS, 2015). Table 1 shows the subcategories of the five areas. The *Criterion* website called the five categories traits.

TABLE 1

Traits and Subcategories of *Criterion* Feedback

Traits	Subcategories
Grammar	Fragments; Run-on Sentences; Garbled Sentences; Subject-Verb Agreement Errors; Verb Form Errors; Pronoun Errors; Possessive Errors; Wrong or Missing Word; Proofread This!
Usage	Determiner Noun Agreement; Article Errors; Confused Words; Wrong Form of Word; Faulty Comparison; Nonstandard Word Form; Negation Error; Wrong Part of Speech
Mechanics	Spelling; Missing Capitalization of Proper Nouns; Missing Initial Capital Letter in a Sentence; Missing Question Mark; Missing Final Punctuation; Missing Apostrophe; Missing Comma; Missing Hyphen; Fused Words; Compound Words; Duplicate Words; Extra Comma
Style	Repetition of Words; Inappropriate Words or Phrases; Sentences with Passive Voice; Long Sentences; Short Sentences; Sentences Beginning with Coordinating Conjunctions
Organization & Development	Introductory Material; Thesis Statement; Main Ideas; Supporting Ideas; Conclusion; Transitional Words and Phrases

The grammar function provides feedback on errors regarding grammar and sentence structure, and the usage check provides feedback on the following errors: Determiner Noun Agreement; Article Errors; Confused Words; Wrong Form of Word; Faulty Comparison; Nonstandard Word Form; Negation Error; Wrong Part of Speech. Even though some subcategories deal with vocabulary errors like misspellings or typos, the feedback in this area is mainly related to grammar errors. Therefore, the grammar and usage checks were considered to provide feedback on grammar errors.

The mechanics function provides feedback on a variety of punctuation errors. The style and organization & development areas are the most difficult for the participants to understand. The style area provides feedback on how learners can revise the following problems: Repetition of Words, Inappropriate Words or Phrases, Sentences with Passive

Voice, Long Sentences, Short Sentences, Sentences Beginning with Coordinating Conjunctions. The *Criterion* website explains the purpose of this section saying, "*Criterion* can provide feedback on how to improve your writing style. Use this section for advice about how to revise words, phrases, and sentences so they are clearer and more effective" (ETS, 2015).

Organization & development offers advice on the organization and development of ideas in essays. This section provides suggestions on each part of an essay: Introduction, Thesis, Main Ideas, Supporting Ideas, Transitional Words and Phrases, Conclusion. *Criterion* users can check the summary of the automated feedback as seen in Figure 1. The *Criterion* website explains each step and provides examples. The screen shot of Figure 1 is one of such examples from the *Criterion* website

FIGURE 1

Screen Shot of *Criterion* Feedback Analysis (ETS, 2015)



Criterion users usually write their essays using a word processing program and copy and paste their essays into the online tool. When they enter their essays in the tool, they can receive immediate feedback. Figure 1 illustrates the summary of the feedback in five traits, and users can also check their errors in each trait. Explanations for each error under each section can be viewed. Figure 2 is another example of feedback in the *Criterion* website.

FIGURE 2

Screen Shot of *Criterion* Feedback (ETS, 2015)



Users can also check the "Writers' Manual" which is on the right side of the upper corner in Figure 2. The red arrow on the right is an icon for the “Writers’ Manual.” The sentences with feedback are indicated by color, and different colors denote different error types. The red arrow on the left side of Figure 2 indicates a menu for five areas, and users can check a summary of their feedback for each category via the drop-down menu.

2. Participants

Seventy-two freshmen at a women’s university in Seoul, Korea participated in this study, and the current data were collected throughout the fall semester. The university required students to take two mandatory English courses before graduation. As the courses were mandatory, the participants’ majors were varied. The university freshmen were required to take the TOEIC test as a placement test before registering for the mandatory courses. Students were placed into three proficiency levels depending on the results of the placement test: advanced, intermediate, and low level.

Students from intermediate and low levels participated in this study, and the participants’ TOEIC scores were used in order to provide a reliable comparison measurement of their English proficiency. At the beginning of the fall semester, the average TOEIC score of the low-level classes was 467, and that of the intermediate-level class was 660. The participants were asked to take the TOEIC test twice in the semester: one at the beginning of the semester (TOEIC_B) and the other at the end of the semester (TOEIC_A). Table 2 illustrates the participants’ detailed TOEIC scores.

TABLE 2

Descriptive Statistics of Participants’ TOEIC Scores

Level	N	Score	Mean (SD)	Min.	Max.
Low	48	TOEIC_B	466.53 (95.36)	295	685
		TOEIC_A	554.99 (89.38)	395	715
		Improvement	88.47 (99.11)	-115	315
Intermediate	24	TOEIC_B	660.21 (40.37)	600	725
		TOEIC_A	709.38 (76.09)	565	850
		Improvement	49.17 (57.40)	-65	125

Students attended classes two times per week for 15 weeks, and each class was 75 minutes in length. Both classes mainly focused on English conversation and preparation for the TOEIC test. Each class for the fall semester consisted of several required tasks, with the main assessments being two essays and follow-up activities.

3. Data Collection and Analysis

In the fall semester, participants were asked to submit two essays on different topics. The participants went through three steps of the writing process before they submitted their final draft. First, they wrote first drafts of essays and used *Criterion* to receive automated feedback. On the basis of the feedback, participants revised their writing and exchanged their revised drafts with their partners for peer feedback. After they rewrote their essays on the basis of the peer feedback, they finally submitted their final versions. The participants were taught how to use *Criterion* and watched a demonstration before they used the tool. They used the tool one time per essay, and school paid the program fee.

After submitting each assignment, each student and the instructor, who was also the researcher, had a writing conference in order to discuss issues in their essays. During the writing conference, the instructor did not provide direct feedback on grammar or sentence structure. Instead, the instructor asked for clarification or guided students in how to revise miscommunications in their essays. During the writing conference, students had a chance to talk about their experience using *Criterion*. At the end of the semester, participants were asked to complete a questionnaire, and their answers were analyzed. Each question followed the five-point Likert scale: strongly agree (5), agree (4), neither agree nor disagree (3), disagree (2), and strongly disagree (1).

In order to examine the internal reliability among the items in each factor, reliability tests were performed. The usage category, which asked about the revision of vocabulary, was excluded in the analysis, as the calculated Cronbach's alpha was lower than .4. Question numbers for each category on the questionnaire are listed in Table 3.

TABLE 3
Questions in Each Factor and Reliability

Factors	Question #	Cronbach's Alpha
Grammar	4, 7, 10, 29	.515
Organization & Development	5, 6, 8, 9, 12, 13, 14, 16, 17, 31, 32, 33, 35, 36	.805
Mechanics	11, 34	.547
Criterion	24, 26, 27, 28, 39	.823
Teacher Conference	40, 41	.704

Independent t-tests were administered in order to explore the mean differences in terms of *Criterion* feedback and writing conferences between intermediate- and low-level participants. A paired sample t-test was used to investigate the mean differences between *Criterion* feedback and writing conferences. Then, bivariate linear regression

analysis was performed in order to examine the relationship among feedback factors and *Criterion* feedback.

IV. RESULTS and DISCUSSION

In accordance with the first research hypothesis, participants were expected to show different levels of satisfaction with *Criterion* feedback and writing conferences. In order to explore whether each proficiency group demonstrated mean differences in terms of effectiveness of *Criterion* and that of writing conferences with the instructor, independent t-tests were performed. Table 4 illustrates the mean differences and standard deviation between the groups, but no statistically significant mean differences were found.

TABLE 4

Results of T-tests between Intermediate & Low Level

	Group	N	Mean (SD)	*t*	*p*
Teacher	Intermediate	24	3.92 (0.55)	-1.49	0.14
	Low	48	4.10 (0.48)		
Criterion	Intermediate	24	2.91 (0.49)	1.59	0.12
	Low	48	2.68 (0.70)		

*p<.05

Even though no statistically significant mean differences were found between the two groups, the means for the effectiveness of the writing conference were much higher than those for the effectiveness of *Criterion*. As group differences were not found in the first t-test, a paired sample t-test was performed in order to investigate the participants' preferences between *Criterion* and writing conference feedback. The t-test results indicated that the mean differences between the perceived effectiveness were statistically significant at the .05 level (t=-10.84, p<.01).

TABLE 5

Results of T-test between *Criterion* & Writing Conference

Group	N	Mean (SD)	*t*	*p*
Teacher	72	4.04 (0.51)	-10.84	0.00***
Criterion		2.76 (0.64)		

*p<.05; **p<.01; ***p<.001

In terms of feedback on their essays, the participants in both groups preferred the writing conference with the instructor to the *Criterion* feedback, and the mean for the writing conference was significantly higher than that for *Criterion*. The mean of the teacher feedback was 4.04 (SD=.51) while that of *Criterion* was 2.76 (SD=.64).

Most English writing textbooks emphasize organization and idea development in the process of revision, but Korean writers still value correction of grammar in revision (Cha, 2007; Kim, 2010, 2011). Some learners still think that revision merely means changing spelling errors or fixing grammatical and mechanical errors. Therefore, three of the five categories were used in order to investigate which element best predicted the participants' perceptions about the effectiveness of *Criterion*. In order to explore what affected *Criterion* users' perceptions on the effectiveness of the tool, three separate linear regression analyses were performed. Normality and homoskedasticity were assumed. Table 6 reports the results of regression analyses between the three factors and *Criterion* effectiveness.

The regression analyses showed that the category of Organization & Development (O&D) and *Criterion* effectiveness were positively related, and the R square was .128. Organization & Development accounted for 12.8% of variance of *Criterion* effectiveness, and it was statistically significant at the .05 level (t=3.223, p<.01). Grammar and Mechanics were positively related with *Criterion* effectiveness, but the R square scores were .021 and .045. Two of them were not statistically significant at p<0.5.

TABLE 6
Regression Analysis Results

	R Square	B	Beta	*t*	*p*
Grammar	.021	.183	.145	1.235	.221
O&D	.128	.519	.357	3.223	.002**
Mechanics	.045	.157	.212	1.830	.071

*p<.05; **p<.01

Previous research (Kim, 2010, 2011) claimed that direct feedback regarding grammatical errors was related to writers' satisfaction levels with *Criterion* feedback. Therefore, the current study hypothesized that feedback on grammar and mechanics might account for the most of the variance of *Criterion* effectiveness. EFL writing textbooks emphasize the importance of organization and development in the revision process, so that category was expected to account for the *Criterion* effectiveness as well. However, the participants' satisfaction levels with *Criterion* feedback were not as high as expected, so the mean of each question on the questionnaire and each question were analyzed. Table 7 shows the means and standard deviations of all questions on *Criterion* feedback.

TABLE 7

Descriptive Statistics of *Criterion* Feedback

	Mean	SD
The *Criterion* feedback was helpful in improving my writing.	2.73	.804
The *Criterion* feedback was helpful in revising my writing.	3.24	.796
The *Criterion* feedback was helpful in increasing my interest in English writing.	2.49	.731
I trust the *Criterion* feedback.	3.22	1.038
The *Criterion* feedback was more helpful than human feedback.	2.11	.820

The mean for helpfulness of *Criterion* in the revision process was the highest among the questions illustrated in Table 7. The participants answered that *Criterion* was helpful for the revision of their essays, and they trusted the feedback. However, the mean for the first question, the *Criterion* feedback helped them to improve their overall writing proficiency, was not as high as that on the helpfulness of *Criterion* in the revision process. Moreover, the mean for the third question, the *Criterion* feedback was helpful in increasing their interest in English writing, was 2.49, and the mean for the comparison between *Criterion* and human feedback was only 2.11 on a five-point scale. The results showed that most of the participants preferred human feedback to the *Criterion* feedback.

TABLE 8

Descriptive Statistics of Teacher Feedback

	Mean	SD
The instructor′s feedback was more helpful than *Criterion*.	4.07	.635
The instructor′s feedback in the 10th week was helpful.	4.01	.517

The descriptive statistics in Table 8 also indicate that the current participants preferred human feedback to the *Criterion* feedback. Means for both questions regarding the helpfulness of the writing conference were over four points on a five-point scale. The participants stated that they could negotiate meaning when they had a question during the writing conference and could reflect on their writing process step by step. In contrast, the *Criterion* feedback was one-way feedback; when the participants disagreed with the feedback or had questions regarding the feedback, they could not ask for clarification.

The participants answered the open-ended questions in the questionnaire regarding the *Criterion* feedback. Even though they answered that *Criterion* was interesting and increased their motivation, they were not happy with using *Criterion*. Most of the participants had never used an automated feedback program before they attended the classes in this study, so they thought that the automated tool was innovative, and it

roused their curiosity. However, the participants responded that they wanted to know what grammatical errors they made and desired explanations about the errors, but the tool did not exactly pinpoint the errors or explain in detail. Sometimes students could not find what was wrong in the sentences. They tried to identify their problems, but they could not. They eventually got frustrated and gave up finding out how to fix their errors. Kim's (2014) participants also had difficulties in using *Criterion*, and thought that "insufficient information" and "lack of context and technological problems" were the weaknesses of the tool.

Another aspect they were disappointed with was that the program indicated word repetition as an error. Students used the first person pronoun “I” several times, and the tool commented that repeating “I” was an “error.” During the writing conference, they asked what was wrong with using “I,” and they did not know how to replace the pronoun. Overall, the *Criterion* feedback made the participants frustrated, and they felt that they could not understand the feedback.

TABLE 9

Sample Feedback of Each Category

Category	Comments
Organization & Development (Supporting Ideas)	*Criterion* has identified three or more supporting ideas in this paragraph. Do these ideas support the topic sentence of your paragraph? Use examples, explanations, and details to support and extend your main ideas. Look in the Writer's Handbook for ways to develop supporting ideas.
Grammar (Fragment)	"Even cell phone rings in class" This sentence may be a fragment. Proofread the sentence to be sure that it has at least one independent clause with a complete subject and predicate.
Mechanics (Spelling)	"Me2day" This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

As seen in the sample feedback of *Criterion* in Table 9, suggestions from the *Criterion* feedback were at times lacking in specificity. The comments for Organization and Development were difficult for low-level writers to understand and correct by themselves. Moreover, the second and third comments were incorrect. The sentence in the grammar feedback was not fragmented but had the subject and the verb. *Criterion* detected "Me2Day" as a spelling error, but it was not misspelled. Even though "Me2Day" stopped its service, it had been a popular website in Korea.

This is not to suggest that the tool should pinpoint every single error and suggest a model answer. However, instead of saying “something is wrong in this sentence,” it could have said that the sentence included a verb error or a verb tense error, for example.

With such detailed feedback, the participants could have understood the feedback more than they did, and it could have increased their satisfaction levels with the tool.

V. CONCLUSION

This study explored participants' preferences between *Criterion* feedback and human feedback and whether or not the participants' proficiency levels influenced the satisfaction levels of *Criterion*. When Shim (2013) investigated the influence of *Criterion* and the most frequently appearing error types, she found that her participants were satisfied with using *Criterion* because of its convenience. Attali (2004) and Kim (2014) also found that *Criterion* was helpful for writers to use when revising their essays; their participants showed overall improvement of writing abilities after using *Criterion*. However, tracking form changes in drafts after using the tool has a limitation in order to be identified as proficiency improvement.

On the other hand, Kim (2010), Moon and Pae (2011), and Otoshi (2005) claimed the opposite regarding the effectiveness of *Criterion*. When Moon and Pae (2011) investigated the effect of the *Criterion* feedback, they found that some feedback was not reflected on students' revisions, and even some of the feedback was incorrect. Otoshi (2005) also found that *Criterion* could not detect errors and could not provide appropriate feedback on several areas.

On the basis of Kang's (2008) study, the current study hypothesized that there might be mean differences between intermediate- and low-level participants regarding the satisfaction levels with *Criterion* feedback. It was also hypothesized that they might prefer *Criterion* feedback to human feedback (Attali, 2004; Kim, 2014). As previous research claimed that EFL writers mainly focused on grammar and mechanical revision, and EFL writing textbooks emphasize the importance of organization and development, Grammar, Organization & Development, and Mechanics factors were expected to be positively related to the participants' satisfaction levels with *Criterion* feedback.

The results indicated that both level participants did not show mean differences in terms of the satisfaction levels with *Criterion* feedback, and they preferred the writing conference to *Criterion* feedback. Furthermore, the regression analyses revealed that Organization & Development was positively related to the participants' satisfaction levels with *Criterion* feedback, but Grammar and Mechanics were not related to the satisfaction level. During the interviews, the participants stated that they were satisfied with the convenience and promptness of *Criterion*, and that the tool was helpful for revising their essays in some aspects. However, it did not help them to improve their overall writing proficiency. They sometimes could not understand the feedback, and

some of the feedback was even incorrect. The most frustrating part was the lack of communication about the automated feedback. They preferred the writing conference because they could ask the instructor and negotiate meaning when they had questions.

Almost ten years ago, Warschauer and Ware (2006) warned that instant evaluation needed to be carefully used in specific teaching and learning contexts. Kim (2010) and Otoshi (2005) concluded that such tools needed not to be used alone in evaluating or providing feedback due to their shortcomings. This research also supported the previous studies. The participants stated that they were so relieved that there was a conference after using the tool. As previous research claimed, *Criterion* is an innovative and effective automated feedback program, but its manner of giving feedback still needs improvement. Moreover, the effectiveness of using the tool by itself still needs additional examination and empirical research.

For further study, exploring the difference between more varied proficiency levels is needed. Kim's (2014) participants were advanced-level writers, and they were satisfied with the automated feedback. Participants in other studies also had different proficiency levels. The previous research results have not reached agreement on the effectiveness of the automated feedback because of the participants' different proficiency levels. Therefore, exploring the differences among varying proficiency levels could provide researchers and teachers with more unified research outcomes. In addition, in order to investigate EFL writers' proficiency improvement, other research methods are needed in addition to tracking changes in drafts.

REFERENCES

Attali, Y. (2004). Exploring the feedback and revision features of *Criterion*. Paper presented at the National Council on Measurement in Education (NCME) conference, April 2004, San Diego, CA.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The *Criterion* online writing service. *AI Magazine, 25*(3), 27–36.

Cha, Y. (2007). A study of peer feedback patterns in CMC modes on Korean EFL students. *Multimedia-Assisted Language Learning, 10*(1), 9-35.

Egbert, J., & Hanson-Smith, E. (Eds.). (1999). *CALL Environments: Research, practice and critical issues*. Alexandria, VA: TESOL, Inc.

Ellis, R. (1997). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language

context. *System, 36*, 353-371.

ETS. (2015). *A student's guide to beginning Criterion*. Retrieved December 10, 2014 from the World Wide Web:https://*Criterion*.ets.org/Content/AStudentsGuidetoBeginning*Criterion*.pdf

Fang, Y. (2010). Perceptions of the computer-assisted writing program among EFL college learners. *Educational Technology and Society, 13*(3), 246-256.

Ferris, D. (1997). The Influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315–339.

Ferris, D. (1999). The case for grammar correction in L2 writing classes. A response to Truscott (1996). *Journal of Second Language Writing, 8*, 1-10.

Ferris, D. (2003). *Response to student writing: Implications for second language students*. Mahwah: Lawrence Erlbaum Associates

Hyland, K., & Hyland, F. (2006). *Feedback in second language writing: Contexts and Issues*. Cambridge: Cambridge University Press.

Ihm, H. (2005). The study of development of multimedia-assisted English courses for college students: On the basis of needs analysis. *English Language Teaching, 17*(3), 235-257.

Kang, D. (2008). Feedback on EFL writing: Teacher, peer, and self-review. *Foreign Languages Education, 15*(1), 1-22.

Kepner, C. (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.

Kim, M. (2007). A study of teacher's feedback on students' errors in writing. *Foreign Languages Education, 14*(3), 429-451.

Kim, J. (2014). The effectiveness of automated essay scoring in an EFL college classroom. *Multimedia-Assisted Language Learning, 17*(3), 11-36.

Kim, J., & Ha, M. (2014). A study on EFL college students' writing responses to different written feedback styles. *English Language & Literature Teaching, 20*(2), 59-77.

Kim, T. (2010). Reflection on using the *Criterion* online writing evaluation service. *Multimedia-Assisted Language Learning, 13*(3), 59-83.

Kim, T. (2011). Corrective feedback that an automated writing evaluation system can and cannot provide. *English Teaching, 66*(1), 111-140.

Lee, S. (2008). Exploring the potential of a web-based writing instruction program and AES: An empirical study using *My Access*. *Multimedia-Assisted Language Learning, 11*(2), 103-125.

Long, M. (1996). The role of linguistic environment in second language acquisition. In W. Ritchie, & T. Bhatia, (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego: Academic Press.

Moon, Y., & Pae, J. (2011). Short-term effects of automated writing feedback and users' evaluation of *Criterion*. *Korean Journal of Applied Linguistics, 27*(4), 125-150.

Na, I., & Kim, T. (2014). Analysis of writing activities in elementary English textbooks and the perceptions of 5th and 6th graders. *English Language & Literature Teaching, 20*(3), 231-256.

Nam, J., & Kim, S. (2009). A case study of incorporating multimedia technology into general college English classes. *New Korean Journal of English Language and Literature, 51*(1), 277-301.

Otoshi, J. (2005). An analysis of the use of *Criterion* in a writing classroom in Japan. *The JALT CALL Journal, 1*(1), 30-38.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Shim, Y. (2013). The effects of online writing evaluation program on writing capacities of Korean students. *Teaching English with Technology, 13*(3), 18-34.

Shin, Y., & Seong, M. (2013). Corrective feedback effects in university English writing clases. *English Language & Literature Teaching, 19*(3), 247-270.

Sim, J., & Heo, J. (2014). Writing feedback: The benefits of conferences. *English Language & Literature Teaching, 20*(3), 165-189.

Simmering, M., Posey, C., & Piccoli, G. (2009). Computer self-efficacy and motivation to learn in a self-directed online course. *Decision Sciences Journal of Innovative Education, 7*(1), 99-121.

Swain, M. (1995) Three functions of output in second language learning. In G. Cook & B. Seidlehofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H.G. Widdowson* (pp.125-144). Oxford: Oxford University Press.

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

Vantage Learning. (2009). *My Access: Efficacy report*. Retrieved September 1, 2014 from the World Wide Web: http://vantagelearning.com

Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *An International Journal, 3*, 22-36.

Warschauer, M., & Meskill, C. (2000). Technology and second language teaching and learning. In Rosenthal, J. (Ed.), *Handbook of Undergraduate Second Language Education*. Lawrence Erlbaum, Mahwah, NJ.

Warschauer, M., Turbee, L., & Roberts, B. (1996). Computer learning networks and student empowerment. *System, 14*(1), 1-14.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research, 10*(2), 157-180.

Examples in: English
Applicable Languages: English
Applicable Level: Advanced Adult Learners

Eun-Hee Lee
130-739 1 Hoegi-dong Dongdaemun-gu
Seoul, S. Korea
Dept. of American Studies
Kyung Hee Cyber University
Phone: 02-3299-8530
Email: ehlee1@khcu.ac.kr

Received in January 15th, 2015
Reviewed in February 26th, 2015
Revised version received in March 5th, 2015