THE EFFECTS OF ONLINE WRITING EVALUATION PROGRAM ON WRITING CAPACITES OF KOREAN STUDENTS

by **Yae Jie Shim**
Seoul Women's University, Korea
hzpb0656 @ naver.com

Abstract

The error-correction program *Criterion* provides students with an immediate essay feedback using tools that can analyze and review writing automatically. This feedback covers grammar, usage, mechanics, style, organization, and development. With its diagnostic tools for scoring essays and offering relevant feedback, the error-correction program provides a way to speed up the otherwise time-consuming process of essay composition and evaluation. The usefulness of the error-correction program is highlighted by analyzing the extent to which it helps second language learners improve their writing abilities, with a focus on the degree of student improvement caused by the program's prompts from the first draft to the final essay submission.

The purpose of this study is to determine the extent to which second language learners can improve their writing capacities using a specific set of online instruction materials, the error-correction program online essay writing tool in a group of 96 university students at the intermediate level of English.

Keywords: *Criterion*; automated error-correction program, writing instruction

1. Introduction

Given the importance of integrating computer technologies in second language learning, especially in learning writing, the issue of interaction between instructors and students has become more significant. The supportable technology and curricula help students become more independent and spontaneous learners. Warschauer (2000) asserts that time and place-independent communication can increase the advantages of interaction in language classrooms. With Web-based online communities and tools, students can participate in self-controlled study programs outside school.

The online feedback tool *Criterion* provides students with immediate feedback about their essays using critique and writing-analysis tools that give feedback on grammar, usage, mechanics, style, and organization and development. (Attali & Burstein, 2006). The most

effective feature of the tool is its ability to automate and speed up the process through its diagnostic feedback and essay scoring capabilities. Thus, the purpose of this article is to reflect upon the success of the *Criterion* program, highlight students' error types, use the tool to outline suggestions for improvement and attempt to find effective learning and teaching method for the L2 classroom.

The purpose of this research is to develop resources that will assist students in dealing with the English language aspects of their subjects more effectively. The research questions are as follows:

(1) Is the evaluation of error types using online instructional writing tool *Criterion* effective in improving Korean students' English proficiency?

(2) What kinds of error types are most frequently shown in Korean student's writing?

2. Literature review

Internet-based learning gives students more freedom and planning time in which to fully identify their thoughts and ideas. This may have increased their motivation as well. Moreover, most of the interviewed students are generally satisfied with the Internet-based learning program as a tool because they believe that it helps them study independently.

As Frizler (1995) argues, the future of education may not offer a choice of whether to teach online, but technical developments may extend language learning beyond the traditional face-to-face class environment. One of the most challenging dilemmas for traditional education is the discrepancy between the knowledge learned in the classroom and the knowledge needed in the real world (Bardine, Bardine & Deegan, 2000). The former is sometimes too distant and abstract to apply to real-life situations. However, in the technology-enriched learning era of today, the classroom has expanded to encompass the world.

Today, in many university contexts around the world, there is a clearly recognizable trend toward a narrow consistency in hardware and software applications. Whereas in the past this may have meant choosing a PC-style computer over a Mac, increasingly this strategy also applies to the software applications that are purchased and supported. Thus, we see *WebCT* or *Blackboard* chosen as the university-wide software development tools. We also find that these are the only tools for which technical support and training is available (Levy & Stockwell, 2006).

Technology-facilitated classroom activities can help students make connections with the real world and equip them with more authentic learning materials (Arms, 1985).

Cho (2001) has described the merits of Internet-facilitated instruction when compared to traditional instruction: the former extends the boundaries of learning and teaching; provides easy access to information; and facilitates the student's active class participation.

Historically, the invention of new technologies has been largely motivated by a desire to extend or overcome innate human limitations, especially those defined by physical and mental capacities. The technologies used in CALL extend well beyond communication tools to include generic tools and devices, such as word processors for writing, online dictionaries for vocabulary work, and MP3 players for intensive listening. Ideally, responsible pedagogy ensures that CALL materials are used in an appropriate, principled, and effective way (Warschauer & Kern, 2000).

Using CALL in English language classrooms is useful as a guideline for instructors and students to accept the class instruction (Breen, 2005). Similarly, Goldberg (2003) indicated that students are more motivated when using computer and online resources and with the reason, they produce higher quality writing and long essays. Giving students feedback is the significant component of a writing class. Interestingly, students look forward to having the feedback to improve their work. It means that "giving students task-specific feedback results in more revisions made to essays" (Folse & Solomon, 2004, p. 52).

Computers have been used to evaluate student writings for several years, and various studies have investigated the validity and reliability of Writing Evaluation Programs. The Educational Testing Service of New Jersey also invests substantial resources in designing products to improve student learning and in evaluating their effects (Burstein & Chodorow, 2004).

Feedback plays an important role in writing education and many researchers have studied the necessity for and various features of error correction and feedback (Cha, 2007). However, many researchers and instructors also agreed with the difficulties of giving feedback as often as students asked. Research also tells us that students are more likely to engage in planning and pre-writing activities if they are provided with tools to facilitate those processes (Harris & Graham, 2006). Furthermore, when students engage in these activities, their writing improves (Graham & Perrin, 2006).

As Figure 1 shows, the error-correction tool can provide instructors with the time to support students in the higher-order features of writing, either whole class or personally, by

changing the instructor's role in writing instruction. The error-correction program can help to revise students' errors and enable the instructors to create more chances to interact with their students on other aspects of writing. If used regularly and appropriately, this tool has the ability to improve the quality of second language writing.

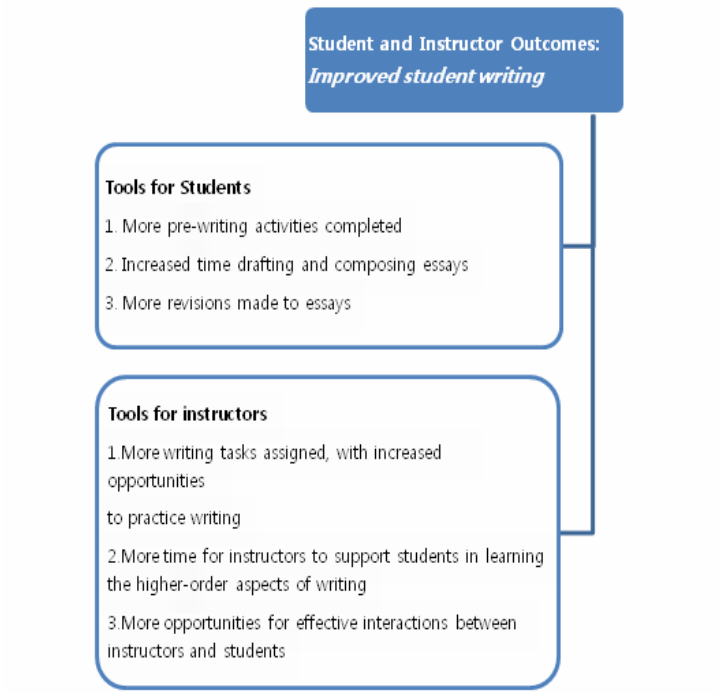

Figure 1. Student and instructor outcomes after using a writing evaluation tool.

3. The study

3.1 Research setting and participants

This study was aimed at measuring the perception of using the error-correction program by students and by instructors and at finding out students' frequent error types using *Criterion.* The study took place in the second semester of 2010. The course was a major English one for general English education; it met for 15 weeks, 3 hours a week. The major objective of the course was to help students develop more effective English writing skills.

The participants for this study consisted of 96 students from 2 classes taking a General English course in 2010. The participants were Class 62, 71 consisting of 53 students who used the program. However, only 43 were included in the corpus because some students did not submit

their essay online. 43 students in Control group from another Class 50, 56 did not use the program. All of them were taking a General English course at university; they were all intermediate level (pre-testing results), demonstrated an interest in English writing, and wanted to improve their writing abilities. They were tested by TOEIC and a standardized writing test. All the students were told they could access the error-correction program site as required in their own time. In addition, all students were asked to complete opinion surveys regarding their use of program in the same week they submitted their assignments, some 16 weeks after being introduced to the resources. It was hoped that students would use the error-correction program several times as the submission date approached. Finally, a series of semi-structured interviews were conducted several times to yield qualitative data.

3.2 Materials: the function of *Criterion*

The tools in the error-correction program identify four main types: grammar, usage, mechanics, and stylistic errors. Each category identifies different error types such as agreement errors, verb formation error, wrong word use, missing punctuation and typographical errors, etc. (Finch, 2004)

Figure 2 shows how some of the most common essay writing errors, such as missing or extra articles, are highlighted automatically when students miss out articles or use the wrong pronouns. The system provides advice via an explanation either in English or Korean. The Korean version of the advice, usually translated, is especially useful to intermediate students. This translated version seems to help students understand their errors, but it is still limited by its simplicity.

Figure 2. Usage: Missing of Extra Article Check (Self-Revision).

3.3 Design and procedure

All students learned from the instructor how to use the tool and how to receive automated feedback. The instructors allowed them to use one time per one submission. The students were required to hand in their drafts alongside their final version.

Each student was pre-tested to check their English levels and selected according to the results of the separate in-class writing test and the TOEIC trial test. This was done to determine their English proficiency levels and their basic writing abilities. The pre-writing test took place in class. The two classes were taught in different classrooms at separate times. They took the regular English course using *Criterion*.

Before starting this study, the researcher administered a series of individual pre-writing tests. The instructor asked participants not to use a dictionary while working on the pre-test because it was important to assess the participants' genuine writing abilities. After they completed the test, they handed the results in to the instructor. They were then tested 4 more times and given feedback by the error-correction program. They were also administered 5 post-tests after they had received 5 treatments by the error-correction program.

3.4 Results and findings

A questionnaire was administered immediately after the tests. All of the participants in the experimental group were asked to comment on whether the error-correction program was useful and whether it was easy or difficult to use. This survey was important as a check on the effectiveness of the error-correction program. The questionnaire comprised 2 questions: the first one concerned the effects of the error-correction program while the second related to its perceived usefulness. The results of the questionnaire were used to reveal the student's attitudes to the program's overall usefulness. The statistical processing utilized SPSS/WIN statistical program 12.0. First, to examine the pre-post writing test results according to time and group, a two-way repeated ANOVA was carried out.

3.4.1. Analysis of writing error number and types

A workable second-language writing program has obvious benefits for English teaching instructors in Korea. A program capable of checking basic writing mistakes will allow instructors and students to free up a much greater portion of time for higher-level issues of content, organization, and argument. Table 1 presents an analysis of the main types of errors made by Korean students over a semester. The main error types are grammar, usage, mechanics, and style issues. The researcher collected student's feedback notes and analyzed the error types and their number. Table 1 presents a breakdown of a set of 43 individual student compositions. In fact, 53 students used the program, but 10 handed in their work without the program's error reports. Each sample represented the student's first draft. There were 2.81 grammar errors, 2.98 usage errors, and 4.19 mechanical errors. Most significantly, there were 28.49 style errors. It would appear, therefore, that style represents the biggest issue for most Korean students.

Table 1. Breakdown of error type in a sample of the work of 43 students

	N	**Mean**	**SD**
Grammar	43	2.81	2.26
Usage	43	2.98	2.78
Mechanics	43	4.19	5.37
Style	43	28.49	14.11

(1) Grammar errors

The error-correction program can isolate eight different categories of grammar errors, including fragments of missing commas, run-on sentences, garbled sentences, subject-verb agreement, ill-formed verbs, pronoun errors, possessive errors, and wrong or missing words. It also highlights some sentences, offering the instruction "proofread this" (Kim, 2007). All errors are individually indicated. Figure 4 and Table 2 show that the most widespread problems in the written compositions of the research sample were fragments of missing commas, run-on sentences, and subject-verb agreement – rated 2.81, 2.98, and 0.33 respectively. The awkward use of the pronoun proved to be unexpectedly infrequent, probably because the students were limited to compositions of approximately 250 to 500 words.

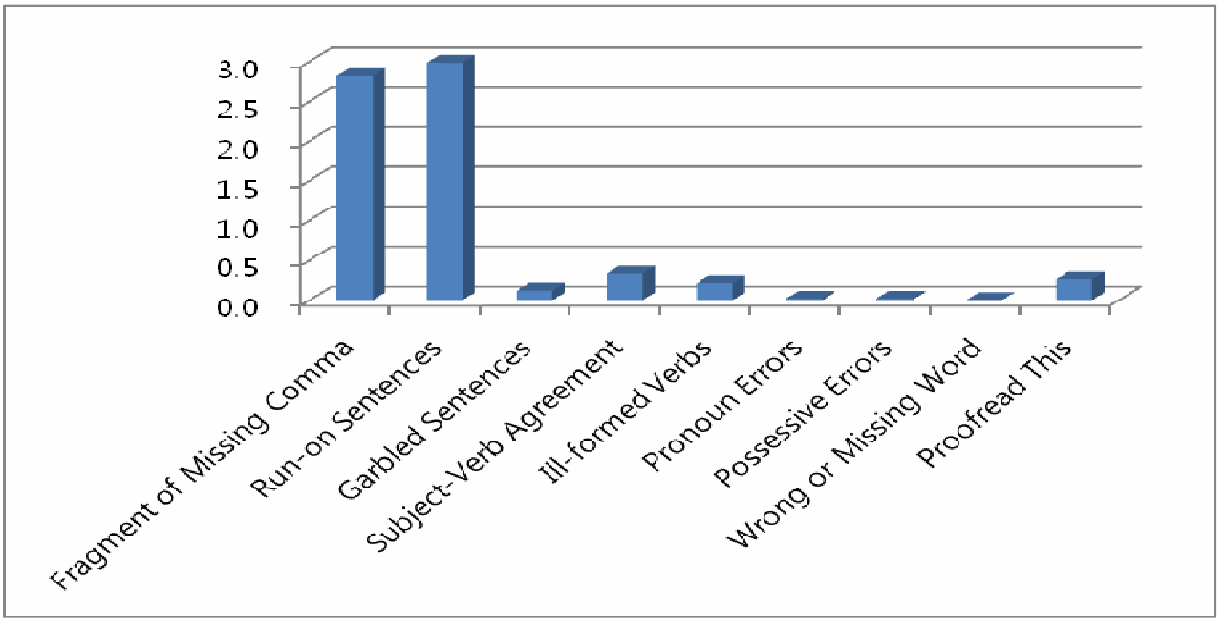

Figure 3. Grammar errors in a corpus of 43 compositions.

Table 2. Grammar errors.

	N	**Mean**	**SD**
Fragment of Missing Comma	43	2.81	2.26
Run-on Sentences	43	2.98	2.78

Garbled Sentences	43	.12	.324
Subject-Verb Agreement	43	.33	.522
Ill-formed Verbs	43	.21	.466
Pronoun Errors	43	.02	.152
Possessive Errors	43	.02	.152
Wrong or Missing Word	43	.00	.539
Proofread This	43	.26	14.11

(2) Usage Errors

The error-correction program also indicates usage errors, including wrong articles, missing articles, extra articles, confused words, wrong forms of the word, faulty comparisons, preposition errors, nonstandard word forms, and negation errors. As Table 3 illustrates, errors of missing or extra article were statistically significant relative to the other types of usage errors. The major reason for this is probably the differences in grammar patterns and principles between Korean and English. To resolve this issue, instructors need to focus more on these language differences. Fortunately, when students are allowed to revise their compositions several times following the instructions of the error-correction program, the error rate in the final draft is often greatly diminished. This leaves only organization and style as outstanding issues, both of which can be addressed by the instructors.

Table 3. Usage errors.

	N	**Mean**	**SD**
Wrong Article	43	.21	.559
Missing or Extra Article	43	2.40	2.574
Confused Works	43	.23	.527
Wrong Form of Word	43	.00	.000
Faulty Comparisons	43	.00	.000
Preposition Error	43	.16	.433
Nonstandard Word Form	43	.00	.000

Negation Error	43	.02	.152

(3) Mechanical Errors

Criterion evaluates ten types of mechanical errors, including spelling, the capitalization of proper nouns, missing initial capitals in sentences, missing question marks, missing final marks, missing apostrophes, missing commas, hyphen errors, fused words, compound words and duplicates. A number of basic functional errors such as missing question marks, missing periods, missing apostrophes, and missing commas are all prominent. As shown in Table 4, spelling errors are highly prominent (scoring 2.05 points), with the failure to capitalize proper nouns next (1.44 points).

Table 4. Mechanical errors.

	N	**Mean**	**SD**
Spelling	43	2.05	2.497
Capitalize Proper Nouns	43	1.44	3.466
Missing Question Mark	43	.49	1.791
Missing Final Punctuation	43	.07	.258
Missing Apostrophe	43	.00	.000
Missing Comma	43	.14	.467
Hyphen Error	43	.00	.000
Fused Words	43	.00	.000
Compound Words	43	.00	.000
Duplicates	43	.00	.000

(4) Stylistic Errors

The error-correction program evaluates six types of stylistic error, including the repetition of words, the use of inappropriate words or phrases, beginning sentences with coordinating conjunctions, too many short or too many sentences, and the inappropriate use of the passive voice. There was lack of balance between short and long sentences among the more significant stylistic errors in the corpus of student essays. For example, many students ended their essays

with very short concluding paragraphs. Although *Criterion*'s advice is not particularly sophisticated, it is able to note the lengths of the short and long sentences and to suggest ways of achieving a better balance.

Perhaps the students' most significant stylistic error is repetition. Other errors include unmotivated lexical repetition, inappropriate lexical or phrase choice, the use of sentences beginning with coordinate conjunctions, the use of too many short or long, sentences, and the improper use of the passive voice. Unmotivated lexical repetition is caused by the students' typically limited lexicon.

Another significant error occurs in the standard lengths of paragraphs. Korean students tend to end their essays abruptly and often use short sentences connected by conjunctions or long sentences with inappropriate clauses.

> **Topic: My Important Event**
>
> November 11thof Korea is a special day on which people give and take Pepero. Sticksnack dipped in chocolate, to and from meaningful people. It is on the day in 2004 that I made the biggest mistake in my whole life. I was busy practicing and being trained as a member of Changwon Youth Choir. That means I didn't have enough time to buy Pepero. I finally decided to give my friends what I took last year. My idea was terrible. When I couldn't expect happened. It absolutely had my people in trouble. First without any consideration I handed the most. Next I left school because of the rehearsal. After that my friends opened what I gave. The class got full of their screaming. From my Pepero hundreds of ants came out of them.
>
> (Essay Sample of Student #1)

Stylistic errors are the most significant errors in Korean second language writing and also the most difficult error to correct. The recognition of writing style is important to all English teachers but is especially important for non-native instructors of English. Faulty stylistic habits result from the typical student's limited vocabulary. The students tend to use many short sentences, with their minimal vocabularies working overtime. This is interspersed with long sentences that occasionally employ inappropriate clauses. As Table 5 shows, the error-correction program scored "word repetition" at 23.37 points, while "too many short sentences" scored 3.56 points.

Table 5. Stylistic errors.

	N	**Mean**	**SD**
Repetition of Words	43	23.37	12.364
Inappropriate Words or Phrases	43	.00	.000
Sentence Beginning with Coordinating Conjunction	43	.47	1.297
Too Many Short Sentences	43	3.56	4.420
Too Many Long Sentences	43	.00	.000
Passive Voice	43	.14	.413

(5) The Analysis of Correlation

Table 6 illustrates the correlation analysis between the error-correction program's error number and the instructor's writing assessment. The results demonstrate a correlation between the error-correction program's indication of mechanical errors and the instructor's first pre-test writing test assessment. Thus, students who tend to make many mechanical errors will also tend to receive a low grade on writing tests. The Critique function of *Criterion* program may be able, then, to assist with student error correction, liberating busy instructors for other tasks.

Table 6. Correlation analysis between *Criterion*'s error number and the instructor's writing assessment.

	TotalErr	**Grammar**	**Usage**	**Mechanic**	**Style**
Instructor's writing assessment	-.244	-.056	-.152	-.356	-.098
	.115	.721	.332	.019**	.533
	43	43	43	43	43

Figure 5 illustrates a correlation analysis of the error-correction program's indication of spelling errors and missing or fragmented commas. The results demonstrate a positive correlation between the two, indicating that students who tend to make many spelling mistakes will also tend to miss commas or use them in a fragmented way. This is a very tentative conclusion, but it warrants

further study in the pursuit of new teaching methods because the instruction of grammar (such as fragments and missing commas) and spelling both play significant roles in writing education; the results might indicate ways of developing a more interactive point.

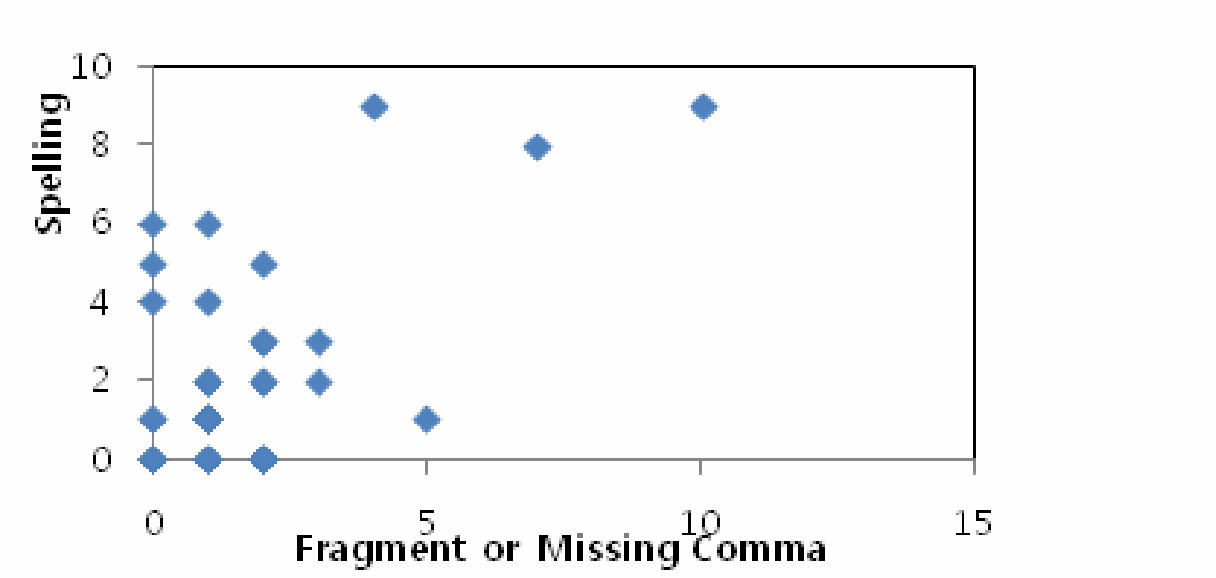

Figure 4. Correlation analysis of *Criterion's* indication of spelling errors and missing/fragmented commas.

3.4.2. Analysis of Grammar Error Number and Types

A repeated measures ANOVA of GA (Grammar score) according to time and group was conducted. The relevant statistics for the mean and standard deviation are noted below.

Table 7. Descriptive statistics of total scores on writing pre- and post-tests according to time and group.

		Control(43)		*Program (43)*		Total (86)	
		Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Wpre	One time	16.08	1.313	16.36	1.642	16.22	1.486
	2 times	18.51	1.295	18.91	1.390	18.71	1.352
	3 times	22.43	1.408	22.79	1.215	22.61	1.321
	4 times	25.21	1.261	25.02	1.611	25.11	1.443
	5 times	29.23	1.235	29.64	1.520	29.43	1.394
Wpost	One time	22.79	1.166	25.23	1.450	24.01	1.791
	2 times	26.23	1.187	27.70	1.202	26.96	1.400

	3 times	29.53	1.265	30.32	1.173	29.92	1.278
	4 times	31.42	1.447	32.68	1.156	32.05	1.450
	5 times	35.87	1.630	38.13	1.161	37.00	1.810

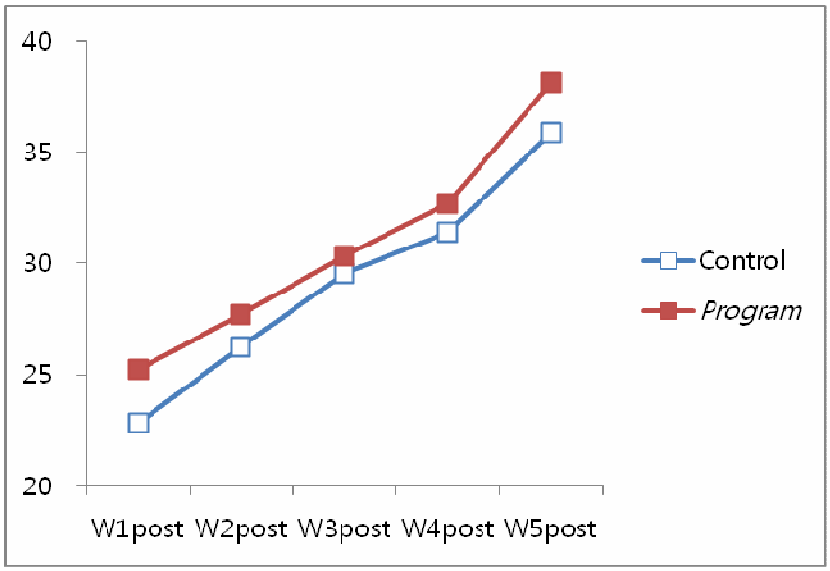

Figure 5. Total score on writing post-test according to time and group.

3.4.3. Results of Questionnaire

(1) "The error-correction program helped me to gain confidence in my writing skills" (because it allows self-editing and proofreading).

35.8% replied "slightly agree," 26.4% replied "slightly disagree," 24.5% replied "agree," 7.5% replied "disagree," and 5.7% answered "strongly agree." Most participants presumably felt confident in their proofreading because they could revise their compositions anytime and by using a convenient method.

(2) "I will use the error-correction program when I study writing."

The error-correction program's most significant function is its self-revision system. To use this system, students can submit their writing without help from their tutors and check their errors independently. Most students are satisfied with this tool and they appreciate the time they can

save while using it. To the statement "I will use the error-correction when I study writing," 22.6% replied "agree," 39.6% replied "slightly agree," 26.4% replied "slightly disagree," 5.7% replied "strongly agree," and 5.7% replied "disagree."

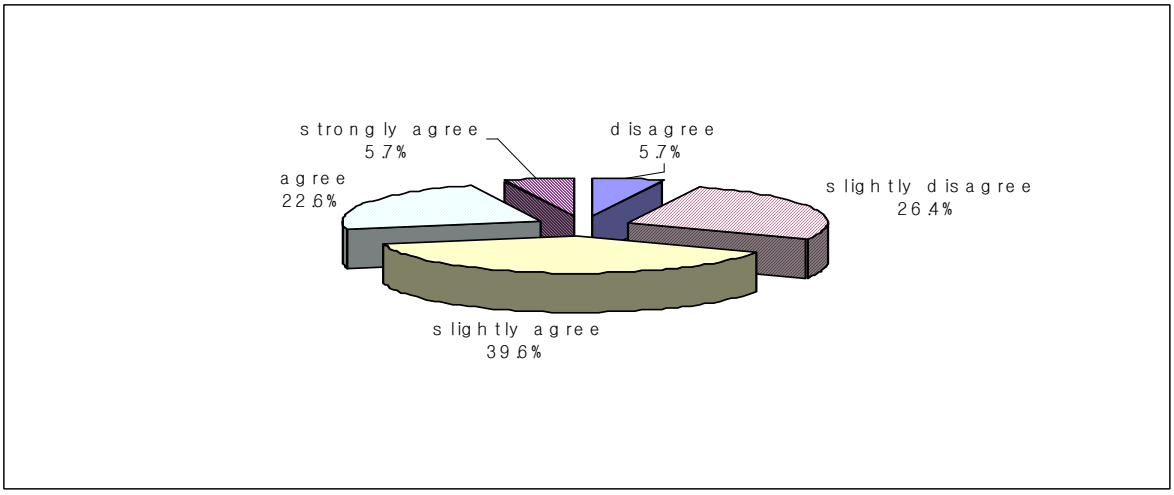

Figure 6. "I will use the error-correction program when I study writing."

4. Discussion and final conclusions

In Korea, English writing is among the least developed of the four main language skills. Korean students regularly perform poorly on cloze, grammar, and vocabulary tests. The level of a learner's writing skills are most likely closely related to his or her knowledge of grammar and vocabulary. This study especially analyzed the number of syntactical and lexical errors. The results demonstrated that the lexical errors were four times as frequent as the syntactical ones. An analysis of some of the more frequently incorrect sentence structures revealed that these often resulted in incomprehensibility.

This study was conducted to observe a detailed case study. The program was tested not on native speakers in English-speaking countries but on Korean L2 learners wishing to study English writing. It can be claimed that most students were satisfied with using the error-correction program while studying. Moreover, it was helpful to intermediate-level students while developing their writing skills; students were happy to be able to check their errors by themselves at any time.

This study was designed to meet its stated purpose. Despite proving the positive effects of the error-correction program, the study has its limitations. First, a small group and few samples were used in the test, and it was brief: a 75-minute class is not enough to cover a composition, so most participants had to do their writing as a home assignment. The writing test was also brief, about 50 minutes. Some participants could not finish it, and the unfinished tests had to be deleted from the results, which is why the sample was so small. Second, the expense of using the error-correction program tool was very high. Most essays contained fewer than 250 to 500 words, and some samples had few mistakes, though most participants were intermediate or lower. Future studies should use a larger participant group and more samples, follow a longer research period, and use participants with different language proficiency levels and tool use frequencies.

Moreover, this study suggested that a well-designed grammar- and vocabulary-based syllabus is probably necessary for teaching English writing and especially to use appropriate technology for teaching will be helpful. In some cases, a comparison between Korean and English may offer helpful insights into the different features of the two languages. Such a comparison may help students improve the accuracy of their writing through practice. This study suggests that a writing class that incorporates the error-correction program can help students obtain the practice they need, and, similarly to what Salomi (2008) suggested, a hybrid of online test for writing class would be necessary nowadays.

References

Arms, V. M. (1985). The computer: An aid to collaborative writing. *Technical Writing Teacher*, *11*(3), 181–185.

Attali, J., & Burstein, J. (2006). Automated essay scoring with E-rater V.2. *The Journal of Technology, Learning and Assessment*, *4*(3), 3–30.

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

Bardine, B. A., Bardine, M. S., & Deegan, E. F. (2000). Beyond the red pen: Clarifying our role in the responses process. *English Journal*, *90*(1), 94–101.

Breen, P. (2005). Two examples of CALL use in the classroom (electronic version). *Asian EFL Journal*, *6*(2), 14-23.

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.

Cho, D. W. (2001). A study on a Web-based English composition class. *English Teaching*, *56*(2), 287–307.

Finch, A. E. (2004). Online peer-assessment of a multimedia project. *Multimedia-Assisted Language Learning*, *7*(2), 65–80.

Folse, K. S., & Solomon, E. V. (2004). *Great Paragraphs: An Introduction to Writing Paragraphs with Chapters*. Boston: Houghton Mifflin Press.

Frizler, K. (1995). The Internet as an educational tool in ESOL writing instruction. In H. Y. Kim, (2002). L2 Pedagogical framework on Internet-Integrated Instruction. *Foreign Languages Education*, 9 (2), 279-302.

Goldberg, R. C. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment*, *2*(1), 1–52.

Graham, S., & Perrin, D. (2006). *Writing Next: Effective Strategies to Improve Writing of Adolescents in Middle and High Schools*. Washington, DC: Alliance for Excellence in Education.

Harris, K. R., Graham, S., & Mason, L. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal*, *43*, 295–340.

Kim, B. R. (2007). Using computer-assisted synchronous peer assessment program in the EFL formal presentation class. *Multimedia-Assisted Language Learning*, *10*(2), 44–69.

Levy, M., & Stockwell, G. (2006). *CALL Dimensions: Options and Issues in Computer-Assisted Language Learning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Salomi, P-S. (2008). A hybrid of a CBT- and a CAT-based new English placement test online (NEPTON). *CALICO Journal*, *25*(2), 276-304.

Warschauer, M. (2000). The changing global economy and the future of English teaching. *TESOL Quarterly*, *34*, 511–535.

Warschauer, M., & Kern, R. (2000). *Network-Based Language Teaching: Concepts and Practice.* Cambridge: Cambridge University Press.