Modern English Education, Vol. 17, No. 4, Winter 2016
http://dx.doi.org/10.18095/meeso.2016.17.4.07

The Effects of Scoring on Korean EFL Student Writing in Formative Assessment*

Sookyung Cho
Hankuk University of Foreign Studies

Cho, Sookyung. (2016). The effects of scoring on Korean EFL student writing in formative assessment. *Modern English Education, 17*(4), 139-157.

While previous research has usually discouraged teacher evaluation, such as scores, in formative assessment based on the result that it demotivates learners, these studies usually focus on end products of student learning, not taking into account the learning process. In order to provide a more comprehensive understanding of the effect of scores in formative assessment, this study investigated written products of 41 Korean learners of English and their revision. Students completed two writing assignments, including a first and second draft, receiving scores and feedback on the first draft of one writing assignment, but only feedback on the first draft of the second writing assignment. The comparison and contrast of these scored and non-scored writing assignments, revision, and perceptions reveal the following results: 1) while their scored writing assignments received higher scores in grammatical accuracy, they did not score higher in other areas; 2) students made more global level changes in scored assignments; and 3) students reported that scores motivated them to aim for higher grades, and that analytic scores in four areas (topic, organization, elaboration, and grammar) provide them with guidance for revision. These findings imply that the effects of scores in formative assessment depend on a variety of factors, such as their format and context, and that writing teachers can take advantage of scores to facilitate student learning.

**[formative assessment/scores/EFL student writing/
형성평가/점수/EFL 학생작문]**

* This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2014S1A5A8018236) and by Hankuk University of Foreign Studies Research Fund of 2016.

I. INTRODUCTION

Formative assessment has been proposed as an alternative to summative assessment (Cumming, 2001; Hamp-Lyons, 1994; Lee, 2007). While summative assessment tends to evaluate students' overall academic achievement based on their final results at the end of a course, formative assessment is usually conducted multiple times during a course to inform students of the gap between their current status and their target goal with the aim of helping students become more independent learners. In formative assessment, feedback is essential because students will be able to improve their abilities only through feedback, receiving suggestions, edits, and questions from readers (Hamp-Lyons, 1994). For this reason, formative assessment is often utilized in process-oriented writing classes where feedback—whether it comes from teacher, peer, or self—intervenes in students' process of writing at every stage. However, adopting formative assessment in a writing classroom poses a practical problem for writing instructors. While most scholars on formative assessment seem to agree that teacher evaluation, such as scores or grades, discourages student learning by focusing their attention on grades rather than on learning itself (Cizek, 2010; Elliot & Dweck, 1988; Harlen & Crick, 2003; Nicol & Macfarlane-Dick, 2004), the fact of the matter is that writing teachers are required by their institutions to evaluate their students' writing at the end of a semester.

Some conflicting results can be found on the effects of scores in formative assessment, depending on contexts and participants. For example, Olina and Sullivan (2002, 2004) studied the effects of teacher evaluation on their Latvian high school students and found that the teacher-evaluation group produced a higher-quality product as measured on a project rating scale than the no-evaluation group. It has also been shown that individual differences seem to affect student performance (Butler, 1988). More importantly, the majority of these studies focus on students' final outcomes (e.g., projects, exams, writing assignments) to determine whether scores have positive or negative effects and do not consider students' learning process or progress throughout a course. Therefore, to offer a more comprehensive examination of the effects of teacher evaluation in formative assessment, this study explores how scoring affects Korean learners of English in a process-oriented writing classroom by comparing their scored and non-scored writing assignments in terms of their revision as well as their writing performance and perceptions.

II. SCORING IN FORMATIVE ASSESSMENT

1. Formative Assessment and Writing

In an attempt to integrate assessment with classroom learning, formative assessment has attracted attention from both teachers and scholars. Since Michael Scriven first used the term "formative assessment" (Scriven, 1967), it has been widely used as an alternative to the conventional evaluation approach, so-called "summative assessment," typically conducted at the end of a course. Formative assessment has been believed to perform both functions of fostering learning and assessing by identifying the gaps between students' current level and the target level while instruction is on-going (Black & Wiliam, 1998, 2003; Wiliam, 2010, 2011). Feedback plays an essential role in providing information about the gap for the purpose of altering and eventually reducing it.

Viewing feedback as an essential part of instruction, writing scholars have also argued for the usefulness of formative assessment during process-oriented writing instruction, which aims to improve students' writing by intervening in their writing process by giving diverse forms of feedback (Cumming, 2001; Hamp-Lyons, 1994; H. Y. Kim, 2016; Lee, 2007). Indistinguishing assessment for learning (AfL) with assessment of learning (AoL), Lee (2007) has emphasized the necessity of adopting formative assessment in writing instruction, particularly in English as a foreign language (EFL) contexts, where writing assessment often focuses on students' final outcomes. In order to implement formative assessment, Lee argues that teachers should "share learning goals with students" (p. 203) and that students should "understand the standards they are working towards" (p. 204). In the vein, Cumming (2001) argues that ESL/EFL writing instructors seem to adopt formative assessment as a way of planning their instruction in reference to groups of students as well as a way of keeping records of individual students.

Despite the benefits of formative assessment in writing classrooms, it is not easy for writing instructors to implement this approach for the practical reason that they must evaluate their students' writing at the end of the semester. Oftentimes, any form of evaluation, such as grading or scoring, has been discouraged in formative assessment because it is believed to raise the anxiety level of students and, as a result, demotivate them (Cizek, 2010; Elliot & Dweck, 1988; Harlen & Crick, 2003; Nicol & Macfarlane-Dick, 2004). After reviewing 18 studies on effects of assessment on student motivation, Harlen and Crick (2003) conclude that frequent marks or grades given to students can lower their motivation to learn. Similarly, Elliot and Dweck (1988) compared the effects of two different goals on student performance—performance goals where students try to receive good judgment and learning goals where students try to improve their own abilities—and state that teacher evaluation tends to make students focus on performance goals like

passing the test rather than learning goals.

In order to overcome the difficulty, writing scholars have suggested that writing instructors should separate their role as readers from their role as teachers by adopting different types of evaluation from the conventional one (Leki, Cumming, & Silva, 2008). For example, Hamp-Lyons (1994) recommends peer evaluation, self-evaluation, or portfolio as ways of interweaving assessment and instruction. For instance, it is argued that portfolio is advantageous because it enables both the teacher and the student to overcome the limited view of one-time tests by taking into account the learner's growth as a writer. On the other hand, Casanave (2004) suggests a writing project that leads students to create an outcome such as a research paper, a novel, or a play through multiple drafts. According to Casanave, these alternative ways of evaluating students enable students to improve their writing on their own and be responsible for their own learning.

2. Teacher Evaluation: Grades and Scores

The question of whether grading should be avoided at all costs in formative assessment is a difficult one to answer definitively. The findings seem to vary depending on the type of feedback given and the participants' individual differences. For example, Olina and Sullivan (2002, 2004) found some positive effects of teacher evaluation in comparison with self- and no-evaluation. In their 2002 study, Olina and Sullivan divided 189 Latvian high school students into three groups: 1) those receiving teacher evaluation, 2) those receiving teacher evaluation and participating in self-evaluation, and 3) those with no evaluation. After comparing these three groups on their final projects and attitudes toward the program, they found that the first two groups—the teacher-evaluation group and the teacher- and self-evaluation group—received better grades than the no-evaluation group, although the no-evaluation group had more positive attitudes toward the program. In the second study, Olina and Sullivan (2004) broaden the scope of their research by examining more students, separating the effect of teacher evaluation from that of self-evaluation, and conducting evaluations twice—in the stage of project design and before revision of the project paper. They randomly assigned 341 Latvian high school students from 16 classes taught by 8 different teachers into four different groups: 1) teacher-evaluation, 2) self-evaluation, 3) combination of teacher- and self-evaluation, and 4) no evaluation. Interestingly, they found group 1 and group 3 received higher ratings on their project, but that group 2 felt more confident in conducting a future experiment on their own.

It has been suggested that individual differences should be considered in the investigation of teacher evaluation in formative assessment. Butler's (1988) study is often cited as a seminal study that proves the negative effect of teacher evaluation in formative assessment. A closer look at the results of his study, however, demonstrates that the effects

of teacher evaluation depend on students' abilities. He classified students as high-achieving (22) and low-achieving (22), placing them into three different groups of comment-only, comment-plus-evaluation, and evaluation-only. Notwithstanding the finding that the comment-only group earned higher grades than the comment-plus-evaluation or the evaluation-only group, the study also shows that high-achieving students were not affected by teacher evaluation. As Black and Wiliam (1998) warn, individual variations such as differences in ability and age should be taken into account when investigating the role of teacher evaluation in formative assessment. In their meta-analysis of studies on formative assessment, Black and Wiliam (1998) state that "close attention needs to be given to the differential effects between low and high achievers, of any type of feedback" (p. 13).

In the same vein, this study attempts to explore how teacher evaluation, particularly scores, affect the writing and attitudes of Korean learners of English in a writing classroom by comparing scoring and non-scoring classes on their writing performance and perceptions. Thus far, studies on the role of scoring in formative assessment have focused more on students' final outcomes (Butler, 1988; Olina & Sullivan, 2002, 2004) or students' attitudes and perceptions toward the scored tasks (Elliot & Dweck, 1988; Harlen & Crick, 2003), rather than the process they engage in through experiencing formative assessment. Even Olina and Sullivan (2004), who asked their participants to revise their project, do not take into account how teacher evaluation affects students' revision. This study, therefore, aims to examine not only the final written products and perceptions of the Korean learners, but also their revision by answering the following questions:

1) How do scored and non-scored writing assignments differ in their quality?
2) How do scored and non-scored assignments differ in their revision?
3) How do students perceive scoring in a writing class?

III. METHOD

1. Participants

The participants were 41 university students who were enrolled in two writing classes taught by the researcher on Monday and Tuesday in the same semester. Out of the 52 students enrolled in the two classes, these 41 students (17 in one class and 24 in the other, 20 male and 21 female) agreed to participate in the study. They were 22.81 years old on average, ranging from 19 to 28. All students were majoring in English linguistics, meaning they were quite advanced language learners, which is also evidenced by their scores on various English proficiency tests: students scored higher than 100 on the TOEFL (Test of

English as a Foreign Language) and higher than 900 on the TOEIC (Test of English for International Communication).

2. Procedure

All the participants were taught based on exactly the same curriculum that aimed to develop students' academic writing skills (e.g., writing paragraphs and essays and using outside sources). The course lasted 16 weeks with four 4-week units focusing on the basics of paragraph writing, expository writing, argumentative writing, and comparison and contrast writing. During the first two units, the participants learned how to organize paragraphs, develop paragraphs to essays, support their opinions using outside sources, and revise essays based on the feedback. In the final two units, the participants were required to do research on their topics and submit two academic essays: an argumentative essay and a comparison and contrast essay, along with a list of all the references they used to support their thesis. For both essays, students wrote their first drafts in the first week, received teacher feedback in the second week, and revised and resubmitted the draft in the third week. The students in both the classes were not provided any specific training on how to revise their drafts in order to control its effects on revision.

In addition to teacher feedback, first and second drafts of the two essays were collected from all the participants. In order to control differences inherent in the two groups and to mainly focus on the effects of teacher scoring, both groups received the same kind of teacher feedback across the two drafts and scores for one of the two essays, and only teacher feedback for the other. For example, Class 1 received teacher feedback and scores on their argumentative essays, but only feedback on their comparison and contrast essay, while Class 2 received only teacher feedback on their argumentative essay, but teacher feedback together with scores on their comparison and contrast essay (see Table 1 for details).

TABLE 1

Data Collection Procedure

	Class 1 (n = 17)	Class 2 (n = 24)
Argumentative Essay		
Week 10	First Draft	First Draft
Week 11	Teacher feedback with Scores	Only Teacher Feedback
Week 12	Second Draft	Second Draft
Comparison and Contrast Essay		
Week 14	First Draft	First Draft
Week 15	Only Teacher Feedback	Teacher Feedback with Scores
Week 16	Second Draft	Second Draft

Teacher feedback is provided in written form and focused on four main areas: topic, elaboration, organization, and grammar. Topic addresses the interest level of the topic, that is, whether it is interesting enough to attract readers' attention; organization refers to whether the essay has essential components, including an introduction, an appropriate number of supporting points, a conclusion, and references; elaboration addresses the development of logic and evidentiary support, and whether the details are persuasive enough to support the thesis; and grammar refers to accuracy in language use and appropriate use of vocabulary for academic papers (see Appendix 1 for an example). Both scored and non-scored feedback was identical except for analytic scores given for each category. Five-point scores were given for topic, organization, elaboration, and references, totaling to a possible 20 points.

At the end of the course, a questionnaire was distributed to understand students' perceptions about the scores they had received during the semester. The participants were asked to provide basic demographic information (e.g., age, gender, and test scores on certified English proficiency exams), and to share their thoughts about the scores they received on their first draft, including whether or not they were helpful in their revision, and if so, in what respect.

3. Data Analysis

Collected writing drafts were analyzed in their quality and the types of revisions. First, in order to understand the effects of scoring on writing quality, the participants' second drafts were graded. Two Korean raters with several years experience teaching English writing at a university level graded each participant's writing assignments separately, and average scores were calculated for each student. An analytic scoring rubric modified from Hyland (1996) was used, consisting of 5 possible points for each of the four main areas (topic, organization, elaboration, and grammar) (see Appendix 2 for details). 5 refers to "very satisfied," 4, 3, 2 refer to the status of "satisfied," "OK," and "dissatisfied" respectively, and lastly 1 means "very dissatisfied."

Second, to examine the student revision, first and second drafts were analyzed based on K. Cho and MacArthur's analysis scheme (2010). Drawing upon Faigley and Witte (1981) and Sommers (1980), K. Cho and MacArthur analyzed student revision in four areas: surface change, micro-meaning change, macro-meaning change, and reference. As seen in Table 2, surface change refers to mechanical changes in spelling and grammar, such as simple word changes and deletion or addition of one or two words; micro-level meaning change refers to correction of points at the sentence- or paragraph-level or elaboration of an existing point or example which does not alter the argument of the essay; macro-level meaning change refers to the addition of new points or major changes in organization

which affect the argument of the essay; finally, reference refers to any addition of evidentiary support or change in the existing supporting materials. The same two raters who graded the writing assignments compared the first and second drafts of both writing assignments for all participants and counted the frequency of each type of revision. The revision involving the same theme or point is counted as one idea unit. The raters reached agreement almost 89 percent of the time. When disagreement occurred, the two raters discussed until they arrived at the agreed-upon frequency count on the particular writing.

TABLE 2

Revision Analysis Scheme

Category	Definition	Examples
Surface Change	Mechanical changes in spelling and grammar; simple word change; deletion or addition of one or two words	Original: South Gyeongsang Province withdrew it for the first time. Revised: South Gyeongsang Province withdrew *free meals* for the first time.
Micro-Level Meaning Change	Correction of points in a sentence or a paragraph level; elaboration of an existing point or example	Original: gaining certificate or doing things to get so-called good 'spec Revised: gaining certificate or doing *beneficial activities* to get so-called good 'spec
Macro-Level Meaning Change	Addition of new points; changes in organization	Original: Lotte world II should not re-open because it threats safety of people and increases inconvenience. Revised: Lotte world II should not re-open *before correcting all the problems.*
Reference	Addition of supportive materials; change in the existing supporting materials	Original: Compared to Korea, American teaching method is "inquiry or discovery based learning." Revised: Compared to Korea, American teaching method is "inquiry or discovery based learning."*Barrow (265) states that inquiry teaching strategy was introduced to United States by John Dewey in 1910.*

Note. Italics indicate the revised parts in the second draft.

Finally, the answers given on the questionnaires were transcribed so they could be compared across participants. As recommended by Leki (2006), all answers were carefully read multiple times until repeating patterns were identified. Once potential themes or categories emerged, the raw data were read again to double-check whether any divergence existed between their answers and the themes. In those cases, additional notes were made and reserved for further analysis.

IV. RESULTS

1. Achievement: Student Writing

The average scores for each student were calculated and grouped into scored and non-scored assignments. As seen in Table 3, scored assignments received higher scores in all four areas (i.e., topic, organization, elaboration, and grammar). The standard deviations show that the scored assignments seem to have received a wider range of scores than the non-scored assignments in organization and elaboration, while non-scored assignments received a wider range in topic and grammar. Table 4 shows the results of the MANOVA test, used to assess whether the differences between scored and non-scored assignments were statistically significant, and demonstrates that scored and non-scored assignments differ significantly only in grammar. That is, while scored assignments were more grammatically accurate than non-scored assignments, their differences in topic, organization, and elaboration are not statistically significant, despite the fact that scored assignments received higher scores than non-scored assignments.[1]

TABLE 3

Descriptive Statistics for Writing Quality

Group	*n*	Topic		Organization		Elaboration		Grammar	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Scored	41	4.1707	.38095	3.9268	.59750	4.0000	.50000	4.0122	.53019
Non-Scored	41	4.0732	.42663	3.8171	.47112	3.8902	.49386	3.6171	.62004

TABLE 4

MANOVA Results of Scoring Effect on Writing Quality

Dependent Variable	*SS*	*df*	*MS*	*F*	*p*
Topic	.195	1	.195	1.193	.278
Organization	.247	1	.247	.853	.358
Elaboration	.247	1	.247	1.000	.320
Grammar	3.200	1	3.200	9.618	.003*

* $p < .05$

2. Student Revision

The results of the four revision categories—surface change, micro-meaning change,

[1] Genre did not make a difference between scored- and non-scored writing assignments in a statistical testing, and thus this study presents only the results of the statistical testing that compares and contrasts scored- and non-scored writing assignments.

macro-meaning change, and reference—are listed in Table 5. The mean frequency of the four categories shows that participants attempted more macro-meaning changes when they received scores on their first drafts, but more surface- and micro-meaning changes when they received only feedback. The participants also seemed to make a few more changes in their references in their scored assignments (3.9429) than in their non-scored assignments (3.2571), but the difference is not great. In order to see whether these differences are statistically significant, matched pair *t*-tests were conducted on each category (see Table 5). Out of the four categories, only surface change and macro-meaning change differ statistically between scored and non-scored assignments. That is, the participants were likely to attempt more surface-level changes in their non-scored assignments, but they tended to make more global-level changes, such as macro-meaning changes, when they received scores on their writing assignments.

TABLE 5

Descriptive Statistics for *t*-test

Revision Codes	Scored		Non-Scored		*p*
	M	*SD*	*M*	*SD*	
Surface Change	7.5278	4.17808	11.0556	7.18309	.005*
Micro-meaning Change	11.2432	7.38394	13.2162	7.45779	.253
Macro-meaning Change	11.4412	7.83617	7.9412	6.85760	.027*
Reference	3.9429	2.82783	3.2571	2.42951	.211

* $p < .05$

For example, Student 10, who did not receive scores on her argumentative essay about the issue of opening university libraries to the public, made primarily surface changes (8) and micro-meaning changes (25) in comparison with macro-meaning changes (2).

> First Draft
> Last year, one civic organization (Gwangju civil group for a society without an academical cliquism) submitted a petition to the three national universities (Seoul National University of Education, University of Seoul, and Gwangju Institute of Science and Technology). The civic organization claimed that the universities should open their libraries for the local residents (Jeong, 76). Recently, opening the university library for the local residents is a hot issue. While students worry about being destroyed their study environment, local residents claim that their right to study is also important. In this situation, I want to support the students since students' right to study can be violated, expansion of the public libraries should be taken first, and the establishment of the original function of university library can be impeded.

Second Draft[2]

Last year, one civic organization (Gwangju civil group for a society *without an educational background*) **(Mic.)** submitted a petition to the three national universities(Seoul National University of Education, University of Seoul, and Gwangju Institute of Science and Technology). The civic organization claimed that the universities should open their libraries to the local residents (Jeong, 76). Recently, opening the university library to the local residents is a hot issue. While students *are worried* **(Surf.)** *that their libraries may be destroyed*. **(Mic.)** local residents claim that their right to study is also important. In this situation, I want to support the students since students' right can be violated, *and the confusion can be caused*. (**Mac**.) *Furthermore* (**Surf.),** the establishment of the original function of university library can be impeded.

As can be seen, from her first to her second drafts, student 10 attempted two surface-level changes (change of voice from active to passive and addition of a transition word) and made a macro-meaning change that does not affect the overall argument of the essay. After receiving the teacher feedback that asks her to check whether her second supporting point, "expansion of the public libraries should be taken first," relates to the topic, student 10 drops it and adds a new supporting point, that is, "the confusion can be caused" to her second draft. As the result of this revision, student 10 received 1.5 higher scores (16/20) than her first draft (14.5/20) from the graders.

However, when she received a score along with teacher feedback on her comparison and contrast essay on differences between British English and American English, student 10 attempted many more macro-meaning changes (36 times) in comparison with surface changes (2 times) and micro-meaning changes (2 times).

First Draft

In the age of globalization, the number of English speaking countries that Korea has cultural and educational interaction is rapidly increasing. However, contrary to the expansion of the range and variety of exchange, Korea's English education is placed too much emphasis on teaching American English. Most of Korean students are much worse in understanding British English than they do in understanding American English. They are unfamiliar with British way of pronouncing words and have difficulty recognizing the words

[2] Changes are indicated in italics and underlines. The bold letters in parentheses indicate a type of revision made. Surf. refers to surface change; Mic. refers to micro-meaning change; Mac. means macro meaning change.

used only in British English (park, 151)[3]. However, to interact with people who feel more at ease with British English such as British, Australians, and Indians, it is important to familiar with British English. Knowing the differences between British English and American English can be helpful to be familiar with it. There are various differences between them such as cultural, linguistic, and historical differences. Among these differences, I will explain the linguistic differences presenting three main points which are spelling, vocabulary, and pronunciation.

Second Draft

When British and American talk together, there is no difficulty communicating each other. They can understand each other completely. **(Mac.)** *However, even though their language is same, there are some subtle differences between British English and American English.* **(Mac.)** *As a student who is majoring English linguistics, I really want to study what those differences are, and the reason why such differences are appeared.* **(Mac.)** *Historically, language have been affected by the cultural·geographical·social backgrounds.* **(Mac.)** *English is the same.* **(Mac.)** *The differences of history and culture made some linguistic differences between British English and American English.* **(Mac.)** *In next paragraphs, I am going to explain the origin of both English, differences of cultural atmosphere, and linguistic differences derived from them.* **(Mac.)**

Although student 10 received the similar kind of teacher feedback—work on her introduction—to the feedback she had received on her first writing assignment along with scores (3.5/5), student 10 made primarily macro-meaning changes by discussing differences in aspects (origin, cultural atmosphere, and linguistic differences derived from them) different from that of her first draft (linguistic differences only). Although student 10 received lower scores on her comparison and contrast essay than her argumentative essay, her score on the second draft increased two points (14.5/20) from the first draft (12.5/20).

3. Student Attitudes Toward Scoring

The results of the questionnaire conducted at the end of the semester demonstrate that the majority of students seem to have positive attitudes toward receiving scores. Out of 41 participants, five students did not evaluate the scores as having been helpful, and two

[3] These parentheses refer to the use of outside sources in the student drafts.

students responded that they were “disappointed” or “depressed” when they received scores. However, those who received high marks on their drafts had positive reactions to receiving scores. More surprisingly, even those who received some of the lower grades reported positive attitudes toward their experiences of receiving scores. While scores are likely to affect high achievers by boosting their “feel-good emotions” like confidence, satisfaction, and pleasantness, scores are found to serve major two functions for the rest of the participants: they serve as motivators and as guidance for further revision.

1) Scores as Motivator

Out of 41, 13 participants mentioned that the scores they received motivated them to work harder. They described the feelings when they received scores with the words “stimulate, motivate, shock, initiate, desire, or anxiety.”For example, student 2 said, “scores let me know my relative status in comparison with the other students and motivated me to work hard.” Student 30 also said, “scores stimulated me to work harder and made me think that I will be able to receive better scores. Without scores, I would feel at a loss what to do.” Student 34 also evaluated scores as being helpful because “lower scores than I expected initiated me to revise more rigorously in order to receive better scores.”

One interesting finding is that nine participants (about 22 percent of the participants) viewed teacher evaluation as being more objective and accurate when scores are reported, even though they received the same kind of feedback on their non-scored assignments. When referring to scores, they often used the terms “accurate, objective, detailed, frank, or visible,” meaning they seem to consider scores as accurate, objective, or concrete indicators of their writing abilities. For example, student 6 stated, “scores were helpful because they identified how writing ability is and what I need to develop more with objective figures, which stimulated me to [work harder].”According to student 22, “scores made me more anxious and desirous for better scores by presenting him with objective goals.” Similarly, student 41 mentioned that “scores made me think that I’m evaluated based on some objective standards, which stimulated me to work harder.” These students are likely to think that scores are correct and objective indicators of their current level of writing ability.

2) Scores as Guidance Toward Revision

The second function the participants frequently attributed to scores is that the scores helped guide them in revision. In particular, analytic scores, rather than holistic scores, seem to make scores look favorable to the participants by informing them which areas they

need to focus on in their revisions. Out of 41, 20 participants used phrasing such as "scores given on four different areas," "weak area," "what to improve," or "what to correct" when referring to the helpfulness of the scores they received on their first drafts.

> Scores on each area, not just total scores, helped me to find out in which areas I am weak. (Student 10)

> Scores are helpful in that they provide scores for individual areas. If the total scores were given to me, they might be very disappointing to me. Categorized scores help me to focus in revision on the areas where I received low grades. (Student 19)

> The scores were certainly helpful. The detailed scores given on each area let me know which area I need to pay more attention to while revising the draft. (Student 23)

As can be seen in the responses of students 10, 19, and 23, the participants seemed to favor partitioned scores over the overall score, as they understood that these scores indicate which areas were weaker and where they should concentrate in their revisions. In particular, they consider the scores helpful because they were analytic, not holistic ones. Student 26 noted that "the bracketed scores are helpful. If overall scores were given, they might disappoint me. But analytic scores given on different areas helped me to revise the draft focusing on the weak areas." As demonstrated here, overall, the participants seem to interpret low scores in a certain areas to be their weak points, which, in turn, led them to spend more energy and time revising the particular area in their second draft.

V. DISCUSSION AND CONCLUSION

The analysis of participants' written drafts reveals that while giving scores along with teacher feedback does not positively affect the quality of the topic, organization, and elaboration, it does positively affect the writing quality in terms of grammar: their scored assignments received better grades in grammar than their non-scored counterparts. This result may relate to the fact that grammar feedback is one of the areas they could easily incorporate into their revision in comparison with other types of feedback. Receiving scores also seems to have affected their revision, given that the participants made more global-level changes (i.e., macro-meaning changes) on their scored assignments, but more local-level changes (i.e., surface changes and micro-meaning changes) on their non-scored

assignments.

These differences between scored and non-scored writing assignments may be associated with their perceptions of scores, demonstrated through their responses to the questionnaire. Most participants perceived scores as having been helpful by motivating them to push themselves harder in order to achieve higher scores and by guiding them toward what to revise. It is quite interesting that while these participants received similar kinds of feedback both in scored and non-scored writing assignments, it is only when scores were given that they considered teacher feedback as being more accurate, honest, objective, and visible. Many students responded that by comparing the analytic scores on the four areas, they were able to identify which areas need more focus in their revisions. This positive perception of scores is quite the opposite to the findings of the previous research on the role of scores in formative assessment (Cizek, 2010; Elliot & Dweck, 1988; Harlen & Crick, 2003; Nicol & Macfarlane-Dick, 2004). Cizek (2010) and Nicol and Macfarlane-Dick (2004) thought that scores discourage students and do not contribute to student learning. These differences may originate from different contexts and cultures where students are located. The Korean participants of this study have been trained to understand success through scores in the highly competitive atmosphere. So far they have been quite successful achievers in this testing culture solely based on scores, and thus, it seems that scores stimulated them to put more effort on their second draft. According to Y. G. Cho (2015), most Korean university students seem to strongly believe that English learning is important, but their strong beliefs do not necessarily motivate their learning behaviors or attitudes. This study implies that scores may play an role of motivating these high achieving students to change their actual learning behaviors and attitudes.

Moreover, the results of this study implies that not merely presence of scores or not, but also their format, that is, how scores are given, analytic or holistic ones may affect learners' perception and performance. In this study, the participants' positive perceptions of scores are mainly drawn upon their experiences of receiving analytic scores, not holistic scores. Panadero and Jonsson (2013) and Jonsson (2010) argue that scoring rubrics can be used formatively—facilitating student learning by making learning goals transparent to them. Likewise, the participants of this study have the tendency to interpret the low scores in some areas as their weak points so that they can mainly put more efforts to revise those particular areas in the revision. Among the Korean learners of English who participated in this study, analytic scores seem to facilitate the original function of formative assessment, that is, enabling learners to identify their current level and place more efforts into achieving higher standards.

However, we must be cautious in interpreting these perceptions, not to generalize them to any types of scores, that is, scores in general or other populations. The participants of this study are very high-achieving students in terms of their English abilities, considering

that their English proficiency test scores are quite high, all of them are majoring in English linguistics, and all are college students at quite a prestigious university located in Seoul. As Butler (1988) did not find any negative effects of scores in formative assessment on high-achieving students in comparison with low-achieving students, high-achievers may be pushed harder to make better achievements when they receive low grades.

As this study shows, scores may perform different functions depending on the context and the characteristics of the population. If writing teachers must serve as both readers and assessors of student writing, the act of scoring should continue to be studied so that scoring can be utilized effectively for student learning instead of simply being avoided. Many areas of exploration exist: for instance, what types of scores should be provided to students in a writing classroom, how often scores should be given, or at what point in the writing process scores should be given to the students. Further studies are needed to investigate how scores should be incorporated into assessing students' writing performance to facilitate their learning.

REFERENCES

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Black, P., & Wiliam, D. (2003). "In praise of educational research": Formative assessment. *British Educational Research Journal, 29*(5), 623-637.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*(1), 1-14.

Casanave, C. P. (2004). *Controversies in second language writing: Dilemmas and decisions in research and instruction.* Ann Arbor: University of Michigan Press.

Cho, Kwangsu., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*(4), 328-338.

Cho, Young-Gyo. (2015). Students' attitudes, goal orientations and imagined success in L2 learning: A case study of Korea EIL context. *Modern English Education, 16*(4), 137-157.

Cizek, G. J. (2010). An introduction to formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 3-17). New York: Routledge.

Cumming, A. (2001). ESL/EFL instructors' practices for writing assessment: Specific purposes or general purposes? *Language Testing, 18*(2), 207-224.

Elliot, E., & Dweck, C. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54*(1), 5-12.

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication, 32*(4), 400-414.

Hamp-Lyons, L. (1994). Interweaving assessment and instruction in college ESL writing classes. *College ESL, 4*(1), 43-55.

Harlen, W., & Crick, R. D (2003). Testing and motivation for learning. *Assessment in Education, 10*(2), 169-207.

Hyland, K. (1996). *Second language writing*. New York: Cambridge University Press.

Jonsson, A. (2010). The use of transparency in the ‘‘Interactive Examination’’ for student teachers. *Assessment in Education: Principles, Policy & Practice, 17*(2), 183–197.

Kim, Hye Yeong. (2016). Changes in EFL students' perceptions and approaches to writing in a process-oriented writing class. *Modern English Education, 17*(1), 49-67.

Lee, I. (2007). Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *The Canadian Modern Language Review, 64*(1), 199-213.

Leki, I. (2006). “You cannot ignore”: L2 graduate students' response to discipline-based written feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing* (pp. 266-285). Cambridge: Cambridge University Press.

Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English*. New York: Routledge.

Nicol, D., & Macfarlane-Dick, D. (2004). Rethinking formative assessment in HE: A theoretical model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218.

Olina, Z., & Sullivan, H. J. (2002). Effects of classroom evaluation strategies on student achievement and attitudes. *Educational Technology Research and Development, 50*(3), 61-75.

Olina, Z., & Sullivan, H. J. (2004). Student self-evaluation, teacher evaluation, and learner performance. *Educational Technology Research and Development, 52*(3), 5-22.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revised: A review. *Educational Research Review, 9*, 129-144.

Scriven, M. (1967). *The methodology of evaluation*. Washington, DC: American Educational Research Association.

Sommers, N. I. (1980). Revision strategies of student writers and experienced writers. *College Composition and Communication, 31*(4), 378-387.

Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18-40). New York: Routledge.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14.

APPENDIX 1

Sample Teacher Feedback

Dear XX

Your argument is quite powerful and successfully supported by convincing evidence. If you work on the following things, your revised draft will look much more improved, I think.

Topic

Your argument is clearly stated, but I think you need to check whether all of your supporting points support your argument quite successfully. To me, the 3rd supporting point looks slightly different from the other supporting points and relatively more general.

Organization

Your essay is quite well-organized. But you need to introduce a counter-argument and rebut it.

Elaboration

1. Some of your sentences need elaboration. I've noticed a logical gap between sentences.
2. Don't forget to add your own interpretations explaining how the evidence relates to your topic.

Grammar

It is not clear what some of your pronouns refer to in the context. See the underlined parts.

APPENDIX 2

Analytic Scoring Rubric

Topic (5 points)

Appropriate topic for the assignment task; substantial concept use; good sense of audience; properly developed ideas; clear thesis statement

Organization (5 points)

Well organized and thorough development through introduction, body, and conclusion; message followed with ease; effective use of transitions and reference

Elaboration (5 points)

Relevant and convincing supporting details; logical progression of content contributes to fluency; unified paragraphs; use of appropriate and relevant outside sources

Grammar (5 points)
Effective use of correct sentences;; no significant errors in agreement, tense, number, person, articles, pronouns and prepositions; effective use of a wide variety of lexical items

Examples in: English
Applicable Languages: English
Applicable Levels: Tertiary

Sookyung Cho
Department of English Linguistics
Hankuk University of Foreign Studies
107 Imun-Ro, Dongdaemun-Gu, Seoul, 02450, Korea
Tel: (02) 2173-3194
Email: sookyungcho@hufs.ac.kr

Received 12 September 2016
Revised 24 October 2016
Accepted 11 November 2016