The Role of Scoring In Formative Assessment of Second Language Writing

Sookyung Cho
sookyungcho@hufs.ac.kr
Hankuk University of Foreign Studies, Korea

Chanho Park (corresponding author)
cpark.irt@gmail.com
Keimyung University, Korea

ABSTRACT

This study examines how scoring with feedback in formative assessment affects learning in an English as a foreign language (EFL) writing classroom. Two EFL writing classes were compared: in one class, teacher feedback was given to students on initial drafts, and scores were given only at the end of the semester; in the second class, teacher feedback and scores were given to students on each draft throughout the semester. This study adopted a mixed-methods approach, including a statistical analysis to explore whether teacher feedback accompanied by scoring makes a difference in student writing, and observation, and interviews of focal students to examine how feedback with scores affects students' perceptions and attitudes towards writing. The results reveal that the scoring class wrote more accurately than the non-scoring class and that the focal students in the scoring class were not only more aware of both their own and their classmates' performances, but that they also made efforts to emulate the students they considered effective writers. This study implies that scoring can fortify the effects of feedback by motivating high achieving students to do their best in their writing assignments.

Keywords: formative assessment; scoring; feedback; second language writing; EFL

INTRODUCTION

While scoring has usually been considered an unwelcome activity by both instructors and students, it is necessary in a classroom setting where grades must eventually be given. Compared to the more traditional summative assessment, conducted at the end of instruction to gauge student learning outcomes, formative assessment, defined as ongoing assessment with the aim of improving student learning through tailored teaching, is gaining popularity (Bloom, Hastings, & Madaus, 1971; Butler, 1988). While scores or grades are often used in summative assessment, they are often discouraged in formative assessment because they are thought to hinder learning by increasing learner anxiety.

However, in writing classes, in which instructors coach students to become increasingly effective and more independent writers, the situation is more complicated. Instructors must provide meaningful and constructive feedback to students to help them improve their writing skills, but they also need to assess student writing, as it is these scores that constitute the students' final grades. While several scholars suggest that instructors separate these two conflicting roles by postponing scoring as late as possible in the semester (Casanave, 2004; Hamp-Lyons, 1994), this study re-examines the effects of scoring and suggests the possibility of consolidating the original role of feedback, that is, to help students progress, with formative assessment. According to Wiliam (2010), studies that "identify more precisely the size of impact on student learning that can be achieved with formative assessment" (p. 37) are no longer helpful. He argues instead that future studies on formative assessment should "relate the kinds of feedback interventions to the learning processes they

engender" (p. 37). To that end, this study examines the learning processes of university students in an English as a foreign language (EFL) writing class in two feedback intervention groups: one group receiving both scoring and commenting and a second group receiving only commenting. A mixed-methods approach is adopted, including quantitative analysis comparing scoring and non-scoring groups in their writing ability and qualitative analysis of student perceptions of the impact of scoring on their learning processes.

SCORING AND FORMATIVE ASSESSMENT IN WRITING INSTRUCTION

Formative assessment has been put forward as an alternative to summative assessment in the context of writing classes. In performance-based formative assessment, feedback is considered an essential component in helping students close the gap between their actual level and their target level (Black & Wiliam, 1998; Wiliam, 2010). Likewise, as the process model has been more adopted by many writing instructors, feedback plays a key role in focusing on the development process of students as writers (Hamp-Lyons, 1994; Mansourizadeh & Abdullah, 2014). Cumming (2001) also points out the usefulness of formative assessment in writing classes by stating that "personalized focus on individual students seemed to prompt instructors to use formative assessment as a basis for record-keeping (in reference to individual students) and instructional planning (in reference to groups of students)" (pp. 215-216).

In spite of the promising results of formative assessment in the context of writing classes, however, the adoption of formative assessment can create conflict for the teacher. Although evaluation through grading has often been discouraged in formative assessment (Cizek, 2010; Nicol & Macfarlane-Dick, 2004), writing teachers serve as both readers and evaluators of student writing. For second language writing teachers to realize formative assessment in their classes, Leki, Cumming, and Silva (2008) suggest they may need to "separate their (a) assessor roles of evaluating students' texts critically from (b) their instructional roles of responding meaningfully to the ideas and content that students are attempting to convey in their written drafts" (p. 84).

One option suggested by scholars in the field of formative assessment is to pre-empt grading, at least until revision is completed. To this end, Hamp-Lyons (1994) recommends peer commenting, process logs (i.e., the exchange of ideas about a piece of student writing between students and teacher), and self-reflective commentary in which students analyze their own writing. As a means of formative assessment in an EFL writing classroom, Ghoorchaei, Tavakoli, and Ansari (2010) support the use of portfolio evaluation based on a collection of writing completed by students throughout the semester. Casanave (2004) introduces a writing project based on Sokolik and Tillyer (1992), in which students complete a final product, such as a research report, a novel, or a play, on one theme or topic across multiple drafts. Casanave (2004) argues that these alternative methods help students initiate improvement of their writing skills, and, as a result, build their autonomy as a writer by taking full responsibility for their learning.

However, the assumption that grading is detrimental to student development may not be unanimously applicable to all levels and types of learners. For instance, Butler (1988) is often cited in the writing literature to support the claim that grading undermines student interest level as well as task performance, but his high-achieving students' interest levels were not negatively affected by receiving grades. He examined the task performance of 22 low achieving elementary students and 22 high achieving students by dividing them into three groups: one group received only comments, a second group received both comments and grades, and the third group received grades only. The results revealed that performance of the comments-only group improved, whereas the comments plus grade group and the grades-only

group decreased across all the sessions on the tasks.

However, as Black and Wiliam (1998) more cautiously state, "close attention needs to be given to the differential effects between low and high achievers, of any type of feedback" (p. 13). In Butler's study (1988), high achievers maintained the same level of interest across the three sessions, whereas low achievers lost interest when receiving grades. Thus, it is important to note that the effects of grading may vary depending on student ability and aptitude. Moreover, students of varying ages and cultures, and across different subject areas may have different perceptions and attitudes towards grading. While Butler (1988) found that grading negatively affected elementary student performance, Martinez and Martinez (1992) found positive effects of frequent grading on the performance of 120 American college students in algebra tests. Therefore, different students will have different perceptions and motivational levels regarding the effects of grading.

Thus, this study addresses the issue of scoring in formative assessment. Although the effects of scoring have been found to be generally negative (Cizek, 2010; Nicol & Macfarlane-Dick, 2004), the effects may not be the same across all students. Of particular interest in this study is the effects of scoring for high achievers, since these students often have high motivation and self-confidence. Therefore, this study explores how scoring affects high achieving college students in Korea by comparing two intermediate writing courses taught by one of the researchers—one course implemented scoring throughout the semester and the other pre-empted scoring until the end of the semester. More specifically, this study adopts a mixed-method approach: Study 1 includes a quantitative analysis of students' writing ability, while a qualitative analysis of their attitudes and perceptions with regard to scoring is conducted in Study 2. Considering the complexity and the difficulty involved in conducting formative assessment, several scholars have emphasized the necessity to take into account its broader context, not just its quality and effectiveness (Black & Wiliam, 1998; Hampy-Lyons, 1990; Wiliam, 2010). Black and Wiliam (1998) state that "the effectiveness of formative work depends not only on the content of the feedback and associated learning opportunities, but also on the broader context of assumptions about the motivations and self-perceptions of students within which it occurs" (p. 17). To address formative assessment in the larger context, as recommended by these scholars, this study investigates student writing processes on one writing task. In particular, this study aims to answer the following research questions:

1. How does scoring affect student writing assignments?
2. How does scoring affect student perceptions and attitudes towards writing assignments?
3. To what extent are student revision processes and styles different between the classes?

METHODS

PARTICIPANTS

The participants are 32 first-year college students enrolled in one of the two classes—one class receiving scoring and written feedback on each paper and the second receiving only feedback. In order to take these intermediate writing courses, all students are required to complete a prerequisite course or to have earned scores exceeding 700 in TEPS (Test of English Proficiency developed by Seoul National University, Korea), which is equivalent to around 94 in the TOEFL iBT. The two courses followed the same curriculum and covered the same contents.

This study was conducted at the most prestigious university in Korea. To prevent grade inflation, this particular university required first-year liberal arts classes, including

writing classes, to follow a strict grading policy: A's should not be given to more than 20 percent of enrolled students, A's and B's should not be given to more than 80 percent, and at least 20 percent should receive grades below C. That is, 20 percent of students must receive a below-average grade no matter their performance.

In this context, receiving good grades is important to students for two reasons. First, all students were first-year students in engineering who needed to select their sub-major the following year, and their selection depended completely on the grades received in the first year. That is, in order to pursue their chosen area of study, they must receive a grade of A+, A0, and A- in most first-year courses. Second, high GPAs are believed to make university graduates more competitive in the Korean job market. As of 2012, Korea's unemployment rate was only 2.8 percent, but the unemployment rate of youths (aged 15 to 29) was almost three times as high as the general unemployment rate (Hwang, 2012). In such a competitive job market, a high GPA may not guarantee college graduates a decent job, but it is usually considered a required qualification (Phy, 2006).

PROCEDURE

The students in both classes were asked to complete four writing assignments: a one-paragraph text exhibiting logical division of ideas, a one-paragraph text explaining a process, a one-paragraph text of comparison and contrast, and an opinion essay. They submitted two drafts of each assignment online so that all students had the opportunity to read their classmates' writing assignments if interested, although it was not required. Between these two drafts, students received both teacher and peer feedback. Drawing on Conrad and Goldstein (1999), teacher feedback was concentrated in four areas: topic, elaboration, organization, and grammar (see Appendix A for a sample teacher feedback) as in Table 1. Peer feedback was provided by three to four group members, with the members changing across writing assignments so students could work with new peers (although it was possible some students may have worked with the same peer in different groups). After receiving feedback from both the instructor and peers, students revised their first drafts and submitted second drafts. It was closely checked by the instructor that the students did not plagiarize in any of their drafts.

TABLE 1. Categories of Teacher Feedback

Category	Definition
Topic	How suitable and interesting the topic and its contents are for the assignment
Elaboration	How successfully the students support their topic using concrete examples and detail
Organization	How well-organized the structure of the writing assignment is
Grammar	How accurately a writing assignment has been written

In the scoring class, along with feedback suggesting elements upon which the writer could improve, the scores of the second draft of each assignment were reported to students within one to two weeks of submission. While each first draft was not graded (for the purpose of encouraging students to actively engage in revision), the second draft was graded. Each second draft of each writing assignment was first reviewed in terms of the four areas mentioned above and ranked from 1 to 16 depending on the overall evaluation of the second draft. The scores ranged from 17.2 to 20 and were distributed evenly in accordance with the departmental grading policy. The 16 students in the non-scoring class, however, received only feedback on each of their first drafts, but no scores on their second drafts. Although all second drafts were graded using the same evaluation guidelines explained above, the students were only able to see their final letter grades (i.e., A, B) after the semester had concluded.

In order to understand better students' attitudes and perceptions between these two

classes, we also had exit interviews with them at the end of the semester (see Appendix B for interview protocol) and collected official records showing how many times each student had visited the writing center, as in both courses, the instructor encouraged the students to receive help by promising extra points to those who visited the center. Writing tutors—all graduate students with very high English proficiency (i.e., higher than 114 in TOEFL iBT)—were available from 9 am to 5 pm at the writing center.[1]

DATA ANALYSIS

Study 1 presents a quantitative analysis of student writing assignments and the number of their writing center visits. In order to understand whether scoring affected student performance on writing assignments, a rater who has taught the same writing course and is familiar with the student population scored the students' final product, the second draft of the opinion essay, on a five-point Likert scale—Very Good, Good, Neither Good nor Bad, Bad, and Very Bad—in the same four areas used by the instructor (i.e., topic, elaboration, organization, and grammar). A second rater independently analyzed the subset of data, and the researchers finalized the coding whenever discrepancies occurred between the raters. In order to control the initial difference in writing ability of the two classes, the first writing of these two groups—first drafts of logical division of ideas paragraph—were scored and compared using the same five-point scale.

Study 2 presents qualitative analysis of four focal participants' orientations and perceptions toward scoring and writing assignments through analysis of their experiences of writing center visits and the interview data. Audio-taped interviews were first transcribed and then reviewed several times as recommended by Leki (2006) to figure out "particularly salient or interesting comments as potential themes or categories to be cued against transcripts" (p. 270). We then compared the transcripts rigorously with the interview responses, "with straightforward responses tabulated and elaborations examined for themes and potential analytic categories to be correlated with themes and categories noted in the oral recordings" (p. 270). In addition, since this study compares the orientations and perceptions of four different participants, their transcripts were contrasted with one another so that we could examine any differences toward their perceptions of scoring and the writing assignments.

Lastly, we compared the four writing assignments of these four participants to see the extent to which each of them elaborated in revision. While we had intended to use a modified version of Cho and MacArthur's analysis (2010) to allow us to see types of revisions each of these four participants made in their revised drafts —such as surface changes, micro-level changes, and macro-level changes—we soon realized that some of their revised versions were so different from their first drafts that it was almost impossible to compare them. In their third and fourth writing assignments, for example, Jun and Jin (the participants) changed topics in the revised drafts and submitted completely different drafts from their first drafts. Therefore, instead of tracing the differences in revisions, we compared how the four participants incorporated teacher feedback into their revisions, in particular, the teacher feedback that they should deal with differences and similarities more in depth rather that at a superficial level.

[1] Students who want to visit the writing center sign up for available time slots on-line and post elements they want reviewed or edited on the site, and on the scheduled day, they visit the center and have an individual conference with the tutor for approximately 30 minutes.

RESULTS

STUDY 1

Table 1 shows the average scores of participants' second drafts of the final writing assignment (opinion essay) for scoring and non-scoring classes in the four areas: topic, elaboration, organization, and grammar. In order to test the differences in average scores between the classes, an analysis of covariance (ANCOVA) was performed. For ANCOVA, scores on the first draft of the first writing assignment (logical division of ideas) were used as a covariate to control for the initial writing ability of the participants. In order to control for the initial ability only on the same area as that of the dependent variable, instead of conducting a multivariate ANCOVA, four separate ANCOVAs were conducted with .0125 (=.05/4) as the nominal type I error rate for Bonferroni adjustment. That is, the first draft scores on topic, elaboration, organization, and grammar, respectively, were used as a covariate when testing the difference between the two classes on the final draft scores on each of the four areas.

TABLE 2. Mean Score by Class

	Scoring Class (16)		Non-Scoring Class (16)	
	Mean	**SD**	**Mean**	**SD**
Topic	3.87	1.13	4.44	0.96
Elaboration	3.86	0.95	3.56	0.51
Organization	4.07	0.83	4.25	0.68
Grammar	3.79	0.89	3.19	0.54

TABLE 3. Analysis of Covariance Summary

Variable	**Source**	**Sum of Squares**	***df***	**Mean Square**	***F***	**Sig.**	**Power**
Topic	Draft1	.40	1	.40	.38	.55	.09
	Class	.99	1	.99	.94	.34	.15
	Error	27.54	26	1.06			
Elaboration	Draft1	1.31	1	1.31	2.53	.13	.33
	Class	1.70	1	1.70	3.29	.08	.41
	Error	12.43	24	.52			
Organization	Draft1	.06	1	.06	.20	.76	.06
	Class	.24	1	.24	.40	.53	.09
	Error	14.34	24	.60			
Grammar	Draft1	.07	1	.07	.16	.69	.07
	Class	2.92	1	2.92	7.27	.01*	.74
	Error	10.04	25	.40			

* $p < .05$

Table 2 shows the results of the ANCOVA[2]. The non-scoring class received higher scores than the scoring class in topic and organization, but the differences were not statistically significant ($F(1, 26) = .94$, $p = .34$ for topic and $F(1, 24) = .40$, $p = .53$ for organization). For elaboration and grammar, the scoring class had higher average scores than the non-scoring class. The difference was statistically significant ($F(1, 25) = 7.27$, $p = .01$) with moderate to high power (.74) for grammar. For elaboration, although the difference was not significant ($F(1, 24) = 3.29$, $p = .08$), it was close to significance and statistical significance could be achieved if more subjects participated in this study with a larger sample.

[2] As the rater did not score two of the first drafts and one of the final drafts due to heavy plagiarism, and partially graded one first draft due to difficulty in deciding on the points, so the *df* varies across areas.

STUDY 2

The results of Study 1 reveal that scoring on a regular basis during the semester does contribute to better performance in grammatical accuracy and possibly in elaboration. In order to better understand how this improvement occurred at an individual level, additionally we conducted Study 2 that compares the attitudes towards and perceptions of the writing assignments as well as the actual revisions made in writing assignments across four participants: June and Jin from the scoring class, and Hyun and Min from the non-scoring class. These four participants were selected for the following reasons: 1) June and Jin improved in grammatical accuracy (from 3 to 4 and 2 to 5, respectively), while Hyun and Min did not (from 4 to 3 and 3 to 3, respectively); 2) June and Jin earned increasingly higher scores over the course of the semester, while Hyun and Min showed either a decrease or low scores; and 3) the improvement/non-improvement of these four students is more easily traceable and identifiable than that of the other participants due to such factors as the number of writing center visits, interviews, and revisions made in writing assignments. Table 4 shows changes in participant scores across writing assignments.

TABLE 4. Scores of writing assignments

Name	**Class**	**1st (rank)**	**2nd(rank)**	**3rd(rank)**
June	Scoring	17.4(13)	19.8(3)	19.2(5)
Jin	Scoring	18.2(9)	18.4(10)	18.6(7)
Hyun	Non-Scoring	19.4(4)	18.6(8)	17.6(15)
Min	Non-Scoring	17.2(15)	19.8(2)	18.2(11)

As students had not seen their scores on the final assignment previous to their interviews, they are not included in the table. As can be seen, while June's and Jin's scores and ranks increase, Hyun's scores and ranks continuously decrease, and Min's scores are consistently low except for the second assignment. In addition to these differences in grammatical accuracy, another differences are noticeable among the four focal students in reading other classmates writing, vising the writing center, and approach to revision.

PERCEPTIONS AND ATTITUDES

The four participants differ in their perceptions and attitudes towards their classmates' writing assignments. The interviews with June and Jin revealed that in addition to reading the assignments of their group members, they also read the writing assignments of other classmates. In fact, the major motivation triggering them to read these writing assignments was scoring.

> Because of *the low scores* I received, I started to read a couple of classmates' writing assignments, like Won and Keun who sit next to me. I did not read their second or third writing assignments, but I read their first ones to know how to write well.…After reading the other students' writing assignments, I understood what an essay should look like, such as what to put in an introduction, and how a conclusion should be formatted. Also, reading others' writing assignments is helpful in understanding how to use source materials. [3]

As can be seen in Table 3, out of 20, June received 17.4 points on his first writing assignment, which ranked thirteenth among 16 students. According to June, these relatively low scores

[3] The interviews were conducted in Korean and translated into English by the researchers.

prompted him to read the writing assignments of his two classmates who received the highest scores in the class: 20 for Won, and 19.6 for Keun. Through reading these writings, he felt he understood better what an essay should look like and was able to improve his draft. Jin, like June, also read other students' writing assignments, in particular, Keun's, since Jin had received a low score on his first writing assignment:

> I always read Keun's writing assignments, because he always received *high scores*. Why did he receive such *high scores* while I received *low scores*… I also read Hoon's writing assignments since Keun told me Hoon got very *high scores* and he really got *very high scores* again. Seeing that he marked very *high scores*, I came to think that if I do as he does, I will be better next time.

While Jin received 18.2 for his first writing assignment, Keun received 19.6 and Hoon received 18.8, which was slightly higher than Jin's. However, as Jin noted, after the first assignment, Hoon received higher ranking, 5th in the second writing assignment, 6th in the third assignment, and 4th in the final assignment. Although the teacher did not make public the students' scores or ranks, Jin happened to know that Keun and Hoon received higher scores than he and believed them to be the best students in class. Thus, he attempted to emulate their writing throughout the semester in addition to consulting tutors at the writing center.

While June and Jin read other classmates' writing assignments because of the low scores they received in the beginning of the semester, neither Hyun nor Min read the writing assignments of classmates other than their group members. They read each other's writing assignments even when they were not in the same group, not because of scores as we observed in June and Jin, but because of friendship. As Min stated:

> I just read my group members' writing assignments and my friend Hyun's. When we don't know something, we ask each other, like how my writing is and how your writing is, but we don't give feedback on it. Especially when we were stuck.

Since Hyun and Min did not receive score reports during the semester, their scores could not lead them to read the other students' writing assignments.

In addition to their attitudes towards other students' writing assignments, the number of writing center visits also shows differences in the amount of effort each participant put into his writing assignments. While Jin and June each reserved and attended appointments at the writing center multiple times, Min and Hyun did not attend, despite having each made one reservation a piece. Jin and June, having followed through with their appointments and having visited the writing center, stated that these visits were beneficial in helping them revise their drafts. Jin stated:

> I consulted tutors twice [while working on the third writing assignment]. As far as I remember, when I worked on the first writing assignment, the writing center was fully booked, so I was not able to visit there. But after that, I went there two or three times per writing assignment. I started going there because I felt that the first writing assignment, which I wrote without any help, was not good and was lacking many things….Looking at your comments and other things, I felt my writing needed to be more sophisticated, but I was too immature to do such writing. However, the tutors helped me with those things.

Unlike Jin, who found the writing center visits essential to revision, Hyun and Min failed to benefit from the visit. Although the official documents indicate that they had each visited the writing center one time during the semester, the interviews with them disclose a different story. When asked about his writing center meeting, Min stated, "I did not go to the writing

center—instead, I usually worked alone. I only took a look at your comments." To the same question, Hyun responded, "I made a reservation for the writing center this afternoon, but because of some unexpected schedule, I won't be able to go there. I have never been there before." Therefore, Min and Hyun only made reservations to visit the writing center, but did not turn up for their appointments.

REVISION

In Study 1, the category of elaboration shows some difference between scoring and non-scoring classes, although this result was not statistically significant. The close textual analysis of the four participants reveals that they are remarkably different in the extent to which they elaborated on a topic or theme in revision. Interestingly, on the third writing assignment, a comparison-and-contrast paragraph, all four participants received similar teacher feedback: they had addressed similarities and differences at a superficial level and needed to consider a common cause among these differences or similarities and relate them to one another. While June and Jin changed the organization of their paragraphs in order to incorporate this feedback into their final drafts, Min and Hyun focused more on sentence level issues without addressing organization. In his first draft, June compares and contrasts the bus and the subway because they are the most popular types of public transportation in Korea:

> Everyday, I commute to college by bus and subway. Considering that I am a college student who gets an allowance from parents, I cannot afford to use taxi everyday. So using the bus and the subway is an only way to go to college except parents' riding. In Seoul, most of citizen as well as I use the bus or the subway almost every day. In other words, the bus and the subway are the most famous sorts of public transportation. Although they are both fully utilized by the public, there are a few differences.

The few differences between the bus and the subway are then discussed regarding three aspects-how various their routes are, whether the passengers face the driver or not, and how punctual they are. As these three aspects are not closely related to one another, which resulted in superficial analysis, he was asked to consider a common cause among these three differences and to relate the other differences to this major cause. Two days later, June submitted the following revised draft:

> In the situation that you have to go to a strange place, firstly, you will search how to get there. If you don't have your own car, which method will you choose? Maybe most of people will go to the destination by bus or by subway. If method of using the bus and method of using the subway are both possible, which method is more efficient? Although the bus and the subway are both the most famous sorts of public transportation, there are a few differences. By looking at the differences between the bus and the subway, you can choose a more practical method suited for each situation.

As can be seen, in his revision, June chooses a clear focus for his comparison between the bus and the subway—efficiency—and, as a result, the introduction becomes more focused than his first draft. His analysis also has more depth, as in the revised draft, June discusses two major differences between the bus and the subway—spatial and temporal differences—due to the common factor that the bus runs on the road while the subway runs on the railroad. Spatially, the bus has more routes than the subway, and temporally, the subway is more punctual than the bus (see Appendix B for a full transcript).

Like June, Jin also made fundamental changes in his revised drafts. In his first draft, Jin compared and contrasted vampires and werewolves shown in movies or novels and

discussed three differences: 1) "werewolves are determined by nature and genetically rather than by their choice," 2) "werewolves have burning hot body and ebullient character," and 3) "werewolves are believed to be shape shifters due to either effect of the full moon or by their own choice." As in June's case, Jin was advised to first narrow his analysis of vampires and werewolves to, for instance, a specific movie, and second, to consider what relationship existed among the differences. In his revised draft, Jin combines the first two differences under the bigger category of difference in origin, and he classifies the third difference into a broader category, that of source of power:

> The fundamental differences between vampires and werewolves derive from their origin. In case of vampires, people who are bitten by other vampires become a vampire. Although their body is dead, they get power and immortal life. . . . On the other hand, werewolves inherited genetic characteristics of werewolf from warrior Taha Aki, a great ancestor of Quileute tribal, who was first werewolf. . . . The source of power is another fundamental difference between them [vampires and werewolves]. Werewolves are shape shifters as I mentioned above. They change their figure by rage and the metamorphosis makes werewolves powerful and fast. . . . On the other hand, vampires are strong without changing their shape. However, they need other source of power: blood. If they didn't drink enough blood, they would feel tired and they would be weak.

As can be seen, both June and Jin reorganized their writings to incorporate the teacher's comments into revision. However, Hyun and Min did not change their drafts notably. For instance, Hyun's first draft compares and contrasts two different types of digital cameras, charge-coupled device (CCD) and complementary metal-oxide semiconductor (CMOS) in three ways: sensitivity to light, complexity in manufacturing, and electricity use. As in the cases of June and Jin, Hyun was asked to find a common cause that brought about these differences. More specifically, the instructor suggested that Hyun recommend either CCD or CMOS for a certain kind of situation and then support his claim by explaining the differences between them. Based on this feedback, Hyun turned in the revised draft (italics mark changes made in revision):

> First, CCD is *highly sensitive* to light, so there are a few image noises-the random variation of *color information in images produced by the sensor*. It is generally regarded as an undesirable by-product of the picture-at pictures which is taken by CCD. Moreover, you can have *a* clear picture with CCD *although* there is *little light* like at *night*. In contrast, CMOS is relatively *less sensitive* to light, so you can see some more image noises. However, recently *the image noises* of CMOS have been reduced by *development in technology*. Second, CCD *requires a very complicated manufacturing process, so it is expensive.* The manufacturing process of CMOS, on the other hand, is relatively simple, so its price is pretty low. *Because of the simple manufacturing process and low price*, CMOS is usually used in many mobile devices such as cell phones. *Conversely, CCD can get more detailed image, so it is used in medical or scientific instruments.* Lastly, CCD uses up a lot of electricity and takes up much space, but CMOS offers lower power dissipation and takes up comparatively little space. These merits of CMOS also make this sensor more suitable for mobile devices, *so most compact digital cameras and even DSLR (Digital Single-Lens Reflex camera, professional photographers mainly use this) are using CMOS.*

Unlike June and Jin, Hyun did not re-conceptualize the differences, but made minor changes

in phrasing that were all teacher-initiated. In response to the suggestion in which he discussed which camera is more suitable for a certain situation and supported his claim by explaining the differences between the two, Hyun maintained his three points in the same order and included additional information about CCD and CMOS in his discussion of the second and third differences.

Min's revision is similar to Hyun's in that changes were made only at the surface level. In response to the teacher feedback that his topic and paragraph should be focused by either discussing similarities or differences between acoustic and electric guitars, Min added several new sentences and changed the original wording bit by bit as follows:

> *Most of people think acoustic and electric guitar are almost same, because they are same type of instrument as guitar. However,* there are several distinct differences between acoustic and electric guitars *for purpose of playing*. *First, they have some similarities that cause people have stereotypes*. The first similarity is that they make sound by vibrating the strings. . . . Second, as they make the string's sound loud, both use the guitar body for the neck to attach to and *frets (block in plate stringed instrument)* for finger replacement. Although these principles of two guitars are similar *and make people confused*, they are *distinguished severe parts. The greatest difference of two is about their desired sound. . . . The reasons of difference are introduced next part. . . .*

As to the teacher's request that he should focus either on similarities or differences, Min removed the word "similarities" from the original thesis statement "there are several distinct similarities and differences between acoustic and electric guitars," while repeating the same similarities of the first draft almost in the half of the whole essay. To the feedback that he should find a common cause leading to the differences, on the other hand, Min added the phrase "for purpose of playing" to the original thesis statement and inserted a new supporting point, that is, "the greatest difference of two is about their desired sound" in the middle of the essay, leaving the rest of the details as they were in the first draft.

While June and Jin re-conceptualized their argument to create a more thoughtful and connected comparison-contrast, which resulted in major changes in organization, Hyun and Min simply added information at the end or in the middle without re-thinking their arguments or organization, despite the similar type of feedback they received from the instructor.

DISCUSSION AND CONCLUSION

The quantitative analysis of the writing assignments (Study 1), reveals that the students in the scoring class did a significantly better job in grammar. Although the difference in elaboration was not statistically significant, the students in the scoring class were likely to elaborate on in a larger scale and in a more global level than those in the non-scoring class. Similar to high achievers in Butler (1988), the participants in this study are considered the best students in Korea. All are attending the most prestigious university in Korea and thus may be highly competitive, with a strong drive to succeed. This high-level of self-confidence may have motivated them to do their best to compete with their classmates and to succeed. That is, the scores reported to the students on a regular basis may have resulted in higher student motivation to improve their writing.

In addition to the difference in grammatical accuracy, Study 2 reveals that differences in attitudes towards writing assignments and efforts to improve writing skills existed between the two classes. The case studies of the four participants—June and Jin from the scoring class, and Min and Hyun from the non-scoring class—help explain how scoring in formative assessment can affect student learning, meaning their actual performance in

writing assignments, in this case. Scoring seems to tap into the three criteria essential for effective feedback suggested by Sadler (1989). According to Sadler (1989), students should "(a) possess a concept of the *standard* (or goal, or reference level) being aimed for, (b) compare the *actual* (or current) *level of performance* with the standard, and (c) engage in appropriate *action* which leads to some closure of the gap" (p. 121). Stimulated by the low scores they received from the teacher, June and Jin constantly related their current performances with what they considered the higher level performances of their classmates; however, Hyun and Min did not. In the cases of June and Jin, their initial low scores made them more willing and eager to read the writing assignments of other classmates, especially those they believed to have better grades than theirs, for the purpose of following their style or approach. On the other hand, given that Min and Hyun, from the non-scoring class, were not given scores during the semester, they may not have had the motivation to read the writing assignments of their peers other than their group members, which was required, or each other's.

These different levels of awareness led the four participants to different actions, i.e., the decision to visit the writing center, the amount of time and effort spent in revision, and the amount of emphasis placed on teacher feedback. In comparison with Hyun and Min, both June and Jin visited the writing center often, seeking help with almost all writing assignments. The more often the students visited the writing center and the more frequently they received help from the tutors, the more likely and more probable it was that their writing assignments were grammatically correct and logically developed. In addition to the difference in the number of writing center visits, the four participants also differed greatly in their response to teacher feedback. While all participants were asked to find a relationship that would connect their supporting points together, June and Jin interestingly worked at a more global level than Hyun and Min, by focusing their arguments and reorganizing their paper contents. We cannot come to the conclusion that these different styles of revision are caused solely by scoring, but it is highly probable that the willingness exhibited by June and Jin to revise their drafts and their initiation in making revisions themselves on areas not mentioned by the teacher could have been affected by their high level of score awareness, which then contributed to better performance in their final products.

Scoring is often believed to be negatively conceived of by learners, but this study provides new insight into the effects of scoring in formative assessment, especially student perceptions of scoring. Many studies on feelings of and attitudes toward scoring found that scores usually have a negative or neutral effect on students (Cheng, 1998; Shohamy, Donitsa-Schmidt, & Ferman, 1996). However, Spratt (2005) concludes that "exams' impact on feelings and attitudes seems clear but how these in turn impact on teaching and learning is much less clear" (p. 18). In response to her question of whether negative attitudes or feelings will necessarily bring about negative effects on learning and teaching, this study implies that scoring can encourage learners to become more fully responsible for their learning and can result in more and better learning. As Hamp-Lyons suggests (1994), "grades, whether on a single paper or at the end of term, are not an unwelcome surprise but simply the formal acknowledgement of what writer and instructor have known all along" (p. 54) when instruction and evaluation are interwoven, as in this study.

Future studies are needed to investigate whether scoring has the same type of positive effects on other learners' perceptions and attitudes and on their learning outcomes as was found here. The participants in this study are top students who are motivated, competitive, and focused, and who are accustomed to and familiar with the practice of scoring in Korea, where the job market has become increasingly competitive for college undergraduates. In such a competitive academia and society, scoring may raise student awareness of a gap existing between their current level and the target level and, as a result, motivate them to

exert more effort to make themselves more competitive. However, the positive attitudes and willingness witnessed in June and Jin may not be present in students from other cultural backgrounds. Students from a less scoring-oriented culture (even if they are competitive, motivated, and focused) may not react like the participants in this study. Therefore, future studies need to examine the effects of scoring across other learners and contexts.

ACKNOWLEDGEMENT

This work was supported by Hankuk University of Foreign Studies 2014 Research Fund.

REFERENCES

Black, P., and Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education. 5*(1), 7-74.

Bloom, B. S., Hastings, J. T., and Madaus, G. F. (Eds.). (1971). *Handbook of Formative and Summative Evaluation of Student Learning.* New York: McGraw-Hill.

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.

Cheng, L. (1998). Impact of a Public English Examination Change on Students' Perceptions and Attitudes toward Their English Learning. *Studies in Educational Evaluation. 24*(3), 279-300.

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.

Cho, K., and MacArthur, C. (2010). Student Revision with Peer and Expert Reviewing. *Learning and Instruction. 20*, 328-338.

Conrad, S. M., and Goldstein, L. M. (1999). ESL Student Revision after Teacher-Written Comments: Text, Contexts, and Individuals. *Journal of Second Language Writing. 8*(2), 147-179.

Cumming, A. (2001). ESL/EFL Instructors' Practices for Writing Assessment: Specific Purposes or General Purposes? *Language Testing. 18*(2), 207-224.

Ghoorchaei, B., Tavakoli, M., and Ansari, D. N. (2010). The Impact of Portfolio Assessment on Iranian EFL Students' Essay Writing: A Process-Oriented Approach. *GEMA Online® Journal of Language Studies. 10*(3), 35-51.

Hamp-Lyons, L. (1990). Second Language Writing Assessment. In B. Kroll (Ed.). *Second Language Writing: Research Insights for the Classroom* (pp. 69-87). New York: Cambridge University Press.

Hamp-Lyons, L. (1994). Interweaving Assessment and Instruction in College ESL Writing Classes. *College ESL. 4*(1), 43-55.

Hwang, S. H. (2012, November, 14). Korea's Jobless Rate Falls while Youth Unemployment Rate Rises. *Arirang News.* Retrieved from http://www.arirang.co.kr

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., and Silva, T. (2008). *A Synthesis of Research on Second Language Writing in English.* New York: Routledge.

Mansourizadeh, Kobra, and Abdullah, Khairi Izwan. (2014). The Effects of Oral and Written Meta-Linguistic Feedback on ESL Students Writing. *3L: Language Linguistics Literature®, Southeast Asian Journal of English Language Studies. 20*(2), 117-126.

Martinez, J. G. R., and Martinez, N. C. (1992). Re-Examining Repeated Testing and Teacher Effect6s in a Remedial Mathematics Course. *British Journal of Educational Psychology. 62*(3), 356-363.

Nicol, D., and 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.

Phy, S. (2006). *Graduate and Employment in the Republic of Korea and Cambodia.* München: GRIN-Verl.

Sadler, D. R. (1989). Formative Assessment and the Design of Instructional Systems. *Instructional Science. 18*, 119-144.

Shohamy, E., Donitsa-Schmidt, S., and Ferman, I. (1996). Test Impact Revisited: Washback Effect Over Time. *Language Testing. 13*(3), 298-317.

Sokolik, M., and Tillyer, A. (1992). Beyond Portfolios: Looking at Student Project as Teaching and Evaluation Devices. *College ESL. 2*(2), 47-51.

Spratt, M. (2005). Washback and the Classroom: The Implications for Teaching and Learning of Studies of Washback from Exams. *Language Teaching Research. 9*(1), 5-29.

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.

APPENDIX A

SCORE REPORT

Dear June,

Your final draft looks much more focused than your first draft. The following are my comments on your final draft.

Topic
Your topic looks much more focused than in the first draft, but I don't think all of your supporting points support it. Except for the first supporting point, the other two supporting points actually talk about the importance of travelling, not the necessity of travelling during the college years.

Elaboration
It is nice of you to put your own examples here. But there exists the same kind of problem pointed out in the above.

Organization
Nice organization except for the problem pointed out in the above.

Grammar
No major mistakes found in the draft, and I only made a couple of suggestions. Although your use of pronouns looks better this time, I noticed one case of mixing "you" and "we." See my suggestions!

Your total: 17.4/20

APPENDIX B

INTERVIEW PROTOCOL

1. Which of the drafts do you think you revised the most? Which of the drafts did you put the most effort into for revision?

2. What kinds of things did you try to focus on when you revised the draft?

3. What kinds of changes did you make in revision?

4. What kinds of comments were the most helpful to you both in peer and teacher feedback?

5. Did you usually read the feedback carefully?

6. Are you willing to make changes in your revision? Do you mind making changes on a large scale?

7. What do you think of revision? Do you think it is necessary?

APPENDIX C

JUNE'S REVISED DRAFT

In the situation that you have to go to a strange place, firstly, you will search how to get there. If you don't have your own car, which method will you choose? Maybe most of people will go to the destination by bus or by subway. If method of using the bus and method of using the subway are both possible, which method is more efficient? Although the bus and the subway are both the most famous sorts of public transportation, there are a few differences. By looking at the differences between the bus and the subway, you can choose a more practical method suited for each situation. The major difference is that the bus moves on the road of city while the subway moves on its own railroad. This difference causes spatial difference and temporal difference. First, spatial difference between the bus and the subway is a variety of route. To make a railroad, large amount of money and time are needed. Also, the railroad cannot be built unless there is a broad space. Therefore, the subway has only several routes, and the place you can go by subway is limited. On the other hand, the road is already located in every place of the city. So the bus has lots of routes, so you can go almost everywhere by bus. That is, you can go to the place by bus, though it's impossible to go there only by subway. Second, temporal difference between the bus and the subway is punctuality. The subway moves on its own rail, so the subway can exactly run on the timetable. In contrast, the bus moves on the road of city. For this reason, the bus' schedule is greatly influenced by traffic. Maybe you have experienced the situation you are late for class or office hour due to traffic jam, when using the bus. If you have important meeting or appointment such as an exam or a blind date, use the subway than the bus. In conclusion, the bus is better than the subway in the aspect of various routes, while the subway has more advantage than the bus in the aspect of punctuality. When you can go to your destination both by bus and by subway, regardless which method you prefer, considering the spatial and temporal differences will be helpful for you to choose which transportation to get on.

ABOUT THE AUTHORS

Sookyung Cho is an Assistant Professor at Hankuk University of Foreign Studies, Seoul, Korea. She is a graduate of the University of Wisconsin, Madison (Ph.D. Second Language Acquisition) and has a strong interest in second language writing, in particular, English language learners' perceptions and attitudes towards peer and teacher feedback.

Chanho Park, Ph.D. is an Assistant Professor at Keimyung University in Daegu, Korea. He holds an MA in ESL from the University of Hawaii at Manoa and a Ph.D. in educational psychology from the University of Wisconsin-Madison. His research interests include language testing, psychometrics, and quantitative research methods.