



Examining the Effectiveness of Learning-Oriented Language Assessment in Second Language Reading-to-Write Tasks: Focusing on Instructor Feedback and Self-Assessment

Hyun Jung Kim

Hankuk University of Foreign Studies, Korea

Ahyoung Alicia Kim

WIDA, Wisconsin Center for Education Research, University of Wisconsin-Madison, USA

Previous research on instructor feedback in second language (L2) writing has been limited to the area of *grammar*. However, it is also important to understand the role of feedback in developing *content*, considering that integrated reading-to-write tasks, which emphasize one's ability to discuss select topics, are gaining popularity in L2 writing. Using the Learning-Oriented Language Assessment framework, this study examined the effectiveness of instructor feedback in the content domain in reading-to-write tasks. It also compared instructor feedback with students' self-assessment of their own writings. Participants were 11 TESOL master's students and their professor in South Korea. During one semester, learners produced eight writings (initial draft and revision on four reading-to-write tasks). For each task, students read an academic research paper of their choice and wrote a one-page review of it. Learners also completed a self-assessment and received instructor feedback in the content domain of their writings. Quantitative and qualitative findings indicate that learners generally improved in the content domain. This could be attributed to repeated self-assessment, combined with instructor feedback, which promoted learners' metacognitive reflection of their writing process and product. The results provide pedagogical implications for using integrated reading-to-write tasks and sustained feedback in L2 academic writing.

Keywords: instructor feedback, feedback in content domain, self-assessment, integrated reading-to-write task, Learning-Oriented Language Assessment

Introduction

In an L2 academic writing context, it is fairly typical for instructors to provide feedback to their students' L2 writing. Students often expect instructor feedback, which is believed to be beneficial for enhancing students' L2 writing ability (e.g., Ashwell, 2000; Bitchener & Knoch, 2009; Ferris & Roberts, 2001; Sheen, 2007). In contrast to instructor feedback in first language (L1) writing settings, in which much emphasis is given to the *content* of student writing (e.g., Lin et al., 2001; Miller, 2003; Nelson & Schunn, 2009; Nilson, 2003), the focus of feedback is frequently placed on *grammar* or *language* in L2 settings (e.g., Han, 2002; Iwashita, 2003; Long et al., 1998; Mackey & Philp, 1998). Although there has been a plethora of research on the topic of grammar or *corrective feedback* in L2 writing (e.g., Han, 2002; Ji, 2015; Long et al., 1998; Plonsky & Brown, 2015), relatively few have examined how feedback in the content domain affects learners' writing development, especially in an L2 academic writing context.



It is critical to examine the role of instructor feedback in the content domain, considering that reading-to-write tasks, which emphasize one's ability to read and then write on a specific topic, are widely used in L2 writing settings (Gebriel, 2009; Hale et al., 1996; Moore & Morton, 1999; Plakans, 2009a). While independent writing tasks allow learners to demonstrate their ability to produce coherent written discourse by referring to their existing knowledge and experience, integrated reading-to-write tasks require more complex cognitive and language abilities. In addition to their written composition, integrated tasks tap into learners' comprehension abilities and their capability to apply the given source text appropriately and meaningfully (Cumming et al., 2006). For instance, in university-level courses, students are required to read texts and produce written reports, such as summaries or critiques. On such reading-to-write tasks, the instructor feedback will be geared toward the content rather than being limited to the grammatical accuracy of the writing.

Instructor feedback in classroom settings not only creates novel learning opportunities, but also functions as a type of classroom-based assessment. Assessments conducted in classroom settings with the purpose of promoting learning are often referred to as *formative assessment* or, more recently, *Learning-Oriented Language Assessment* (LOLA; Green & Hamp-Lyons, 2015; Hamp-Lyons & Green, 2014). Within the LOLA framework, learning can be maximized by bridging language instruction and assessment. It is necessary to connect the two because instruction alone does not necessarily lead to learning nor vice versa. The language assessment component in the LOLA framework can be administered by the teacher in the form of instructor feedback, but it can also include peer- or self-assessment. Due to its novelty up to now, however, few empirical studies have been conducted using the LOLA framework, and little is known about how the components of LOLA, including instructor feedback and student involvement, affect L2 development.

To this end, this study examined the effectiveness of LOLA in an L2 academic writing context. Specifically, the study examined the effectiveness of LOLA in the content domain of L2 academic reading-to-write tasks. It also compared learners' self-assessment with instructor feedback in terms of summative (numeric score) and formative information (written comments) to examine the effectiveness of student involvement in LOLA. While self-assessment was internally generated by the learners themselves, instructor feedback functioned as a type of classroom-based assessment of the students. For these purposes, the following research questions were examined:

1. To what degree is LOLA effective in improving L2 students' reading-to-write ability in the content domain?
2. What are the differences between students' self-assessment and instructor feedback in the content domain of L2 reading-to-write tasks?

Literature Review

Learning-Oriented Language Assessment Framework

Learning-Oriented Language Assessment or *Learning-Oriented Assessment* (LOA) refers to an assessment framework that bridges instruction with assessment. In the LOLA framework, the instructional aspect is crucial in promoting student learning (Kim & Yang, 2018). For learning to occur, individuals need instructional support. Therefore, teachers need to provide individualized feedback, preferably extended feedback for a long time period (Turner & Purpura, 2016). Students' learning is measured by drawing inferences from both their learning processes and outcomes. This assessment aspect requires understanding how students process the information they receive, for example, in the form of instructor feedback.

By drawing upon the theories that connect instruction and assessment (e.g., Carless, 2007; Keppell & Carless, 2006; Turner & Purpura, 2016), Hamp-Lyons and Green (2014) and Green and Hamp-Lyons

(2015) proposed three main elements of LOLA: (1) *learning-focused tasks* that simulate real-life language use, (2) student involvement in *self- or peer-assessment*, which promotes learner engagement, and (3) feedback as *feed-forward* that enhances learning. While the first LOLA element of learning-focused tasks creates a foundational context for LOLA, the second and third elements promote learning to occur. The second element of self-assessment is valuable, as it fosters self-regulation, autonomy, and motivation (Patri, 2002; Saito, 2008). While self-assessment is internally generated by the learner, the third element of feed-forward is externally provided, usually in the form of instructor feedback. Feedback, by definition, indicates the difference between the current and ideal state of learners' performance (Wiggins, 1993). Although these two elements of self-assessment and instructor feedback are key elements for learning to occur in the LOLA framework, there is little research on how their interaction affects the process and product of learning. The two elements are discussed further below.

Self-Assessment in L2 Writing Research

Self-assessment, by definition, is a type of self-regulated learning, which offers students an opportunity to set goals, monitor progress, and evaluate their own performance. This nature of self-assessment offers students a sense of control over their learning and may increase motivation (Paris & Paris, 2001). Self-assessments are implemented in classroom settings for varying reasons. For instance, self-assessment can function as a measurement tool in addition to a facilitation tool for learning (Boud, 1995).

The rising interest in self-assessment reflects a shift from teacher-centered to learner-centered instruction (Boud, 1995). Contrary to teacher-assessment, self-assessment offers students complete agency of assessment. Thus, self-assessments are generally believed to be effective for assessing L2 proficiency levels (Ganji, 2009; Ross, 1998). In an empirical study, Butler and Lee (2010) found self-assessment to have a positive influence on students' English performance, although the effect sizes were small. In addition, students were able to improve their ability of self-assessment over time. Context was important, as well. In an earlier study, Butler and Lee (2006) found that self-assessment was more effective when it was done after completing a specific task rather than without any context. Also, self-assessment after a task was less influenced by attitude/personality, in comparison to that of before or without a task.

Due to the unique characteristics of self-assessment, research suggests similarities, yet differences between students' self-assessment and other assessments, including peer- or teacher- assessments (Babaii et al., 2016). Studies have found higher correlations between self-assessment and teacher-assessment in receptive skills (e.g., listening, reading) than in expressive skills (e.g., speaking, writing; Ross, 1998). In another study, Matsuno (2009) examined how self- and peer-assessments worked in comparison to teacher-assessment in university writing courses. Self-raters often perceived their writing to be lower than it actually was. This was particularly true for high-proficient students than low-proficient ones. Interestingly, she found peer-assessments to be more consistent than other assessments, and therefore, more applicable to writing classes. These inconclusive results suggest the need for more research in self-assessment and its interaction with other types of assessment.

Feedback in L2 Writing Research

Previous research on L2 writing feedback has generally been limited to corrective feedback in the area of grammar or form (e.g., Han, 2002; Iwashita, 2003; Long et al., 1998; Mackey & Philp, 1998). In their extensive meta-analysis of research on the effectiveness of feedback for L1 and L2 writing development, Biber et al. (2011) found that feedback focusing solely on form (i.e., grammar) was less effective than feedback focusing on both content and form. In fact, gains in spelling accuracy decreased after receiving feedback. Likewise, outcome measure gains in content were small and not significant, but larger gains were attained for grammar and holistic measures (presumably reflecting both form and content). Although meta-analysis findings show that student made no significant gains in the content domain,

general training in the revision process may lead to improvements in the content area. When comparing outcome measures from receiving feedback, students displayed more gains if they were asked to revise their writings than receiving a single score.

In addition to the nature of the feedback, other variables (e.g., scope of feedback, students' writing proficiency) influence the effectiveness of feedback in L2 writing development. Biber et al. (2011) examined how the scope of feedback affects writing ability by comparing the general comments or explanation of specific errors versus simple error location. The findings indicated the former to be more effective than the latter. The effectiveness of feedback also depended on students' language proficiency. Students with lower proficiency made more gains in their writing ability based on feedback than those with higher proficiency.

Despite the plethora of research in L2 writing feedback, studies have generally focused on the role of feedback in grammar. Thus, it is questionable as to whether such findings can be generalizable for reading-to-write tasks in which content plays a crucial role, due to their heavy reliance on reading texts. Although there is existing research on reading-to-write tasks, such as learners' processes when engaging in reading-to-write tasks (e.g., McCulloch, 2013; Plakans, 2009a, 2009b), few have examined the role of feedback in the content domain in the L2 reading-to-write task context. Thus, very little is known regarding how instructor feedback affects students' ability to integrate source materials into their writing or how it is incorporated into learners' writing, suggesting the need for more research on this topic.

Methodology

Context of the Study

The study took place within a master's Teaching English to Speakers of Other Languages (TESOL) graduate program at a university in Seoul, Korea. The university is known for its leading role in foreign language/foreign language education. The TESOL graduate program aims to train students to become competent in English language teaching and content development. As part of the Research Methods course, students read empirical studies from academic journals in the field of TESOL, and summarized and critiqued them. While this reading-to-write task was integrated mainly to help students narrow down their research topic and find research gaps for their master's thesis writing, the course also aimed to improve students' academic writing ability, making it an L2 academic writing course. The medium of instruction was in English.

Participants

Students

The participants were 11 second-year graduate students in TESOL (10 females and one male) in their 20s and 30s. The students were generally fluent in English and had high proficiency in their L2 writing ability. However, they were not familiar with writing literature reviews or critiquing empirical studies because most courses in their first year focused on basic TESOL theories and practice, and did not require extensive reviews or critiques.

Instructor

The instructor was a middle-aged female professor with a Ph.D. in Applied Linguistics. She was familiar with LOLA and was enthusiastic in implementing it in her research methods course, which she

had taught for over five years. Through this study, she wanted to understand how effective the assessments were in promoting students' academic writing ability.

Instruments

L2 reading-to-write tasks

As part of the course, students were required to complete four reading-to-write tasks during the semester. For each task, they submitted both an initial draft and a revision, totaling eight writings. For the reading portion of each task, students were given the freedom to select a journal article (e.g., empirical research articles) on the topic of their interest, which was related to their anticipated master's thesis topic. In this sense, the tasks were relevant LOLA tasks, which were authentic and learning-focused. The writing portion of the task required students to produce a one-page review, consisting of a *summary* and *critique* of their reading.

Assessment sheet

The assessment sheet, which indicated both the student's self-assessment and instructor assessment, was completed for both the initial and revised drafts. As seen in Appendix, the sheet comprised three sections: (1) a holistic *rubric* used as a reference for rating students' reading-to-write performance, (2) the students' *self-assessment* of their performance, and (3) the *instructor assessment* of the students' performance. The rubric specifically focused on assessing the content domain of the students' writing (consisting of a summary and a critique), and ranged from a score of 1 (very unclear) to 5 (excellent). The criteria of the summary section focused on the clarity and meaningfulness of the elements of *purpose*, *procedures (methods)*, and *findings*, discussed in the journal articles the students read. These elements are typically important in comprehending the content of the articles. The criteria of the critique section were on the logic and meaningfulness of the writing.

The self-assessment section allowed students to first indicate the numeric score of their reading-to-write performance by referring to the rubric. Also, the section had space for students to provide a brief rationale of the score for each major element of the writing (i.e., purpose, procedures, findings, and critique). Therefore, students' self-assessment comprised both quantitative and qualitative information. Likewise, the instructor assessment section allowed the instructor to provide feedback by marking the numeric score and explaining the rationale for it. This was offered in addition to the local and general comments she directly provided on the students' writing.

Procedures

Students were asked to select a topic of their interest in the field of TESOL. Within the topic, they chose four empirical studies from varying academic journals. During the first week of the data collection (Week 1), students submitted an initial draft of their first task (Task 1 - Draft 1) along with the assessment sheet. Prior to submission, the instructor provided a training for completing the assessment sheet by explaining the descriptors of the scoring rubric and how to assign a rating for the self-assessment section with concrete examples.

Upon submission, the instructor (1) provided detailed feedback (both local and general) directly on the students' writing and (2) completed the instructor assessment section on the assessment sheet. Both the instructor feedback and the assessment sheet were returned to the students the following week (Week 2). Then in Week 3, the students submitted a revision of their first writing (Task 1 - Draft 2), along with their self-assessment of the revision. Students' revisions were reviewed by the instructor, on which the instructor provided detailed feedback; she also completed the instructor assessment section on the assessment sheet. This reading-to-write and feedback cycle was repeated for the other three tasks.

Therefore, during one semester, the students submitted a total eight writings (including initial and revised drafts), conducted eight self-assessments, and received instructor feedback for each writing.

TABLE 1
Schedule for the Assignment, Self-assessment, and Instructor Feedback

Week	Reading-to-write task (assignment)	Self-assessment (classroom activity)	Instructor feedback & assessment
1	Task 1 - Draft 1	Task 1 - Draft 1	
2	Task 2 - Draft 1	Task 2 - Draft 1	Task 1 - Draft 1
3	Task 1 - Draft 2	Task 1 - Draft 2	Task 2 - Draft 1
4	Task 2 - Draft 2	Task 2 - Draft 2	Task 1 - Draft 2
5	Task 3 - Draft 1	Task 3 - Draft 1	Task 2 - Draft 2
6	Task 4 - Draft 1	Task 4 - Draft 1	Task 3 - Draft 1
7	Task 3 - Draft 2	Task 3 - Draft 2	Task 4 - Draft 1
8	Task 4 - Draft 2	Task 4 - Draft 2	Task 3 - Draft 2
9			Task 4 - Draft 2

All collected data were analyzed, both quantitatively and qualitatively. Students' self-assessment and the instructor's assessment scores across all of the writings were analyzed using descriptive statistics. To check for the reliability of the instructor ratings, a second rater with a doctoral degree in Applied Linguistics independently scored a sample of the student writing (approximately 20%) by focusing on the content domain. The inter-rater reliability (Spearman rank correlation coefficient) between the two raters was high, at 0.92. Moreover, students' performance was qualitatively analyzed; preliminary results indicated that students either made steady progress from the beginning or showed gradual enhancement from mid-way. Thus, the performance of two representative students—Chung (who progress from the beginning) and Park (who made progress mid-way)—were analyzed and presented in depth in this study. Their initial drafts and revisions, self-assessments, and instructor feedback were analyzed together. Specifically, the degree of the students' incorporation of instructor feedback (e.g., correct uptake or incorrect uptake) was examined. Furthermore, in order to compare the students' self-assessment with instructor feedback, their written comments on the assessment sheet were compared across each of the four sections (purpose, procedure, findings, and critique).

Findings

Effectiveness of LOLA

The effectiveness of LOLA was first quantitatively examined by analyzing the students' progress in their reading-to-write tasks, using scores from the instructor assessment and the students' self-assessments. From the instructor assessment, scores from the initial drafts and revisions were respectively examined. Ratings on the *initial* drafts for each task gradually increased from Task 1 - Draft 1 ($M = 2.91$ out of 5) to Task 4 - Draft 1 ($M = 3.70$) (see Table 2). Likewise, the ratings on the *revised* drafts increased from Task 1 - Draft 2 ($M = 3.64$) to Task 4 - Draft 2 ($M = 4.18$). Moreover, for each of the four tasks, the instructor gave higher scores on the revision than on the initial draft.

As seen in Table 2, similar to instructor assessment, participants' self-assessments of the *initial* drafts increased from Task 1 - Draft 1 ($M = 3.36$) to Task 4 - Draft 1 ($M = 3.67$). Likewise, their self-assessment scores of the *revised* drafts across the four tasks increased from Task 1 - Draft 2 ($M = 3.75$) to Task 4 - Draft 2 ($M = 4.18$). For each task, students gave higher scores on the revision than on the initial draft.

TABLE 2

Descriptive Statistics of Instructor and Students' Self-Assessment Scores

Week	Task	Instructor assessment		Students' self-assessment	
		Mean	SD	Mean	SD
1	Task 1 - Draft 1	2.91	0.54	3.36	0.81
2	Task 2 - Draft 1	3.09	0.54	3.00	0.63
3	Task 1 - Draft 2	3.64	0.67	3.75	0.46
4	Task 2 - Draft 2	3.73	0.79	3.91	0.54
5	Task 3 - Draft 1	3.50	0.53	3.70	0.48
6	Task 4 - Draft 1	3.70	0.67	3.67	0.50
7	Task 3 - Draft 2	3.80	0.63	4.10	0.57
8	Task 4 - Draft 2	4.18	0.75	4.18	0.60
	Total	3.56	0.73	3.70	0.68

Note. Due to the time required for the instructor to provide feedback to the students, on Week 2, students worked on Task 2 - Draft 1 and completed Task 1 - Draft 2 on Week 3.

Further in-depth qualitative analyses of the two students—Chung and Park—were conducted to supplement the quantitative findings. The two students were selected because they reflected the trends in students' score improvements—either showing improvement from the beginning or mid-way during the study. Chung demonstrated steady progress early on whereas Park showed gradual enhancement from mid-way in their development of L2 reading-to-write ability (refer to Table 3).

TABLE 3

Chung and Park's Over-time Progress

Week	Reading-to-write task	Chung		Park	
		Self-assessment score	Instructor score	Self-assessment score	Instructor score
1	Task 1 - Draft 1	3	3	4	3
2	Task 2 - Draft 1	3	3	3	3
3	Task 1 - Draft 2	4	4	4	3
4	Task 2 - Draft 2	4	5	4	3
5	Task 3 - Draft 1	4	3	4	4
6	Task 4 - Draft 1	4	5	NA*	4
7	Task 3 - Draft 2	4	5	5	4
8	Task 4 - Draft 2	5	5	5	5

Note. Park submitted an initial draft of Task 4 – Draft 1, but did not complete a self-assessment due to his absence.

According to the findings from the instructor assessment, Chung's scores for the first two initial drafts of Tasks 1 and 2 were the same, with a score of 3 (out of 5). When she revised the two drafts based on the instructor's feedback, the score increased to a score of 4 and 5, respectively. However, she did not continue to show such improvement; when she wrote the next initial draft of Task 3 in Week 5 (Task 3 - Draft 1), she received a score of 3, the same score as in the first two initial drafts. Although she successfully addressed the instructor's feedback on the revisions, she could not apply the feedback when writing a new draft. For example, as shown in the following excerpts from Task 3, she could not provide concrete information while summarizing the procedure section of the journal article she read (the instructor's feedback is inserted in bold).

(Chung, Task 3 - Draft 1, Procedure part)

A 42 year-old Japanese business man who had not studied English for the past 20 years was given a selection of graded readers to read once a day for a year. He was told to keep record of the books he read and to write a brief summary of each story in Japanese. The participants did not engage in any other grammar instruction or writing in English, and there was no error correction <what errors?> during the treatment. <What about data (TOEIC scores) collection and analysis?> The results showed that...

However, she easily integrated the instructor's feedback in the revision (Task 3 - Draft 2), as shown below, and she received a score of 5.

(Chung, Task 3 - Draft 2, Procedure part)

A 42 year-old Japanese man who had not studied English for the past 20 years was given a selection of graded readers to read once a day for a year. He was told to keep record of the books he read and to write a brief summary of each story in Japanese. The participants did not engage in any other grammar instruction or writing in English during the treatment but did attend TOEIC listening comprehension classes for a total of 30 hours. After reading for six months, the participants took the TOEIC test...

By Task 4, she received a full score of 5 in both the initial and revised drafts. Contrary to the previous initial drafts, she finally seemed to have internalized the instructor feedback and understood how to write a review.

Moreover, Chung had almost no problem with writing the critique section from the beginning. In order to evaluate a research article critically, students should be able to understand it first. It seems that Chung had no difficulty comprehending the journal articles, which might have attributed to her good performance in the reading-to-write tasks. For instance, the following excerpt demonstrates her thorough understanding of the article, as well as her critical thinking of the research design.

(Chung, Task 2 - Draft 1, Critique part)

Many other studies such as Tse (1996) and Shin (1998) have also revealed that ER can improve attitude toward reading, which suggest that ER might well catch two birds with one stone; improving reading proficiency while reducing the emotional burden learners carry when learning a new language. However, the researcher's role as a teacher in the current study may have contributed to the increase in the student's motivation and reduction in anxiety level. The full support provided by the researcher would have created a safe and comfortable environment for the participants, and this kind of support may not always be present during ER.

In the excerpt, Chung demonstrated her understanding of the role of the researcher as a teacher by mentioning "the researcher's role as a teacher." In addition, she commented on how such a role might have affected the research findings with reference to other previous studies ("... may have contributed to the increase in the student's motivation and reduction in anxiety level. The full support provided by the researcher would have created a safe and comfortable environment for the participants, and this kind of support may not always be present during ER."). Therefore, the instructor left a brief positive comment ("good") on this critique part.

The other student, Park, exhibited a different developmental pattern from Chung. It took longer for Park to reveal the effect of self-assessment and the instructor feedback (refer to Table 3). Specifically, Park received a score of 3 for the initial drafts of both Tasks 1 and 2 (Task 1 - Draft 1 and Task 2 - Draft 1) from the instructor. Contrary to Chung, he failed to address the instructor's feedback in the revisions (Task 1 - Draft 2 and Task 2 - Draft 2) and ended up receiving the same score of 3. In other words, with respect to instructor assessment, there was no change for the first four weeks.

For instance, the following two excerpts show how Park failed to incorporate the instructor's feedback in the earlier stage of his performance, specifically in describing the purpose of the study.

(Park, Task 2 - Draft 1, Purpose part)

The purpose of this study is to encourage students to participate in extensive listening and to give them more exposure to English <English in general? Or listening?> outside the classroom. The other goal is to encourage students to concentrate on form of adverb placement and passives as well

as meaning. **<meaning of what?> <Are these (to encourage) the goal of the study or goal of the instruction included in the study?>**

(Park, Task 2 - Draft 2, Purpose part)

The important goal of this study is to encourage students to participate in extensive listening and to give them more exposure to English outside the classroom. The second purpose is to encourage students to concentrate on form of adverb placement and passives as well as meaning of the content students were supposed to listen to.

The instructor left comments on the initial draft of Task 2, which asked for more concrete information and clarification: “English in general? Or listening?”; “meaning of what?”; and “Are these (to encourage) the goal of the study or goal of the instruction included in the study?” Although Park provided more concrete information about the second comment (“meaning of what”) by including “meaning of the content students were supposed to listen to” (as underlined in the revision), he failed to address the other two comments.

However, the effects of LOLA began to emerge from Week 5 on the remaining tasks. His scores from the instructor feedback on the initial drafts of Tasks 3 and 4 (Task 3 - Draft 1 and Task 4 - Draft 1) were 4, which indicate improvement of his performance after four weeks of practice. As seen below, he wrote a very clear, concrete purpose on the initial draft of Task 3 (Task 3 - Draft 1):

(Park, Task 3 - Draft 1, Purpose part)

The purpose of the study was to inspect the effectiveness of four different types of listening support on listening comprehension of EFL learners at two listening proficiency levels: high level proficiency (HLP), low level proficiency (LLP). The four types of support were...

It appears that Park finally understood the concrete and specific information required for writing the purpose part when composing a review. Although he demonstrated better performance on Task 3 than on Tasks 1 and 2, his scores remained at a score of 4 for both the initial and revised drafts of Task 3 (Task 3 - Draft 1 and Task 3 - Draft 2). Therefore, it seems that he had some difficulty in understanding the instructor’s feedback or addressing it in his revision. He finally addressed the instructor’s feedback successfully on Task 4 (Task 4 - Draft 2) and received a full score of 5.

One noticeable aspect of Park’s performance over time was his difficulty in writing critiques. He received similar instructor comments (e.g., “lack of details”) in the critique part across the four initial drafts, which were not successfully addressed in the revisions. Although the overall scores improved from Tasks 1 and 2 to Tasks 3 and 4, the qualitative analysis of Park’s writing revealed his lack of progress in the critique part. It appears that he was able to improve his ability to summarize the research article more easily and quickly than he could write a critique. In his revisions of Tasks 3 and 4, however, he tried to be more specific when writing a critique and finally did not receive any further comments.

Comparing Students’ Self-assessment with Instructor Feedback

Students’ self-assessment with instructor assessment scores were compared across the eight writings. On the earlier tasks, students perceived their reading-to-write performance as higher in comparison to the instructor’s evaluation (see Table 2). For instance, students’ average self-assessment score on Task 1 - Draft 1 (3.36) was higher than the average instructor assessment score on Task 1 - Draft 1 (2.91). Toward the latter tasks, students’ self-assessments generally matched the instructor’s assessments of the writing samples. For example, the average score of both students’ self-assessment and the instructor’s score was 4.18 in the final writing.

For more in-depth analysis, the explanations (i.e., qualitative comments) of the scores provided by the two students—Chung and Park—were compared with those of the instructor across the eight writings. Chung’s self-assessment scores were generally similar to the instructor’s scores each time she performed

a self-assessment, and this pattern was also observed in the written comments made by Chung and the instructor. From the onset (i.e., Task 1 - Draft 1), as seen in Table 4, Chung made similar comments about her own writing as the instructor. Both Chung and the instructor had the same opinions about the clarity of purpose and a lack of information about the test used in the procedure part. More specifically, Chung not only picked the same point (i.e., test) that the instructor felt was problematic, but she also had a very similar evaluation regarding the point (i.e., a lack of elaboration/information). Therefore, it seems that she was able to self-assess the strengths and weaknesses of her own writing from the beginning. The similarities between Chung's and the instructor's comments became more noticeable in Chung's revisions. Chung successfully understood and mostly addressed the instructor's feedback in her revisions. Therefore, her self-assessment was more aligned with the instructor assessment in the revisions.

TABLE 4

Comparison of Chung's Self-assessment and Instructor Feedback (Task 1 - Draft 1)

	Chung's Self-assessment	Instructor Assessment
Purpose	<ul style="list-style-type: none"> • Purpose is fairly clear. 	<ul style="list-style-type: none"> • Good
Procedure	<ul style="list-style-type: none"> • Maybe more information about the test should be included. • Did not include whether the test was multiple-choice or not (Test type). 	<ul style="list-style-type: none"> • Instead of the name of the project, explain what tests were used (e.g., type, constructs).

Notably, Chung's comments in her self-assessment revealed her own weakness in summarizing the procedure part of a study. This was explicitly written in her self-assessment:

(Chung, Task 2 - Draft 1, Procedure part)

To be honest, it was difficult for me to write the procedure section because I couldn't find the right expressions, so it may be unclear at times how the data was analyzed.

(Chung, Task 4 - Draft 1, Procedure part)

I think the procedure section is too long, but I couldn't find a more effective way to include the descriptions of the tests.

In the excerpts above, Chung explicitly explained why she struggled to summarize the procedure part. It was due to language problems (e.g., "I couldn't find the right expressions") required for developing a succinct summary. These instances show that Chung reflected on her review writing and found her own weakness while performing repeated self-assessments. Therefore, they provide evidence of the effects of LOLA in general, and self-assessment in particular.

The other student, Park, showed both similarities and differences in his self-assessment from the instructor feedback. He had opinions similar to those of the instructor, mainly for the purpose and critique parts of his review writing, while he had different comments regarding the procedure part. For example, while assessing the revised draft of Task 2 (Task 2 - Draft 2), Park and the instructor made similar comments on the purpose part, but opposing comments on the procedure part. As seen in Table 5, both Park and the instructor thought that the purpose was unclear. Regarding the procedure, Park evaluated it as being detailed, whereas the instructor wanted to see more information.

TABLE 5

Comparison of Park's Self-assessment and Instructor Feedback (Task 2 - Draft 2)

	Park's Self-assessment	Instructor Assessment
Purpose	<ul style="list-style-type: none"> • Two purposes of the study are stated, but not clear enough to see what is exactly assessed for the results of this study. 	<ul style="list-style-type: none"> • It is not clear what the purpose of the study was, instead of the purpose of lessons embedded in the study.
Procedure	<ul style="list-style-type: none"> • Procedures are mentioned in detail. 	<ul style="list-style-type: none"> • More information is required, especially in relation to the materials.

In addition, Park aptly recognized whether his critique of the review was appropriate or not; thus, his comments on the critique part were often aligned with the instructor’s comments. While self-assessing his initial draft of Task 3 (Task 3 - Draft 1), he pointed out the inappropriateness of his critique by writing, “The critique I wrote might not be correct,” and the instructor also provided a negative comment: “Try to make the critique more concrete.” In another instance when he addressed the instructor’s feedback in the revision (Task 1 - Draft 2), he recognized the appropriateness of his critique (“I think the critique is developed more accurately than the previous draft”), and he made similar comments to those of the instructor (“Feedback addressed, good”).

Although his comments in the self-assessment were generally similar to those of the instructor in terms of the purpose and critique parts of review writing, he mentioned his weaknesses in writing the critique part in his self-assessments. Similar to Chung, Park became aware of his weaknesses by reflecting on his writing through repeated self-assessments:

(Park, Task 1 - Draft 1, Critique part)

I am not confident with my critique I mentioned.

(Park, Task 3 - Draft 1, Critique part)

The critique I wrote might not be correct.

These excerpts partly explain why Park remained at the same level almost toward the end of the study period in terms of the critique part of writing. It indicates that he could not easily overcome his weaknesses, even though he noticed them. Therefore, he might need focused support or scaffolding in the critique part, as well as repeated practice to overcome his weaknesses.

Park’s comments in his self-assessment showed further evidence of LOLA, in relation to the effect of the instructor’s feed-forward. Several instances of his revision showed his efforts to address the instructor’s comments, which directly and indirectly guided his revisions. As shown in Table 6, in the initial draft of Task 2 (Task 2 - Draft 1), the instructor commented on the problem of the findings part (i.e., “The findings are not presented in relation to the research purposes.”). Park revisited this later in the revision process and addressed it in his self-assessment (e.g., “The results of the study are stated based on the two purposes of this study.”). The instructor’s comment on the initial draft led him to present the research findings in relation to the two purposes of the study, and he reported this effort in his self-assessment. Such an effect of LOLA (the instructor feedback as feed-forward) was found several times in his self-assessments of the revised drafts.

TABLE 6
Transition of Park’s Self-assessment and Instructor Feedback (Task 2)

Task	Section	Park’s Self-assessment	Instructor Assessment
Task 2 - Draft 1	Findings	There were some problems in terms of second purpose of this study, but I couldn’t write down details because of the space limit.	The findings are not presented in relation to the research purposes.
Task 2 - Draft 2	Findings	The results of the study are stated based on the two purposes of this study.	Okay. Sometimes not very clear due to the language (word choice) problems.

Discussion & Conclusion

Students’ reading-to-write performance generally improved within the LOLA framework, as supported by both quantitative and qualitative data. Students’ writing scores generally increased during the eight weeks, which was reflected in the quality of the writing samples, self-assessment, and instructor feedback. Contrary to the existing literature (e.g., Biber et al., 2011), which did not find a significant effect of

instructor feedback on students' performance in the content domain, the instructor feedback built into the LOLA framework in the current study appears to have been effective in improving L2 learners' content domain in writing. Notably, the instructor feedback influenced individual students differently, as indicated by the varying developmental patterns of two representative students. While Chung demonstrated improvement in her reading-to-write ability early on, Park demonstrated a delayed effect of receiving instructor feedback.

In addition to instructor feedback, students' improvement could be attributed to their self-assessment. Similar to prior research, which found a positive effect of self-assessment (Butler & Lee, 2010; Ross, 1998), the self-assessment in this study allowed students to reflect upon (Paris & Paris, 2001) and realize their strengths and weaknesses in writing. For example, while Chung found that her weakness lay in clearly and succinctly expressing her thoughts by using appropriate expressions, Park realized that his greatest weakness was in critiquing the articles he read.

The fact that students' self-assessment scores progressively matched the instructor's scores suggests that students' self-assessment could be a reliable source for evaluating learners' reading-to-write performance (Boud, 1995; Ross, 1998). Therefore, implementing self-assessment could assist in enhancing students' reading-to-write ability in L2 academic writing settings (Paris & Paris, 2001). This could be particularly useful when the instructor cannot provide individualized feedback to all students.

Nevertheless, some qualitative differences were found between the two sample students' self-assessment and instructor feedback. Chung's self-assessment comments were overall similar to the instructor feedback; this was more so in her revisions than in her initial drafts because she dutifully incorporated the instructor feedback into her revisions. Meanwhile, Park's self-assessment comments and the instructor's comments indicated differences as well as similarities. The comments were the most different in the procedure section of the writing in comparison to other parts, such as the purpose and the critique part of the writing.

Overall, the study findings support the effectiveness of LOLA. This could be attributed to the combined effect of the two key elements of LOLA—self-assessment and instructor feedback. Repeated reading-to-write practice, supported by self-reflection (in the form of self-assessment) and external input (in the form of instructor feedback), proved to be an effective way of enhancing students' L2 reading-to-write ability.

However, a couple of limitations need to be addressed in future studies. To better understand the effectiveness of LOLA, it would be helpful to integrate students' voices. It would be beneficial to examine what students think is helpful in the feedback cycle. For example, stimulated recall interviews (Gass & Mackey, 2000) can be used to recall students' writing process while they review the documents they created beforehand. This process would reveal why students' self-assessment might differ from the instructor feedback, and could suggest ways to narrow the gap, if needed.

This study was also limited in terms of the number of classes that participated. Because only one reading-to-write course with one instructor was examined, it is difficult to generalize the findings to a larger context. For instance, it is difficult to claim that similar findings would be found in courses taught by other teachers. There could be variables that affect teacher and student interaction that have not been explored in this study, such as an instructor's teaching experience. Moreover, future studies could benefit from observing the long-term effect of LOLA on students' reading-to-write performance. It would be meaningful to investigate whether the effect of LOLA can be sustained, even after students stop engaging in self-assessment or receiving instructor feedback.

Despite these limitations, this study is meaningful in terms of examining both the effects of students' self-assessment and instructor feedback in an L2 academic writing context. The study results suggest the effectiveness of incorporating students' self-assessment and instructor feedback into L2 academic writing instruction (e.g., Babaii et al., 2016). Repeated writing practice and relevant feedback helped the students to effectively tackle their weaknesses and promote their L2 reading-to-write ability. This study demonstrates that combining assessment practices with regular writing instruction could create a synergy

effect for enhancing students' L2 writing ability (Green & Hamp-Lyons, 2015; Hamp-Lyons & Green, 2014; Turner & Purpura, 2016).

Another pedagogical implication of this study is the necessity of sustained writing practice. The results from qualitative analysis showed that students' revisions, followed by instructor feedback, did not necessarily lead to immediate improvement in their subsequent writing. In some instances, students did not demonstrate any progress for weeks, although they eventually internalized the instructor feedback much later. Therefore, the current study suggests the importance of giving multiple, repeated opportunities to students in order to help them internalize learning at their own pace.

Acknowledgements

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

The Authors

Hyun Jung Kim is a professor of the Graduate School of TESOL at Hankuk University of Foreign Studies in Seoul, Korea. Her research interests include second and foreign language assessment, learning-oriented assessment, and language test validation. She holds an Ed.D. in Applied Linguistics from Teachers College, Columbia University.

Graduate School of TESOL
Hankuk University of Foreign Studies
107, Imun-ro, Dongdaemun-gu, Seoul, 02450, Korea
Tel: +82-2-2173-3154
Email: hkim@hufs.ac.kr

Ahyoung Alicia Kim (corresponding author) is a researcher at WIDA at the University of Wisconsin-Madison where she examines the language development of K-12 multilingual learners. Her research interests include language assessment, child bilingualism, second language literacy development, and computer-assisted language learning. Alicia holds an Ed.D. in Applied Linguistics from Teachers College, Columbia University.

WIDA at the Wisconsin Center for Education Research
University of Wisconsin-Madison
1025 West Johnson Street, MD#23, Madison, WI, 53706-1706, USA
Email: alicia.kim@wisc.edu

References

- Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing, 9*, 227-258.
- Babaii, E., Taghanddomi, S., & Pashmforoosh, R. (2016). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing, 33*(3), 411-437.
- Bitchener, J., & Knoch, U. (2009). The value of a focused approach to written corrective feedback. *ELT Journal, 63*(3), 204-211.

- Biber, D., Nekrasova, T., & Horn, B. (2011). *The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis* (ETS Research Report RR-11-05). ETS. <https://www.ets.org/Media/Research/pdf/RR-11-05.pdf>
- Boud, D. (1995). *Enhancing learning through self assessment*. Kogan Page.
- Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessment among Korean elementary school students studying English. *The Modern Language Journal*, 90(4), 423-446.
- Butler, Y. G., & Lee, J. (2010). The effect of self-assessment among young learners. *Language Testing*, 17(1), 1-27.
- Carless, D. (2007). Learning-oriented assessment: Conceptual basis and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated tasks for the new TOEFL* (ETS Research Report RM-05-13). ETS. <https://www.ets.org/Media/Research/pdf/RR-05-13.pdf>
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10, 161-184.
- Ganji, M. (2009). Teacher-correction, peer-correction and self-correction: Their impacts on Iranian students' IELTS essay writing performance. *The Journal of Asia TEFL*, 6(1), 117-139.
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Routledge.
- Gebril, A. (2009). Score generalizability of academic writing tasks: Does one test method fit it all? *Journal of Language Testing*, 26, 507-531.
- Green, T., & Hamp-Lyon, L. (2015, March 16-20). *Learning oriented language assessment for the classroom: A primer* [Workshop presentation]. The 37th Language Testing Research Colloquium, Toronto, Canada.
- Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1996). *A study of writing tasks assigned in academic degree programs*. (ETS Research Report RR-95-44). ETS. https://www.ets.org/research/policy_research_reports/publications/report/1996/codn
- Hamp-Lyon, L., & Green, T. (2014, October). *Applying a concept model of learning-oriented language assessment to a large-scale speaking test* [Conference paper]. The 3rd Teachers College Columbia University Roundtable in Second Language Studies (TCCRISLS), New York, NY.
- Han, Z.-H. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36, 543-572.
- Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction. *Studies in Second Language Acquisition*, 25, 1-36.
- Ji, X. (2015). Error correction in college EFL writing instruction: Students' expectations and correction effects. *The Journal of Asia TEFL*, 12(1), 117-140.
- Keppell, M., & Carless, D. (2006). Learning-oriented assessment: A technology-based case study. *Assessment in Education: Principles, Policy and Practice*, 13(2), 179-191.
- Kim, H. J., & Yang, S. S. (2018). Evaluating the validity of classroom assessments in the secondary EFL curriculum. *The Journal of Asia TEFL*, 15(4), 1021-1035. <http://dx.doi.org/10.18823/asiatefl.2018.15.4.9.1021>
- Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, 17(4), 420-432.
- Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal*, 82, 357-371.
- Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal*, 82, 338-356.

- McCulloch, S. (2013). Investigating the reading-to-write processes and source use of L2 postgraduate students in real-life academic tasks: An exploratory study. *Journal of English for Academic Purposes*, 12(2), 136-147.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75-100.
- Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment & Evaluation in Higher Education*, 28(4), 383-394.
- Moore, T., & Morton, J. (1999). *Authenticity in the IELTS academic writing & test: A comparative study of Task 2 items & university research* (IELTS Research Reports, pp. 64-106). IELTS Australia.
- Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 37(4), 375-401.
- Nilson, L. B. (2003). Improving student peer feedback. *College Teaching*, 51(1), 34-38.
- Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychology*, 36(2), 89-101.
- Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, 19(2), 109-131.
- Plakans, L. (2009a). Discourse synthesis in integrated second language writing assessment. *Language Testing*, 26(4), 561-587.
- Plakans, L. (2009b). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes*, 8, 252-266.
- Plonsky, L., & Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have different results). *Second Language Research*, 31(2), 267-278.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1-20.
- Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, 25, 553-581.
- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41(2), 255-283.
- Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 255-273). De Gruyter Mouton.
- Wiggins, G. P. (1993). *Assessing student performance*. Jossey-Bass Publishers.

(Received August 17, 2021; Revised November 20, 2021; Accepted December 18, 2021)

Appendix

Assessment Sheet

Name: _____ Date: _____

Title of the Article:

Rubric

	<ul style="list-style-type: none"> • To what extent does a student summarize an article successfully including the purpose of the study, overall procedures, and major findings? • To what extent does a student read an article critically and write a logical critique?
5	Excellent summary / Excellent critique
4	Clear summary / (and/or) Logical critique
3	At times unclear summary / (and/or) At times illogical critique
2	Often unclear summary / (and/or) Lack of/illogical critique
1	Mostly unclear summary / (and/or) Severe lack of/illogical critique

Self-Assessment

Score: _____

Evaluation:

Purpose	
Procedure	
Findings	
Critique	

Teacher Assessment

Score: _____

Evaluation:

Purpose	
Procedure	
Findings	
Critique	