Correlation of Proficiency with Complexity, Accuracy, and Fluency in Spoken and Written Production: Evidence from L2 Korean

Youngjoo Kim
Kyung Hee University

Juyeon Nam
Kyung Hee University

Sun-Young Lee
Cyber Hankuk University of Foreign Studies

Abstract

This study investigates the relationship between L2 proficiency and production of 130 L2 Korean learners with four different L1s. We measured their Korean language proficiency and evaluated their writing and speaking on complexity, accuracy, and fluency (CAF) from two story-retelling tasks. To measure complexity, we used the number of dependent clauses per T-unit for writing and C-Unit for speaking; for accuracy, we compared the number of error-free clauses against the total number of clauses; for fluency, we calculated the number of syllables per minute. The results showed (a) a strong correlation between proficiency and CAF and (b) strong correlations among CAF across speaking and writing, indicating that CAF improve both in speaking and in writing alongside the development of L2 proficiency.

1.   Introduction

Speaking and writing in a second language (L2) is very different from speaking and writing in a native or first language (L1). Obviously, an L2 learner must take on a different cognitive load in order to process a task linguistically during the course of speaking and writing. Researchers have conducted studies on how the complexity, accuracy, and fluency (CAF) of a learner's language production vary depending on the learner's cognitive load or the task

characteristics (Crookes, 1989; Foster, 1996; Foster & Skehan, 1996; Oh & Lee, 2012; Robinson, 2003; Rutherford, 2001; Skehan & Foster, 2005).

Skehan's (1996) limited capacity hypothesis claimed that human beings are limited in the amount of information that they can process. Accordingly, he asserted that because individuals are unable to concentrate equally on CAF during language production when they concentrate on one resource, concentration on the other resources drops in a trade-off effect. Consequently, in a complex task, the complexity of the language may increase, but the accuracy may decrease; moreover, in a simple task, the complexity of the language may decrease, but the accuracy may increase.

However, as Robinson (2001, 2003, 2005) argued in his cognitive hypothesis, a competitive relationship does not exist between an L2 learner's language complexity and accuracy; therefore, he hypothesized that L2 learners can instead improve these capabilities simultaneously during language task performances. This implies that a human's cognitive resources are unlimited and independent of each other; hence, a variety of these resources could be used at the same time. In this view, this relationship is not necessarily an inverse relationship; in fact, language complexity and accuracy can show a positive relationship in a complex task if the learner is given advance planning time. However, Robinson (2001) argued that, while complexity and accuracy can improve simultaneously, they don't have a positive relationship with fluency and that, contrarily, in a simple task, accuracy and complexity decline as fluency increases.

The goal of this research is to investigate how the aspects of CAF in L2 production are related to the learner's development of proficiency and how the aspects of CAF in L2 production interact with each other in speaking and writing. Do the complexity, accuracy, and fluency of a learner's language production develop in an organically connected way? Do complexity, accuracy, and fluency demonstrate an inverse relationship as Skehan (1996) proposed, or do complexity and accuracy develop together but have no positive

relationship with fluency as Robinson (2001) suggested? On the one hand, as the learner's L2 proficiency improves, the learner might focus on either complexity or accuracy (e.g., U-shaped development in either case), using different amounts of cognitive load depending on the type of task. On the other hand, these two abilities might develop simultaneously as the learner's L2 proficiency improves.

In addition, this study compares spoken data and written data—which require different amounts of cognitive load due to different time pressure and production processes—in order to provide empirical evidence for evaluating the two hypotheses above. The research in this area has produced contradictory findings about the difference in L2 production depending on whether it is spoken or written. By comparing both L2 speaking and writing with the general proficiency of L2 learners of Korean, this study investigates the developmental patterns of L2 learners' speaking and writing abilities and the relationships among CAF in different production modes.

2.   Literature Review

2.1   Constructs for L2 Production

Linguistics research has identified three major constructs for the evaluation of L2 production: complexity, accuracy, and fluency. *Complexity* relates to how many clauses the learner connects or includes within a sentence. This construct in L2 production shows the development of the restructuring process within the L2 learners' inter-language systems (Skehan, 1996). The units most frequently used for measuring complexity are the T-unit (e.g., Larsen-Freeman, 2006; Sangarum, 2005; Yuan & Ellis, 2003;) and the C-unit (e.g., Foster, 1996; Foster & Skehan, 1996; Rutherford, 2001). In analyzing complexity, the subordinate clause becomes an important criterion. This is because, as a learner's L2 proficiency increases, the learner uses subordinate clauses more frequently than coordinating clauses. When studying spoken language, which includes many incomplete and subject-omitted sentences, researchers typically use the C-unit for

analyzing proficiency because it can eliminate the meaningless repetitions and adjust to the omitted elements in a sentence. In studying written language, scholars have generally analyzed the number of clauses per T-unit to measure proficiency (Benevento & Storch, 2011; Larsen-Freeman, 2006).

Accuracy refers to the learner's ability to exercise the maximum level of control to prevent errors during a language performance. In other words, accuracy refers to a learner reducing errors by avoiding challenging structures that can cause errors (Ellis, 2009). Thus, in order to evaluate the accuracy of the learner's language production, researchers have analyzed the L2 production data in terms of how much of the learner's utterances or compositions were produced without errors; that is, how many error-free clauses they produced. For example, studies have evaluated accuracy by comparing the number of error-free clauses in speaking with error-free clauses in writing (Oh & Lee, 2012; Rutherford, 2001; Yuan & Ellis, 2003). In L2 acquisition, accuracy is an important construct for evaluating the development of the learner's L2 grammar. One cannot judge a learner's L2 language production as accurate simply based on the partial lexicon the learner shares with the researchers or the grammatically correct connection of words and clauses. Consequently, many studies, including the present study, consider error-free clauses—as a measurement unit that is compared against the total number of clauses—to be important evidence for accuracy evaluation.

Fluency refers to the learner's ability to use the language with a high number of words and without extensive pauses and corrections during a limited time. During that time limit, language users must draw from their lexicalized system through a maximum use of affixes or phrases in order to emphasize their intended meaning (Ellis, 2009). Studies have used a variety of methods to measure speaking fluency, such as calculating the following items: the number of meaningful syllables per total syllables (Yuan & Ellis, 2003), the number of

syllables per minute (de Jong & Perfetti, 2011; Sangarum, 2005), the word count per C-unit (Robinson, 2003), the word count per T-unit (Larsen-Freeman, 2006), and the word count per minute (Oh & Lee, 2012). Other studies have counted how often hesitations or pauses occur in an utterance or how many pauses occur and how long the pauses last in an utterance (Foster & Skehan, 1996; Skehan & Foster, 2005). For measuring the fluency of written language, scholars will generally count the number of words per composition or per T-unit (Larsen-Freeman, 2006).

2.2   Analysis Units for L2 Production

Evaluating a learner's language production must begin with a syntactic analysis of the learner's language. The most widely used unit of measurement for spoken and written language analysis has been the T-unit (Foster, Tonkyn, & Wigglesworth, 2000). Proposed by Hunt (1965), T-unit refers to the "minimally terminal unit." A T-unit comprises one independent clause and all subordinate clauses that embellish this clause. Subordinate clauses, connected to the independent clause via subordinate conjunctions (e.g., *that*, *when*, *if*, *whether*, *though*, *although*), function as noun clauses, adjective clauses, or adverb clauses. Even if two independent clauses are connected with a coordinating conjunction (e.g., *and*, *but*, *or*, *nor*), the total number of T-units is two, not one. A text analysis using the T-unit typically evaluates the number of words per T-unit, number of T-units per sentence, number of clauses per T-unit, and number of words per clause.

However, sometimes it is necessary to include other types of units in an analysis to accommodate the varying characteristics of the two production modes. For example, in spoken language analysis, in addition to using the T-unit, researchers have used the C-unit as mentioned above. C-unit refers to a word, a phrase, a sentence, or a grammatical/non-grammatical utterance unit (Pica et al., 1989). While analyzing the existing T-units allows for a distinction between the boundaries of independent clauses, subordinate clauses, and relative clauses, the T-unit is more appropriate for analyzing written language than spoken language because accurately differentiating such

boundaries in spoken language is difficult. In order to overcome this difficulty, scholars adapted the idea of the T-unit to spoken language, where sentence elements are often omitted, and named this unit the C-Unit. Consequently, the C-unit allows the analysis to include items that could not be evaluated through the T-unit, such as subject or predicate omissions, incomplete sentences, and even interjections (Foster, 1996; Lee, 2012; Owens, 2004).

Another type of unit is the (analysis of speech unit)[1] (analysis of speech unit) for spoken language analysis as used in Foster, Tonkyn, and Wigglesworth (2000) and Skehan and Foster (2005). It refers to an individual's utterance unit that consists of an independent clause with or without a subordinate clause. The AS-unit even includes pauses exceeding 0.5 seconds and the rise and fall in intonation of a sentence's initial syllable. For example, Skehan and Foster (2005) used the AS-unit to study the effects of pre-task guidance on how L2 learners planned their spoken performance and how much time they received for the task performance.

On the other hand, as an alternative to using the T-unit to analyze written language, scholars also use punctuation as an analysis unit. In the Korean language, for example, in not only the spoken form but also in the written form, learners often omit sentence subjects, and scholars often encounter challenges in distinguishing the elements of complex sentences, which in turn makes it difficult to conduct a T-unit analysis. This is because the T-unit often is problematic as an analysis unit when it comes to evaluating writing from low-proficiency learners who generally produce short sentences or coordinated sentences (McKay, 2006). Homburg (1984) proposed using punctuation as demarcation for analyzing writing from low proficiency learners. Punctuation units can be more easily applied in cases where the learner's writing has low syntactic completeness and fewer complex sentences.

---

[1] According to Foster, Tonkyn, and Wigglesworth (2000, p. 366) the AS-unit is analyzed as below.
| I have no opportunity to visit | (1 clause, 1 AS-unit)
| it is my hope :: to study crop protection | (2 clauses, 1 AS-unit)
| and you would be surprised :: how he can work | (2 clauses, 1 AS-unit)

2.3   Development of L2 Production

Scholars have studied the development of L2 production from several different perspectives: its relationship to L2 proficiency, to the influence of task type and difficulty, and to the mode of production (either spoken or written). First, the findings from previous studies are not consistent in regards to the relationship between L2 proficiency and CAF in L2 production. Some studies showed positive correlations between the two; that is, higher proficiency level learners were superior in writing when compared to lower proficiency level learners (Pennington & So, 1993). Other studies found that, as proficiency improved, accuracy declined because the proportion of syntactic errors in connective endings increased (Seo & Eo, 2011). Regarding accuracy in particular, for example, Larsen-Freeman (2006) analyzed the spoken and written language development patterns of five Chinese learners of English in four stages over a six-month period. She found more irregularity in accuracy than in fluency or complexity in the development of the learners' written production. While this research showed that the development of fluency and complexity produced a relatively smooth, upward curve in four out of five subjects, the development of accuracy showed sharp, irregular ups and downs. On the other hand, Larsen-Freeman also found that the developmental pattern of complexity is different depending on the production modes—spoken and written—which we will discuss in detail later in this section.

Secondly, the different aspects of L2 production were also influenced by how the study conducted the elicitation task: whether through task planning, task type, task difficulty, or another method. For example, a one-minute task planning session had a positive impact on fluency (Foster, 1996; Foster & Skehan, 1996; Skehan & Foster, 1997). Crooks (1989) also found that the students with a ten-minute planning time produced more complex sentences than those without planning time. However, research has also shown that the effects of task planning are different depending on the task type. For example, Foster and Skehan (1996) learned that accuracy was higher among students with planning time than students without, but

accuracy was higher only in the decision-making task, not in the narrative task. Moreover, accuracy and complexity did not increase together when students received planning time. Crookes (1989) discovered that learners delivered more complex productions for harder tasks, while they delivered more accurate productions for simpler tasks. In other words, the complexity of the learners' production increased with preparation time, but it did so without much improvement to accuracy; in short, the study revealed a trade-off between these two constructs. Similarly, in Foster and Skehan (1996), comparison of CAF in the three tasks of information exchange, picture description, and decision-making confirmed the inverse relationship between accuracy and complexity.

Thirdly, L2 production is also influenced by the type of production: either spoken or written. The previous studies on the difference between spoken and written production have mainly focused on complexity and accuracy (Kuiken & Vedder, 2012), and the findings are somewhat mixed. In grammatical complexity, Ferrari and Nuzzo (2009) found the learners' written mode to be superior, whereas other studies uncovered no such difference between the learners' spoken and written modes in regards to complexity (e.g., Granfeldt, 2007; Kormos & Trebits, 2009). Ferrari and Nuzzo (2009) compared spoken and written productions based on an oral retelling task and a written narrative task with L2 learners of Italian and with native speakers. They ascertained that the L2 learners' productions were slightly more complex in written production than in spoken production, even though the L1 speakers' productions were clearly superior in complexity for the written mode. Ferrari and Nuzzo suggested that grammatical competence might be in play between the two groups. But in Granfeldt (2007), which investigated Swedish learners of French, the study found no such difference in complexity between the spoken and written modes. Instead, Granfeldt observed differences at the individual (i.e., participant) level, which suggests that an individual's preference of mode might account for differences among L2 learners. Regarding grammatical accuracy, Kormos and Trebits (2009) measured L2 learners' accuracy using error-free past tense verbs, relative clauses, and error-free clauses; they discovered

that accuracy was higher in the written mode than in the spoken mode.

The studies on the difference between the spoken and written productions relate to the cognitive process of the task (Kuiken & Vedder, 2012). Speaking is different from writing in many aspects, including verbalization time, preplanning, memory, and the use of implicit versus explicit knowledge. Gathering one's thoughts can take less time for speaking than for writing, and preplanning is easier for writing than for speaking. Individuals must remember previous messages and plans in the process of speaking in order to speak coherently, whereas in writing, individuals can easily retrieve those messages and plans by referring to what they had written down during the preplanning stage. In other words, speaking involves much more processing of implicit knowledge about L2 grammar than writing does; furthermore, learners can more easily retrieve explicit knowledge of L2 grammar for writing because they are under less time pressure (Towell, Hawkins, & Bazergui, 1996). Therefore, speaking is a more cognitively demanding task than writing. By comparing spoken and written L2 production, we can understand how different cognitive demands influence a learner's L2 production. This will enable us to test the two aforementioned (and competing) hypotheses about how task difficulty influences L2 production.

2.4   Korean L2 Production

Previous studies on production in L2 Korean have mainly focused on the lexical and grammatical complexity of the L2 productions and their relationships with a learner's global proficiency. Lee (2003) investigated verb tense based on the spoken productions from Chinese and Japanese learners of Korean. He learned that the accuracy of verb tense decreased in speaking as compared to writing, with intermediate learners showing a U-shaped developmental pattern. Ji (2006) also analyzed the grammatical accuracy of spoken interview data from intermediate and advanced Japanese, English, and Chinese learners of Korean. Ji ascertained that the learners had difficulty with particles, verb endings, and honorifics. Others have conducted similar error analyses using written production data. For

example, Cha and Song's (2006) analysis of compositions from intermediate Chinese learners of Korean showed that morphological errors were the most frequent type of errors. More specifically, Lee (2008) observed that an L2 Korean learner's lexicon usage was most accurate at the beginning level and that a learner's accuracy decreased both morphologically and semantically at the intermediate level. Seo and Eo (2001), who investigated the written productions from Chinese, Japanese, and Russian learners of Korean, also identified an increase of syntactic errors and semantic errors at the intermediate level compared with the beginning level.

Despite these results, Seo (2009) found that L2 Korean learners' syntactic accuracy increased at the intermediate level when she analyzed the frequency of error-free clauses both quantitatively and qualitatively. For the quantitative analysis, she examined the number of error-free clauses per sentence; for the qualitative analysis, she examined the types of frequent errors. For grammatical and lexical complexity, she compared the number of words and clauses per sentence and the number of morphemes and subordinate clauses per clause. The results revealed that the higher the learner's proficiency level, the more diverse and numerous the learner's grammatical forms were; grammatical accuracy also improved as the learner's proficiency increased. Further, the number of words, clauses, and morphemes per sentence showed significant differences depending on the learners' proficiency levels.

Even though many studies have been conducted in the field of L2 Korean regarding the development of L2 production and its relationship with proficiency, we are still in need of more data in this research field. The findings of the previous studies are not consistent; for example, Lee (2003) found a U-shaped development of accuracy (2003), but Seo (2009) did not. In addition, each study dealt with only limited types of grammatical elements, such as tense, verbal endings, honorifics, and so on. Moreover, researchers considered only two

limited aspects of production: complexity and accuracy. Fluency should be also investigated in order to obtain the full picture of the development of L2 production.

Consequently, the previous studies on L2 production including Korean are lacking in certain areas, necessitating further research. First, previous studies on L2 production focused on complexity and accuracy as mentioned above. We need to consider L2 production from a more diverse perspective that jointly considers complexity, accuracy, *and* fluency. Next, previous studies on the L2 production typically investigated L2 learners of English, French, and Italian. We need more data from typologically different languages such as Korean in order to verify the findings of previous studies; this will allow us to better understand the development of L2 production in general. More importantly, few studies looked at the relationship between L2 proficiency and the CAF of L2 production with the same learners from four different L1 backgrounds. Moreover, to obtain a broader understanding of how L2 production develops, we should investigate comparisons between L2 spoken and written data to see how these different aspects of L2 production abilities interact with each other. Based on the literature review presented above, this study investigates the relationship between proficiency and CAF in speaking and writing using the data from L2 learners of Korean. The following research questions directed this study:

R1   How does proficiency relate to CAF in the spoken production of L2 Korean? How do CAF interact with each other in the development of L2 speaking?

R2   How does proficiency relate to CAF in the written production of L2 Korean? How do CAF interact with each other in the development of L2 writing?

R3   How does the mode of production (spoken and written) influence the performance of L2 production? In other words, what are the similarities and differences between speaking and

writing in the development of CAF in L2 production?

3.   Method

3.1   Participants

We recruited a total of 130 participants with various L1 backgrounds, including French (20), English (24), Japanese (41), and Chinese (45). All of them were university students. We enlisted Chinese- and Japanese-speaking students from Korea, whereas we enlisted English- and French-speaking students from the United States and France, respectively. We allowed the participants' male-to-female ratio, age, and length of L2 study to fall naturally without adjustments to control them. However, we included only non-heritage learners of Korean in order to eliminate the possible influence of heritage speakers' different linguistic backgrounds as recently found in SLA research. All participants had received intermediate-level Korean language education, but we divided them into nine groups of general proficiency based on the mini-TOPIK scores (which will be described in detail later) using the 9-point Stanine scale.[2] Of the participants, 37 were in the upper ranks of proficiency (7–9), 51 were in the middle ranks (4–6), and 42 were in the lower ranks (2–3). No participant fell into the Stanine 1 rank. Table 1 shows the detailed Stanine proficiency levels for the participants.

---

[2] The Stanine score—a standard score allocated on a scale of 1 to 9 based on the original score's location in the normal distribution—shows the learner's ranking within the specific group. The lowest level within the group is scaled to Stanine 1 and the highest level to Stanine 9. Stanine 1, 2, and 3 are below average; Stanine 4, 5, and 6 are average; and Stanine 7, 8, and 9 are above average. Stanine scores make it easy to compare the scores of a variety of learners by using scores to group similar individuals into sections. Stanine's score section per scale is divided by half of the standard deviation (i.e., it would be divided into 0.5).

Stanine Level	No.	Mini-TOPIK Mean	Gender		L1			
			M	F	ENG	FRE	CHI	JAP
9	5	107.40	1	4	1	0	4	0
8	12	94.25	3	9	3	0	7	2
7	20	84.15	6	14	4	0	8	8
6	15	70.53	2	13	1	1	7	6
5	20	57.30	8	12	7	1	5	7
4	16	45.94	7	9	4	3	5	4
3	30	33.03	6	24	2	10	6	12
2	12	22.67	5	7	2	5	3	2
1	0	0	0	0	0	0	0	0

Note: Stanine 1, 2, and 3 are below average; Stanine 4, 5, and 6 are average; and Stanine 7, 8, and 9 are above average.

Table 1: Participants' Stanine Proficiency Levels

3.2   Materials

3.2.1   General proficiency test

For the proficiency evaluation, we used the nationally accredited Test of Proficiency in Korean (TOPIK), which has won global recognition for reliability and validity. Since the actual TOPIK test takes a total of 180 minutes to complete, we developed a mini-TOPIK for this research by reducing the total number of questions to one-third of the original so that students could cover all areas on the test within 60 minutes. Accordingly, the mini-TOPIK consisted of four subparts, including vocabulary and grammar (10 multiple-choice questions), writing (4 multiple-choice questions and 2 sentence-composition questions), listening (10 multiple-choice

questions), and reading (10 multiple-choice questions). Also, because this research addressed intermediate-level learners, we based the questions on the intermediate (Third to Fourth Level) exam questions from the TOPIK tests conducted from 2010 to 2011. We also adopted specific grading criteria for objective evaluation of the short answer questions in the writing part.

Two Korean language education professionals prepared the mini-TOPIK questions. To verify the test's validity, four Korean language instructors participated in evaluating the difficulty level; moreover, three Chinese students who scored in the Third Level of an actual TOPIK also tested the validity of the mini-TOPIK. All judged it to be appropriate for the intermediate level.

3.2.2   Story-telling Tasks

We employed two story-retelling tasks to elicit the learners' spoken and written productions. The participants watched two 10-minute clips from silent episodes of *Mr. Bean* and then retold one clip's plot in spoken form and the other clip's plot in written form. This video format had two advantages: (a) all learners could view and understand the clips, regardless of their mother tongue or their Korean proficiency level, and (b) the content was familiar to the learners. To minimize the effects of the task's difficulty level on the language production process, we selected two clips with similar difficulty levels in terms of composition and expressions. The video for the speaking task was about an incident during an episode called "Do It Yourself," and the one for the writing task was about the protagonist's participation in a school event called "Back to School." Mock evaluations using students enrolled in actual intermediate-level Korean language courses confirmed that the two clips did not pose difficulties to intermediate-level learners.

4.   **Procedure**

We administered the experiment in a quiet classroom. Participants completed the writing task in four groups based on the learners' native language. Participants completed the speaking task individually. For the writing task, for example, a group of English-speaking learners first watched the video together and then wrote about the story for 30 minutes. For the speaking task, each participant sat in a separate room to view the video on a computer screen, after which he or she retold the story while a computer recorded these retellings for later analysis; each participant received 10 minutes for this task. We briefed the participants on the procedures in their mother tongue before showing the video but gave them no preplanning time.

5.   **Analysis**

5.1   Transcription

All data was transcribed. Three teams conducted the transcription and analysis in shifts. Each team included two Korean language education professionals. In each team, one professional transcribed a certain amount of data, and then the other professional reviewed that transcription. Finally, both professionals met together and conducted a third review to address any discrepancies between the results of the first two reviews and to consolidate the results.

5.2   Scoring

While researchers have a variety of methods for analyzing spoken and written production depending on the purpose of the evaluation, for this research, in which we wanted to compare a learner's proficiency and productive ability, we evaluated the CAF for both spoken and written language. Taking into consideration the various factors influencing the measures for CAF as reviewed in the

above section, we made the following decisions on the measuring units for each aspect.

5.3   Complexity

We used the number of subordinate clauses as the analysis unit for complexity because the use of subordinate clauses is more closely related to the complexity level of the sentence structure than is the use of coordinate clauses. We counted the number of subordinate clauses per T-unit for writing complexity and the number of subordinate clauses per C-unit for speaking complexity.

5.4   Accuracy

We calculated the number of error-free clauses against the total number of clauses as the analysis unit for accuracy. The types of errors our study measured included misspellings, mispronunciations, inappropriate vocabulary use for the context, the learner's use of his or her mother tongue instead of Korean, incorrect use of adjectives and verbs, case markers with inappropriate meaning and form, and ending errors. Despite the number of errors within one clause, we considered error-free clauses to be only those clauses that were accurate and completely error-free in meaning and form. However, for the spoken language test, due to the natural and numerous utterances in speech that do not interfere with message conveyance, we did not count errors such as clause duplications from repetitions, adjustments, and re-forming.

5.5   Fluency

We chose the number of syllables per minute as the analysis unit for fluency, measuring how much spoken output versus written output the learners produced in the same amount of time (10 minutes). For speaking, however, our criteria involved counting only *meaningful* syllables because, as Foster and Skehan's research

illustrated (Foster, 1996; Foster & Skehan, 1996; Skehan & Foster, 2005), self-correction occurs regularly in utterances through repetition and error correction of the sentence initials.

6.   **Results**

We obtained the mean scores for CAF in spoken and written productions from the learners' proficiency level in order to answer our research questions. The results are presented below and organized by research question.

6.1   Spoken L2 production

R1: Does proficiency relate to CAF in the spoken production of L2 Korean, and how do CAF interact with each other in the development of L2 speaking? In order to answer our first research question, we calculated means for CAF scores for each proficiency level; the results are presented in Table 2.

Stanine Level	No	Complexity		Accuracy		Fluency	
		M	*SD*	*M*	*SD*	*M*	*SD*
9	5	4.29	4.61	0.61	0.09	128.1	18.96
8	12	2.92	1.27	0.49	0.21	105.55	35.98
7	20	2.79	1.77	0.53	0.12	83.57	34.19
6	15	2.18	1.14	0.48	0.14	72.31	26.11
5	20	1.76	0.78	0.42	0.1	80.25	31.54
4	16	1.97	1.33	0.44	0.15	62.7	23.25
3	30	1.21	0.75	0.42	0.14	47.56	26.75
2	12	1.04	0.79	0.32	0.11	50.98	33.93
Total Mean		**2**	**1.42**	**0.46**	**0.14**	**72.57**	**30.07**

Note. Stanine 1, 2, and 3 are below average; Stanine 4, 5, and 6 are average; and Stanine 7, 8, and 9 are above average.

Table 2: CAF in Spoken Production by Proficiency

Table 2 shows that accuracy for L2 spoken production was 0.32 for learners at proficiency level 2 and about twice that amount at 0.61 for learners at proficiency level 9, which means that accuracy

does indeed have a positive correlation to proficiency in spoken L2 production.

The mean scores for spoken fluency also increased as proficiency improved. The learners at level 2 verbally produced 50.98 meaningful syllables per minute on average, while the learners at level 9 produced over twice that: an average of 128.1 meaningful syllables per minute. This demonstrates that the learners' spoken L2 production increases as their proficiency levels increase.

Complexity also has a positive correlation to proficiency in spoken L2 production, with learners at level 2 scoring 1.04 and learners at level 9 scoring an impressive 4.27. The learners produced only 1 subordinate clause per C-unit at level 2, but learners at level 9 produced 4.27 subordinate clauses per C-unit, which indicates that L2 learners' speech becomes more linguistically complex as their proficiency improves. The higher overall mean scores for all three aspects of production at the higher proficiency level indicates a tendency for positive correlations between proficiency and each of the production aspects: complexity, accuracy, and fluency.

We statistically tested this tendency for positive correlations by comparing the Pearson correlation coefficient between proficiency and each production element. The results are given in Table 3.

	Pearson Correlation Coefficient (N = 130)		
	Complexity	Accuracy	Fluency
Spoken production	0.464**	0.391**	0.559**

Note. **$p < 0.01$

Table 3: Correlation between Language Proficiency and Spoken Production

Table 3 reveals that the individual learners' proficiency strongly correlated with complexity (r =0.464, p < 0.01), with accuracy (r

=0.391, p < 0.01), and with fluency (r =0.559, p < 0.01). This reveals that CAF all improved in correlation with learners' speaking proficiency; fluency showed the strongest relationship with proficiency.

Next, in order to find the relationships among CAF in the development of L2 spoken production, we drew correlations among these three elements by again using the Pearson correlation coefficient. The results are shown in Table 4.

	Complexity	Accuracy	Fluency
Complexity	--	0.394**	0.380**
Accuracy	0.394**	--	0.407**
Fluency	0.380**	0.407**	--

Note. **p < 0.01

Table 4: Correlations between CAF in Spoken Production

Table 4 shows that all three aspects of spoken production are strongly correlated with each other: The correlation coefficient ranks highest between accuracy and fluency (r=0.407, p < 0.01), lowest between fluency and complexity (r =0.380, p < 0.01), and in the middle between accuracy and complexity (r =0.394, p < 0.01). These results, along with those from Table 3, indicate that the CAF of the learners' spoken production develop along with their L2 Korean proficiency.

6.2   Written L2 production

R2: Does proficiency relate to CAF in the written production of L2 Korean, and how do CAF interact with each other in the development of L2 writing? We calculated the mean CAF scores for written production for each level of proficiency in order to answer our second research question. The results are presented in Table 5.

Stanine Level	No.	Complexity		Accuracy		Fluency	
		M	*SD*	*M*	*SD*	*M*	*SD*
9	5	2.37	0.76	0.68	0.14	45.52	11.96
8	12	2.09	0.76	0.59	0.11	41.64	10.81
7	20	2.42	0.92	0.6	0.18	31.8	8.57
6	15	2.02	0.85	0.51	0.21	28.57	14.04
5	20	1.97	0.8	0.52	0.13	33.76	9.86
4	16	1.35	0.66	0.48	0.18	23.88	7.81
3	30	1.15	0.8	0.48	0.16	22.62	10.29
2	12	0.87	0.45	0.35	0.19	22.96	9.5
Total Mean		**1.67**	**0.86**	**0.52**	**0.17**	**29.99**	**10.24**

Note. Stanine 1, 2, and 3 are below average; Stanine 4, 5, and 6 are average; and Stanine 7, 8, and 9 are above average.

Table 5: CAF in Written Production by Proficiency Level

Table 5 shows that the accuracy in L2 written production was 0.35 for learners at proficiency level 2 and 0.68 for learners at proficiency level 9, indicating just a 35% error-free clause rate against the total number of clauses for level 2 learners and a 68% error-free clause rate for level 9 learners. This means that accuracy correlates positively with L2 proficiency in written production as well as spoken production.

Mean scores for fluency also showed a significant difference in correlation to spoken proficiency. Learners at level 2 uttered an average of 22.96 meaningful syllables per minute, yet learners at level 9 uttered an average of over twice that at 45.52 meaningful syllables per minute. This suggests that learners' written L2 production will increase as their proficiency improves.

The mean scores for complexity show a similar correlation with proficiency. Learners at level 2 scored 0.87 in complexity, while learners at level 9 scored 2.37. In other words, the learners at level 2 produced 0.87 subordinate clauses per T-unit, but this number more than doubled to 2.37 for learners at level 9. Therefore, the more proficient the L2 learners are, the greater the number of linguistically complex sentences they will produce in their writing. Overall, the greater the mean scores for all three elements of written L2 production, the greater the proficiency level.

We used the Pearson correlation coefficient to conduct a statistical analysis of correlations between proficiency and each element of CAF. The results are shown in Table 6.

	Pearson Correlation Coefficient		
	Complexity	Accuracy	Fluency
Written Production	0.536**	0.389**	0.494**

Note. **$p < 0.01$

Table 6: Correlation between Language Proficiency and Written Production

Table 6 shows that the individual learners' proficiency strongly correlated with accuracy (r =0.389, p < 0.01), with fluency (r =0.494, p < 0.01), and with complexity (r =0.536, p < 0.01). This suggests that the CAF of learners' L2 writing will improve along with proficiency, with complexity showing the strongest relationship to proficiency.

Next, we used the Pearson correlation coefficient in order to find the relationships among CAF in the development of L2 written production. The results are shown in Table 7.

	Complexity	Accuracy	Fluency
Complexity	--	0.450**	0.309**
Accuracy	0.450**	--	0.286**
Fluency	0.309**	0.286**	--

Note. **$p < 0.01$

Table 7: Correlations among CAF in Written Production

Table 7 shows that the CAF of learners' L2 writing are strongly correlated with each other: The correlation coefficient ranks highest between accuracy and complexity (r =0.450, p < 0.01), lowest between fluency and complexity (r =0.286, p < 0.01), and in the middle between accuracy and complexity (r =0.309, p < 0.01). The high correlations among complexity, accuracy, and fluency themselves as well as the correlations between proficiency and each of these three elements (as shown in Table 6) tell us that the CAF of a learners' L2 Korean writing should develop simultaneously with their L2 Korean proficiency.

6.3   Comparison between spoken and written productions

R3: How does the mode of production (spoken and written) influence the performance of the L2 production; in other words, what are the similarities and differences between spoken and written productions in the development of CAF in L2 production? First, we calculated the total mean scores for CAF in both speaking and writing. The results are shown in Table 8.

L2 Production	Spoken Production		Written Production	
	M	SD	M	SD
Complexity (subordinate clauses per sentence)	2	1.58	1.71	0.93

Accuracy (error-free clauses per T/C-unit	0.45	0.15	0.52	0.18
Fluency (no. of syllables per minute)	71.62	35.84	29.25	12.07

Table 8: Mean Scores for CAF in L2 Production (N=130)

Table 8 shows that the combined L2 production accuracy in writing and speaking for intermediate-level learners of Korean was about 50%. Accuracy was slightly higher in written productions (52%) than in spoken productions (45%), which means that about half of the learners' written productions were error-free. Regarding fluency scores, the learners produced 71.62 spoken syllables per minute on average, whereas they produced 29.25 written syllables per minute. The big difference between these two modes is understandable considering the dissimilar speed of the two production modes; speaking is faster than writing. The mean score for speaking complexity was 2.00 whereas the mean score for writing complexity was 1.71. The learners produced more complex sentences in speaking than in writing, which contrasted with the results of accuracy, where the learners produced more accurate language in writing than in speaking. We also compared the mean scores for CAF in the two modes according to proficiency level in order to examine the developmental patterns in both modes. See Figures 1, 2, and 3 for the comparisons, which are based on the data from Tables 2 and 5 above.
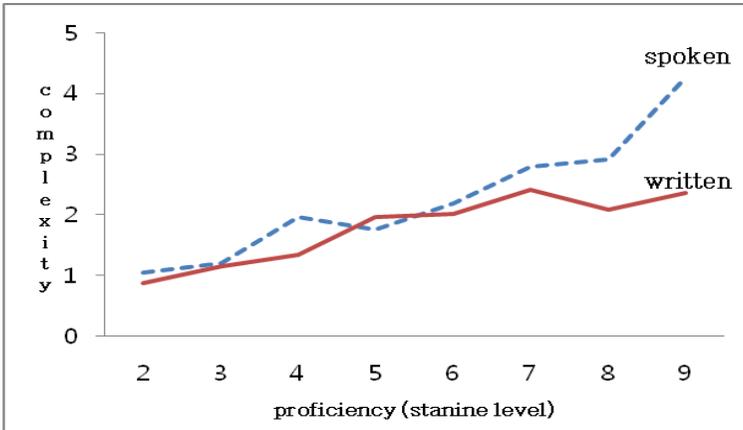
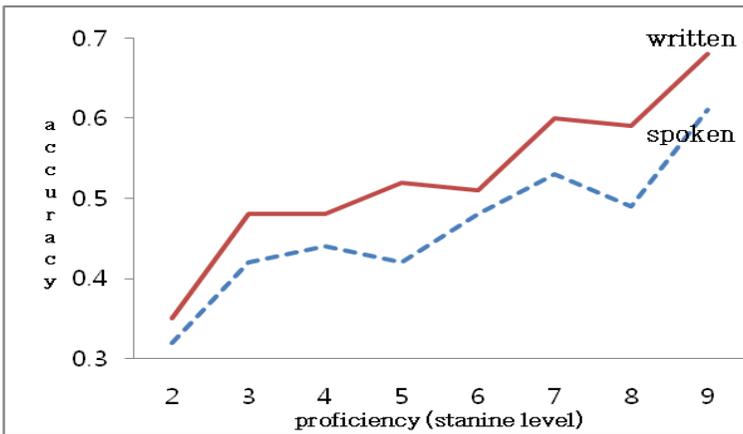*Figure 1.* Complexity development by proficiency in L2 production.

Figure 2. Accuracy development by proficiency in L2 production.
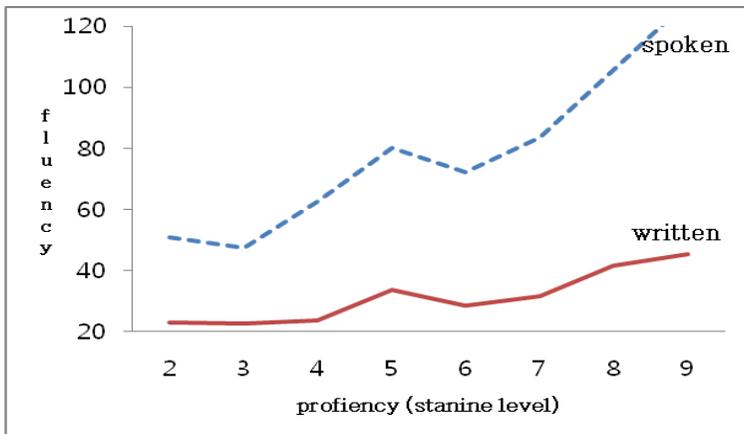
Figure 3. Fluency development by proficiency in L2 production.

First, all the figures depict positive correlations between CAF and spoken productions and CAF and written productions. This means that CAF all develop together both in speaking and writing as the learners' global proficiency improves from level 2 (beginning-intermediate) to level 9 (intermediate-high). Within these two levels of proficiency, CAF show more than a 100% increase: from 1 to 4.5 in speaking complexity, and from 1 to 2 in writing complexity, from 30% to 60% in written and spoken accuracy, from 50 to 120 syllables per minute in speaking fluency, and from 20 to 40 syllables in writing fluency. Still, the data illustrates some differences between spoken and written productions across the three aspects. Second, writing proficiency is superior only in accuracy, not in complexity or fluency. As mentioned above, speaking proficiency is superior in fluency because verbal output is physically faster than written output. Excluding fluency, speaking proficiency is superior in complexity only at levels 8 and 9 (the higher levels of proficiency). Second, the developmental patterns of speaking and writing are more alike in accuracy than in other aspects. Accuracy increases along with proficiency in a parallel fashion, with it being consistently higher in writing than in speaking. However, regarding complexity, the difference between the two production modes is found only at levels 8 and 9, where speaking is about two times more complex than writing.

In order to examine the relationships of CAF in L2 productions across production modes, we conducted a correlation analysis among the three elements; the results are presented in Table 9.

		Written Production		
		Complexity	Accuracy	Fluency
Spoken	Complexity	0.606**	0.365**	0.363**
Production	Accuracy	0.348**	0.314**	.234**
	Fluency	0.337**	0.231**	0.688**

Note. **$p < .01$

Table 9: Correlations between CAF of Spoken and Written
Productions

Table 9 shows very strong correlations among CAF across both modes. The correlation coefficient between spoken and written productions was much higher in the aspects of fluency (r =0.688, p < 0.01) and complexity (r =0.606, p < 0.01) than in accuracy (r =0.314, p < 0.01) and in other combinations as well. This is consistent with the results of a correlation analysis between proficiency and each of the three aspects in both modes; that is, L2 proficiency was more strongly correlated with complexity and fluency than accuracy in both speaking and writing, as presented in Tables 3 and 6 above.

Therefore, the results of our experiment can be summarized like so: the CAF of L2 spoken and written productions all develop together as L2 proficiency increases, but different modes of production are less likely to influence fluency and complexity than accuracy.

7.   **Discussion**

The goal of this study is twofold: to investigate the relationship between L2 proficiency and L2 production abilities in the aspects of CAF, and to compare spoken and written production in order to find out how different modes of production influence CAF in L2 production. The results of our experiment with L2 learners of Korean are discussed below, along with implications of the findings for L2 production.

Our first research question related to the relationship between proficiency and the CAF of speaking, and the data indicates that proficiency and spoken production are strongly correlated with each other in CAF. In addition, the relationships among CAF for speaking were highly correlated. Taken together, these results indicate that the CAF of L2 speaking all increase with the development of higher proficiency. Our data revealed similar results regarding our second research question about the relationship between proficiency and the CAF of written production; that is, the learners' L2 proficiency was strongly correlated with CAF in their writing. In addition, the relationships among CAF for writing were highly correlated in these results. The findings indicate that L2 learners' CAF in writing develop as their proficiency improves. This study's third research question was how L2 production is influenced by both modes: speaking and writing. The results of the experiment revealed strong correlations among CAF across these two modes, with the strongest correlations between fluency and complexity across spoken and written modes. This implies that the learners who speak complex sentences more fluently are likely to produce even more complex sentences in any sort of writing task.

The findings of our study can be discussed from various perspectives. First, it seems safe to say that the CAF of L2 production are interconnected in speaking and writing, indicating that

as speaking ability improves, writing ability improves as a result or vice versa. Moreover, such development can happen in a short span in the development of proficiency and within the intermediate level of L2 Korean. In other words, the findings provide supporting evidence for the claims that CAF can be valid constructs for measuring the development of L2 production in L2 acquisition.

Second, the findings of this study differ from some previous findings on the relationship between accuracy and complexity. For example, Benevento and Storch (2011) observed learners' L2 writing assignments over a six-month period and reported significant improvements over time in language complexity and discourse, while accuracy did not show a distinct improvement. The variance seems related to the different ranges of proficiency levels; the proficiency of this study's participants was intermediate (i.e., from beginning-intermediate to high-intermediate levels). We still need to further investigate the entire range of the proficiency in order to get a more complete picture of the developmental patterns of L2 production.

Third, taking into consideration that our proficiency test included four sub-areas of language skills—vocabulary and grammar, writing, reading, and listening—we can assume that one of the language proficiency sub-skills might show a greater correlation with a particular mode of production than do the other sub-skills. In order to find out the relationships between the sub-skills of language proficiency and the CAF in both L2 production modes, respectively, we conducted a brief correlations analysis between the proficiency levels and their language productions. The results revealed statistically significant correlations between all the language sub-skills and language production in both speaking and writing.[3] We observed a

---

[3] For example, speaking fluency correlated strongly with vocabulary and grammar (0.499, $p < 0.01$) and reading (0.552, $p < 0.01$); and writing fluency showed strong correlations with vocabulary and grammar (0.431, $p < 0.01$) and reading (0.469, $p < 0.01$).

similarly strong correlation between speaking and writing complexity and language sub-skills. Consequently, the higher the learner's L2 proficiency level, the more clearly that proficiency is reflected in the learner's L2 fluency and complexity. Additionally, proficiency's stronger correlations with fluency and complexity than with accuracy in L2 productions seem to suggest that we should include not only accuracy but also fluency and complexity in the evaluation of L2 development.

Finally, our findings also enable us to evaluate the previous theories regarding the influence of task difficulty on L2 production. In Skehan's (1998) limited capacity hypothesis, he noted that accuracy increases while complexity decreases when many cognitive tasks are required due to our inability to pay attention to multiple processes simultaneously. However, the comparison of spoken and written production results in our research—where the cognitive loads between modes are different—shows that CAF develop in a significantly coordinated manner. This finding conforms to the predictions of Robinson's (2003, 2005) cognition hypothesis that the complexity and accuracy of L2 production can improve simultaneously in more cognitively challenging tasks compared to less challenging tasks. On the other hand, contrary to Robinson's (2001) claim that improvements in complexity and accuracy have little relationship with that of fluency, the findings of the current study indicate that complexity, accuracy, and fluency all develop interactively in L2 productions.

8.   Conclusion

This study examines the relationship between L2 proficiency and production in the acquisition of Korean as the L2. We analyzed the complexity, accuracy, and fluency of the spoken and written L2 production data from two story-retelling tasks with 130 learners of Korean from four different L1 backgrounds. We used a mini-TOPIK to collect and score the results from both production tasks with the

learners' L2 proficiency scores. Results of the correlation analysis reveal that learners' proficiency is strongly correlated with CAF in both L2 spoken and written production, suggesting that CAF develop simultaneously in both spoken and written production among intermediate learners of Korean. This means that, as the learners' proficiency improves, they are more likely to produce accurate and complex outputs fluently, regardless of the production mode within the intermediate level of proficiency in L2 Korean. This finding provides supporting evidence for the use of CAF as valid constructs for evaluation of L2 production, and in particular, for the validity of the criteria used in this study to measure each construct.

Moreover, the similarity between the CAF of spoken and written production, where different cognitive processes and difficulties are involved, does not conform to the predictions of Skehan's (1998) limited capacity hypothesis. The findings of the present study identify no inverse relationship between complexity and accuracy in spoken production and written production. Instead, both accuracy and complexity improved in the spoken and written production along with the learners' proficiency levels, even though the proficiency of spoken and written productions correlated more strongly with complexity and fluency rather than accuracy.

This study provides different types of L2 production data from rarely studied L2 Korean. By comparing the spoken and written production of the same learners through similar tasks, this study makes a direct comparison between spoken and written productions. However, due to the limited range of proficiency dealt with in this study, more studies are needed to consider a broader range of proficiency. In addition, further accumulation of results from not only cross-sectional studies such as this one but also longitudinal studies of the same learners would be necessary to confirm the findings of this study and to better understand the development of L2 production.

References

Benevento, C., & Storch, N. (2011). Investigating writing development in secondary school learners of French. *Assessing Writing 16*(2), 97–110.

Bulte, B., & Housen, A. (2009). The development of lexical proficiency in L2 speaking and writing tasks by Dutch-speaking learners of French in Brussels. Presented at the *Task-Based Language Teaching Conference*. Lancaster, England.

Cha, S. J. & Song, H. K. (2006). An analytic study on the Korean lexical errors by Chinese speaking learners. *Journal of Korean Language Education 17*(1), 415–436.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition 11*(4), 367–383.

De Jong, N., & Perfetti, C. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning 61*(2), 533–568.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics 30*(4), 474–509.

Ferrari, S. & Nuzzo, E. (2009). Meeting the challenge of diversity with TBLT: Connecting speaking and writing in the classroom. Presented at the *Task-Based Language Teaching Conference*. Lancaster, England.

Foster, P. (1996). Doing the task better: How planning time influences students' performance. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 126–135). Oxford, England: Heinemann.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition 18,* 299–323.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*

     21(3), 354–375.

Granfeldt, J., Schlyter, S., & Kihlstedt, M. (2007). French as cL2, 2L1 and L1 in pre-school children. In Jonas Granfeldt (Ed.), *Studies in romance bilingual acquisition age of onset and development of French and Spanish* (Vol 21, Extra Seriem, pp. 6–41). Lund, Sweden: Petites études de l'Institut d'études romanes de Lund.

Homburg, T. J. (1984). Holistic evaluation of ESL composition: Can it be validated objectively? *TESOL Quarterly 18,* 87–107.

Hunt, K. (1965). *Grammatical structures written at three grade levels.* Urbana, IL: National Council of Teachers of English.

Hunt, K. (1966). Recent measures in syntactic development. *Elementary English 43*, 732–739.

Ji , H. S. (2006). The error analysis of spoken grammar in Korean interview tests. *Journal of Korean Language Education 17*(3), 301–323.

Kim, E. H. (2007). *An Analysis on the cohesive devices of Korean language of KFL (Korean as a foreign language) learners.* (Unpublished master's thesis). Ewha Woman's University, Seoul, South Korea.

Kuiken, F., & I. Vedder (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency complexity, accuracy, and fluency in SLA* (pp. 143–169). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics 27*(4), 590–619.

Lee, H. Y. (2003). A research on the developmental patterns of grammatical items for tense expression produced by Japanese and Chinese learners of the Korean language. *Bilingual Research 22*, 269–298.

Lee, J. E. (2012). *Effects of task complexity and task familiarity on complexity, accuracy and fluency of oral production in L2 Korean.* (Unpublished master's thesis). Kyung Hee University, Seoul, South Korea.

Lee, J. H. (2008). Error analysis of Chinese learners of Korean language: Focus on source analysis of content-based errors. *Journal of Korean Language Education 19*(3), 403–425.

McKay, S. M. (2006). *Researching second language classrooms*. Mahwah, NJ: Lawrence Erlbaum Associates.

Oh, M., & Lee, H. (2012). The effects of task complexity and task condition on learner language. *Korean Journal of Applied Linguistics 28*(4), 39–68.

Owens, R. E. (2004). *Language disorders: A functional approach to assessment and intervention* (4th ed.). New York, NY: Allyn and Bacon.

Park, J. S., & Seo, S. J. (2009). Syntactic proficiency measurement of Korean language learner's writing. *The Language and Culture 5*(2), 151–173.

Pennington, M. C., & So, S. (1993). Comparing writing process and product across two languages: A study of six Singaporean university student writers. *Journal of Second Language Writing 2*, 41–63.

Pica, T., Holliday, L., Lewis, N., & Morganthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition 11*, 63–90.

Polinsky, M. (2007). American Russian: Language loss meets language acquisition. In W. Browne et al. (Eds.), *Formal approaches to Slavic linguistics: The Cornell Meeting* (pp. 370–406). Ann Arbor, MI: Michigan Slavic Publications.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics 22*(1), 27–57.

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies 21*(2), 45–105.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics 43*, 1–32.

Rutherford, K. (2001). *An investigation of the effects of planning on oral production in a second language* (Unpublished master's thesis). University of Auckland.

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. *Language Learning and Language Teaching 11*, 111–142.

Seo, S. J. (2009). *Study on the interlanguage development of Korean language learners by syntactic proficiency assessment: Based on analyzing syntactic features of writings.* (Unpublished master's thesis). Yonsei University, Seoul, South Korea.

Seo, S. J., & Eo, J. H. (2011). Study on the accuracy variation of connective endings by Korean proficiency level. *Journal of Korean Language Education 22*(1), 123–143.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics 17*(1), 38–62.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research 1*(3), 185–211.

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task-performance in a second language* (pp. 193–218). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics 17*(1), 84–119.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity, and accuracy in L2 monologue oral production. *Applied Linguistics 24*(1), 1–27.