THE JOURNAL OF ASIA TEFL Vol.11, No. 1, pp. 125-148, Spring2014

Linguistic Differences in the Writing Performance of Adolescent EFL Learners: The Influence of Independent and Integrated Tasks

Hye Won Shin
Teachers College, Columbia University, USA

Hyun Jung Kim*
Hankuk University of Foreign Studies, Korea

This study investigates the different linguistic effects that independent (writing-only) and integrated (reading-to-write) tasks have on adolescent English as a Foreign Language (EFL) learners' writing performance. An automated tool, Coh-Metrix, is employed to examine how productive skills are affected by the two types of tasks. In a 2x2x2 study design, 122 randomly assigned 11th-grade EFL students took both independent and integrated writing assessments. The results reveal that the two types of writing tasks produced submissions from students that were significantly different in their lexical sophistication, cohesion, and syntactic complexity. The present study highlights the importance of task type and of a learner's language proficiency when examining writing test performance. In particular, on the integrated task, high-proficiency test-takers were able to produce texts that were more lexically and syntactically sophisticated, as well as more cohesive, than low-proficiency test-takers.

Keywords: writing test, independent task, integrated task, Coh-Metrix

* Corresponding author

INTRODUCTION

The communicative approach to English language teaching and learning was first introduced to Korea in 1992 with the development of the Sixth National Curriculum for middle and high schools. The new curricula aimed to replace grammar-translation and audiolingual methods, which had been dominant in secondary school English teaching, with the fostering of students' communicative competence in English (Choi, Park, & Kim, 1986; Development Committee of the Sixth Curriculum for High School English, 1992). With the release of Communicative Language Teaching (CLT) curricula, a series of English textbooks were newly developed and published. These textbooks focused on the balanced development of language skills and components. The current English textbooks approved by the Ministry of Education for public schools are based on a "multi-syllabus," in which a wide variety of skills, situations, and topics are explicitly and systematically addressed (McDonough & Shaw, 2003). Since the implementation of the Sixth National Curricula in Korea, the country's English educational system has emphasized CLT and the development of communicative language ability.

Despite these CLT-based curricula, however, language learning and teaching have not actually changed much. A balanced emphasis on different language skills has not been a priority in public education. In particular, students in secondary schools have not been provided with opportunities to learn or practice productive skills until recently (Bae, 2012; Yook, 2011). This lack of opportunity to practice is mainly due to the influence of the nationwide, high stakes College Scholastic Ability Test (CSAT). The English section of the CSAT still centers on receptive skills: it assesses only listening and reading, using only multiple-choice questions (Jin, 2013; Kim, 2007; Yang & Sohn, 2009). Both students and teachers have little incentive to work on productive language skills and improve their ability to communicate (Jeon, Lee, & Kim, 2011; Park, Chang, Park, & Paek, 2012).

In order to align language teaching and learning with CLT-based curricula, the National English Ability Test (NEAT) has recently been developed (Jin,

2013). One of the distinct characteristics of the NEAT is the inclusion of sections that test speaking and writing. First administered in 2012, the NEAT has been used for admissions at selected universities. As expected, the new testing system is gradually bringing about positive impacts in classroom teaching and learning. For example, different teaching-learning models for speaking and writing have been suggested that consider the unique contexts of secondary schools (e.g., regions, students' proficiency levels, facilities, and human resources) in order to help teachers effectively implement speaking and writing lessons and promote CLT in the classroom (Korea Institute for Curriculum and Evaluation, 2011). Recent studies have investigated the needs of students and teachers, and their perceptions about the NEAT (e.g., Kim, 2009; Kim, 2012; Lee & Yu, 2012). Few studies have focused exclusively on the English writing and speaking ability of secondary school students. Without an accurate understanding of student ability, it is difficult for teachers to address those students' strengths and weaknesses while teaching, or to provide students with appropriate feedback. The present study will therefore explore the English writing ability of adolescent Korean students. More specifically, we examine how different types of tasks might reveal different facets of their writing performance. The study also examines whether task type effects differ for students with varying English proficiency levels. This kind of analysis has pedagogical implications for foreign language writing in Korea, as well as in other Asian countries that promote CLT-style approaches to writing in a foreign language.

LITERATURE REVIEW

Task Effects in Performance Assessment

In second language (L2) performance assessment, which requires test-takers to actually perform/produce language, various factors are necessarily involved. According to McNamara's (1995) schematic representation, test-

takers' performance and their test scores are influenced by various factors and their interactions in (1) the testing situation in which test-takers perform a task; and (2) the following rating situation in which their performance is scored. As presented in Figure 1, test-takers' test performance is affected by their language ability and backgrounds (candidate factor) in the first place, and also by a given test task (task factor). Their test performance and scores are then influenced by raters' rating behavior (rater factor) and scoring criteria (scale/criteria factor). That is, test-takers' test performance is not an isolated feature in L2 performance assessment, but rather the result of complex interactions among various elements involved in the testing context. Therefore, these various elements should be taken into consideration when evaluating test-takers' language performance and making inferences about their language ability.

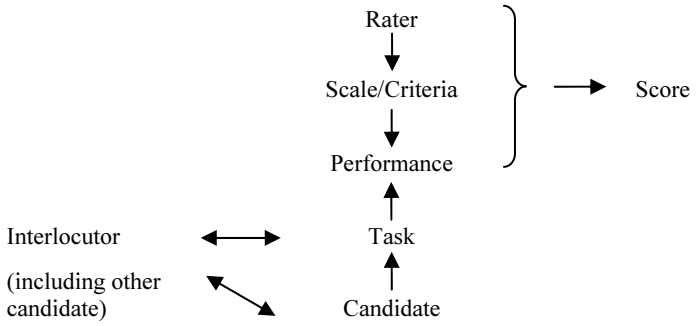

FIGURE 1
Interactions in Performance Assessment
(McNamara, 1995, p. 173)

Among many elements, considerable attention has been paid to task factors in L2 performance assessment because test-takers' performance can vary to a large extent across tasks, making it difficult to generalize test results (Brennan, 1992; Shavelson & Webb, 1991). Although McNamara (1995) did not specify each factor with sub-elements in his model, empirical studies have identified what the task factor includes, and how it interacts with

learners in L2 writing performance assessment. For example, previous studies involving task effects have investigated (1) *the effects of discourse modes,* such as narrative vs. expository (e.g., Carlson, Bridgeman, Camp, & Wanders, 1985; Park, 2007; Quellmaz, Capell, & Chih-Ping, 1982; Wiseman, 2008); (2) *the effects of topics,* such as familiar vs. non-familiar (e.g., Ferris & Hedgecock, 1998; Powers & Fowles, 1999; Spaan, 1993; Tedick, 1990); and (3) *the effects of task types,* such as independent vs. integrated (e.g., Brown, Hilgers, & Marsella, 1991; Lewkowicz, 1994). However, most of these previous studies have compared writing performances statistically, using mean ratings. As a result, it is still unclear as to how each of the different task aspects influences test-takers' performance (Park, 2007). In other words, test-takers' performance features have rarely been examined in relation to varying conditions of tasks, beyond comparing differences in ratings.

Independent and Integrated Writing Tasks

As one aspect of the task factor, the use of independent and integrated tasks has been investigated in writing performance assessments. Independent writing tasks are designed to measure only test-takers' writing ability. Therefore, test-takers are provided with a short instruction or prompt and are asked to write on a given topic using their background knowledge or prior personal experience. On the other hand, integrated writing tasks aim to measure test-takers' ability to understand input (e.g., listening or reading) and further use it when writing on a given prompt (Lee, 2006). Test-takers can approach an integrated writing task with ease because they are provided with sources/contents used for their response. Another positive effect of inputs provided as part of a task is that limited or controlled contents from listening or reading input can reduce the effects of task topics and test-takers' background knowledge on writing. Other than these reasons, integrated writing tasks have also been favored because they are more authentic and diverse than independent tasks and have more positive washback effects

(Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Hamp-Lyons & Kroll, 1997; Read, 1990; Rosenfeld, Leung, & Oltman, 2001; Weir, 1993). However, the integration of different skills for writing assessments often makes it difficult to separate test-takers' writing ability from their listening or reading ability, and thus threatens the validity of test results (Weigle & Parker, 2012). Another shortcoming of integrated tasks is that test-takers may extensively use and rely on given listening and reading input for their writing responses (Lewkowicz, 1994; Watanabe, 2001).

Therefore, numerous studies have investigated whether the integration of other language skills has facilitating effects on test-takers' writing performance, or whether it puts another burden on test-takers. Analyses of differences in test performance across independent and integrated writing tasks have yielded conflicting results. For example, statistical analyses of test-takers' scores on two types of writing tasks found no significant differences (e.g., Brown, Hilgers, & Marsella, 1991; Gebril, 2009; Lewkowicz, 1994). In contrast, an analysis of discourse features in independent and integrated tasks (reading-to-write and listening-to-write) found differences in lexical complexity, syntactic complexity, rhetorical features, and pragmatic qualities (e.g., Cumminget al.,2006).

Overall, the impact of writing task types (i.e., independent and integrated tasks) still remains unexploredin relation to test-takers' writing performance features, in particular. There may be numerous reasons for this deficiency. First of all, most comparisons of writing performance between two types of tasks have been conducted based on statistical analyses, which have centered on differences in overall ratings. Such statistical analyses of test score comparisons cannot provide detailed information about test-takers' strengths/weaknesses or similarities/distinctness in two types of writing. Second, analyses of writing response features (e.g., lexical and syntactic features) often rely on the numbers of features appearing in the responses (e.g., Cumming et al., 2006; Engber, 1995; Polio, 1997). Frequency-based comparisons may not reflect the actual diversity and complexity of test-takers' language use (Guo, Crossley, & McNamara, 2013). Third, writing

response features have been analyzed separately for each of independent and integrated tasks, without comparisons between two task types (e.g., Plakans, 2009; Plakans & Gebril, 2013; Weigle & Parker, 2012). For these reasons, it is necessary to systematically examine the differences of specific linguistic features in writing between the two types of tasks.

Coh-Metrix

Coh-Metrix is an automated computational tool designed to evaluate linguistic features of a text (Grasesser, McNamara, Louwerse, & Cai, 2004; McNamara & Graesser, 2012). It generates over 100 indices. Summarized below are three constructs related to Coh-Metrix indices that we used to measure the quality of test-takers' essays: lexical sophistication, cohesion, and syntactic complexity. (See Graesser et al. (2004) for an extensive overview of the indices supported by Coh-Metrix.)

Lexical sophistication. Lexical sophistication typically refers to lexical diversity. Coh-Metrix calculates the diversity parameter by combining the Measure of Textual Lexical Diversity (MTLD) (McCarthy & Jarvis, 2007) and D (Malvern & Richards, 1997; Jarvis, 2002). These indices correct for highly correlated text-length problems across a corpus of text due to strikingly different word counts.

Cohesion. Cohesion refers to explicit cues that help to link ideas in a text (Crismore, Markkanen, & Steffensen, 1993). An index of text cohesion comes from connectives. Coh-Metrix estimates incidence scores for connectives that are associated with particular classes of cohesion as identified by Louwerse (2001), with the higher numbers the index contains indicating more connectives. Coh-Metrix provides an additional linguistic index for causal relations by measuring the ratio of causal particles to causal verbs (Graesser et al., 2004).

Syntactic complexity. Coh-Metrix assesses syntactic complexity by calculating the mean number of words before the main verb, measuring the mean number of high-level constituents per word, and measuring the uniformity and consistency of the syntactic constructions in the text. Structurally dense sentences, for example, or those that have many embedded constituents, are syntactically complex, and therefore more difficult to process and comprehend (Perfetti, Landi, & Oakhill, 2005). Coh-Metrix's syntactic analyses are based on the ApplePie parser (Sekine & Grishman, 1995) and the POS tagger (Brill, 1995), as cited in Graesser et al. (2004).

The Current Study

The current study examines the effects of independent (writing-only) and integrated (reading-to-write) writing tasks on adolescent EFL learners' writing performance. Comparisons of linguistic features are made across two types of tasks; we also compare differences in the response features across varying proficiency levels. This study is guided by the following questions: (1) To what extent do the linguistic features of adolescent EFL students' test performances differ between independent and integrated tasks? (2) Does English proficiency skill moderate adolescent EFL students' writing test performance?

METHODS

Participants

The participants were 122 11th-grade students from a large urban school district in the southeast part of Seoul, Korea. These students were mostly female (78 percent), with a mean age of 16 at the time of the study (see Table 1). Fifty-three of the participants self-identified as science majors, whereas the remaining students self-identified as being on the humanities track (n=69).

The high school was selected based on convenience sampling.

TABLE 1
Student Demographics by Classrooms (*n*=122)

	n	Male	Female	Science	Humanities
Class A	27	27	--	27	--
Class B	26	--	26	26	--
Class C	34	--	34	--	34
Class D	35	--	35	--	35

**Note. n*=number of participants

All of the students had begun learning English as a compulsory subject at age nine (third grade of elementary school), as prescribed by Korea's Ministry of Education; however, two-thirds of the participants reported being introduced to English even before they began formal language instruction in primary school. While the majority of the students reported that they had had zero to six months of experience living in English-speaking countries, 13 students had one to two years of experience, while six students had three to five years of experience.

Ministry of Education policy advocates the teaching of all four language skills (i.e., listening, reading, speaking, writing), and it encourages the use of writing in English classes. Unfortunately, many teachers do not have the opportunity to implement the teaching of writing. Therefore, students are not generally exposed to productive writing skills in the classroom.

Instruments

The researchers developed two types of writing tasks, independent and integrated, to most effectively measure participants' writing skills. The only difference between the independent and integrated tasks was the presence of reading passages. Passages covered topics that would appeal to young Korean adults: K-pop and sports. Since the researchers' goal was to investigate linguistic features of the participants' compositions, these two topics were also chosen to control for background knowledge. For both task types, test-takers were given 20 minutes to write an essay in English. In other

words, time to complete the tests was controlled, but the length of writing was not limited.

Independent Tasks. The independent tasks were designed to evaluate each student's ability to write an essay on a given prompt. The first of the two prompts asked for opinions about a musical group's decision to join an entertainment agency, and the latter focused on whether women should compete against men in sports.

Integrated Tasks. These tasks required students to integrate information from a prompt into their essays. The total number of words in the K-pop and sports reading passages were 196 and 208 words, and the Flesch Reading Ease scores were 61 and 56, respectively. Flesch Reading Ease is readability metric that assesses the difficulty of written English texts (Flesch, 1948). By this metric, the two texts were roughly the same difficulty. Researchers used the same writing prompts (and questions) as they did for the two independent tasks. Naturally, each reading passage included examples and details that could be used for the students' arguments.

Data Collection Procedures

Students attended two successive sessions within a one-week period. As presented in Figure 2, the order of the independent and integrated tasks and the writing topics varied across the four groups (A, B, C, and D). Students in Group A wrote their responses after reading a text about K-pop, and then they wrote about sports without reading any related text (reading-to-write to writing-only). Group B's members wrote about sports after reading about the same topic, and then they wrote about K-pop without any text (reading-to-write to writing-only). Groups C and D did the tasks in the opposite order, from writing-only to reading-to-write). The tasks were randomly assigned and administered across the four intact classes.

For each task type, students were given 20 minutes to write an essay.

During the integrated tasks, however, researchers initially gave the man additional five minutes for reading. In other words, the time given for the writing tasks in the two treatment conditions, whether integrated or independent, was equivalent.

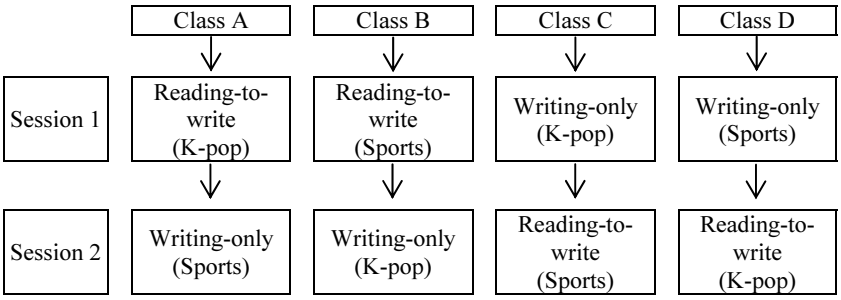

FIGURE 2
Procedures for the Study

Data Analysis

Students' essay scores from Coh-Metrix indices were randomly subjected to a mixed two-(task type) by-two (topic type)-by two (order effect) analysis of variance (ANOVA). This design counterbalanced task type (i.e., independent or integrated), topic type (i.e., K-pop or sports), and order of tasks (i.e., first session or second session) during the study. We computed descriptive statistics as well as a series of ANOVA and analysis of covariance (ANCOVA) to better understand the differences between the two task groups in terms of the linguistic features. To determine students' English proficiency skills, they were classified into either low- or high-skill groups based on the scores in their English courses. Furthermore, effect-size estimates, eta-squared (η^2), were manually calculated; these values can be interpreted as follows: η^2=0.010 is a small association, η^2=0.059 is a medium association, and η^2=0.138 is a large association (Cohen, 1998). All effects are reported as statistically significant at 0.05 level and higher.

RESULTS

To ensure that the participants in the two conditions were equivalent, one-way ANOVA was conducted between the independent and integrated groups. Analysis revealed that the differences in English course scores between the independent group (M=59.740, SD=20.360) and the integrated group (M=64.970, SD=18.606) were statistically not significant, $F(1,137)=2.471$, $p=0.118$.

Table 2 displays the means and standard deviations of the independent and integrated tasks. Overall, when the independent and integrated task groups were compared, there was a significant main effect for the following indices: lexical diversity, verb incidence, positive and negative connectives, verb overlap, number of modifiers per noun phrase, minimal edit distance, and agentless passive voice density (see Table 3). For the sake of conciseness, we describe only those results that were statistically significant, and present the results by each construct.

Lexical Sophistication

There was a significant difference between independent and integrated tasks in the following two parameters: lexical diversity and verb incidence. An ANOVA analysis yielded a significant model at a 99 percent level of confidence, $F(1,120)=10.707$, $p=0.001$, $\eta^2=0.082$ and $F(1,120)=7.345$, $p=0.008$, $\eta^2=0.058$, respectively. This indicates that test-takers in the integrated task group produced essays that exhibited more diversity in word and verb usage than their peers in the independent task group. Further analysis revealed that interaction between lexical diversity and language proficiency was significant, $F(2,1)=19.060$, $p=0.000$. Similarly, we confirmed a significant interaction between verb incidence and language proficiency, $F(2,1)=3.587$, $p=0.031$. Consequently, test-takers with higher levels of proficiency produced essays with greater differences in lexical diversity and verb incidence than did lower-skill students in either the

integrated or independent task group.

Cohesion

Analysis showed that there was a main effect for positive connectives, $F(1,120)=25.844$, $p=0.000$, $\eta^2=0.177$; and negative connectives, $F(1,120)=12.389$, $p=0.001$, $\eta^2=0.094$; reflecting more connectives in the integrated condition group than in the independent condition group. Additionally, there was a significant effect for verb overlap, $F(1,120)=11.563$, $p=0.001$, $\eta^2=0.088$, indicating that test-takers' essays in the reading-to-write condition ($M=0.45$, $SD=0.20$) contained more verb overlap across sentences than those of their peers in the other group. Put differently, it was easier to process the reading-to-write essays than it was to process the essays from the writing-only condition group ($M=0.34$, $SD=0.17$).

Subsequent analyses showed significant interaction effects between language proficiency and positive connectives, $F(2,1)=14.312$, $p=0.000$; language proficiency and negative connectives, $F(2,1)=7.693$, $p=0.001$; and language proficiency and verb overlap, $F(2,1)=7.975$, $p=0.001$. This suggests that participants with better language skills used more connectives associated with the positive additives (*plus, furthermore*), negative additives (*but, yet*), and verb overlap indices. As a result, it was easier to process these essays, given that there were added connections tying the ideas together, than it was to process the lower-proficiency participants' essays.

Syntactic Complexity

The findings show that the integrated condition demonstrated much richer syntactic composition than their peers in the other group. Their sentences contained more modifiers per noun phrase, had minimal edit distance (i.e., part of speech, all words, lemmas), and had lower agentless passive voice density. Specifically, our results showed that (a) sentences that contained more modifiers per noun phrase were seen significantly more in the

integrated condition (*M*=0.770, *SD*=0.250) than in the independent condition, (*M*=0.630, *SD*=0.320), *F*(1,120)=7.122, *p*=0.009, η^2=0.056; (b) the minimal edit distance in parts of speech was significantly higher in the integrated condition essays (*M*=0.67, *SD*=0.17) than the independent condition essays (*M*=0.580, *SD*=0.260), *F*(1,120)=5.206, *p*=0.024, η^2=0.042; (c) the minimal edit distance of all words was significantly higher in the integrated condition (*M*=0.850, *SD*=0.200) than in the independent condition (*M*=0.750, *SD*=0.310), *F*(1,120)=4.819, *p*=0.030, η^2=0.039; (d) the minimal edit distance of lemmas was significantly higher in the integrated condition (*M*=0.830, *SD*=0.200) than in the independent condition (*M*=0.730, *SD*=0.310), *F*(1,120)=5.433, *p*=0.021, η^2=0.043; and (e) agentless passive voice density was significantly lower in the integrated condition (*M*=7.640, *SD*=10.630) than in the independent condition (*M*=13.520, *SD*=20.310), *F*(1,120)=4.119, *p*=0.045, η^2=0.033.

The interaction of syntactic complexity and language proficiency was reliable for four of the five indices: the number of modifiers per noun phrase, *F*(2,1)=5.850, *p*=0.004; minimal edit distance in parts of speech, *F*(2,1)=13.476, *p*=0.000; minimal edit distance of all words, *F*(2,1)=14.716, *p*=0.000;and minimal edit distance of lemmas, *F*(2,1)=14.817, *p*=0.000. Only one index, agentless passive voice density, was not significant, *F*(2,1)=2.149, *p*=0.121. In short, the proficiency levels did have some bearing on the fact that the integrated condition demonstrated much richer syntactic composition.

TABLE 2
Comparison of Means and Standard Deviations by Task Type

Construct	Linguistic index	Independent Task	Integrated Task
		Mean (*SD*)	Mean (*SD*)
Lexical sophistication	Lexical diversity	47.14 (26.54)	64.63 (31.94)
	Verb incidence	102.73 (37.05)	121.60 (39.59)
Cohesion	Positive connectives	0.00 (0.00)	31.36 (46.96)
	Negative connectives	0.00 (0.00)	5.75 (12.44)
	Verb overlap	0.34 (0.17)	0.45 (0.20)
Syntactic complexity	Number of modifiers per noun phrase	0.63 (0.32)	0.77 (0.250)
	Minimal edit distance (parts of speech)	0.58 (0.26)	0.67 (0.17)
	Minimal edit distance (words)	0.75 (0.31)	0.85 (0.20)
	Minimal edit distance (lemmas)	0.73 (0.31)	0.83 (0.20)
	Agentless passive voice density	13.52 (20.31)	7.64 (10.63)

TABLE 3
Summary of ANOVA Results for Selected Linguistic Indices (*n*=122)

	Coh-Metrix index	*df*	*F*	η^2	*p*
Lexical sophistication	Lexical diversity	1	10.707	0.082	0.001
	Verb incidence	1	7.345	0.058	0.008
Cohesion	Positive connectives	1	25.844	0.177	0.000
	Negative connectives	1	12.389	0.094	0.001
	Verb overlap	1	11.563	0.088	0.001
Syntactic complexity	Number of modifiers per noun phrase	1	7.122	0.056	0.009
	Minimal edit distance (parts of speech)	1	5.206	0.042	0.024
	Minimal edit distance (words)	1	4.819	0.039	0.030
	Minimal edit distance (lemmas)	1	5.433	0.043	0.021
	Agentless passive voice density	1	4.119	0.033	0.045

DISCUSSION

The study was designed to examine the effects of task type (in this case, independent and integrated) on student writing. An ANOVA was conducted on the Coh-Metrix indices to investigate which variables produced differences between the two task groups. The results indicated main effects, mostly with small to medium effect sizes, for task type. Of the indices, 10 showed significant differences at the lexical sophistication, cohesion, and syntactic complexity levels in the writing compositions of the 122 test-takers in the independent and integrated task groups.

The test-takers in the integrated task group, in comparison to those in the independent task group, employed a wider range of words, wrote more cohesive essays with many connectives, and produced syntactically richer sentences. These discrepancies highlight the differences in writing quality that emerge from the two task types. This is in line with the claim of Cummingand his colleagues (2005) that different types of tasks produce, in many respects, different written discourse. Essentially, this study adopted Cumming et al.'s approach to distinguishing texts, but through coding of the sampled compositions. We also examined whether these task effects differed for students with low and high English language ability. Not surprisingly, the more proficient English speakers in our sample demonstrated higher levels of language sophistication (in terms of the three constructs) than did test-takers who were less proficient in English.

Notably, although test-takers in the independent and integrated task groups generally showed significant differences in the linguistic features they used, their performance did not differ in two aspects: text information and text ease. The analysis of linguistic indices indicated that task type did not influence the length of the students' writing (text information) or its difficulty (text easability). This finding is noteworthy because previous studies that compared written performance features in different task-type groups reported the opposite result (Guo et al., 2013). One of the shortcomings of integrated tasks is that test-takers tend to rely on given listening or reading input for

their writing (Lewkowicz, 1994; Watanabe, 2001). Consequently, it was assumed that the test-takers in the integrated task group, utilizing the given reading texts, might find writing easier and would write lengthier essays than those in the independent task group. Contrary to this expectation, all of the test-takers' writing was very similar in terms of length and ease of reading.

How could this be explained? The test-takers who had not regularly practiced writing argumentative essays in schools might have found it difficult to write an English essay within the limited time. The time constraint might also have not allowed for the test-takers to fully express their ideas when performing the integrated tasks. However, participants in the two task groups did reveal significant differences in other constructs (i.e., lexical sophistication, cohesion, and syntactic complexity). It also appears that the test-takers referred to words, expressions, and cohesive devices in the texts they read in the course of their writing. Although they could not utilize the given information to make their essays longer, they might have tried to integrate more diverse, cohesive, and complex linguistic features in their writing.

Although further research is warranted, our findings lend support to the notion that there are linguistic differences in lexical sophistication, cohesion, and syntactic complexity that emerge between those performing independent and integrated tasks. Further, the students' English proficiency moderated differences in the same three aspects of writing performance (i.e., lexical sophistication, cohesion, and syntactic complexity) in the two task groups. The presence of these task effects provides a useful framework forassessing adolescent EFL learners' writing performance, which has rarely been explored in spite of a growing emphasis on CLT and performance assessment in the Korean national curriculum for secondary schools. Theoretically, such a framework contributes to explaining task effects in relation to varying task conditions. Based on the systematic text analysis, it provides specific, detailed linguistic features used differently for different writing tasks, beyond comparisons of ratings between tasks and task types. Moreover, it contributes to a better understanding of English learners' writing ability based on L2

performance assessment.

Writing performance has typically been evaluated based on scoring rubrics that operationalize test constructs. For practicality reasons, holistic scoring rubrics have often been preferred in various testing contexts (Xi, 2007). Even more intricate analytic rubrics often tend to include only brief descriptors, leaving room for raters' own interpretations (Alderson & Banerjee, 2002; Kim, 2001; Pollitt & Murray, 1996). It has been difficult to define differentiate individual writing features in scoring rubrics, or to use them for making score decisions. The analysis adopted in the present study, as well as the suggested framework, could help to compensate for such difficulties.

In addition to its theoretical implications, such a framework could pedagogically contribute to English language teaching in EFL contexts, including Korean secondary schools. A framework derived from empirical evidence allows teachers to understand and diagnose adolescent learners' strengths and weaknesses across different writing tasks and topics, as well as across varying proficiency levels. For instance, teachers might use the indices when assessing student' writing performance before, during, and after teaching writing (e.g., throughout a semester), as well when providing students with formative feedback at each stage. This will not only help teachers understand the nature of writing ability and performance for different groups, but will also prepare them to teach. Subsequently, teachers taking level- and task-appropriate approaches to the teaching of writing and the providing of feedback may lead to positive washback effects on student learning. Ultimately, this framework could help students to build positive attitudes towards writing in English, and contribute to the improvement of their English writing ability.

CONCLUSION

The results of this study indicate that the two writing task types produced statistically significant differences in linguistic features related to lexical

sophistication, cohesion, and syntactic complexity. Further, these features were moderated by students' English proficiency. The more proficient test-takers showed a notable difference in lexical sophistication, cohesion, and syntactic complexitybetween groups, while these increases were notas pronounced among the test-takers who were less proficient in English.

Given a post-test only, quasi-experimental group design, there still remain a number of limitations to the findings reported here. First, the reading prompts used to analyze the differences in writing performance only covered two topics (music and sports). Although the topics were sampled based on both students' interests and the curriculum guide, prompts covering a variety of topics beyond the two given here might elicit a different performance from study. In addition, participants in the present study were recruited based on convenience sampling, and the four intact classes were used to collect student writing performance data. This consequently, limits the external validity of the results. The involvement of a larger selected sample of participants, based on a systematic sampling procedure (e.g., stratified random sampling or cluster sampling), would help to verify the linguistic differences that learners from varying proficiency levels produce across different tasks.

THE AUTHORS

Hye Won Shin is a doctoral candidate in Applied Linguistics at Teachers College, Columbia University. Her interests include L2 vocabulary learning, L2 corrective feedback in writing, and instructed SLA.

Teachers College, Columbia University
525 West 120th St. New York, NY, 10027, USA
hs2509@columbia.edu
82-10-9245-9525

*Hyun Jung Kim*earned an Ed.D. in Applied Linguistics from Teachers College, Columbia University. Currently, she is an assistant professor of the Graduate School of TESOL at Hankuk University of Foreign Studies, Korea. Her research interests include second and foreign language assessment, language test validation, and discourse analysis.

Hankuk University of Foreign Studies
107, Imun-ro, Dongdaemun-gu, Seoul, 130-791, Korea
hkim@hufs.ac.kr
82-10-2819-5302

ACKNOWLEDGEMENTS

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

REFERENCES

Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching, 35*, 79-113.

Bae, H. J. (2012). The effectiveness of genre-based L2 writing instruction on Korean middle school students' writing ability. *English Teaching, 67*(3), 147-180.

Brennan, R. L. (1992). *Elements of generalizability theory*. Iowa City, IA: ACT, Inc.

Brill, E. (1995). Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics, 21*, 543-566.

Brown, J. D., Hilgers, T., & Marsella, J. (1991). Essay prompts and topics: Minimizing the effect of mean differences. *Written Communication, 8*, 533-556.

Carlson, J. G., Bridgeman, B., Camp, R., & Wanders, J. (1985). *Relationship of admission test scores to writing performance of native and nonnative speakers of English* (TOEFL Research Report No. 19). Princeton, NJ: Educational Testing Service.

Choi, C., Park, K., & Kim, C. (1986). *On how to improve foreign language teaching*.

Seoul, Korea:Korean Educational Development Institute.

Cohen, J. (1998). *Statistical power analysis for the behavior sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Crismore, A., Markkanen, R., & Steffensen, M. S. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communications*, *10*(1), 39-71.

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing, 21*, 159-197.

Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, R. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the New TOEFL* (TOEFL Monograph Series MS-30). Princeton, NJ: Educational Testing Service.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005).Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing, 10*, 5-43.

Development Committee of the Sixth Curriculum for High School English. (1992). *The report on the revision of the English curriculum for high school*. Seoul, Korea.

Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, *4*, 139-155.

Ferris, D., & Hedgecock, J. S. (1998). *Teaching ESL composition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 23*, 221-233.

Gebril, A. (2009). Score generalizability of academic writing tasks: Does one test method fit it all? *Journal of Language Testing, 26*, 507–531.

Grasesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers, 36*, 193-202.

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing, 18*, 218-238.

Hamp-Lyons, L., & Kroll, B. (1997). *TOEFL 2000-Writing: Composition, community, and assessment* (TOEFL Monograph Report No. 5). Princeton, NJ: Educational Testing Service.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diveristy.

Language Testing, 19(1), 57-84.

Jeon, J., Lee, W., & Kim, J. (2011). Investigating the English speaking proficiency level Korean people want to achieve. *English Teaching, 66*(2), 273-305.

Jin, K.-A. (2013). *National-level on-line assessment of English four skills: NEAT*. Symposium paper presented at the 35th Language Testing Research Colloquium, Seoul.

Kim, E. (2012). An evaluation of the NEAT from the perspective of secondary school teachers. *English Language Teaching, 24*(2), 205-224.

Kim, H. (2009). An antecedent study of the state-wide English language ability test. *Modern English Education, 10*(2), 44-59.

Kim, M. (2001). Detecting DIF across the different language groups in a speaking test. *Language Testing, 18*, 89-114.

Kim, T.-E. (2007). Exploring teachers' beliefs and knowledge about English writing andtheir writing instruction in ESL context. *English Language & Literature Teaching, 13*(4), 87-108.

Korea Institute for Curriculum and Evaluation. (2011). *The report on English teaching and learning methods in preparation of speaking and writing assessments*. Seoul, Korea: Korea Institute for Curriculum and Evaluation.

Lee, Y. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing, 23*, 131-166.

Lee, S., & Yu, H. (2012). The perception of teachers and students on the NEAT as an alternative English test to the current scholastic one: A case study. *Studies in Foreign Language Education, 26*(2), 93-112.

Lewkowicz, J. (1994). *Writing from sources: Does source material help or hinder students' performance?* Paper presented at the Annual International Language in Education Conference, Hong Kong.

Louwerse, M. (2001). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics, 12*, 291-315.

Malvern, D., & Richards, B. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58-71). Clevedon, UK: Multilingual Matters.

McCarthy, P. M., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing, 24,* 459-488.

McDonough, J., & Shaw, C. (2003). *Materials and methods in ELT: A teacher's guide* (2nd ed.). Malden, MA: Blackwell Publishing.

McNamara, T. F.(1995). Modelling performance: Opening pandora's box. *Applied Linguistics, 16*, 159-179.

McNamara, D. S., & Grasesser, A. C. (2012). Coh-Metrix. In P. M. McCarthy & C.

Boonthum (Eds.), *Applied natural language processing and content analysis:Identification, investigation, and resolution*. Hershey, PA: IGI Global.

Park, T. (2007). *Investigating the construct validity of the community language program (CLP) English writing test*. Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

Park, J., Chang, K.-S., Park, J.-E., & Paek, I.-W. (2012). A survey on Korean high school English teachers' perception of English educationpolicies. *English Teaching, 67*(2), 367-391.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 227-247). Oxford: Blackwell.

Plakans, L. (2009). Discourse synthesis in integrated second language assessment. *Language Testing, 26*(4), 561–587.

Plakans, L., & Gebril, A. (2013). A close investigation into source use in integrated second language writing tasks. *Assessing Writing, 17*, 18-34.

Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, *47*, 101-143.

Pollitt, A., & Murray, N. (1996). What raters really pay attention to. In M. Milanovic & N. Saville (Eds.), *Performance testing, cognition and assessment* (Studies in Language Testing 3, pp. 74-91). Cambridge: University of Cambridge Local Examinations Syndicate and Cambridge University Press.

Powers, D. E., & Fowles, M. E. (1999). Test takers' judgments of essay prompts: Perceptions and performance. *Educational Assessment, 6*, 3-22.

Quellmaz, E. S., Capell, F. J., & Chih-Ping, C. (1982). Effects of discourse and response model on the measurement of writing competence. *Journal of Educational Measurement, 19*, 241-258.

Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes, 9*, 109-121.

Rosenfeld, M., Leung, S., & Oltman, P. (2001). *The reading, writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels* (TOEFL Monograph Report No. 21). Princeton, NJ: Educational Testing Service.

Sekine, S., & Grishman, R. (1995). A corpus-based probabilistic grammar with only two nonterminals. In *Fourth International Workshop on Parsing Technologie*s (pp. 260-270). Prague: Karlovy Vary.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Spaan, M. (1993). The effects of prompt on essay examination. In D. Douglas & C.

Chapelle(Eds.), *A new decade of language testing research* (pp. 98-122). Alexandria, VA: TESOL.

Tedick, D. (1990). ESL writing assessment: Subject-matter knowledge and its impact on performance. *English for Specific Purposes, 9*, 123-143.

Watanabe, Y. (2001). *Read-to-write tasks for the assessment of second language academic writing skills: investigating text features and rater reactions*. Unpublished doctoral dissertation, University of Hawaii, Manoa.

Weigle, S. C., & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing, 21*(2), 118-133.

Weir, C. (1993). *Understanding and developing language tests*. London: Prentice Hall.

Wiseman, C. S. (2008). *Investigating selected facets in measuring second language writing ability using holistic and analytic scoring methods*. Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

Xi, X. (2007). Evaluating analytic scoring for the TOEFL academic speaking test (TAST) for operational use. *Language Testing, 24*, 251-286.

Yang, M. S., & Sohn, Y. G. (2009). A study on the current English writing education insecondary schools. *The Journal of English Education, 38*, 106-135.

Yook, C. (2011). Interactions between beliefs, practices, and perceptions of Korean EFL teachers. *English Teaching, 66*(4), 3-26.