

## **Can a Summary Task Be Valid Writing Assessment for Less-Proficient EFL Students?**

**Nan-Young Ji**  
Korea University

**Ji, Nan-Young. (2011). Can a summary task be valid writing assessment for less-proficient EFL Students? *Modern English Education*, 12(3), 46-64.**

In recent times, integrated reading/writing tasks including summarization have rigorously been incorporated into high-stakes proficiency tests, i.e., iBT TOEFL, TOEIC, TEPS. The present study investigated whether or not summary tasks reflect writing ability rather than reading comprehension of Korean EFL learners with an aim to validate summary tasks as a measure of writing ability and to seek diverse composition tasks more appropriate for EFL learners with low proficiency. A multiple-choice reading comprehension task and an essay task were adopted to assess their reading and writing competence respectively since they are the most common type of assessment. The scores of the three tasks: summary, reading comprehension, and essay were obtained by 50 college students and analyzed through Pearson correlation. The results demonstrated the scores of the summary task are significantly correlated with those of the essay task, while they are little correlated with those of the reading comprehension. Another finding is the degree of relationship between the summary scores and the essay scores is greater within the less proficient group than within the more proficient group. It is concluded that a summary task is valid as a measure of writing competence, particularly for test-takers with low proficiency.

[language assessment/writing assessment/summary task/ summarizing skill/  
/언어평가/영작문평가/요약문과제/요약능력]

### **I. INTRODUCTION**

Summarization has long been considered a critical exercise pedagogically and academically (Dalglish, 1888; Fielden, 1950; Seidlhofer, 1995). In an ESL/EFL setting, summarization has also been considered as something necessary to acquire and develop (Johns, 1988; Johns & Mayes, 1990). In the domain of language assessment, summary

tasks are regarded as effective to measure test-takers' "organizational competence involved in controlling the formal structure of language for producing or recognizing grammatically correct sentences, comprehending their propositional content, and ordering them to form texts" (Bachman, 1990). However, they have become inconsistent in terms of the content they purport to measure. They have been used to measure the degree of reading comprehension in some instances. At other times, they have served as a measure of writing competence. The compounding effect of reading and writing on summarization caused the International English Language Testing System (IELTS) to delete the task from their test battery in 1995. The new iBT Test of English as a Foreign Language (TOEFL), on the other hand, has recently included the task in the writing section.

In recent times, summary tasks or integrated reading/ writing tasks have rigorously been incorporated into the writing sections of high-stakes proficiency tests, i.e., the Test of English for International Communication (TOEIC), and the Test of English Proficiency developed by Seoul National University (TEPS). This has been done for the purpose of enhancing the reliability of measurement (Inn-Chull Choi, 2000) and boosting the wash back effects of the assessment (Cumming et al., 2005; Cumming, Kantor, & Powers, 2001, 2002). Given the current interest in incorporating integrated reading/writing tasks into high-stakes language tests, the need has arisen for conducting empirical research to validate the use of a summary task as a measure of writing ability in the Korean EFL context. An attempt is also needed to seek types of writing assessments adequate for specific groups of test-takers depending on their proficiency. It would be of value to investigate whether a particular type of, or which type of, writing task is more appropriate for test-takers with a certain level of proficiency.

Using summary task as a measure of writing, rather than an essay task, appears to be a better fit for the less proficient group of a majority of Korean secondary EFL students. It requires less cognitive load in comparison with an essay task. Summary task demands the ability to "organize or reorganize what's heard or read" (Vahapassi, 1982) by mainly controlling lexical and syntactic accuracy and diversity. Essay task, however, requires the control of more demanding cognitive tasks such as arguing, opposing, etc., in addition to linguistic ability (Weigle, 2004). It would be sensible to measure test-takers' writing competence in gradual progression: from less demanding toward more demanding. Accordingly, a summary task seems to better fit the less proficient test-takers, since it involves less demanding tasks to control linguistic and organizational elements of writing, while an essay task requires more demanding work to orchestrate all of the linguistic, organizational, and rhetorical components.

Little research, however, has been conducted on the validation of a summary task as a measure of writing competence targeted at Korean EFL students. Historically, some researchers (Taylor, 1986; Yu, 2008) have challenged the validity of a summary task as a

measure of reading comprehension, but few studies have been conducted to investigate directly which competence is more warranted for summary task to measure: reading comprehension or writing competence. This investigation can be made by employing correlation studies. If the scores of a summary task are more correlated to those of either an essay task or a reading comprehension task, it will be concluded that a summary task is more valid as a measure of that particular competence based on the assumption that it shares more features of measuring competence with that task. If one group with a certain level of proficiency demonstrates comparatively higher correlation, it will be concluded that summary task is more valid for that particular group of students.

Therefore, in an attempt to investigate whether or not a summary task is valid for Korean EFL test-takers with intermediate or below-level proficiency, the present study will answer two questions: 1) whether the total scores of a summary task are more correlated with those of an essay task than with those of reading comprehension or vice versa, and 2) in the case that a summary task shows significantly high correlation with either of the two tasks, whether and how the degree of the relationship is different between the two groups: the less proficient group and the more proficient group.

With regard to the research questions, this study hypothesized as follows: 1) if correlation coefficient ( $r$ ) between summary task (one independent variable) and either essay task (one dependent variable) or reading comprehension (another dependent variable) is statistically significantly great, summary task is considered to be more valid to measure the particular competence, 2) if the coefficient between the two correlated tasks is greater in one group than in the other group, it will be concluded that a summary task is more valid for that particular group. This study adopted the Pearson correlation analysis to examine whether and the degree to which the scores of a summary task are correlated with those of an essay task or those of a reading comprehension task.

## II. LITERATURE REVIEW

### 1. Aspects of Summarization

Summarization has been defined in many ways: “to report information using a lot fewer words than were used in the original communication (Wohl, 1978),” “the reproduction of a large amount of information to its most important points (Langan, 1993),” or “a condensed version, in your own words, of the writing of someone else (McAnulty, 1981).” According to these definitions, summarization apparently involves two different types of process: receptive process and productive process. The understanding and identification of important ideas will be regarded as the receptive process, whereas reporting what has been

identified as important ideas using original language involves the productive process.

The summarizing process has been categorized in varying terms. However, it also involves basically two types of process—reception and production. Kintsch and van Dijk (1978) modeled the summarization process in three stages: deletion, generalization, and construction. Brown and Day (1983) specified the process and divided it into six categories: deletion of trivial material, deletion of redundant material, generalization, integration, selection, and invention. Winograd (1984) removed the deleting and generalizing processes and scaled it into four types: reproductions, combinations, run-on combinations, and inventions. Coffman (1994) articulated the productive process into three types: reproduction (restatement or paraphrase), transformation (the combination of two or more content units into one superordinate content unit), and intrusion (a statement that represents the reader's prior knowledge). As research on summarization progresses, the summarizing process or strategy has been categorized and articulated with a focus on the tangible and productive process—i.e., reproduction and transformation, rather than on the conceptual and receptive process—i.e., deletion, selection, and generalization.

## 2. Summary as Reading Assessment

The compound aspect of summarization enables a summary task to serve as a measure of two different skills: receptive skills such as reading comprehension, and productive skills such as writing. On one hand, researchers (Brown & Day, 1983; Kintsch & van Dijk, 1978; Taylor, 1984) believe summarization to be a valid measure of reading comprehension since they assume summary—equivalent to gist—represents what readers have grasped about a text. Moreover, summarization has commonly been used as a measure of reading comprehension in addition to a means to improve reading comprehension (Johns, 1988; Sung-Ae Kim, 2001).

In an ESL/EFL context, on the other hand, summarization in the target language has been regarded to be problematic as a measure of L2 text comprehension. Head, Readence, and Buss (1989) demonstrated that there exists little correlation between a multiple-choice task and a summary task in assessing reading comprehension. Even if learners understand all of the relevant macro-productions, they may fail to reproduce them in the summary. The finding shows the gap between readers' understanding of a text, and expressing what they have understood. It is also argued that there is an overlapping effect of reading and writing competence on summarization performance (Alderson, 1996; Alderson, Clapham, & Wall, 1995). This compounding effect is related to L2 learners' proficiency in their target language (Rost, 1990). In particular, the production process involving L2 writing skills is said to be crucial in summarization (Ellis, 1985).

Therefore, summarization in the native language rather than the target language is

argued to be a better measure of students' reading abilities, since summarization in the target language will limit the learners' ability to demonstrate what they comprehended. By conducting empirical research, Lee (1986) demonstrated that L2 learners showed greater comprehension when they summarized in their native language than in their target language. Kozminsky and Graetz (1986) also concluded that Hebrew-speaking EFL learners' summaries in their native language are of better quality compared to those in the target language.

These findings support that summarization in the native language can better assess readers' text comprehension in that it can remove the effect of L2 writing skills on summarization. Although little empirical research addressed directly the issues of using summary tasks as a measure of foreign language reading comprehension, efforts have been made to seek alternate forms of summary tasks such as summary-choice (Huhta & Randell, 1996) and summary-cloze (Courchene & Bayliss, 1995). This also provides evidence that summarization involves L2 writing skills to manipulate and express the information, and a lack of these skills leads to semantic distortions, inability to paraphrase and other problems (Kirkland & Saunders, 1991).

### 3. Summary as Writing Assessment

The notion that the effect of L2 writing competence on summarization is critical appears to influence high-stakes language tests. For instance, a new iBT TOEFL has recently included integrated reading/writing tasks involving summarization in the writing section. As the incorporation of summarization into high-stakes tests as a measure of writing is diffusing, debates continue on the elements of writing competence that a summary task measures. Hidi and Anderson (1986) contend that summarization is fundamentally different from the other writing tasks in that it involves "operations based on already planned and generated discourse," while the other general writing tasks involve all the processes of planning and generation as well as operations. In response to it, Head et al (1989) claim that the writing ability a summary requires is similar to that required by other composition tasks since "the ability to plan and use important text information in a summary may be a refinement of general writing ability" (Sung-Ae Kim, 2001, p. 570). Thus, a low level of general writing ability would account for a low level of summary writing.

Differences in written discourse between essay and summary were examined (Cumming et al, 2005) as a means to validate incorporating integrated reading/writing tasks into the new iBT TOEFL, since it is argued that one type of writing task such as the argumentative essay fails to fully represent learners' writing abilities (Cumming, 1997; Huot, 1990; Purves, 1992; Raimes, 1990). The findings show that there exists a rhetorical difference

between an integrated reading/writing task and an essay task; in dependent essay tasks elicit production of arguments and the integrated tasks prompt production of textual information as originally expected. However, it demonstrates the difference in their ability to use language between high and low groups consistently across the differing task types. Across the task types, test-takers with a higher level of proficiency tend to produce longer and more complex sentences, and to demonstrate more varied vocabulary and greater grammatical accuracy, compared to those with lower proficiency. This tendency provides support for the existence of shared features in writing qualities between an integrated reading/writing task and an essay task.

Moreover, it is noteworthy that test-takers are more likely to use longer and a wider range of words and write longer and more clauses in the integrated tasks than they do in the independent tasks (Cumming et al, 2005). This finding allows us to presume that test-takers may focus better on their use of language with freedom from being concerned about rhetorical issues such as argumentation, and they benefit from the source text in terms of their use of language. This allows us predict that an integrated summary task may be a more appropriate measure of writing for groups of lower-proficiency students than an essay task, whereby their composition skills can be measured with their cognitive load alleviated.

Analysis of Korean EFL learners' summary protocols by Sung-Ae Kim (2001) illustrates that they depend greatly on deletion rules and struggle to transform source text into their own text. This suggests Korean EFL students' lack of summary skills, which brings about the importance of instruction and practice. Sung-Ae Kim points out Korean teachers seldom pay attention to the development of summarization skills, and Korean L2 students rarely have the opportunity to develop summary writing skills. She also claims that summarization skills should be taught directly and explicitly. Considering the reality Korean EFL students and teachers face, validation of a summary task as a writing assessment for students with low-level proficiency will be of more significance.

### **III. METHOD**

#### **1. Participants**

A total of 50 college students who enrolled in an English grammar course at a university located in Seoul participated in this study as test-takers. Their grade levels range from freshmen to seniors. They were allowed to use pseudonyms instead of authentic names. Another group of participants consisted of two native-speaking raters. Both raters were selected based on the criteria: different gender (one male and the other female), English

major, EFL teaching experience. One female rater has a Master's degree in English and American Literature and a substantial amount of teaching experience at a language institute in Seoul and at a high school in Tampa, Florida. The other male rater majored in English Education and has been teaching English at a college in Seoul.

## 2. Instrument

A language proficiency test was given to all the test-takers for the purpose of group placement: the less proficient group and the more proficient group. For a proficiency test, one mock test of TEPS was administered. The full version of the mock test was reduced to half of the questions to fit the total exam period of 40 minutes including three written sections: vocabulary, grammar, and reading comprehension, due to the fact that this study involves only the written domain of language ability. The test was composed of 60 multiple-choice question items: 24 grammar, 24 vocabulary, and 12 reading comprehension questions. Ten minutes after the proficiency test, one summary and one essay task were administered. Each task has one prompt. The essay topic was about whether recommendation letters should be open to students or kept confidential (see Appendix A). The source text for summarization was a four-paragraph-long passage with a topic similar to the essay topic (see Appendix B). The topics were selected based on the test takers' interest and background knowledge. By selecting the topic relevant and interesting to education majors at college level, it was intended to increase their attention and willingness to read the text (Hidi, Renninger, & Krapp, 1992; Langer, 1984). The test-takers were given a total time limit of 40 minutes—15 minutes for the summary and 25 minutes for the essay task.

The proficiency test measuring the three domains of language competence—grammar, vocabulary, and reading comprehension—was objectively scored based on the answer key provided by Wolgan TEPS. Two writing tasks (summary and essay) were scored by two native English raters based on the two different sets of rubrics: Summary Scoring Rubric (see Appendix C) and Essay Scoring Rubric (see Appendix D), both of which are provided by the American Library Association, and both rubrics have six categories and four scoring scales. In an attempt to minimize raters' subjective interpretation of scoring rubrics and accordingly increase the inter-rater reliability, the two raters were oriented and trained via e-mails and phone calls until their pilot scoring showed consistency between the two raters' scores. To be specific, the raters were required to score the first three students' writings based on the rubrics, providing their reasons for scoring. To diminish the split scores in the first pilot scoring set between the two raters, e.g., 2/4, 3/1, the raters were provided with detailed guidelines, re-trained using them, and asked to conduct another pilot scoring using a different set of writings. This process continued until few split scores were shown in the

working pilot scoring set.

### 3. Procedure

This study adopted the Pearson correlation analysis to examine whether and the degree to which the scores of a summary task are correlated with those of an essay task or those of a reading comprehension task. Additionally, correlation coefficients were calculated within the two groups, the less proficient group and the more proficient group, in order to investigate whether the degree of correlation of the two sets of scores is different between the two groups and which group shows greater correlation. Thus, this correlation study involved the following steps. 1) Cronbach's alpha was calculated to confirm whether or not the scores of each task are consistent between the two raters since adequate inter-rater reliability will make this study legitimate. 2) Correlation coefficients obtained between 50 scores of the summary task and 50 scores of the reading comprehension task were compared with those gained between the summary task and the essay task. 3) The two groups: more proficient group and less proficient group were formed based on the mean score of the total 50 scores of the proficiency test.

As shown in Table 1, the more proficient group whose average scores are above the mean ( $M = 130.92$ ) and the less proficient group whose average scores fall below the mean. While both groups show similar variability in scores, a t-test proved that the two groups are significantly different in proficiency level,  $t(24) = 15.85$ ,  $p < .01$ . 4) Lastly, correlation coefficients were obtained between 25 scores of the summary task and 25 scores of the essay task within the two different proficiency groups and they were compared.

Table 1  
*More Proficient Group vs. Less Proficient Group*

	N	Minimum	Maximum	Mean	Std. Deviation
Total test-takers	50	60.00	198.00	130.9200	25.10455
More proficient group	25	132.00	198.00	149.9200	16.79762
Less proficient group	25	60.00	130.00	111.9200	15.89004
Valid N (listwise)	25				

## IV. RESULTS

In analysis of inter-rater reliability, the two raters showed consistent agreement in

scoring both of the two tasks. Alpha value indicating inter-rater reliability was .7 in the summary task and .8 in the essay task. The values greater than .7 mean adequate level of reliability and those greater than .8 mean very good according to a statistician, Kline (2011). Therefore, the scores of the two tasks by the two raters proved to be reliable and this result constitutes to legitimating the subsequent analysis. It is notable that the inter-rater reliability of the summary task was lower than that of the essay task. This is probably due to the fact that the raters were less familiar or less experienced with the summary rubric than with the essay rubric. When they were asked how long it took them to score summaries versus essays, they stated that they needed more time for summaries, because in the beginning they struggled to familiarize themselves with the summary rubric. Their testimonies reflect Korean EFL context where summarization is rarely instructed or evaluated.

Correlation coefficients calculated between the total scores of the summary task and the total scores of the reading comprehension test was only  $r = .24, p < .10$ , which means weak correlation (see Table 2). In contrast, correlation between the total scores of the summary task and the essay task was  $r = .68, p < .01$ , which indicates statistically significant at the probability level of 0.01 (see Table 3). This result demonstrated that the summary task is correlated with the essay task while it is little associated with the reading comprehension test. Thus, it can be concluded that the summary task is more valid to assess writing ability rather than reading comprehension, based on the hypothesis that if correlation coefficient ( $r$ ) between the summary task and either the essay task or reading comprehension is statistically significantly great, the summary task is considered to be more valid to measure either ability.

Table 2

*Correlation: Summary vs. Reading Comprehension*

	Summary	Reading Comprehension
Summary (N=50)	1	.235
Reading Comprehension (N=50)	.235	1

Table 3

*Correlation: Summary vs. Essay*

	Summary	Essay
Summary (N=50)	1	.684(**)
Essay (N=50)	.684(**)	1

\*\* Correlation is significant at the 0.01 level (2-tailed).

In order to study the discrepancy between the two groups in light of the degree of correlation, correlation coefficients were calculated between the two tasks within each group: the more proficient group and the less proficient group. As for the more proficient group, the correlation coefficient between the two sets of scores decreased to  $r = .57$ ,  $p < .01$  (see Table 4). For the less proficient group, the correlation coefficient between the two sets of scores rose to  $r = .71$ ,  $p < .01$  (see Table 5). In other words, the more proficient group showed a moderate level of correlation, whereas the less proficient group demonstrated a higher level of correlation, as illustrated in Figure 1 and 2. Given the hypothesis, it can be concluded that the summary task is a more valid writing assessment for the less proficient group than for the more proficient group.

Table 4

*More Proficient Group: Summary vs. Essay*

	Summary	Essay
Summary (N=25)	1	.570(**)
Essay (N=25)	.570(**)	1

Table 5

*Less Proficient Group: Summary vs. Essay*

	Summary	Essay
Summary (N=25)	1	.709(**)
Essay (N=25)	.709(**)	1

\*\* Correlation is significant at the 0.01 level (2-tailed).

In sum, following are the answers to the research questions devised to investigate whether or not a summary task is valid for EFL test-takers with intermediate or below proficiency level: (1) The total scores of the summary task are significantly correlated with those of the essay task, while they are little correlated with those of reading comprehension. (2) The degree of relationship between the summary scores and the essay scores is different between the two groups examined and it is greater within the less proficient group than the more proficient group. This suggests that the summary task is more relevant to assess writing than reading comprehension, and that it better fits the less proficient group.

## V. DISCUSSION

The results of the present study support that a summary task is more valid as a writing

assessment than as an assessment of reading comprehension. The weak relationship between the summary scores and those of reading comprehension may be indicative of a crucial effect of L2 writing skills on summarization (Ellis, 1985; Kirkland & Saunders, 1991; Vahapassi, 1982). Insufficient L2 writing skills will make readers with good comprehension fail to produce what is read and even distort what is accurately understood by choosing inappropriate synonyms and/or sentence structures. Likewise, this critical effect of L2 writing skills on summarization accounts for the strong relationship between the summary scores and the essay scores. This finding is also supportive of the claim that to summarize requires a writing ability similar to that required by other composing tasks (Head, Readence, & Buss, 1989).

Another notable finding is that the degree of the relationship between the summary scores and the essay scores differs depending on the proficiency level. It increases for the less proficiency group while it decreases for the more proficient group. The lower degree of correlation for the more proficient group can be probably due to a rhetorical difference existing between an integrated reading/writing task and an essay task (Cumming, 1997; Huot, 1990; Purves, 1992; Raimes, 1990). Summarization involves “operations based on already planned and generated discourse,” unlike an essay involving “all the process of planning and generation as well as operations” (Hidi & Anderson, 1986). The more proficient test-takers may be aware of and be able to control textual requirements for the two different types of writing demand. This may account for the weakened correlation between summary and essay scores.

On the other hand, the higher degree of correlation for the less proficient group might reflect that test-takers with insufficient L2 proficiency may be ignorant of rhetorical requirements or unable to control them while managing to use the target language. This may account for the consistency of their scores between the two writing tasks. We can thus speculate that a summary task may allow them to focus better on their use of language as they are relieved of concerns about rhetoric for argumentation and other illocutionary elements, and provided with the source text facilitating their use of language. This speculation corresponds to another research finding that test-takers tend to show better language use in the integrated reading/writing tasks than in the independent tasks (Cumming et al, 2005). Thus, it can be concluded that an integrated reading/writing task, or a summary task, is more warranted as a measure of writing for the lower proficiency test-takers, since it allows them to focus more on grammatical components involving lexical/syntactic accuracy and diversity rather than textual/rhetorical elements demanding greater cognitive load.

This research-based validation of a summary task as a measure of writing ability is expected to influence Korean EFL instruction and curriculum. Although a large amount of opportunity has been given to improve their text comprehension, little chance has been

given for Korean EFL learners to summarize or retell what they comprehend. Lack of practice and instruction will sufficiently account for their lack of summarization skill: failure to transform what is identified as a gist in their own words and overdependence on deletion rules (Sung-Ae Kim, 2001) or copying the sentences identified as main ideas. However, popular high-stakes proficiency testing batteries such as TOEFL, TOEIC, and TEPS have already included performance tasks to measure speaking and writing ability. Moreover, performance tasks are to be added to the government-developed high-stakes English test which is referred to as the National English Ability Test (NEAT)/National English Proficiency Test (NEPT). According to the Ministry of Education, Science and Technology, the NEAT will replace the current English language section of the College Scholastic Ability Test (CSAT) in 2016 if pilot projects are successfully completed ([http://www.koreatimes.co.kr/www/news/nation/2010/09/113\\_72541.html](http://www.koreatimes.co.kr/www/news/nation/2010/09/113_72541.html)).

Considering that more societal attention is being paid to performance skills, the secondary school classroom and curriculum has imminent responsibility to prepare students in response to those societal needs. Sung-Ae Kim (2001) recommends teaching summarization skills explicitly in a school setting. The teaching practice of summarization will be motivated and facilitated by incorporating a summary task into assessment for the Korean secondary school curriculum. Inclusion of summarization practice and assessment in the secondary curriculum will become a promising solution to fulfill the responsibility particularly due to its unique and valuable characteristic of reading-writing connection. Summary procedures naturally allow students to link writing to reading and vice versa (Hirvela, 2004; Williams, 2003). In other words, students are encouraged to read text from the writer's viewpoint and apply the same organizational/rhetorical structures encountered in their reading of the assigned texts in their writing summaries. At the same time, they use their knowledge of rhetorical structure of writing to better understand their reading by looking for topic sentences and supporting ideas. The reciprocal effect of summarization will constitute a robust development of basic composition skills which is a prerequisite to the cognitively more demanding essay task.

## VI. CONCLUSION

In sum, this study revealed that significant correlation exists between a summary task and an essay task, whereas little correlation exists between a summary task and reading comprehension. The finding indicates that L2 writing skills play a vital role in summarizing L2 text. Another revelation is that the correlation between a summary task and an essay task is greater in the less proficient group. It suggests that test-takers with insufficient L2 proficiency may not be able to respond to the rhetorical difference between

summary and essay, being busy operating the target language, or they are ignorant of the rhetorical difference. Accordingly, it can be concluded that a summary task is valid as a measure of writing competence, not of reading comprehension, and that it is more valid for test-takers at a proficiency level of intermediate or below.

It is hoped that this study be a catalyst for diversifying and expanding the domain of direct writing assessment for Korean EFL students, and particularly for seeking diverse types of writing tasks corresponding EFL test-takers' varying range of proficiency and readiness. It is also expected that this study positively affect learning and teaching practices of writing in the school setting – in particular the secondary school setting - since the students will benefit from summarization in terms of reading and writing connection. They will improve composition skills as well as reading skills in an integrated and reciprocal way: learning how to write from reading and how to read from writing. In order to achieve it in a Korean EFL setting, secondary school teachers' role is essential and they should be taught how to teach writing (Hyun-Jin Kim, 2009). Likewise, instruction of how to teach summarization should also be implemented in the training program for pre- and in-service L2 writing teachers.

This study, however, is restricted to only 50 subjects and they are rather homogeneous since they are enrolled in the same department of the same university. Studies involving a greater number of and a wider range of participants, in particular secondary level students, are recommended. As Eunsook Shim (2010) points out, a majority of studies on Korean L2 writing is restricted in the tertiary settings and only sixteen percent of studies were conducted in a secondary school setting. Involving primary and secondary school students will enhance validity and reliability of the result.

Another restriction of this study is the limited number of question items and prompts. More numbers of comprehension question items and multiple prompts for summary and essay tasks will enhance reliability of the results. Furthermore, employing different types of task for reading comprehension is recommended for future studies. In this study, a multiple choice task was adopted to assess reading comprehension since it is the most prevailing type of task to measure receptive skills such as reading comprehension. On the other hand, summary and essay skills were assessed using performance tasks. Inconsistency of task types might have been influential to the result. Apart from topical knowledge and personal characteristics, test method including task types - multiple choice or open-ended—is a critical element to affect test-takers' performance (Bachman, 1990). That is, a reading comprehension task in a productive mode, summarizing in L1 per se, could reveal different pattern of their reading comprehension scores and might transform the results of this study. Thus, involving varied types of task are recommended to validate a summary task for assessing writing.

## REFERENCES

- Alderson, J. C. (1996). The testing of reading. In C. Nuttall (Ed), *Teaching reading skills in a foreign language* (pp. 212-228). London: Heinemann.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. New York: Oxford University Press.
- Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, 22(1), 1-14.
- Choi, Inn-Chull. (2000). Construct validation of English simulated oral proficiency interview test method facets. *Journal of the Applied Linguistics Association of Korea*, 16(1), 215-246.
- Coffman, G. A. (1994). The influence of question and story variations on sixth graders' summarization behaviors. *Reading Research and Instruction*, 34(1), 19-38.
- Courchene, R., & Bayliss, D. (1995). Summary cloze: What is it? What does it measure? In R. Courchene, S. Burger, C. Cornaire, R. LeBlanc, S. Paribakht & H. Seguin (Eds.), *Twenty-five years of second language teaching at the university of Ottawa* (pp. 305-327). Ottawa: Second Language Institute, University of Ottawa.
- Cumming, A. (1997). The testing of second-language writing. In D. Corson (Series Ed.) & C. Clapham (Vol. Ed.), *The encyclopedia of language and education: Vol. 7 Language assessment* (pp. 51-63). Dordrecht: Kluwer.
- 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(1), 5-43.
- Cumming, A., Kantor, R., & Powers, D.E. (2001). *Scoring TOEFL essays and TOEFL 2000 prototype writing tasks: An investigation into raters' decision making and development of a preliminary analytic framework (TOEFL monograph series, report no. 22)*. Princeton, NJ: Educational Testing Service.
- Cumming, A., Kantor, R., & Powers, D. E. (2002). Decision making while rating ESL/EFL writing tasks: A descriptive framework. *Modern Language Journal*, 86(1), 67-96.
- Dalgleish, W. S. (1888). *Introductory textbook of English composition: Based on grammatical synthesis, with appendices treating of figures of language and prosody*. Halifax, Nova Scotia: A & W Mackinlay.
- Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.
- Fielden, F. J. (1950). *A guide to précis-writing*. London: University Tutorial Press.

- Head, M. H., Readence, J. E., & Buss, R. R. (1989). An examination of summary writing as a measure of reading comprehension. *Reading Research and Instruction, 28*(4), 1-11.
- Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instructions. *Review of Educational Research, 56*, 473-493.
- Hidi, S., Renninger, K. A., Krapp, A. (1992). The present state of interest research. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 433-446). Hillsdale, NJ: Lawrence Erlbaum.
- Hirvela, A. (2004). *Connecting reading and writing: In second language writing instruction*. Ann Arbor, MI: The University of Michigan Press.
- Huhta, A., & Randell, E. (1996). Multiple-choice summary: A measure of text comprehension. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 94-110). Clevedon: Multilingual Matters.
- Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research, 60*, 237-263.
- Johns, A. M. (1988). Reading for summarizing: An approach to text orientation and processing. *Reading in a Foreign Language, 4*(2), 79-90.
- Johns, A. M., & Mayes, P. (1990). An analysis of summary protocols of university ESL students. *Applied Linguistics, 11*(3), 253-271.
- Kim, Hyun-Jin. (2009). Beliefs about language teacher competence and microteaching in the in-service teacher training program. *Modern English Education, 10*(3), 42-61.
- Kim, Sung-Ae. (2001). Characteristics of EFL readers' summary writing: A study with Korean university students. *Foreign Language Annals, 34*(6), 569-581.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.
- Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly, 25*(1), 105-121.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford.
- Kozminsky, E., & Graetz, N. (1986). First vs second language comprehension: Some evidence from text summarizing. *Journal of Research in Reading, 9*(1), 3-21.
- Langan, J. (1993). *Ten steps to advancing college reading skills* (2nd ed.). Marlton, NJ: Townsend Press.
- Langer, J. A. (1984). Examining background knowledge and text comprehension. *Reading Research Quarterly, 19*(4), 468-481.
- McAnulty, S. J. (1981). Paraphrase, summary, précis: Advantages, definitions, models. *Teaching English in the Two-year College, 8*(1), 47-51.

- Lee, J. F. (1986). On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition*, 8(2), 201-211.
- Purves, A. (1992). Reflections on research and assessment in written composition. *Research in the Teaching of English*, 26, 109-123.
- Raimes, A. (1990). The TOEFL test of written English: Causes for concern. *TESOL Quarterly*, 24, 427-442.
- Rost, M. (1990). *Listening in language learning*. Harlow: Longman.
- Seidlhofer, B. (1995). *Approaches to summarization: Discourse analysis and language education*. Tübingen: Gunter Narr.
- Shim, Eunsook. (2010). Research trends in writing teacher education. *Modern English Education*, 11(1), 36-50.
- Taylor, K. K. (1984). The different summary skills of inexperienced and professional writers. *Journal of Reading*, 27(8), 691-699.
- Taylor, K. K. (1986). Summary writing by young children. *Reading Research Quarterly*, 21(2), 193-208.
- Vahapassi, A. (1982). On the specification of the domain of the school writing. *Evaluation in Education: An International Review*, 5, 265-289.
- Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9(1), 27-55.
- Williams, J. D. (2003). *Preparing to teach writing: Research, theory, and practice* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Winograd, P. N. (1984). Strategic difficulties in summarizing texts. *Reading Research Quarterly*, 19(4), 404-425.
- Wohl, M. (1978). *Techniques for writing: Composition*. Rowley, MA: Newbury.
- Yu, G. (2008). Reading to summarize in English and Chinese: A tale of two languages? *Language Testing*, 25(4), 521-551.

## APPENDIX A

### Essay Prompt

**Read the essay question below and express your opinion clearly and persuasively in about 300 words.**

Some people argue that recommendation letters written by teachers need to be open to students or parents. Some people assert that recommendation letters should be kept confidential. Which side are you on? Provide your rationale using relevant examples.

## APPENDIX B

## Summary Prompt

**Summarize the passage in about 50 words. Make sure you include the main idea and the necessary details. Remember to paraphrase.**

In this age of information most of us probably have a personal history tucked away somewhere on computer disks. For students, the history may consist of school records, test scores, and ratings by teachers on everything from citizenship to punctuality. Teachers and other staff members judge a student's character and potential, and others use those judgments to decide whether the student should go to this school or get that job. Certainly we need some system of exchanging information about one another; otherwise, we would hire only our friends or attend only those schools where enough people knew us to vouch for us. However, the kind of information in school records may be very imperfect, and the danger that it will be misinterpreted or fall into the wrong hands is great.

In the early 1970s, a series of situations came to light in which information was poorly used or parents and students were denied access to records (for example, when a diagnosis was used to justify sending a child to a class for students with mental retardation). In response, Congress passed the Family Educational Rights and Privacy Act in 1974. The act, also known as the Buckley amendment, outlines who may and who may not see a student's record and under what conditions. A clear winner from this legislation is parents, who previously were kept from many of the officially recorded judgments that affected their children's futures. The amendment states that federal funds will be denied to a school if it prevents parents from exercising the right to inspect and review their children's educational records. Parents must receive an explanation or interpretation of the records if they so request. However, the Buckley amendment does not give parents the right to see a teacher's or an administrator's unofficial records. For instance, a teacher's private diary of a class's progress or private notes about a particular child may not be inspected without the teacher's consent.

Although the Buckley amendment has undoubtedly reduced the potential for abuse of information, it has had a somewhat chilling effect on teachers' and others' willingness to be candid\* in their judgments when writing student recommendations for jobs or colleges. Because students may elect to see a teacher's letter of recommendation, some teachers choose to play it safe and write a vague, general letter that lacks discriminating judgments, pro or con, about the student. In effect, some faculty members and other recommenders have adopted the attitude "Well, if a student doesn't trust me enough to let me write a confidential recommendation, I'll simply write an adequate, safe recommendation." (Adapted from Ryan and Cooper, *Those Who Can, Teach*, pp. 434, 436.)

## APPENDIX C

## Summary Scoring Rubric

	4	3	2	1
<b>IDEA &amp; CONTENT</b>	-Clearly states the main idea of the passage. -Clearly states significant details.	-States the overall idea of the passage. -States significant details.	-Somewhat states the overall idea of the passage. -Lists some details.	-Rarely states the overall idea of the passage. -Lists one or none of the details.
<b>ORGANIZATION</b>	-Clearly follows the logical sequence of the original passage.	-Most of the sequencing follows the original passage.	-Some of the sequencing follows the original passage.	-Sequencing is erratic and does not follow the original passage.
<b>VOICE</b>	-Formal voice is clearly objective and accurate for the passage.	-Formal voice is clearly objective and accurate for the passage most of the time.	-Limited formal voice is somewhat objective and accurate for the passage.	-Lack of formal voice appropriate for the passage.
<b>SENTENCE FLUENCY</b>	-Sentences flow naturally. -Sentence length and type vary.	-Most sentences flow. -Sentence length and type vary.	-Limited number of sentences flow. -Limited number of sentences vary in length and type.	-Sentences are choppy and uneven -Sentence length and type are mainly identical.
<b>WORD CHOICE</b>	-Clearly paraphrase with vivid descriptive language.	-Usually paraphrase with some descriptive language.	-Sometimes paraphrase with some descriptive language.	-Uses the original language of the passage. Or uses inappropriate words or phrases.
<b>CONVENTIONS</b>	-Contains few, if any, errors in punctuation, capitalization, spelling, and grammar.	-Contains some errors in punctuation, capitalization, spelling, and grammar that do not interfere with the meaning.	-Contains several errors in punctuation, capitalization, spelling, and grammar that may interfere with the meaning.	-Contains many errors in punctuation, capitalization, spelling, and grammar that interfere with the meaning.

## APPENDIX D

## Essay Scoring Rubric

	4	3	2	1
<b>FOCUS ON TOPIC</b>	-There's clear, well-focused topic. -Main idea stands out and is supported by detailed information.	-Main idea is clear but the supporting information is not specific.	-Main idea is somewhat clear but there is need for more supporting information.	-Main idea is not clear. There is seemingly random selection of information.

<b>ORGANIZATION (INTRODUCTION)</b>	-The introduction is inviting, states the main topic, and previews the structure of the paper.	-The introduction clearly states the main topic, and previews the structure of the paper, but it is not inviting to the reader.	-The introduction states the main topic, but does not adequately preview the structure of the paper nor is it inviting to the reader.	-There is not clear introduction of the main topic or structure of the paper.
<b>ORGANIZATION (SEQUENCING)</b>	-Details are placed in a logical order and the way they are presented effectively keeps the interest of the reader.	- Details are placed in a logical order but the way they are presented sometimes makes the writing less interesting.	-Some details are not logical or expected order, this distracts the reader.	- Many details are not logical or expected order. There is little sense that the writing is organized.
<b>SENTENCE FLUENCY</b>	-All sentences sound natural. Each sentence is clear and has an obvious emphasis.	- Almost all sentences sound natural but 1 or 2 are awkward or difficult to understand.	- Most sentences sound natural but several are awkward or difficult to understand.	-Sentences sound awkward and difficult to understand.
<b>WORD CHOICE</b>	-Uses various vivid words and phrases accurately.	-Uses various vivid words and phrases, but occasionally the words are used inaccurately	-Uses some vivid words and phrases but lacks variety and accuracy.	-Uses limited vocabulary, which blocks communication.
<b>CONVENTIONS</b>	-Contains few, if any, errors in punctuation, capitalization, spelling, and grammar.	-Contains some errors in punctuation, capitalization, spelling, and grammar that do not interfere with the meaning.	-Contains several errors in punctuation, capitalization, spelling, and grammar that may interfere with the meaning.	-Contains many errors in punctuation, capitalization, spelling, and grammar that interfere with the meaning.

Nan-Young Ji  
 English Education Department, Korea University  
 Anam-dong, Seongbuk-gu, Seoul, 136-701, Korea  
 C.P.: 010-3695-0509  
 Email: jinnymecc@korea.ac.kr

Received 15 July 2011

Revised 3 September 2011

Accepted 14 September 2011