



Investigation Into Validity of Paraphrasing Task as a Writing Performance Test Item for EFL Learners*

Nan-Young Ji

Korea Polytechnic University

ARTICLE INFO

Received 25 September 2017

Revised 7 May 2018

Accepted 17 May 2018

Examples in: English

Applicable Languages: English

Applicable Levels: Secondary

KEYWORD

language assessment

writing assessment

performance test items

retelling

paraphrasing

영어평가 / 작문평가

수행평가 / 문장변환

ABSTRACT

Paraphrasing quality has proven to be highly associated with two major linguistic components: lexical competence and syntactic competence. Nonetheless, few attempts have been made to view the skill as a window through which learners' interlanguage can be estimated. How learners manipulate sentence structures and vocabulary in retelling may represent their current level of language proficiency. Therefore, with an aim to verify whether paraphrasing tasks are legitimate as a writing test item to accurately identify the learners' productive language abilities, a correlation study was conducted with 364 test-takers ranging from grade 7 to university freshmen. The scores the learners earned from paraphrasing tasks were compared with those obtained from self-assessments of their English abilities in the case of the secondary school students and from TOEIC in the case of the university students. Paraphrase rating scales adopted in this study were developed by school teachers, considering the range of Korean secondary students' English proficiency. It has been revealed that paraphrasing task has the potential as a valid writing test item as proven by statistically significant correlation coefficients between two sets of scores.

I. INTRODUCTION

Types of writing performance test items are highly limited in secondary school settings in Korea. Limited kinds of assessments have been employed to test middle and high school students' writing performance: to copy sentences from a textbook, to rearrange given words to compose a sentence, to identify and correct inaccurate parts of sentences, etc. Some adventurous teachers adopt more challenging modes of assessment, such as paragraph building or essay tasks.

These test items, however, have limitations that should not be ignored. Memorize-and-copy modes of performance test items, which are commonly selected by teachers on the ground of facile scoring and objective scores,

can hardly assess the degree to which students can utilize their grammatical and lexical knowledge to compose sentences. Alternately, paragraph level performance test items are difficult to score fairly and objectively. In addition, raters' subjectivity, low score reliability, time and cost inefficiency, and impracticality tend to cause teachers to be reluctant to administer these tasks.

This triggers a demand for more diverse assessment tools, which teachers can choose from depending on their evaluation purposes and students' levels of proficiency. Moreover, the introduction of a criteria-referenced KSAT (Korean SAT) and the increasing impact on their GPAs of in-class performance tests make it more vital to seek a variety of writing performance test items.

Few researchers have viewed paraphrasing skill as an

* This research has been supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF--2015S1A-5B5A07042791).

Nan-Young Ji

Department of Nano-Optical Engineering Korea Polytechnic University 237 Sangidaehak-ro, Siheung-si, Gyeonggi-do, Korea, 15973

Tel: (031) 8041-1375 / Email: Jinnymecc@kpu.ac.kr

indicator of learners' interlanguage proficiency although some researchers have observed and analyzed the degree to which learners use their grammatical and lexical knowledge (N. Ji, 2012; Keck, 2006; McCarthy, Guess, & McNamara, 2009; Shi, 2004). With an intent to broaden the range of writing assessment tasks, this study was conducted to confirm the potential of paraphrasing tasks as a valid assessment form for Korean EFL learners' productive language skills.

II. LITERATURE REVIEW

1. Paraphrasing Task

The importance of paraphrasing has been emphasized by many researchers and teachers as a way to avoid plagiarism (Keck, 2006; McInnis, 2009), to improve comprehension, to clarify ideas, and even to improve memory (Reid, Lienemann, & Hagamann, 2013). Similarly, emphasis has been on teaching paraphrasing skills since English speakers lack sentence reformulation techniques, whether they use English as L1 or L2 (McInnis, 2009). Research conducted by Connor and McCagg (1983) demonstrated that L2 writers have a stronger tendency to be reluctant to transform sentence structures. They conducted a study with two aims: first, to investigate whether or not L2 writers' native language and their cultural background and influence the argumentation structure of their writing in L2 when required to write what they remember after reading an expository prose and second, to investigate what traits of paraphrasing experienced ESL writing teachers regard as positive or negative. It is revealed that their L1 does not apparently influence their L2 writing and that L2 writers struggle to reconstruct the source text without distorting its original meaning in paraphrasing task.

Attempts have also been made to identify characteristics of paraphrases and to specify the process in which sentences are rephrased. Russo and Pippa. (2004) regarded productive competence of synonyms as a strong predictor of paraphrasing skills. Based on a widely recognized notion that 'synonymy is a fundamental semantic relation in natural language' (Žolkovskij & Mel'čuk, 1967), some researchers such as Polguère (2004) and Milićević (2008, 2009) proposed incorporating the Meaning-Text theory (Mel'čuk, 1997) to model paraphrasing, which involves lexical functions and paraphrasing rules. According to the theory, meaning is defined as 'the invariant of paraphrases' and the production of speech is regarded as 'virtual paraphrasing'. That is, delivery of meaning is 'a series of choices between possible synonymous expressions of a starting linguistic meaning'.

With regard to L2 writers, studies have been carried out to better understand the paraphrasing strategies they adopt (Keck, 2006; McCarthy et al., 2009; Shi, 2004). Shi (2004) introduced a systematic and simplified method of analyzing paraphrasing strategies. The method outlines

how to identify a unit of textual borrowing and code it into categories: exactly copied, slightly modified, and closely paraphrased, in order to examine the effect of both L1 and task types—summarization versus opinion essays—on ESL college students' word usage from source readings of their English writing. It was found that a summarization task elicits more textual borrowing than an essay task, and that Chinese students use source texts usually without citing references for either task. Keck (2006) further elaborated on the classification of paraphrasing strategies in order to compare L1 and L2 writers' use of paraphrasing strategies in summarization tasks. This study showed that L2 writers strongly depend on direct copying or minimal modification of words, unlike L1 writers, and suggested that L2 writers' paraphrasing strategies may be influenced by their linguistic competence.

McCarthy et al. in an attempt to examine the potential of an automated assessment of the quality of paraphrasing, identified two critical components constituting quality paraphrasing: the degree of variety in sentence structure and semantic completeness closely interrelated with lexical dissimilarity. This demonstrates that paraphrasing techniques can be a function of syntactic and lexical knowledge. N. Ji (2012) also proved that the paraphrasing techniques L2 writers adopt are correlated with their linguistic proficiency. Substantial modifications of sentence structure are more often attempted in a high proficiency group than in a low proficiency group. Given this, it would not be an exaggeration that the quality of paraphrases can be a determiner by which L2 writers' linguistic proficiency can be estimated. Nonetheless, little research has been conducted to investigate into the potential of paraphrasing tasks as test items to assess L2 writers' levels of proficiency.

2. Self-evaluation

This research involves a correlational study between self-evaluated writing ability and paraphrasing ability evaluated by a rater. In this sense, prior research on the accuracy of self-assessment needs to be reviewed. Many studies show self-assessments are reliable and consistent. Ross, Rolheiser and Hogaboam-Gray (1999) demonstrated that students in grades 4-5 could assess their own English proficiency across a range of tasks with highly reliable correlation coefficients ($\alpha=.84$). Fitzgerald, White, and Gruppen (2003) examined the self-assessments of medical students across two tasks and found that students' self-assessments were consistent over different skills and tasks. Sung, Chang, Chiou, and Hou (2005) examined how accurately 14-15 year-old students assess the quality of their web-designs and found no significant differences across occasions: after completing their designs, after viewing the designs of others in their own group and after viewing the best and worst designs in the class. Blatchford (1997) also reported that self-assessments from the group of older students (ages 11 to 16) were stable. However, the group of younger students showed mixed evidence and low performing students tend to show inconsistent,

inflated, unreliable scores (Claes & Salame, 1975; Keil, McClintock, Kramer, & Platow, 1990; Kwok & Lai, 1993; Laveault & Miles, 2002; Mitman & Lash, 1988; Ng & Earl, 2008; Sung et al., Sung, Chang, K. E., Chang, T. H., & Yu, 2010).

3. Research Questions

This study aims to confirm whether paraphrasing tasks can be valid test items to accurately measure Korean EFL learners' writing ability. In response to the question, it is hypothesized that if the scores from paraphrasing tasks are correlated with the scores from Korean EFL learners' self-assessments of their English ability, the tasks can be potential writing performance test items to measure Korean EFL learners' English proficiency. This hypothesis is based on prior research findings that learners' self-assessment is reliable (Fitzgerald, White, & Gruppen, 2003; Ross, Rolheiser, & Hogaboam-Gray, 1999; Sung, Chang, Chiou, & Hou, 2005).

Additionally, it was inevitable to adopt learners' self-assessment as an alternative test item and a criterion against which paraphrasing tasks are tested for validation for practical reasons: (1) there are few writing performance test items fit for young learners in a Korean EFL setting, and (2) student participants aged 14 to 19 were so busy preparing for their exams that they could not find extra time to take another writing test. Alternately, freshmen of the college had recently obtained TOEIC proficiency test scores since they need to have them for graduation. Thus, rather than having the university students participate in the self-assessment questionnaire, their TOEIC scores were adopted as a criterion that represents their English proficiency.

A secondary purpose of this research is to identify what types of paraphrasing tasks would be the most suitable for each grade level. Three different types of paraphrasing tasks have been designed by the researcher based on level of difficulty, which means the degree of response open-endedness. The score correlation from each type of task is hypothesized to demonstrate which type is the most warranted for a certain grade level. Thus, detailed research questions to address in this study are as follows:

- 1) Are paraphrasing tasks valid for assessing EFL learners' English ability?
- 2) What types of paraphrasing tasks are the most adequate for each grade level if there is any difference among task types?

III. METHOD

1. Participants

Three hundred and sixty four students participated in this study as test-takers ranging from middle school

through college: 111 middle school students, 169 high school students, and 80 college freshmen. The four schools they belong to are all located in Kyounggi Province: one middle school in Yongin, one high school in Ilsan, another high school in Bundang, and one college in Shiheung. Another group of participants are developers of the paraphrase scoring scale: one middle school teacher, two high school teachers, and the researcher teaching at a college. They were recruited by the researcher based on their eagerness to participate and time availability. The scoring scale developers are all females at ages ranging from 38 to 52 with more than 5 years of teaching experience.

2. Instrument

Ten paraphrasing test items were designed by the researcher for this study (See Appendix A). They are divided into three sections: (1) four items for gap-filling paraphrased sentences, (2) three items for partial paraphrasing (3) three items for entire paraphrasing. The maximum score of the ten paraphrasing tasks is 44, consisting of 16 from the first section, 10 from the second section and 18 from the third section. The paraphrasing tasks have been developed based on the researcher's teaching experience, as well as her research background, with a focus on paraphrasing skills targeting college students. In order to score the test-takers' responses to the third section of paraphrasing tasks (entire paraphrasing), a paraphrase scoring scale was developed by the three teachers recruited and the researcher (See Appendix B). Another set of instruments is the survey questionnaire whereby the secondary school students evaluate their own general English proficiency and English writing ability as in the questions 5 and 6, respectively (See Appendix C). The 5-point Likert scales for each questionnaire were converted into comparable scores with those from paraphrasing test items. Considering only scores from the two relevant questions 5 and 6, the maximum score of the self-assessment is 10. The remainder of the questionnaires are intended for extra information, such as learners' satisfaction with performance test items currently implemented in secondary schools.

3. Procedure

The three school teachers recruited administered the paraphrasing test for randomly-selected groups of their students. Before the paraphrasing test began, the teachers were required to conduct the survey on the students' self-assessment of their own English ability.

After the administration of the test and the survey, the paraphrasing scoring scale was developed by the same group of teachers during a number of conferences at the researcher's office or a café conference room. A pilot scoring was conducted using one set of sample responses, which were selected by the researcher in consideration of grade levels and response quality. Then, a discussion of their rationales for their scoring followed. Another set of

sample responses was scored individually by developers and in-depth discussion followed. This process continued until all developers agreed on the final version of scoring criteria. Upon completing development of the scoring scale, the developers were required to submit their personal opinions about the scoring scale.

Using the newly-developed scoring scales, all of the paraphrasing tasks were scored by the researcher and another rater who is an ABD in the field of English education. When scoring the third section of the paraphrasing tasks, the inter-rater reliability between the two raters proved to be high with coefficient of .92. The discrepancy between the raters was discussed until they reach an agreement.

4. Analysis

The 5-point Likert scales on the survey questionnaire, requiring participants to self-assess their English skill and English writing ability, were converted into scores and tallied to be compared with the scores from the paraphrasing tasks. Using SPSS version 20, correlation coefficients were obtained between the scores from the paraphrasing tasks and those from the survey questionnaire.

In the case of the university students, the scores from the paraphrasing tasks were compared with their TOEIC scores instead of their self-evaluation. They had TOEIC scores, officially representing their proficiency, and could not find extra time to respond to the survey. Correlation coefficients were obtained against the scores from the paraphrasing tasks.

Correlation coefficients were obtained using different analysis programs for different data. Spearman rank correlation was adopted for the middle and high school groups, while Pearson correlation was employed for the college group. These differing methods of analysis were used because self-assessments were measured on an ordinal scale and Spearman rank correlation, not requiring the assumption that the data should be normally distributed, proves to be a legitimate method of analysis (Chen, & Popovich, 2002). On the other hand, TOEIC scores are interval scales and the data distribution proved to be bell-shaped. This warrants the use of Pearson correlation coefficients.

IV. RESULTS AND DISCUSSION

Table 1 shows the mean and standard deviation of self-assessments by 284 students from two groups (115 middle school and 169 high school students) and those of TOEIC scores obtained by 80 first-year college students. The three-digit TOEIC scores have been proportionately converted to two-digit scores, making it easy to compare with other scores with two digits in descriptive analysis.

It is noticeable that the mean of self-assessments from the middle school group is higher than that from the high

school group. The middle school is reported to be one of the most highly competitive schools in Yongin-si, Kyonggi-do. According to a teacher at the school, a quarter of the students enter a special-purpose high school or a foreign language high school, which may explain why the mean of their self-assessments (5.33 out of 10) is higher than that of the high school students (4.82).

Likewise, it can also be accounted for that their mean from paraphrasing tasks (3.83) is similar to, even slightly higher than, that of the high school students (3.82) as seen in Table 2. This represents consistent results in both self-assessment and paraphrasing tasks. Therefore, it allows a logical conclusion that their self-assessment reflects their English ability proven as their paraphrasing task scores.

Eighty university students demonstrated the highest mean (17.58) on the paraphrasing tasks. The result might be due to their prior learning experience. The textbook they used for college English, which covers the importance and uses of paraphrasing, allowed them to practice the skill and yielded a much higher mean score, compared with the other groups of students who have had little exposure to the paraphrasing skill.

TABLE 1
Descriptive Statistics (Self-Assessment or TOEIC)

	<i>n</i>	Min	Max	<i>M</i>	<i>SD</i>
Self-Assessment (Mid)	115	.0	10.00	5.33	2.53
Self-Assessment (High)	169	.0	10.00	4.82	1.88
TOEIC (Univ.)	80	.0	15.00	10.56	3.60

TABLE 2
Descriptive Statistics (Paraphrasing Tasks)

	<i>n</i>	Min	Max	<i>M</i>	<i>SD</i>
Paraphrasing (Mid)	115	.0	25.00	3.83	5.51
Paraphrasing (High)	169	.0	29.50	3.83	6.23
Paraphrasing (Univ.)	80	.0	38.00	17.59	11.12

In response to the first question, “Are paraphrasing tasks valid for assessing Korean EFL learners’ English ability?” correlation coefficients obtained from middle school, high school, and university students will be reported respectively. It is noteworthy that all three groups showed statistically significant correlation between the two scores at 0.01 level. As seen in Table 3, paraphrasing tasks represent middle school students’ English ability in statistically meaningful way (i.e., $r = .41, p < .01$). Also, as indicated in Table 4, paraphrasing tasks represent high school students’ English ability in statistically meaningful way (i.e., $r = .33, p < .01$). Finally, as Table 5 shows, paraphrasing tasks represent college students’ English ability in statistically meaningful way (i.e., $r = .74, p < .01$).

TABLE 3
Correlations (Middle School)

	Self-Assessment	Paraphrasing
Self-Assessment (Mid)	1.00	.41**
Paraphrasing (Mid)	.41**	1.00

** $p < .01$

TABLE 4
Correlations (High School)

	Self-Assessment	Paraphrasing
Self-Assessment (High)	1.00	.33**
Paraphrasing (High)	.33**	1.00

** $p < .01$

TABLE 5
Correlations (University)

	TOEIC	Paraphrasing
TOEIC (Univ.)	1	.74**
Paraphrasing (Univ.)	.74**	1

** $p < .01$

A statistically significant degree of correlation in all three groups means a strong potential of paraphrasing tasks as an alternative writing performance assessment. In particular, an obvious correlation between TOEIC scores and paraphrasing task scores proves that paraphrasing tasks can be a valid assessment tool for university students.

On the other hand, the secondary school groups show weak or moderate correlations between the two variables. This may allow for two interpretations: One is that self-assessments are not as reliable as the TOEIC scores to adopt as a criterion against which their English proficiency is assessed. The other is that paraphrasing tasks may not measure their English proficiency accurately. Considering the high correlation proven by the tertiary group, the first interpretation would be more plausible.

The weak correlation between student self-evaluation and paraphrasing tasks can be due to methodological limitations. According to Ross (2006), students need to be actively involved in developing criteria and be taught how to apply those criteria for improved accuracy of their self-assessment. In this study, however, few training sessions on how to self-evaluate were provided for students. Incorporating a training program on self-assessment for students could reveal higher correlation coefficients.

Another culprit for the weak correlation could be overly high task difficulty. As seen in Table 2, the mean scores of paraphrasing tasks are much lower (about 3.8) in the middle and high school groups compared to that (about 17.5) in the university group. Considering that middle and high school students are hardly exposed to paraphrasing tasks, the excessively challenging and unfamiliar writing tasks could possibly lower discrimination power and neg-

atively influence the correlation. Thus, it can be argued that methodological improvements (i.e., by implementing training programs on how to self-assess and how to paraphrase) could yield higher correlation.

Another interesting finding is that the middle school students, who evaluated themselves more positively, show a higher correlation between the two sets of scores than the high school students, who evaluated themselves less positively. This positively proportional relationship existing between self-assessments and paraphrasing tasks in comparing the mean scores of the secondary school groups was also demonstrated in the descriptive statistics (Table 1 & 2). This tendency may imply that the more confident group in terms of English proficiency shows a stronger relationship between their (self-assessed) English proficiency and paraphrasing competence whereas the less confident group shows a weaker relationship. That is, it can be argued that paraphrasing task could better measure English proficiency of more confident group of students, and further that its capacity as a measurement could be enhanced if the secondary school students are more frequently exposed to paraphrasing skill.

In response to the second research question, "What type of paraphrasing task is the most adequate for each grade level if there is any difference among task types?," middle and high school students yielded improved correlation coefficients in Part 1 (.55 and .53, respectively) and Part 2 (.45 and .42, respectively). In contrast, the university student group yielded higher correlation in Part 3 (.79). The secondary school students, who have had little exposure to paraphrasing skills, produced scores closer to their self-assessments in a guided type of paraphrasing task such as gap-fill and partial paraphrasing. Alternately, university students, who have been exposed to paraphrasing skills, yielded higher relationship (with their TOEIC scores in an open-ended paraphrasing task with no prompt given).

This confirms that guided types of paraphrasing tasks are more adequate for secondary school students or lower performing paraphrase writers. Although it seems to be a somewhat reasonable conclusion, it would be meaningful in that paraphrasing task reflects test-takers' proficiency and/or learning effect, and further that paraphrasing task can be fractionized and specified in consideration of their prior learning experience as well as level of proficiency.

V. CONCLUSION

With an aim to seek more diverse writing performance assessment tools, particularly targeting Korean EFL learners ranging from middle school to college students, it was examined whether paraphrasing task has the potential as a writing test item. The middle and high school groups yielded statistically significant yet weak or moderate correlations between their paraphrasing ability and self-assessed English ability at 0.01 level. On the other hand, the college students showed a high correlation between their paraphrasing ability and language proficiency measured

by TOEIC. This suggests that paraphrasing task can reflect test-takers' English proficiency although the degree of reflection varies among groups. Therefore, it can be concluded that paraphrasing task has the potential as a valid writing test item.

Interestingly, there exists discrepancy across three different groups depending on paraphrasing task types. Guided types of paraphrasing tasks proved to be more suitable for the secondary school students, showing a stronger relationship with their self-assessed English ability. Conversely, conventional types of paraphrasing task requiring entire reformation are more likely to be appropriate for university students who have been exposed to paraphrasing skill. This implies a need for diverse paraphrasing task types which can be adopted in accordance with test-takers' prior learning experience, English proficiency, and degree of confidence they have.

VI. LIMITATIONS

To begin with, it should be noted that paraphrasing tasks adopted in this study reflect test-takers' sentence-level writing performance, not paragraph-level or essay writing performance. This means that validity of paraphrasing task as a writing test item is restricted to assessment of students' lexical and syntactic knowledge and application of the knowledge in writing.

This study as a preliminary investigation into the potentiality of a brand-new writing performance assessment has a few limitations. First, the number of participants is limited and they are concentrated only in Kyeonggi-do. Demographically and geographically more extensive studies may yield different results. Second, the number of test items is limited to 10 and this could hinder more valid results. Third, in developing paraphrasing tasks as test items, a pilot test was not administered with the secondary school students although it was developed based on the researcher's long-term experience of teaching paraphrasing skills to a tertiary group. More sophisticated process in developing test items could improve reliability of the results.

Last but not the least, inconsistent criteria have been used to validate paraphrasing task against for practical reasons. In comparison with scores from paraphrasing task, self-assessments were adopted for the secondary school group and TOEIC scores for the college group. If further studies adopt the same criteria consistently across different groups (i.e., TOEIC scores or other types of writing tests for both groups), it could yield more dependable and robust results.

Despite the limitations, this study suggests meaningful implications for teachers and test administrators. Paraphrasing task can be a good addition to a restricted reservoir of writing performance test items targeting test-takers whose proficiency is at the level of sentence composition. In fact, it would be overly demanding, time-consuming and unpractical to assess their textual competence involving content and organization beyond sentential

competence. Moreover, integration of paraphrasing tasks into writing performance test battery is expected to elicit positive wash-back effect which in turn promotes L2 writers' active application of lexical and syntactic knowledge in teaching and learning environment.

REFERENCES

- Blatchford, P. (1997). Students' self assessment of academic attainment: Accuracy and stability from 7 to 16 years and influence of domain and social comparison group. *Educational Psychology, 17*(3), 345-360.
- Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage Publications.
- Claes, M., & Salame, R. (1975). Motivation toward accomplishment and the self-evaluation of performances in relation to school achievement. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement, 7*(4), 397-410.
- Connor, U., & McCagg, P. (1983). Cross-cultural differences and perceived quality in written paraphrases of English expository prose. *Applied Linguistics, 4*(3), 259-268.
- Fitzgerald J. T., White C. B., & Gruppen L. D. (2003). A longitudinal study of self-assessment accuracy. *Medical Education, 37*(7), 645-649.
- Ji, Nan Young. (2012). Modes of paraphrasing attempted by Korean L2 writers. *English Teaching, 67*(4), 131-148.
- Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing, 15*(4), 261-278.
- Keil, L. J., McClintock, C. G., Kramer, R., & Platow, M. J. (1990). Children's use of social comparison standards in judging performance and their effects on self-evaluation. *Contemporary Educational Psychology, 15*(1), 75-91.
- Kwok, D. C., & Lai, D. W. (1993, May). *The self-perception of competence by Canadian and Chinese children*. Paper presented at the annual convention of the Canadian Psychological Association, Montreal, QC.
- Laveault, D., & Miles, C. (2002, April). *The study of individual differences in the utility and validity of rubrics in the learning of writing ability*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- McCarthy, P. M., Guess, R., & McNamara, D. S. (2009). The components of paraphrase evaluations. *Behavior Research Methods, 41*(3), 682-690.
- McInnis, L. (2009). *Analyzing English L1 and L2 paraphrasing strategies through concurrent verbal report and stimulated recall protocols*. Unpublished

- master's thesis, University of Toronto, Toronto.
- Mel'čuk, I. (1997). *Vers une Linguistique Sens-Texte, Le»con inaugurale au Collège de France, Paris: Collège de France.*
- Milicevic, J. (2008). Paraphrase as a tool for achieving lexical competence in L2. In Van Dalee, S. et al (Eds), *Proceedings of the Symposium 'Complexity, Accuracy and Fluency in Second Language Use, Learning and Teaching'* (pp. 153-167). Brussels: Royal Belgium Academy of Science and Arts.
- Milićević, J. (2009). Lexical functions and paraphrasing rules as a bridge between L1 and L2. *Proceedings of first and second languages: Exploring the relationship in pedagogy-related contexts.*
- Mitman, A. L., & Lash, A. A. (1988). Student's perceptions of their academic standing and classroom behavior. *Elementary School Journal*, 89(1), 55-68.
- Ng, J. R., & Earl, J. K. (2008). Accuracy in self-assessment: The role of ability, feedback, self-efficacy and goal orientation. *Australian Journal of Career Development*, 17(3), 39-50.
- Polguère, A. (2004). La paraphrase comme outil pédagogique de modélisation des liens lexicaux. In E. Calaque & J. David (Eds.), *Didactique du lexique: contextes, démarches, supports* (pp. 115-125). Bruxelles: De Boeck.
- Reid R., Lienemann, T. O., & Hagemann, J. L. (2013). *Strategy instruction for students with learning disabilities* (2nd ed.). New York: Guilford Publications.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation*, 11(10), 1-13.
- Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1999). Effect of self-evaluation on narrative writing. *Assessing Writing*, 6(1), 107-132.
- Russo, M., & Pippa, S. (2004): *Aptitude and simultaneous interpretation: a proposal for a testing methodology based on paraphrase*. Amsterdam, Philadelphia: John Benjamins Publishing.
- Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication*, 21(2), 171-200.
- Sung, Y. T., Chang, K. E., Chang, T. H., & Yu, W. C. (2010). How many heads are better than one? The reliability and validity of teenagers' self-and peer assessments. *Journal of Adolescence*, 33(1), 135-145.
- Sung, Y. T., Chang, K. E., Chiou, S. K., & Hou, H. T. (2005). The design and application of a web-based self-and peer-assessment system. *Computers and Education*, 45(2), 187-202.
- Žolkovskij, A., & Mel'čuk, I. (1967). O semantičeskom sinteze. *Problemy Kibernetiki*, 19, 177-238.

APPENDIX A
Paraphrasing Task

이름: _____ 학교: _____ 학년: _____

(소요시간: 30분)

[예시]

A great number of people use Mandarin.

→ Mandarin *is spoken by* a lot of people.

→ Mandarin *is the language used by* a substantial number of people.

[Q1-4] 주어진 문장과 의미가 일치하도록 어휘 및 문장구조를 변경한 (→ 이후 paraphrase) 문장 속 빈칸을 채우시오. 괄호() 속 철자가 주어진 경우는 반드시 그 철자로 시작하는 단어를 한 개 쓰시오.

1. Although TV can be relaxing, it is only occasionally entertaining and very rarely involving.

→ TV can give us (rel) _____, _____, it is sometimes (enj) _____ and rarely _____.

2. In their play, kids often imitate people entirely different from them.

→ _____ children play, they frequently (mi) _____ people _____ are (op) _____ to them.

3. Instead of buying a brand I've never heard of, I pay more for a brand I've seen on TV.

→ I _____ to (pu) _____ a familiar brand than to buy (un) _____ one.

4. With educational television, it is much more critical to be able to attract and hold an audience, while still being concerned with the acquisition of educational information.

→ Educational television _____ it even more (im) _____ to fascinate and (sus) _____ students' attention _____ it is still concerned with the way they (ac) _____ educational messages.

[Q 5-7] 주어진 문장과 의미가 일치하도록 어휘 및 문장구조를 변경한 (→ 이후 paraphrase) 문장 속 빈칸을 채우시오. 주어진 단어를 참고하여 빈 칸에 여러 단어로 채우시오.

5. The termination of projects has reduced the value of resources.

→ The value of resources _____ after projects _____.

6. Fruits and vegetables are believed to help prevent cancer.

→ It _____ that _____.

7. People of any age will benefit from eating a healthy diet.

→ A healthy diet _____.

[Q8-10] 주어진 문장과 의미가 일치하도록 어휘 및 문장구조를 변경하여 빈칸에 완전한 문장으로 채우시오.

8. The weather forecast says it will be mostly windy and rainy over the next few days.

→ _____.

9. Cell phone usage has exploded over the past decade and continues to rise.

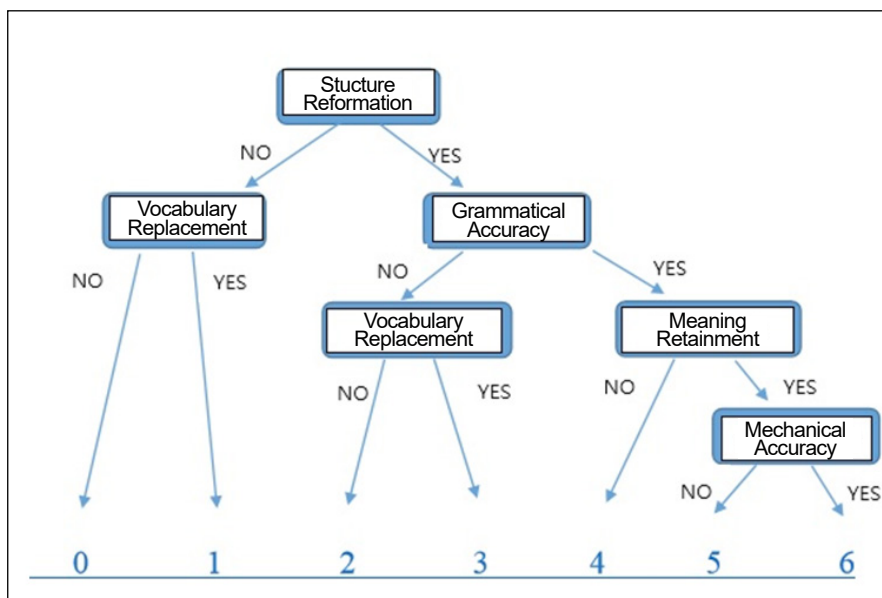
→ _____.

10. Brain cancer rates have risen since cell phones were introduced, leading some people to wonder if cell phone usage is the reason for the increase.

→ _____.

수고하셨습니다

APPENDIX B Paraphrase Rating Scale



Guidelines for Rating Scale Interpretation:

- 1) ‘Structure Reformation’ means that fundamental modification of a sentence structure is made such as: from a simple to a complex sentence and vice versa; from active to passive and vice versa; using causative verbs, emphatic structures, ‘it’ as a dummy subject or object, ellipsis and inversion, etc. Note that relocating adverbial phrases is not included.
- 2) ‘Vocabulary Replacement’ means successful replacement of words with their synonyms, antonyms, meronyms, etc. and accurate uses of derivatives and inflected word forms.
- 3) ‘Grammatical Accuracy’ means that few global errors are made. Verb tenses are accurately used.
- 4) ‘Meaning Retainment’ means that a paraphrased sentence has the same meaning as its original text, including key words.
- 5) ‘Mechanical Accuracy’ means that few errors are made in using punctuation, spelling, use of article, etc.

APPENDIX C
Survey Questionnaires

학생용 설문조사서

1. 이름: _____
2. 학교: _____
3. 학년: _____

4. 성별:
 Male Female

5. 학년 대비 나의 영어능력 수준은?
 매우 잘한다 잘하는 편이다 평균수준이다 부족한 편이다 매우 부족하다

6. 학년 대비 나의 영어작문능력 수준은?
 매우 잘한다 잘하는 편이다 평균수준이다 부족한 편이다 매우 부족하다

7. 학교수업 외에 영어작문 연습 및 수업을 추가로 한 경험이 있는가?
(사교육 및 유학포함)
 경험 있다 경험 없다

8. 학교 수업 외에 경험 한 영어작문 연습 및 수업 기간은 얼마인가?
(사교육 및 유학 포함)
 경험 없다 1년 이하 2-3년 4-5년 6년 이상

9. 중/고등학교에서 경험한 수행평가가 나의 영작문 능력을 평가하기에
충분하다고 생각하는가?
 매우 그렇다 그렇다 잘 모르겠다 그렇지 않다 매우 그렇지 않다

10. 위 9번 질문에 대한 답변의 이유를 구체적으로 써주시기 바랍니다.

Your participation is greatly appreciated. Thank you.