MEESO

Modern English Education

https://doi.org/10.18095/meeso.2021.22.3.12

A Comparative Analysis of NAEA English Reading Passages for Middle and High School Students Using Coh-Metrix*

Jiyoung Lee
Pusan National University

ARTICLE INFO

Received: 16 March 2021
Revised: 12 July 2021
Accepted: 17 August 2021

Examples in: English
Applicable Languages: English
Applicable Levels: Secondary

KEYWORDS

National Assessment of Educational Achievement/ English reading passage/ Coh-Metrix
국가수준 학업성취도 평가/ 영어 읽기지문/코메트릭스

ABSTRACT

Lee, Jiyoung. (2021). A comparative analysis of NAEA English reading passages for middle and high school students using Coh-Metrix. *Modern English Education*, *22*(3), 12-23.

This study aims to identify differences of English reading passages in the National Assessment of Educational Achievement (NAEA) for middle and high school students. For this study, 208 NAEA English reading passages from the five consecutive years from 2015 to 2019 were extracted. The two research questions consider the differences of two tests, and correlations among the selected Coh-Metrix indices such as descriptive measures, lexical characteristics, syntactic complexity, discourse patterns, and readability level. The extracted reading passages were analyzed based on 17 selected Coh-Metrix indices related to each selected index. As for the differences of two tests, descriptive measures, lexical difficulty and readability level showed clear differences, whereas lexical diversity, syntactic complexity, and discourse patterns did not present critical gaps between the two tests. As for the correlations, the positive correlations between readability level and descriptive measures were inevitable in that the way to estimate the Flesch Reading Ease score and the Flesch-Kincaid Grade Level is based on the average number of syllables per word and words per sentence. Lexical diversity was negatively correlated to all the indices of discourse patterns. The implications and limitations of the current study are included.

I. INTRODUCTION

The National Assessment of Educational Achievement (NAEA) plays an important role to improve teaching, learning, and even the national curriculum. The NAEA is designed to figure out students' basic scholastic abilities systematically by the Korea Institute for Curriculum and Evaluation (KICE) in order to improve educational programs (KICE, 2014). The NAEA could be classified as the criterion-referenced assessment, which assesses whether the test-takers achieve the given educational goal without comparing test-takers' academic achievement (Brown, 2007; I. Y. Park et al., 2017). However, the NAEA has been employed as research materials less than the College Scholastic Ability Test (CSAT).

The targets of the NAEA are 9th grade, which is the third grade of middle school, and 11th grade, which is the second grade of high school. The subjects of the NAEA are different according to the school level. On behalf of middle school students, 9th grade students take an exam in Korean language, math, English language, science, and social studies. On the other hand, on behalf of high school students, 11th grade students take an exam in Korean language, English language, and math. For 9th grade students, the NAEA covers what the students have learned in the first and second grade of middle school. For 11th grade

* This paper has been developed from the presentation, *A Coh-Metrix Analysis of English Reading Passages in the National Assessment of Educational Achievement for 9th and 11th Grade*, delivered at the 2020 MEESo International Conference.

Jiyoung Lee (PhD)
Department of English Language Education, Pusan National University, 2, Pusandaehak-ro, Geumjeong-gu, Busan, 46241, Korea
Email: jylee1915@gmail.com

students, the exam covers what the students have learned in middle school and the first grade of high school.

Especially, the English language test consists of two sections: the listening section and the reading section. Each section has two types of items: multiple-choice items and constructed-response items. Technically, the constructed items in the listening section are designed to assess students' speaking ability. Likewise, the constructed items in the reading section are designed to assess students' writing ability. Just like the CSAT, reading comprehension is one of the most important parts to assess in the NAEA in that reading is the key source of language input (Krashen, 2004). The current study focuses on the reading section of English language test.

It is really important that the test has to have appropriate text difficulty in order to assess students' language proficiency precisely. If the difficulty is not proper, the test cannot provide test-takers and teachers right feedback. Therefore, analyses on English reading passages of the NAEA are meaningful. However, little research has compared the NAEA reading passages for middle and high school students in one study.

The current study aims to identify differences between the NAEA middle and high school English reading passages in order to shed light on whether the text difficulty systematically rises as the grade increases. Descriptive measures, lexical diversity, syntactic complexity, discourse patterns and readability are analyzed using Coh-Metrix. These Coh-Metrix indices can provide information related to the level of text difficulty. The analysis could denote whether the difficulty of English reading passages is different as the two different target graders have different readability. In this regard, this study is necessary. To accomplish this goal, 208 English reading passages are extracted from the 2015-2019 NAEA English tests for 9th graders and 11th graders.

The two research questions are as follows:

1) Are there any differences in descriptive measures, lexical characteristics, syntactic complexity, discourse patterns, and readability level in NAEA reading passages for middle and high school students?
2) Do measured variables have correlations?

II. THEORETICAL FRAMEWORK AND LITERATURE REVIEW

1. National Assessment of Educational Achievement (NAEA)

Previous studies of the NAEA have mainly analyzed student responses of the test and student-related variables. Student responses were one of the main topics of research on the NAEA. Especially, constructed-response items were usually selected for analysis (S. H. Kim, 2016; J. K. Pae, 2018). S. H. Kim (2016) investigated the responses of 11th graders in one speaking response item of each test and one writing response item of each test from 2010-2014 NAEA. J. K. Pae (2018) analyzed students' responses on constructed-response items in the 2015 NAEA for 3rd grade of middle school students. These two studies identify characteristics of different English proficiency levels of students with their responses. J. K. Pae and P. Hwang (2015) analyzed English test items of 2010-2013 NAEA for 9th grade using distribution curves of response rates to identify specific and detailed characteristics of students' educational achievement.

Another research focus of the NAEA is student-related variables. Y. J. Lee (2012) examined the effect of 6th, 9th, and 11th grade students' background characteristics on their performance on the test using hierarchical regression. T. I. Pae (2012) identified the predictors of the variance in English achievement for middle and high school students with multiple regression. Namely, this kind of research identifies the relationship between students' characteristics and achievement. This is for supporting low-level students and improving curriculum. Analyses on student responses could identify whether the test was well-designed.

Recently, researchers started to compare the exam and textbooks. Vocabulary could be the subject to an analysis (H. Lee, 2021). This research focused on how vocabulary of the 2015 Revised National Curriculum of English had been applied to high school English textbooks and the NAEA English test. Lexical diversity, lexical density and word frequency were analyzed in both listening and reading texts in eleven high school English textbooks and the 2019 NAEA English test for 11th grade using Range32 and KICE WordProfiler 2015. As a result, both textbooks and the NAEA English test showed a significant difference in lexical diversity by comparing passages to previously learned ones. Lexical density did not show any significant difference between textbooks and NAEA. Lexical difficulty and readability were investigated in O. I. Ha and Y. Cho's (2016) study in the reading passages of the NAEA and middle school English textbooks. Readability was estimated using the Flesch-Kincaid Grade Level, whereas lexical difficulty including the standardized type-token ratio and lexical frequency was assessed with VocabProfile and Oxford WordSmith Tools 7.0. The NAEA reading passages had more various words and less repetitive lexical elements than the textbooks. The characteristics of speech acts in High School Practical English 1 textbooks and the 2014-2015 NAEA English tests were compared by M. K. Cha and H. Lee (2016). The comparative research on NAEA and textbooks can identify the relationship between learning and testing, and suggest whether the test items are well-organized.

2. Coh-Metrix Research

Coh-Metrix, a computing tool for textual analysis, was developed to assess the linguistic features of given texts (Graesser et al., 2004; McNamara et al., 2010; McNamara et al., 2014). The program provides 106 indices for ana-

lyzing descriptive measures, discourse patterns, readability level, syntactic complexity, and word information (Crossley & McNamara, 2009; McNamara et al., 2014).

These days, many researchers have utilized Coh-Metrix in English language education. Coh-Metrix research has conducted on textbooks (M. Jeon, 2011; M. Jeon & I. Lim, 2009; J. Kim & J. Yang, 2012), and tests (J. Y. Chang, 2018; J. Lee, 2020). Reading passages in textbooks are one of the primary research materials in English education in EFL settings. Coh-Metrix is a useful tool to compare and contrast textbooks by dividing two different groups in order to know the level of continuity. The text difficulty has to rise systematically in accordance of students' level or grade.

Reading passages of middle school English textbooks showed inconsistent continuity. Under the revised 2009 curriculum, the continuity of textbooks for the first and second graders was well-established in descriptive measures, readability level, syntactic complexity, connectives, and word information, but not in word frequency, syntactic density, pronoun information, referential cohesion, and semantic cohesion (M. Jeon, 2011). Textbooks did not show the well-established continuity among three graders in overall Coh-Metrix measures instead of descriptive measures and syntactic complexity (M. Jeon, 2015). Under the revised 2015 curriculum, the overall difficulty of textbooks systematically increased with three grade levels for most measures, whereas word frequency, lexical diversity, and NP density were not consistent with the general expectation (J. Ryu & M. Jeon, 2020).

Similarly, reading passages of high school English textbooks did not become gradually difficult according to grade levels. The level of text difficulty between high school English 1 and 2 textbooks under the revised 2009 English curriculum did not gradually increase (S. H. Lee, 2013). High school textbooks of the revised 2009 English curriculum for the first, second and third graders did not show the well-established continuity in descriptive measures, FKGL, lexical diversity, referential cohesion, and semantic cohesion (M. Jeon, 2014). Although high school English textbooks were newly published under the 2015 revised English curriculum, the problem still remained (H. Park & C. K. Jung, 2020). This kind of textbook research could give useful information to manipulate the text difficulty within the same school level, but as J. Kim and J. Yang (2012) indicated, previous research hardly compared reading passages of textbooks across different school level using Coh-Metrix.

In addition to analyses on textbooks, Coh-Metrix research has focused on analysis on reading passages in tests such as the civil-service examination and the CSAT (College Scholastic Ability Test). The comparative research did not demonstrate the significant differences of English reading passages in the test for 7th and 9th grade civil-servants in terms of descriptive measures, lexical diversity, syntactic complexity, cohesion, and readability (Y. Shin & Y. Jong, 2018). Coh-Metrix analysis on the CSAT has conducted to identify whether the text difficulty changed after the testing type changed. The research did not show significant changes of text difficulty between before and after the change of the testing type in text difficulty indices including the descriptive measures, readability level and text easability (J. Y. Chang, 2018; J. Lee, 2020), and linguistic features including lexical diversity, word frequency, syntactic complexity, and cohesion (Y. Shin, 2019).

Researchers have rarely adopted NAEA reading passages as materials for Coh-Metrix analyses. More specifically, J. Kim (2018) is the only Coh-Metrix analysis on NAEA reading passages. He examined the homogeneity of 2015-2017 NAEA English reading passages for 9th grade with descriptive measures, lexical diversity, word frequency, syntactic complexity, and cohesion. The study denoted that the homogeneity of three consecutive tests remained in almost all selected Coh-Metrix indices. However, sentences became statistically longer and more complex.

Among many sources of materials for Coh-Metrix analyses, little research has been conducted on the comparison of the NAEA reading passages for middle and high school students in one study. It is necessary to draw educational implications about whether the text difficulty systematically rises as the grade increases.

III. METHOD

1. Corpus Construction

NAEA English reading passages of the five consecutive years from 2015 to 2019 were extracted in order to construct corpora. The targets of the National Assessment of Educational Achievement are 9th grade, which is the third grade of middle school, and 11th grade, which is the second grade of high school. The corpora consist of 208 reading passages. Among 208 passages, 110 passages are NAEA reading passages for 9th grade, 98 passages are for 11th grade. Table 1 presents the basic information of two corpora.

As shown in Table 1, English tests of 2015-2019 NAEA for 9th grade consist of 22 reading passages. English tests of 2015-2018 NAEA for 11th grade consist of 20 short

TABLE 1
Basic Information of Corpora

Year	NAEA reading passages for middle school students Passages (Words)	NAEA reading passages for high school students Passages (Words)
2015	22 (1,874)	21 (2,493)
2016	22 (2,040)	21 (2,589)
2017	22 (2,006)	21 (2,655)
2018	22 (2,048)	21 (2,759)
2019	22 (1,915)	14 (1,848)
Total	110 (9,883)	98 (12,344)

reading passages and one longer text, whereas there are 14 passages in 2019 NAEA.

Only perfect sentences in the corpora were analyzed. Namely, segments of sentences were not included. Marks, pictures, charts and grapes were deleted. Blanks in the passages were filled with the answer from the key or options.

2. Coh-Metrix

The on-line version of Coh-Metrix 3.0 (http://141.225.61.35/cohmetrix2017) is utilized for this study. The current study tries to analyze the linguistic features of NAEA English reading passages including descriptive measures, lexical characteristics, syntactic complexity, discourse patterns, and readability level. For this purpose, 17 indices were selected among 106 indices. These Coh-Metrix indices can provide information directly related to the level of text difficulty. Table 2 displays Coh-Metrix indices which are selected for this study.

1) Descriptive Measures

Coh-Metrix provides basic information about the given text such as word count, sentence count, word length and sentence length to analyze the patterns of corpora (McNamara et al., 2014). The word count and the sentence count are calculated based on the average number of words in the given text and the average number of sentences in the given text respectively. Word length and sentence length are estimated based on the average number of syllables per word and the average number of words per sentence respectively. More words, more sentences and longer sentences increase text difficulty (M. Jeon, 2011).

2) Lexical Characteristics

For measuring lexical characteristics, this study focuses on lexical diversity, word frequency, and word information. Coh-Metrix measures lexical diversity with a type, which is the variety of different lexical items, and a token, which is the number of every single word in a text regardless of redundancy (Crossley & McNamara, 2011a; Kennedy, 2014). Namely, the lexical diversity index presents the diversity of unique vocabulary items. Coh-Metrix provides two indices such as the type-token ratio for all words and for content words. This study only focuses on the type-token ratio for content words. If the value of the type-token ratio for content words is high, the corpus includes diverse content words. The type-token ratio is negatively correlated with text length in that the more word tokens are, the less unique the words are (McNamara et al., 2014).

Word frequency measures high frequency words on the basis of the CELEX corpus (Baayen et al., 1995). High frequency words are easy to comprehend and process because they often occur in English language (M. Jeong, 2015; M. Jeon & I. Lim, 2009). To identify word frequency, the current study utilizes the CELEX word frequency for content words.

Coh-Metrix gives additional lexical information with indices including age of acquisition, word concreteness, and word familiarity on the basis of MRC Psycholinguistics Database (Coltheart, 1981). The age-of-acquisition

TABLE 2
Coh-Metrix Indices

Linguistic feature	Category	Coh-Metrix index
Descriptive measures	Descriptive statistics	Word count Sentence count Word length Sentence length
Lexical characteristics	Lexical diversity	Type-token ratio for content words
	Word frequency	Word frequency for content words
	Word information	Age of acquisition Familiarity Concreteness
Syntactic complexity	Syntactic complexity	Words before the main verb Modifiers for NP
Discourse patterns	Referential cohesion	Argument overlap in adjacent sentences Noun overlap in adjacent sentences Content word overlap in adjacent sentences
	Semantic coreferentiality	LSA overlap in adjacent sentences
Readability level	Readability level	Flesch Reading Ease score Flesch-Kincaid Grade level

measure specifies the age when children acquire particular words, so the higher age-of-acquisition values denote that people learn the specific word later (McNamara et al., 2014). Word familiarity is an index of how familiar vocabulary is to adults (McNamara et al., 2014). Word concreteness denotes how concrete a word is (Crossley & McNamara, 2011a). The measured values in age of acquisition, word familiarity, and word concreteness range from 100 to 700. These three indices are closely related to text difficulty (McNamara et al., 2014).

3) Syntactic Complexity

Coh-Metrix measures the mean number of words before the main verb, and the average number of modifiers for NP in order to present the syntactic complexity of the given text (Crossley & McNamara, 2011a). The higher measured value of syntactic complexity denotes the more difficult text in that texts with fewer words and shorter phrases are easier to process on working memory (M. Jeong, 2015; M. Jeon & I. Lim, 2009; McNamara et al., 2014). As complex syntactic patterns are difficult to process cognitively, the level of text complexity rises (M. Jeon, 2011; McNamara et al., 2014).

4) Discourse Patterns

For this study, referential cohesion and semantic coreferentiality are analyzed to identify the discourse patterns of the NAEA English reading passages. Coh-Metrix indices related to discourse patterns measure the overlapped elements within adjacent sentences or in the whole text (McNamara et al., 2014; McNamara et al., 2012). Indices which analyze adjacent sentences are selected in this study.

For the referential cohesion, noun overlap, argument overlap, and content word overlap are selected in this study. Noun overlap, argument overlap, and content word overlap are estimated by counting the same nouns, arguments such as nouns, pronouns or noun phrases, and content words in adjacent sentences respectively (McNamara et al., 2014). Text difficulty could decrease as the text is referentially cohesive because same elements are repeatedly used in a text (Crossley & McNamara, 2011b; McNamara et al., 2014).

Semantic coreferentiality is assessed on the basis of Latent Semantic Analysis (Landauer et al., 1998), a mathematical technique. Latent Semantic Analysis (LSA) computes text coherence with a large corpus to represent general world knowledge (McNamara et al., 2014). Just like referential cohesion, LSA indices measure the overlapped elements within adjacent sentences or in the whole text. The only Coh-Metrix index of semantic coreferentiality is LSA overlap in adjacent sentences in this study. Text difficulty can be predicted as high LSA overlap scores indicate easy texts (M. Jeon, 2011; McNamara et al., 2014).

5) Readability Level

In order to measure text difficulty or ease, two traditional readability approaches, Flesch-Kincaid metrics: Flesch Reading Ease (Flesch, 1948; Klare, 1974-1975) and Flesch-Kincaid Grade Level (Kincaid, et al., 1975), are provided by Coh-Metrix. These two indices use word length and sentence length to measure readability level. The range of the Flesch Reading Ease index is from the lowest score 0 for the most difficult text to the highest score 100 for the easiest text. On the other hand, the range of the Flesch-Kincaid Grade Level is from the lowest score 1 to the highest score is 12 (McNamara et al., 2014).

Coh-Metrix provides comprehensive linguistic measures. The selected measures for the study are related to text difficulty. They are suitable to figure out whether the text difficulty of NAEA English reading passages corresponds to the level of grade.

3. Statistical Analysis

The results from Coh-Metrix were coded, and then analyzed with R 4.0.1, which is a statistical software. An independent two sample *t*-test was conducted. Pearson's product moment correlation was used. The selected 17 Coh-Metrix indices are the dependent variables, and the NAEA English reading passages for middle and high school students are the independent variables.

IV. RESULTS AND DISCUSSION

1. Differences of NAEA English Reading Passages for Middle and High School Students

Research question 1 is about whether differences exist in descriptive measures, lexical characteristics, syntactic complexity, discourse patterns, and readability level of 208 NAEA English reading passages for middle and high school students. In order to investigate differences between two tests, an independent two sample *t*-test was conducted.

1) Descriptive Measures

Table 3 displays the result of an independent two sample *t*-test for the descriptive measures of 208 NAEA English reading passages such as the number of words and sentences, the number of syllables per word, and the number of words per sentence. All the indices of descriptive measures show rising trends as the level of the grade is higher.

As indicated in Table 3, the number of words ($t = -9.61$, $p < .001$), and sentences ($t = -3.55$, $p < .001$) are significantly different between two grades. In addition, there were statistically significant differences in the number of syllables per words ($t = -6.81$, $p < .001$) and the number of

TABLE 3
T-test for Descriptive Measures

Variable	Middle school Ss (n = 110) M (SD)	High school Ss (n = 98) M (SD)	t	df	Effect size	p
Number of words	89.85 (24.65)	125.96 (29.54)	-9.61	206	-1.33	.000***
Number of sentences	7.73 (2.36)	9.36 (4.11)	-3.55	206	-.49	.000***
Number of syllables per word	1.39 (.13)	1.53 (.18)	-6.81	206	-.95	.000***
Number of words per sentence	12.03 (3.20)	14.84 (4.43)	-5.28	206	-.73	.000***

*** $p < .001$

TABLE 4
T-test for Lexical Characteristics

Variable	Middle school Ss (n = 110) M (SD)	High school Ss (n = 98) M (SD)	t	df	Effect size	p
Type-token ratio	.80 (.10)	.79 (.09)	.84	206	.12	.401
Word frequency	2.37 (.24)	2.28 (.17)	2.95	206	.41	.003**
Age of acquisition	307.92 (42.78)	348.13 (55.55)	-5.88	206	-.82	.000***
Word familiarity	580.51 (11.46)	575.87 (8.76)	3.25	206	.45	.001**
Word concreteness	401.10 (38.97)	375.33 (35.15)	4.98	206	.69	.000***

** $p < .01$, *** $p < .001$

words per sentence (t = -5.28, $p < .001$). It means that text length rises as the grade increases.

The results are similar with comparative research on textbooks within the same school level (M. Jeon, 2011; M. Jeon, 2015; J. Kim & J. Yang, 2012; J. Ryu & M. Jeon, 2020). This is because the number of words and sentences are the explicit factors that manipulate the text difficulty to adjust students' level. However, the results are different from the comparative analysis between 7th and 9th civil-service examination (Y. Shin & Y. Jong, 2018) in that the existing study did not show any significant difference in the word count, the sentence count, and sentence length. The target of the tests could affect the results. The continuity of the NAEA could be much more important than that of the civil-service examination because the result of the NAEA influences the teaching and curriculum.

2) Lexical Characteristics

Table 4 shows the result of an independent two sample *t*-test on the lexical characteristics of 208 NAEA English reading passages including the type-token ratio, word frequency, age of acquisition, word familiarity, and word concreteness. Word frequency, familiarity, and concreteness display downward trends, whereas age of acquisition has an upward trend as the school level moves from middle to high school.

As demonstrated in Table 4, there are statistically significant differences of NAEA English reading passages for middle and high school in word frequency (t = 2.95, $p < .01$) and age of acquisition (t = -5.88, $p < .001$), word familiarity (t = 3.25, $p < .01$), and word concreteness (t = 4.98, $p < .001$). The high school English reading passages include less high frequency words than the middle school English reading passages. English reading passages for high school students include significantly less high-frequency, familiar and concrete words. High frequency words are easy to comprehend and process because they often occur in English language (M. Jeong, 2015; M. Jeon & I. Lim, 2009). Text difficulty decreases with familiar and concrete words and younger age of acquisition. In this regard, the results indicate that vocabulary in the reading passages of high school students is significantly more difficult than vocabulary in the reading passages of middle school students.

However, NAEA English reading passages for middle and high school students do not show statistical differences in the type-token ratio index. The result indicates that reading passages of high school students do not have more diverse content words than those of middle school students. The result corresponds to O. I. Ha and Y. Cho's (2016) study in that the NAEA reading passages of middle school students had various words and less repetition.

3) Syntactic Complexity

Table 5 shows the result of an independent two sample *t*-test for the syntactic complexity of 208 NAEA English reading passages. The sentence structures of the NAEA English reading passages of high school students are

TABLE 5
T-test for Syntactic Complexity

Variable	Middle school Ss (n = 110) M (SD)	High school Ss (n = 98) M (SD)	t	df	Effect size	p
Words before the main verb	3.11 (1.53)	3.63 (2.37)	-1.92	206	-.27	.056
Modifiers for NP	.76 (.28)	.80 (.22)	-1.33	206	-.19	.184

TABLE 6
T-test for Discourse Patterns

Variable	Middle school Ss (n = 110) M (SD)	High school Ss (n = 98) M (SD)	t	df	Effect size	p
Noun overlap	.31 (.28)	.37 (.25)	-1.73	206	-.24	.085
Argument overlap	.52 (.27)	.54 (.23)	-.52	206	-.07	.606
Content word overlap	.13 (.09)	.12 (.07)	.82	206	.11	.415
LSA overlap	.23 (.13)	.23 (.13)	-.22	206	-.03	.826

slightly more complex. However, there were no statistically significant differences between two groups of reading passages in the number of words before the main verb index (t = -1.92, p > .05) and the number of modifiers for NP index (t = -1.33, p > .05).

The complexity of the sentence structures does not rise as the grade increases in that the higher measured value of syntactic complexity denotes the more difficult text. Texts with fewer words and shorter phrases are easier to process on working memory (M. Jeong, 2015; M. Jeon & I. Lim, 2009; McNamara et al., 2014). These results could be drawn from the fact that sentences in the NAEA reading texts for middle school students were getting significantly longer and more complex (J. Kim, 2018). The results match up with the comparative analysis between 7th and 9th civil-service examination (Y. Shin & Y. Jong, 2018) in that the previous study showed no differences in the two syntactic complexity indices. It suggests that the text difficulty does not change in accordance with the level of grade. It needs to carefully manipulate sentence structures within the text for the proper text difficulty.

4) Discourse Patterns

Table 6 shows the result of an independent two sample *t*-test for the discourse patterns of 208 NAEA English reading passages such as noun overlap, argument overlap, content word overlap, and LSA overlap. Slightly more nouns and arguments in the reading passages of high school students are overlapped in adjacent sentences than those of middle school students. However, the difference shows no statistical significances in all the indices of discourse patters such as noun overlap (t = -1.73, p > .05), argument overlap (t = -.52, p > .05), content word overlap (t = .82, p > .05), and LSA overlap (t = -.22, p > .05).

In other words, the two tests have similar level of repetition of same elements. English reading passages for middle and high school students have the similar text difficulty in textual level in that the repetition of the same element in a text can decrease text difficulty (Crossley & McNamara, 2011b; McNamara et al., 2014). The results correspond to O. I. Ha and Y. Cho's (2016) study in that the NAEA reading passages for middle school students were less repetitive. The results are also similar with the comparative analysis between 7th and 9th civil-service examination (Y. Shin & Y. Jong, 2018) in that the previous research showed no differences in the discourse pattern indices. It suggests that the text difficulty does not change in accordance with the level of grade. It needs to carefully manipulate the overlap of the same elements within the text for the proper text difficulty with considering the targets.

5) Readability Level

Table 7 shows the result of an independent two sample *t*-test for the readability level of 208 NAEA English reading passages. There are significant differences between two groups of tests in the Flesch Reading Ease index (t = 6.20, p < .001), and the Flesch-Kincaid Grade Level (t = -6.64, p < .001).

The two readability indices show the upward trend of the text difficulty in accordance with the grade level in that more difficult texts have lower Flesch Reading Ease scores and higher Flesch-Kincaid Grade level. Namely, the reading passages for high school students are significantly more difficult than reading passages for middle school students. The two readability indices are estimated on the basis of the mean number of syllables per word and the mean number of words per sentence (McNamara et al., 2014). The way of calculation affects the statistical differences between the two tests in that the current study demonstrates word length and sentence length rise as the

TABLE 7
T-test for Readability Level

Variable	Middle school Ss (n = 110) M (SD)	High school Ss (n = 98) M (SD)	t	df	Effect size	p
Flesch Reading Ease score	76.93 (12.00)	64.67 (16.39)	6.20	206	.86	.000***
Flesch-Kincaid Grade Level	5.53 (2.10)	7.91 (3.01)	-6.64	206	-.92	.000***

*** $p < .001$

grade increases in the NAEA English tests. The results suggest that unlike research on textbooks (M. Jeon, 2014, 2015; S. H. Lee, 2013; H. Park & C. K. Jung, 2020), the NAEA English reading passages become gradually difficult according to grade levels.

2. Correlations

Research question 2 considers whether measured indices are correlated. The 17 Coh-Metrix were selected because they are known to affect the level of text difficulty directly or indirectly. In order to examine correlations among measured 17 Coh-Metrix indices, Pearson's product moment correlation was conducted.

1) Correlations Between the Readability Indices and Other Indices

This section tries to identify whether the measured Coh-Metrix indices including descriptive measures, lexical characteristics, syntactic complexity, and discourse patterns really affect the standard readability indices including the Flesch Reading Ease (FRE) and the Flesch-Kincaid Grade Level (FKGL) in this study. Table 8 shows the result of Pearson's product moment correlation between readability and other indices.

Readability measures are affected from the word count, word length, sentence length, age of acquisition, word familiarity, and word concreteness in NAEA reading passages. However, discourse patterns are not related to the readability level of NAEA reading passages. In these indices, the FRE and FKGL indices have opposite values because more difficult texts have lower Flesch Reading Ease scores and higher Flesch-Kincaid Grade Level. Both FRE and FKGL indices have significant correlations with word length and sentence length. The correlations with word length and sentence length are inevitable because the way to estimate the FRE and FKGL indices is based on the average number of syllables per word and words per sentence. In addition, the two readability measures have correlations with the three word-information indices such as age of acquisition, word frequency, and word concreteness. It suggests that difficult words increase the level of text difficulty in NAEA reading passages.

Only one syntactic complexity index, the number of words before the main verb, has significant correlations

TABLE 8
Pearson's Product Moment Correlation Between Readability and Other Indices

Category	Variable	FRE	FKGL
Descriptive measures	Word count	-.275***	.315***
	Sentence count	-.031	.043
	Word length	-.294***	.303***
	Sentence length	-.266***	.293***
Lexical characteristics	Lexical diversity	-.049	.031
	Word frequency	.147*	-.132
	Age of acquisition	-.296***	.290***
	Word familiarity	.179**	-.167*
	Word concreteness	.182**	-.207**
Syntactic complexity	Words before the main verb	-.148*	.172*
	NP modifiers	.011	-.021
Discourse patterns	Noun overlap	.036	.009
	Argument overlap	-.032	.058
	Content word overlap	.106	-.096
	LSA overlap	.012	.007

* $p < .05$, ** $p < .01$, *** $p < .001$

with the two readability indices. It denotes that more words before the main verb could make a text more difficult. The results of syntactic complexity indices partially correspond to the explanation that the greater number of words before the main verb, and the greater number of NP modifiers increase the level of text difficulty (M. Jeon, 2011; McNamara et al., 2014). All four discourse pattern indices do not show any significant correlations with the readability indices. The result of discourse patterns is contrary to the point that text difficulty can be predicted as high overlap scores including noun, argument, content word and LSA overlap indicate easy texts (McNamara et al., 2014).

The results are different from the Coh-Metrix analysis on the civil-service exam in that Y. Shin and Y. Jong (2018) indicated the correlations between argument overlap and FRE, and between word frequency and FKGL. The difference comes from the different aspects of the test. The results partially correspond to previous studies that word length, sentence length, lexical difficulty, and sentence structures affect readability (Klare, 1984), and vocabulary, sentence length, sentence structure, and cohesion influence text difficulty (Graves & Graves, 2003).

2) Correlations among Descriptive Measures, Lexical Characteristics, Syntactic Complexity, and Discourse Patterns

This section tries to demonstrate whether the measured Coh-Metrix indices including descriptive measures, lexical characteristics, syntactic complexity, and discourse patterns are really correlated one another in this study. Table 9 shows the result of Pearson's product moment correlation between lexical characteristics and other indices including descriptive measures and discourse patterns.

Lexical diversity has significant negative correlations with the word count (r = -.400, p < .001) and the sentence count (r = -.285, p < .001). It means that the lexical diversity of the NAEA reading passages is affected from the word count and the sentence count. The result is natural in that lexical diversity is measured with the number of unique vocabulary items. Word frequency has a significant positive correlation with sentence count (r = .170, p < .05), but a significant negative correlation with word length (r = -.403, p < .001). The results are contradictory to the comparative analysis between 7th and 9th civil-service examination (Y. Shin & Y. Jong, 2018) in that lexical diversity and word frequency were not significantly correlated to any descriptive measure in the previous study. Age of acquisition, word familiarity and word concreteness have significant correlations with word length and sentence length. It suggests that more syllables in a word increase the level of lexical difficulty.

Lexical diversity has significant negative correlations with all four discourse pattern indices. These negative correlations between lexical diversity and discourse patterns are inevitable since the notion of the type-token ratio, which refers to lexical diversity, is contradictory to noun overlap, argument overlap, content word overlap, and LSA overlap.

Table 10 shows the result of Pearson's product moment correlation between syntactic complexity and other indices such as descriptive measures, lexical characteristics, and referential cohesion. The two syntactic complexity indices have negative correlations with sentence counts and positive correlations with word length. According to their characteristics, the number of words before the main verb has a positive correlation with sentence length (r = .642, p < .001), and the number of NP modifiers has a negative correlation with word count (r = -.137, p < .05). All four discourse pattern indices are positively correlated with the number of words before the main verb. It suggests that the more nouns, arguments, content words and semantic elements are overlapped, the more numbers of words before the main verb appear.

Table 11 shows the result of Pearson's product moment correlation between discourse patterns and descriptive measures. Word counts are not significantly correlated to all four indices in this study. The result agrees with the comparative analysis between 7th and 9th civil-service examination (Y. Shin & Y. Jong, 2018) in that word counts did not have any correlations with argument overlap and

TABLE 9
Pearson's Product Moment Correlation Between Lexical Characteristics and Other Indices

Category	Variable	Lexical diversity	Word frequency	Age of acquisition	Word familiarity	Word concreteness
Descriptive measures	Word count	-.400***	.102	.104	.046	-.277***
	Sentence count	-.285***	.170*	-.095	.080	-.066
	Word length	.119	-.403***	.582***	-.350***	-.270***
	Sentence length	-.075	-.117	.292***	-.086	-.267***
Discourse patterns	Noun overlap	-.599***	-.161*	.075	-.038	.063
	Argument overlap	-.388***	-.009	.028	.010	-.064
	Content word overlap	-.600***	.126	-.025	.208**	-.040
	LSA overlap	-.479***	-.175*	.075	-.043	.051

* p < .05, ** p < .01, *** p < .001

TABLE 10
Pearson's Product Moment Correlation Between Syntactic Complexity and Other Indices

Category	Variable	Words before the main verb	NP modifiers
Descriptive measures	Word count	.130	-.137*
	Sentence count	-.323***	-.192**
	Word length	.214**	.314***
	Sentence length	.642***	.080
Lexical characteristics	Lexical diversity	-.081	.035
	Word frequency	-.028	-.437***
	Age of acquisition	.127	.059
	Word familiarity	.030	-.313***
	Word concreteness	-.139*	.250***
Discourse patterns	Noun overlap	.277***	.140*
	Argument overlap	.310***	-.111
	Content word overlap	.242***	-.096
	LSA overlap	.251***	.075

* $p < .05$, ** $p < .01$, *** $p < .001$

TABLE 11
Pearson's Product Moment Correlation Between Discourse Patterns and Descriptive Measures

Category	Variable	Noun overlap	Argument overlap	Content word overlap	LSA overlap
Descriptive measures	Word count	.120	.069	.011	-.070
	Sentence count	-.140*	-.181**	-.111	-.284***
	Word length	.140*	-.005	-.120	.113
	Sentence length	.296***	.306***	.172*	.281***

* $p < .05$, ** $p < .01$, *** $p < .001$

LSA overlap. On the other hand, sentence length is significantly correlated to all four indices. The result is contradictory to the research on 7th and 9th civil-service examination (Y. Shin & Y. Jong, 2018) in that sentence length did not have any correlations with argument overlap and LSA overlap in the previous study. It indicates that the number of sentences in the text decreases and the number of words in a sentence increases the overlap of the same linguistic elements in the NAEA reading passages.

V. CONCLUSION

The current study aims to identify differences between the NAEA English reading passages for middle and high school students in order to shed light on whether the text difficulty systematically rises as the grade increases. Descriptive measures, lexical diversity, syntactic complexity, discourse patterns and readability were analyzed using Coh-Metrix. Few studies investigated differences between the NAEA English reading passages for middle and high school students. In this regard, this study is necessary.

For this study, the NAEA English reading passages of the five consecutive years from 2015 to 2019 were extracted. The extracted reading passages were analyzed based on 17 selected Coh-Metrix indices related to descriptive measures, lexical characteristics, syntactic complexity, discourse patterns, and readability level.

The two research questions were raised to compare the NAEA English reading passages for middle and high school students. The first research question focused on the differences in descriptive measures, lexical characteristics, syntactic complexity, discourse patterns, and readability level. The second research questions regarded the correlations among descriptive measures, lexical characteristics, syntactic complexity, discourse patterns, and readability level.

As for the differences of two tests, descriptive measures, lexical difficulty and readability level showed clear differences, whereas lexical diversity, syntactic complexity, and discourse patterns did not present critical gaps between two tests. More specifically, regarding descriptive measures, text length of the test increased as the grade was raised. The results from the lexical characteristics indicated that the high school English reading passages consisted of more difficult words than middle school reading passages, but the two tests had the almost same level of lexical diversity. The results could be explained by the

previous study conducted by O. I. Ha and Y. Cho (2016) in that the NAEA middle school reading passages had various words. The complexity of the sentence structures did not rise as the grade increased in that the higher measured value of syntactic complexity denotes the more difficult text although their targets were supposed to have different English proficiency. The results of discourse patterns denoted that English reading passages for middle and high school students had the similar text difficulty in that the repetition of the same element in a text can decrease text difficulty (Crossley & McNamara, 2011b; McNamara et al., 2014). The readability level in the current study indicated that high school English reading passages were significantly more difficult than middle school English reading passages. The way of calculation could affect the statistical differences between two tests in that the current study demonstrated words and sentences became longer as the grade increased in the NAEA English tests.

As for the correlations among the selected Coh-Metrix indices, it was inevitable that the type-token ratio was negatively correlated to all the indices of discourse patterns in that lexical diversity, which the type-token ratio measures, is contradictory to noun overlap, argument overlap, content word overlap, and LSA overlap. The results partially corresponded to previous studies that word length, sentence length, lexical difficulty, and sentence structures affect readability (Klare, 1984), and vocabulary, sentence length, sentence structure, and cohesion influence text difficulty (Graves & Graves, 2003). The positive correlations with word length and sentence length are inevitable because the way to estimate the Flesch Reading Ease score and the Flesch-Kincaid Grade Level is based on the average number of syllables per word and words per sentence.

The educational implications are as follows. Word length and sentence length are the main key to distinguish the text difficulty of the two tests with lexical difficulty. However, the two tests had the same level related to lexical diversity, sentence structures and cohesive devices although their target had different English reading level. The NAEA high school English reading passages need to be carefully manipulated in order to increase the level of text difficulty in accordance with their target's school year. For the right difficulty, the same content words have to be less overlapped, and sentence structures need to be more complex in the NAEA high school English reading passages.

The limitation of this study is that the result could be different with more reading passages or with listening texts. Especially with more previous tests like 2010 or 2013 NAEA, research could draw different results as the sentences in the NAEA reading texts for middle school students were getting significantly longer and more complex than before (J. Kim, 2018). Further studies need to extract more samples. In conclusion, this study provides educational implications to improve the quality of the NEAE English reading tests, which could eventually help to enhance teaching and learning.

REFERENCES

Baayen, R, H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania.

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy* (3rd ed.). Pearson Education.

Cha, Min Kyoung, & Lee, Hyunoo. (2016). Analysis of speech acts in the high school English textbooks and the NAEA high school English conversational materials. *Secondary English Education*, *9*(3), 123-148.

Chang, Ji-Yeon. (2018). A comparison of 2017-2018 CSAT reading passages via Coh-Metrix: Focusing on descriptive, readability, and easability measures. *Foreign Languages Education*, *25*(4), 81-106.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33*(4), 497-505.

Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, *18*(2), 119-135.

Crossley, S. A., & McNamara, D. S. (2011a). Shared features of L2 writing: Intergroup homogeneity and text classification. *Journal of Second Language Writing*, *20*(4), 271-285.

Crossley, S. A., & McNamara, D. S. (2011b). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life Long Learning*, *21*(2-3), 170-191.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221-233.

Graesser, A., McNamara, D., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, *36*(2), 193-202.

Graves, M. F., & Graves, B. B. (2003). *Scaffolding reading experiences: Designs for student success*. Christopher-Gordon.

Ha, Ok I, & Cho, Yunkyoung. (2016). A comparative analysis of the reading passages of National Assessment of Educational Achievement and middle school English textbooks: Focus on lexical difficulty and readability. *Foreign Languages Education*, *23*(3), 191-213.

Jeon, Moongee. (2011). A corpus-based analysis of continuity of the reading materials in middle school English 1 and 2 textbooks with Coh-Metrix. *The Journal of Linguistic Science*, *56*(1), 201-218.

Jeon, Moongee. (2014). An analysis of the continuity of the reading materials in high school English textbooks. *Journal of the Korean Data Analysis Society*, *16*(2), 925-938.

Jeon, Moongee. (2015). An analysis of the continuity among middle school English textbooks with an automated language analysis program. *Modern English Education*, *16*(1), 195-218.

Jeon, Moongee, & Lim, Injae. (2009). A corpus-based analysis of middle school English 1 textbooks with Coh-Metrix. *English Language Teaching*, *21*(4), 265-292.

Jeong, Mikyung. (2015). *A comparative study on the English argumentative essays of Korean and native English university students with Coh-Metrix: Focused on lexical, syntactic, and discourse features* [Unpublished doctoral dissertation]. Kangwon National University.

Kennedy, G. (2014). *An introduction to corpus linguistics*. Routledge.

Kim, Jeong-ryeol, & Yang, Jiyun. (2012). An analysis of the continuity of elementary and middle school English textbooks using Coh-Metrix. *English Teaching*, *67*(2), 319-341.

Kim, Joonkoo. (2018). A corpus-based investigation of reading passages in the National Assessment of Educational Achievement English test using Coh-Metrix. *Secondary English Education*, *11*(2), 27-51.

Kim, Sung Hye. (2016). Analysis of student responses to constructed response items in English of the National Assessment of Educational Achievement. *Modern English Education*, *17*(3), 245-262.

Kincaid, J., Fishburne, R., Rogers, R., & Chissom, B. (1975). *Derivation of new readability formulas for navy enlisted personnel*. Branch Report 8-75. Naval Technical Training Command Millington TN Research Branch.

Klare, G. R. (1974-1975). Assessing readability. *Reading Research Quarterly*, *10*(1), 62-102.

Klare, G. R. (1984). Readability. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 168-744). Longman.

Korea Institute for Curriculum and Evaluation. (2014). *The analysis of result from the 2013 National Assessment of Educational Achievement: English*. https://www.kice.re.kr/resrchBoard/view.do?seq=41018&s=-kice&m=030101

Krashen, S. D. (2004). *The power of reading*. Paw Prints.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259-284.

Lee, Hyunoo. (2021). Distribution of graded vocabulary in the high school English textbooks and K-11 NAEA English test based on the 2015 Revised National Curriculum. *Secondary English Education*, *14*(1), 22-44.

Lee, Jiyoung. (2020). A comparison of text difficulty in 2015-2020 CSAT English reading passages between testing types and among question types. *Journal of the Korea English Education Society*, *19*(2), 21-43.

Lee, Seung-Hwan. (2013). A corpus-based analysis of the difficulty of the reading materials in high school English 1 and 2 textbooks using Coh-Metrix. *Studies in Foreign Language Education*, *27*(2), 131-148.

Lee, Young-Ju. (2012). The effect of students' background characteristics on English performance in the National Assessment of Educational Achievement. *English Teaching*, *67*(2), 393-415.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1), 57-86.

McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Source of text difficulty: Across the ages and genres. In J. P. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 89-116). Rowman & Littlefield Education.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Pae, Jue-Kyoung. (2018). An analysis of English writing achievement characteristics of middle-school students: Focusing on the response to constructed-response items of the 2015 National Assessment of Educational Achievement (NAEA). *Journal of the Korea English Education Society*, *17*(1), 85-105.

Pae, Jue-Kyoung, & Hwang, Pil-ah. (2015). An analysis of test items for the National Assessment of Educational Achievement (NAEA) in Grade 9 English using the distribution curves of response rates. *Journal of the Korea English Education Society*, *14*(3), 161-182.

Pae, Tae-Il. (2012). A comparative study of the factors affecting English achievement for Korean middle and high school students. *English Language Teaching*, *24*(1), 167-186.

Park, Hyeonsuk, & Jung, Chae Kwan. (2020). An analysis of the continuity among reading passages in high school English textbooks in the 2009 and 2015 Revised English Curriculums. *Studies in Foreign Language Education*, *34*(4). 179-202.

Park, In-Yong, Kim, Wan-Soo, Jung, Hye-Kyung, Seo, Min-Hee, Han, Jung-Ah, & Lee, Hyun-Sook. (2017). *Analysis on academic achievement and affective domains in the results of the National Assessment of Educational Achievement*. Korea Institute for Curriculum and Evaluation. https://www.kice.re.kr/resrchBoard/view.do?seq=329&s=kice&m=030101

Ryu, Jisu, & Jeon, Moongee. (2020). An analysis of text difficulty across grades in Korean middle school English textbooks using Coh-Metrix. *The Journal of Asia TEFL*, *17*(3), 921-936.

Shin, Yousun. (2019). Analyzing CSAT reading passages by using Coh-Metrix and VocaProfile: Focusing on four years from 2016 to 2019. *Journal of Language Sciences*, *26*(4), 109-127.

Shin, Yousun, & Jong, Youngkyong. (2018). Analyzing reading passages in the English tests for civil servants with Coh-Metrix: Focusing on lexical diversity, syntactic complexity, cohesion, and readability. *Studies in British and American Language and Literature*, *129*, 143-161.