



## An Analysis of Text Difficulty across Grades in Korean Middle School English Textbooks Using Coh-Metrix

Jisu Ryu

*Konkuk University, South Korea*

Moongee Jeon

*Konkuk University, South Korea*

This study aims to investigate the text difficulty of the reading materials of Korean middle school English textbooks with Coh-Metrix, a software developed by the Institute for Intelligent Systems at the University of Memphis to analyze the linguistic and psycholinguistic features of English text and textbooks with a wide range of indices on cohesion and language. In this study, the textbook corpus consisted of the text files extracted from 13 English textbooks. These files were used for analyzing the text difficulty among grades with Coh-Metrix. The Coh-Metrix indices selected for this study contained basic counts, word frequency, word features, lexical diversity, pronouns, connectives, readability, syntax complexity, syntax similarity, reference cohesion, semantic cohesion, and situation model measures. The results showed that there were significant differences among grades for basic counts, word features, first pronouns, causal and temporal connectives, readability, reference and semantic cohesion, the number of words before main verbs, syntactic similarity, and situation model measures. The differences among grades, however, were not significant for word frequency, lexical diversity, second and third person pronouns, additive connectives, and NP density measures. The findings have educational implications for textbook design and language learning for English learners.

**Keywords:** text difficult, reading materials, Korean middle school English textbooks, Coh-Metrix, corpus

### Introduction

The significance of English textbooks has been widely acknowledged, particularly in EFL settings in which L2 learners have an extensive, albeit not exclusive, reliance on textbooks (Almujaiwel, 2018; Chao, 2011; Chen, 2016; Jeon & Lim, 2009, 2010; Plakans & Bilki, 2016; Ryu & Jeon, 2020; To, 2018). Nevertheless, relatively little attention has been paid to the appropriateness of textual input per se, despite its pivotal role in promoting linguistic competence (Chen, 2016). Recently an increasing number of researchers, however, have begun to address the relevance of textual input in L2 English textbooks focusing on text difficulty. Specifically, they assume that text difficulty should systematically increase as proficiency levels progress for the effective development of linguistic competence. The idea is theoretically motivated by the Zone of Proximal Development (Vygotsky, 1980) and the input hypothesis (Krashen, 1985) which imply that the input slightly higher than learners' current state is optimal for maximum learning outcomes.

Previous studies in the text difficulty of L2 textbooks primarily employed readability indices and text

analysis tools in which their unit of analysis was predominantly limited to surface features such as word or sentence level features (Goh & Park, 2007; Im, Cho, & Jong, 2015; Kim & Chun, 2008; Kim & Song, 2007). The growing skepticism regarding the adequacy of such shallow features in determining overall text difficulty (Morales, 2019; Plakans & Bilkis, 2016; Spencer & Wagner, 2016) led to the development of more sophisticated tools such as Coh-Metrix (Graesser, Jeon, Yan, & Cai, 2007) to analyze more diverse features of text. For example, Plakans and Bilkis (2016) employed 17 indices provided by Coh-Metrix to examine whether L2 English textbooks of beginning, intermediate, and advanced levels systematically differ in text difficulty; their findings showed that systematic arrangement was not observed for cohesion-related indices, whereas text difficulty linearly increased for surface features such as sentence length.

Despite their contributions for textbook studies, prior research in the text difficulty of L2 textbooks is limited in the number of indices used for measuring overall text difficulty (around 10 indices for many studies). This indicates that more comprehensive indices need to be considered for manipulating the difficulty of textbooks across grade levels effectively. In addition, various linguistic or psycholinguistic measures help teachers to select appropriate textbooks based on textbook difficulty in school education.

Recently, a series of Korean middle school English textbooks have been newly developed by the 2015 national curriculum. The primary purpose of this study is to determine whether the newly developed textbooks conform to the systematic arrangement of the text difficulty across grade levels. In order to achieve this goal, this study analyzed L2 English textbooks for first, second, and third graders in junior high school by a total of 31 critical indices of Coh-Metrix for examining a wider range of lexical, syntactic, cohesion-related, and situation model related features reflected in the textbooks. Coh-Metrix is a language analysis software that has been widely used in the areas of English education and psycholinguistics to examine the linguistic and psycholinguistic features of textbooks, texts, and essays based on various measures (Graesser et al., 2007). Coh-Metrix is suitable for analyzing the properties of a wide range of texts because it provides surface and deep level measurements. For this reason, in this study, Coh-Metrix was used to analyze the text difficulty of middle school English textbooks.

## Previous Studies

### English Textbooks in EFL Settings

Given the importance of English textbooks in English education in L2 settings where the target language exposure is constrained (Chen, 2016; Morales, 2019; Tai, 2016), the textbooks serve as a primary source of input (Jeon & Lim, 2010). Furthermore, it has also been suggested that textbooks may have an even greater influence on learning outcomes than teachers (Nicol & Crespo, 2006). Noting the vital role of textbooks in L2 language development, previous studies have investigated diverse aspects of English textbooks.

Researches have conducted on each of four skills (Lee & Kim, 2007 for speaking; Lim & Jeon, 2013, Song, 2013 for listening; Kim & Jeon, 2013; Kim & Yang, 2012, Jeon & Lim, 2009 for reading; Son & Lee, 2005 for writing), cultural aspects reflected in textbooks (Jang & Kim, 2018), the students' perception of textbooks (Jeong & Kim, 2017), the comparison of textbooks across different countries (Choi, 2015; Ahn, 2016), and checklists for a strict evaluation of textbooks (Tomlinson, 2012; Tsagari & Sifakis, 2014).

Among many components of textbooks, relatively little attention has been paid to the textual input *per se*. Emerging interests that (textual) input serves as a central role in promoting linguistic competence (Chen, 2016, Plakans & Bilki, 2016), however, triggered a series of research examining the linguistic aspects of text materials, particularly in terms of text difficulty in textbooks. The basic premise of this line of research is that selecting the appropriate level of text is essential for maximum learning outcomes (Paul, 2003). More specifically, text difficulty should linearly increase as a function of target readers'

proficiency levels (Berendes et al. 2019; Chen, 2016; Gupta, 2013; Morales, 2019). This idea is theoretically motivated; according to the Zone of Proximal Development (Vygotsky, 1980) and Krashen's input hypothesis (Krashen, 1985), effective learning occurs when the difficulty of input is slightly higher than learners' current level, thereby emphasizing the significance of a careful matching of text difficulty with readers' competence. Without this deliberate arrangement of text difficulty across readers' progressively increasing levels, this would lead to unexpected frustration or boredom (Young & Riegeluth, 1988), overload in working memory (Berendes et al., 2009), and/or decreased motivation and engagement (Feng, D'Mello, & Graesser, 2013).

## **Text Difficulty in Textbooks**

For promoting the optimal development of linguistic competence, the difficulty of textbooks should systematically increase across readers' level. Textbook development and evaluation process, however, are often driven by developers' intuition (see Dubin, 1995; Gupta, 2013; Plakans & Bilkis, 2016) which has been criticized for the relative lack of objectivity and validity. Another major source that affects textbook development and evaluation is readability indices (Crossely, Allen, & McNamara, 2011), which present the numerical level of text difficulty, mostly as a function of word length as a substitute for semantic complexity and sentence length as a substitute for syntactic complexity (Berendes et al., 2019). A number of readability indices have been put forth over the several decades, in order to select, develop, and evaluate text materials of various sorts (e.g., Flesch-Kincaid Grade Level Readability Formula, The Gunning's Fog Index, SourceRater, Person Reading Maturity Metric, and Lexile Index).

Readability indices have also been used in textbook research for analyzing and evaluating as well as developing and modifying text materials. More specifically, researchers have been concerned with addressing such issues as whether English textbooks follow the systematic progression of text difficulty along academic years by using readability indices (Kim & Song, 2007; Im et al., 2015). For example, Im et al. (2015) examined English textbooks of Grades 6 and 7 to evaluate the systematicity in difficulty progression across grade levels via Flesch-Kincaid Grade Level and TTR (type-token ratio) for lexical diversity. Their findings showed that text difficulty increased with grades in terms of both readability and lexical difficulty; the gap, however, was overly wide so this failure of matching text difficulty with readers' expected level (or slightly higher than the current level) might likely cause learning difficulties.

In line with the increasing attention paid to text difficulty measures, a similar level of interests has arisen as to the development of various text analysis tools such as WordSmith Tools, NLP Tools, Range32H (Heatley & Nation, 2006), and VocabProfile. Using these tools, a number of studies have been conducted to investigate the text difficulty of textbooks across grade levels (Goh & Park, 2007; Kim & Chun, 2008). For instance, Goh and Park (2007) employed WordSmith Tools to investigate the text difficulty of L2 English textbooks across Grades 7 to 9 as well as across units within each grade. The major findings indicated that textbooks, in general, are consistent with the linear complexification principle (Berendes et al., 2019). Namely, both the total number of tokens and sentence length gradually increased with grade levels; in contrast, some textbooks were shown not to conform to this principle for the type token ratio (TTR) and the presentation/repetition of novel words.

The analytical tools used in these studies along with existing readability indices mainly address word-level or sentence-level features in analyzing overall text difficulty. These measures, however, have been criticized due to its exclusive focus on surface level features, with an emerging awareness that reading is multidimensional process, not properly captured by such shallow level features (Morales, 2019; Plakans & Bilkis, 2016; Spencer & Wagner, 2016).

Driven by the concern for more objective and valid measurement tools for text analysis, advanced language analysis tools have been developed to measure the various aspects of texts. One of the widely used is Coh-Matrix which provides over 100 indices based on the WordNet (Miller, Beckwith, Feldbaum, Gross, & Miller, 1990), the MRC Psycholinguistic Database (Coltheart, 1981), the Latent Semantic Analysis (Landauer, 2007), and parsing and tagging analysis tools (Brill, 1995). Due to their

comprehensive analysis of text features, such tools have been employed as alternatives for previous intuition-based or surface-level based analyses in textbook research.

The validity of Coh-Metrix has been tested by various text and textbook studies. Subsequent researchers in textbooks have explored various linguistic and psycholinguistic features to examine whether textbooks are systematically organized with increasing text difficulty across expected proficiency levels such as grade levels (Bae, 2019; Berendes et al., 2019; Chen, 2016; Gupta, 2013; Jeon, 2011, 2014, 2015; Kim, 2014; Kim & Jeon, 2013; Kim & Yang, 2012; Lee, 2013; Plakans & Bilki, 2016; Song, 2013; To, 2018). For example, Plakans and Bilki (2016) examined L2 college level English textbooks to see if three levels of textbooks (i.e., beginning, intermediate, and advanced) are systematically distinct in text complexity. They analyzed 17 Coh-Metrix features including descriptive indices, syntactic indices, cohesion-related indices, and others. Their results showed that the systematic arrangement of text difficulty was well established with respect to surface features such as sentence length, text length, and readability scores; other features, particularly cohesion-related features, however, were not significantly different among the different levels of textbooks. They concluded that only shallow features were considered for the textbook development, and suggested that more diverse features (particularly cohesion-related ones) needed to be reflected in the development process as well.

Previous textbook studies in text difficulty are significant to the extent that they specified how existing textbooks reflect orderly arrangement of text difficulty in some aspects and not in others, thereby affording more concrete guidelines for the future textbook development. Nevertheless, past studies in L2 textbooks only employed a limited number of indices (more or less 10 indices for many studies; except a few studies examining around 20 indices). These results suggest that more indices are required for measuring more comprehensive aspects of text. In line with these findings, this study analyzed the text difficulty of the reading materials in the middle school English textbooks with a wide range of linguistic and psycholinguistic measures.

## Method

### English Textbook Corpus for the Study

In this study, the entire list of reading materials in middle school English 1, 2, and 3 textbooks was selected to analyze the text difficulty among grades (Table 1). These textbooks were written by the 2015 National Curriculum. Specifically, a total of 13 English textbooks for each grade level by 10 publishers were analyzed. The reading materials extracted from a total of 39 English textbooks were constructed as a corpus to analyze the linkage among grades.

TABLE 1  
*Composition of Instruction Groups*

English Textbook List	Publisher
Middle School English 1, 2, 3	Kumsung Publishing Co. (Choi et al.)
Middle School English 1, 2, 3	Darakwon Inc. (Kang et al.)
Middle School English 1, 2, 3	Dong-A Publishing Co. (Yoon et al.)
Middle School English 1, 2, 3	Dong-A Publishing Co. (Lee et al.)
Middle School English 1, 2, 3	Mirae-N Co. (Choi et al.)
Middle School English 1, 2, 3	Visang Education Co. (Kim et al.)
Middle School English 1, 2, 3	NE Neungyule Inc. (Kim et al.)
Middle School English 1, 2, 3	NE Neungyule Inc. (Yang et al.)
Middle School English 1, 2, 3	YBM Co. (Park et al.)
Middle School English 1, 2, 3	YBM Co. (Song et al.)
Middle School English 1, 2, 3	Jihaksa Publishing Co. (Min et al.)
Middle School English 1, 2, 3	Chunjae Education Inc. (Lee et al.)
Middle School English 1, 2, 3	Chunjae Education Inc. (Jeong et al.)

## Corpus Construction and Analysis Tool

Text files (txt files) in the Unicode format were used to analyze the text difficulty among grades by using the Coh-Metrix system developed by the Institute for Intelligent Systems at the University of Memphis in the United States. Specifically, the text files were extracted from 13 English textbooks and constructed as a corpus. When building the corpus, it was constructed for each chapter included in each textbook. The entire corpus consisted of 71 files for the 1st grade, 139 for the 2nd grade, and 105 files for the 3rd grade. So, a total of 315 reading material files were used for analyzing the text difficulty among grades with Coh-Metrix. The Coh-Metrix measures selected for this study included basic counts, word frequency, word features, lexical diversity, pronouns, connectives, readability indices, syntax complexity, syntax similarity, reference cohesion, semantic cohesion, and situation model.

## Results

In this study, a one-way ANOVA was performed to analyze the text difficulty for the reading materials in middle school English 1, 2, and 3 textbooks using SPSS 16.0 (Statistical Package for the Social Sciences). Specifically, the independent variable was grade level and the dependent variable was each measure of Coh-Metrix. Each analysis was performed at a significance level of 5% ( $p = .05$ ).

### Basic Counts

The results of basic count analysis were summarized in Table 2. As shown in Table 2, the text difficulty among the middle school English textbooks was well established for the number of words, the number of sentences, the average sentence length, and the average word length. Specifically, as the grade level went up, the scores of all these measures increased significantly. Tukey HSD post hoc analyses also showed there were significant differences among the three grade levels for the number of words and the average sentence length ( $p < .05$ ). The difference between middle school English 1 and middle school English 2, however, was not statistically significant for the number of sentences ( $p = .627$ ) and the average sentence length ( $p = .098$ ).

TABLE 2  
*Results of Basic Counts*

Indices	ME 1 (n = 71)	ME 2 (n = 139)	ME 3 (n = 105)	F	p
Number of words	213.5(45.7)	270.9(63.3)	408.8(78.2)	220.5	.000
Number of sentences	29.9(7.6)	31.1(7.9)	37.3(9.9)	21.66	.000
Average sentence length	7.35(1.47)	8.92(1.87)	11.25(1.88)	107.7	.000
Average word length	1.30(.08)	1.33(.08)	1.37(.08)	13.47	.000

Sentence length and the number of sentences are important linguistic factors that can influence text comprehension or difficulty (Graesser et al., 2007). So, these results suggest that textbook developers need to control these elements (i.e., the number of words, the average sentence length) more strictly for English language learners.

### Word Frequency

Word frequency is an important ingredient influencing the difficulty of text (Graesser et al., 2007; Graesser, McNamara, Louwerse, & Cai, 2004). So, in developing English textbooks, it is pivotal to control the frequency of words appearing in textbooks to manipulate the difficulty level of textbooks gradually (Jeon, 2011, 2015). The Coh-Metrix system calculates word frequency measures using the CELEX database created by the Dutch Center for Lexical Information for word studies (Baayen,

Piepenbrock, & Gulikers, 1995). In this study, word frequency measures converted by the logarithmic transformation were used for analyzing the text difficulty among grades. When raw frequency scores are log-transformed, the distribution of the log-transformed scores becomes more approximate to the normal distribution. So, this kind of converted score is widely used for word frequency analyses (Graesser et al., 2004).

TABLE 3  
*Results of Word Frequency*

Indices	ME 1 (n = 71)	ME 2 (n = 139)	ME 3 (n = 105)	F	p
Word frequency (content words)	2.47(.13)	2.47(.13)	2.45(.13)	1.21	.301
Word frequency (all words)	3.10(.10)	3.13(.09)	3.12(.09)	1.51	.223

As presented in Table 3, the differences among the grades were not statistically significant for the word frequency scores for both content words and all words, thereby indicating that the text difficulty among the three grade levels was not well established. The results suggest that these lexical elements should be systematically reflected in developing textbooks in order to gradually manipulate the text difficulty along with the results of the number of sentences and the average sentence length.

## Word Features

Word features such as imageability, concreteness, age of acquisition, familiarity, and meaningfulness are essential factors affecting the difficulty of texts (Graesser et al., 2004; Jeon, 2015; Ryu & Jeon, 2020). The Coh-Metrix system calculates the word feature measures based on the MRC Psycholinguistics Database (Coltheart, 1981). The values of these measures are in the range of 100 to 700.

TABLE 4  
*Results of Word Features*

Indices	ME 1 (n = 71)	ME 2 (n = 139)	ME 3 (n = 105)	F	p
Imageability	441.3(23.44)	431.9(23.94)	420.9(24.34)	15.81	.000
Concreteness	406.1(28.03)	398.5(27.83)	388.1(27.96)	9.25	.000
Age of acquisition	269.2(27.68)	283.0(24.37)	301.0(23.82)	35.96	.000
Familiarity	583.5(6.69)	582.2(6.47)	580.1(7.25)	5.66	.004
Meaningfulness	444.1(14.23)	439.9(13.90)	437.9(14.19)	5.32	.005

As presented in Table 4, there were significant differences among the three grade levels for all the word feature indices of Coh-Metrix, indicating that the text difficulty for the three grade levels was well constructed for these measures. Tukey HSD post hoc analyses also showed the differences among the three grades were statistically significant for the imageability and age of acquisition measures. The difference between the middle school English 1 and middle school English 2 textbooks, however, was not significant for the concreteness ( $p = .153$ ), familiarity ( $p = .400$ ), and meaningfulness measures ( $p = .103$ ). The difference between the middle school English 2 and middle school English 3 textbooks was also not significant for the meaningfulness ( $p = .262$ ). These results imply that textbook developers need to focus on these specific word features in greater detail when they write textbooks to manipulate the text difficulty more properly.

## Lexical Diversity

The lexical diversity measure provided by the Coh-Metrix system is presented as a type-token ratio value for content words (Graesser et al., 2004; Ryu & Jeon, 2020). A type refers to the first occurrence frequency of an individual word, whereas a token indicates the total occurrence frequency of the word. So,

the type value is always 1. For example, if the word “corpus” appears two times, the type value of this word is 1 and the token is 2. In this case, the type-token ratio value is 0.5 ( $1/2 = 0.5$ ). When a type-token ratio value of a text is high, the possibility that various words are used in the text increases accordingly because the type value is relatively high and the token value is relatively low. So, if a type-token ratio value of a text is high, the difficulty of the text tends to increase because the processing burden of new words gets larger (Graesser et al., 2007).

TABLE 5  
*Results of Lexical Diversity*

Indices	ME 1 (n = 71)	ME 2 (n = 139)	ME 3 (n = 105)	<i>F</i>	<i>p</i>
Type-token ratio	.70(.07)	.69(.07)	.68(.06)	1.66	.191

As presented in Table 5, there was no significant difference among the three grade levels for the type-token ratio for content words, thereby indicating that the text difficulty for the grade level was not well established. The current findings suggest that textbooks for lower grades include as diverse words as those in textbooks for higher grades. This can pose substantial problems for the L2 learners, because new words cannot be systematically learned as the grade increases. This result suggests that a computer tool such as Coh-Metrix can be useful for developing textbooks because it is difficult to explicitly examine the lexical diversity of the textbooks without the aid of computer tools.

## Personal Pronouns

As the occurrence of pronouns in a text increases, the difficulty level of the text tends to increase accordingly because the processing burden of handling the coreference resolution for those pronouns also increases (Graesser et al., 2004). As shown in Table 6, the differences among the grade levels were statistically significant for the first person pronouns.

TABLE 6  
*Results of Personal Pronouns*

Indices (incidence score)	ME 1 (n = 71)	ME 2 (n = 139)	ME 3 (n = 105)	<i>F</i>	<i>p</i>
First person pronouns	31.90(25.91)	21.95(21.83)	20.96(18.68)	6.26	.002
Second person pronouns	9.62(11.28)	11.29(13.09)	9.87(9.22)	.685	.505
Third person pronouns	17.56(14.87)	18.68(15.09)	17.29(14.34)	.318	.728

On the other hand, there were no significant differences among the grade levels for the second and the third person pronouns. Tukey HSD post hoc analyses showed the difference between the middle school English 2 and the middle school English 3 textbooks was not significant for the first person pronouns.

It is also suggested that how to control the use of the second and the third person pronouns should be considered as important factors affecting the text difficulty when textbook writers develop textbooks for English language instructors and learners.

## Connectives

Connectives generally tend to promote the connection between sentences. In particular, causal connectives play an important role in facilitating the semantic connections between sentences (Millis & Just, 1994; Murray, 1997). Coh-Metrix provides causal connectives, additive connectives, and temporal connectives.

TABLE 7  
*Results of Connectives*

Indices	ME 1 (n = 71)	ME 2 (n = 139)	ME 3 (n = 105)	F	p
Causal connectives	21.34(11.96)	24.34(12.80)	28.70(9.41)	9.11	.000
Additive connectives	7.58(7.08)	7.71(6.35)	9.28(5.43)	2.36	.096
Temporal connectives	24.30(14.24)	21.21(12.31)	19.32(8.54)	3.84	.023

As presented in Table 7, there were statistically significant differences among the grades for the causal and temporal connectives. The differences, however, were not significant for the additive connectives. Tukey post hoc analyses showed that the difference between the middle school English 1 and 2 textbooks was not significant ( $p = .180$ ) for the causal connectives. Tukey post hoc analyses also indicated that the difference between the English 1 and 2 textbooks was not significant ( $p = .168$ ) for the temporal connectives, and that the difference between the English 2 and 3 textbooks was not significant either ( $p = .425$ ).

## Readability Indices

With regard to the standard readability indices, the Coh-Metrix system presents Flesch Reading Ease (FRE), Flesch-Kincaid Grade Level (FKGL), and L2 Reading scores as shown in Table 8 (Graesser et al., 2007). FRE scores are expressed in the range 0 to 100 and FKGL scores are presented in the range 1 to 12. A higher FRE score indicates easier reading, whereas a higher FKGL score means more demanding reading (Graesser et al., 2004). The L2 Reading score provided by Coh-Metrix is a readability index for second language learners (McNamara, Graesser, McCarthy, & Cai, 2014). A higher L2 Reading score shows easier reading.

TABLE 8  
*Results of Readability Indices*

Indices	ME 1 (n = 71)	ME 2 (n = 139)	ME 3 (n = 105)	F	p
Flesch Reading Ease	88.88(6.77)	85.26(7.64)	79.80(7.75)	33.40	.000
Flesch-Kincaid Grade Level	2.67(1.18)	3.57(1.40)	4.92(1.40)	62.74	.000
L2 Reading Score	27.63(5.37)	25.88(4.77)	23.85(4.40)	13.57	.000

As presented in Table 8, there were significant differences among the three grades for the Flesch Reading Ease, Flesch-Kincaid Grade Level, and L2 Reading scores. Tukey HSD post hoc analyses also showed that all the differences between the comparison groups (i.e., ME 1 vs. ME 2, ME 2 vs. ME 3, ME 1 vs. ME 3) were statistically significant for the readability measures selected for this study. These results suggest that the text difficulty among the grade levels were well established for the readability indices. The findings are also consistent with the previous findings on the word and sentence length, indicating that these indices are calculated based on the word and sentence length.

## Reference Cohesion

Reference cohesion is an important linguistic factor affecting text comprehension (Cirilo, 1981; Haberlandt & Bingham, 1978). Reference cohesion is formed when arguments are used between two adjacent sentences in a text. Those arguments consist of nouns, pronouns, and noun phrases (Graesser et al., 2004). If the reference cohesion of a text is high, the comprehension of the text increases, thereby indicating that the difficulty level of the text can decrease (Graesser et al., 2007). In Coh-Metrix, reference cohesion is measured by argument overlap scores for adjacent and all sentences.



TABLE 9  
*Results of Referential Cohesion*

Indices	ME 1 (n=71)	ME 2 (n=139)	ME 3 (n=105)	<i>F</i>	<i>p</i>
Overlap for adjacent sentence	.38(.15)	.42(.16)	.51(.15)	17.92	.000
Overlap for all sentences	.27(.11)	.31(.13)	.37(.12)	14.75	.000

As shown in Table 9, there were statistically significant differences among the grade levels for the argument overlap scores for adjacent and all sentences. Tukey HSD post hoc analyses also showed that the differences between the comparison groups were statistically significant for these measures except for the comparison between the middle school 1st grade and 2nd English textbooks ( $p = .283$  for the argument overlap for adjacent sentences,  $p = .058$  for the argument overlap for all sentences). These results suggest that reference cohesion increases as the grade level increases, thereby indicating that the difficulty level of the textbooks analyzed for this study decreases with the grade level. These results are contrary to general expectations for the manipulation of text difficulty. So, as suggested in Jeon's study (2015), these findings suggest that the linguistic properties of reference cohesion should be well understood by textbook developers to manipulate the difficulty of textbooks more systematically in the right direction.

### Semantic Cohesion

The Coh-Metrix system applies the LSA mathematical model to measure the semantic cohesion of a text (Graesser et al., 2004). This model calculates the semantic similarity between words or sentences based on a huge size of corpus data. Like the reference cohesion measures, if the value of this measure is high, the degree of comprehension of a text decreases.

As presented in Table 10, there were statistically significant differences among the grade levels for the LSA cosine values for adjacent and all sentences. Tukey HSD post hoc analyses also showed that there were significant differences between the comparison groups for these measures except for the comparison between the middle school 1st grade and 2nd English textbooks ( $p = .385$  for the LSA cosine for adjacent sentences,  $p = .776$  for the LSA cosine for all sentences). These results are consistent with the results of the reference cohesion analyses. So, these results also suggest that textbook developers need to understand the linguistic features of LSA measures to control the difficulty of textbooks effectively for language learners.

TABLE 10  
*Results of Semantic Cohesion*

Indices	ME 1 (n=71)	ME 2 (n=139)	ME 3 (n=105)	<i>F</i>	<i>p</i>
LSA for adjacent sentences	.16(.06)	.18(.07)	.20(.07)	5.25	.006
LSA for all sentences	.07(.09)	.07(.10)	.16(.07)	41.13	.000

### Syntactic Complexity

Coh-Metrix measures the syntactic complexity of a text by a noun density score and the number of words before main verbs (Graesser et al., 2004). The noun phrase density score represents a value obtained by dividing the number of words that modify headwords by the number of words that make up the noun phrase. The number of words before main verbs represents the number of words occurring before the main verbs. When the values of these measures of a text are high, the text difficulty tends to increase (Graesser et al., 2004).

TABLE 11  
*Results of Syntactic Complexity*

Indices	ME 1 (n=71)	ME 2 (n=139)	ME 3 (n=105)	F	p
NP density	.65(.14)	.69(.15)	.69(.13)	2.06	.129
Words before main verbs	1.68(.53)	2.18(.73)	2.83(.80)	57.54	.000

According to Table 11, there were no significant differences among the three grade levels for the noun phrase density score. The difference among the grades, however, were statistically significant for the number of words before main verbs. Tukey HSD post hoc analyses also showed that all the differences between the comparison groups (i.e., ME 1 vs. ME 2, ME 2 vs. ME 3, ME 1 vs. ME 3) were statistically significant for the number of words before main verbs ( $p < .05$ ). These results indicate that the text difficulty among grades was well constructed for the number of words before main verbs, which was relatively easy to control, but it was not well established for the noun phrases used in the sentences. So, text developers need to pay attention to the linguistic features such as noun phrases to effectively manipulate the difficulty of textbooks.

### Syntactic Similarity

Coh-Metrix provides two types of syntactic similarity measures based on the syntactic structures between sentences in the text (McNamara et al., 2014). The syntactic similarity for all sentences indicates the syntactic similarity for all combinations of sentences in the text, whereas the syntactic similarity for adjacent sentences represents the syntactic resemblance for adjacent sentences in the text. The higher values of these measures result in easier understanding of texts for readers (McNamara et al., 2014).

As shown in Table 12, there were statistically significant differences among the grade levels for both the syntactic similarity for all sentences and the syntactic similarity for adjacent sentences. Tukey HSD post hoc analyses also showed that all of the differences between the comparison groups (i.e., ME 1 vs. ME 2, ME 2 vs. ME 3, ME 1 vs. ME 3) were statistically significant for these measures (all  $ps < .05$ ), thereby indicating that the text difficulty for the middle school English textbooks was well established for language learners. These results indicate that the L2 learners can benefit from the English textbooks to the extent that they are exposed to more similar (i.e., syntactically similar) sentences in their early course of learning, and they gradually can learn diverse syntactic structures as their proficiency goes up.

TABLE 12  
*Results of Syntactic Similarity*

Indices	ME 1 (n=71)	ME 2 (n=139)	ME 3 (n=105)	F	p
Similarity for all sentences	.20(.05)	.17(.04)	.13(.02)	62.38	.000
Similarity for adjacent sentences	.20(.05)	.17(.04)	.14(.03)	52.65	.000

### Situation Model

Coh-Metrix provides two types of situation model indices (Graesser et al., 2004; McNamara et al., 2014). The causal cohesion indicates the number of causal particles (e.g., *because, in order to*) divided by the number causal verbs (e.g., *break, impact, hit, move*), and the intentional cohesion refers to the number of intentional particles (e.g., *by means of, so that*) divided by the number intentional verbs (e.g., *contact, drop, walk*). Basically, language learners tend to spend more cognitive resources to understand highly causal and intentional writings (Graesser et al., 2004; McNamara et al., 2014). These cognitive burdens can make it more difficult for the learners to understand those writings.

TABLE 13  
*Results of Situation Model*

Indices	ME 1 (n=71)	ME 2 (n=139)	ME 3 (n=105)	<i>F</i>	<i>p</i>
Causal cohesion	.19(.19)	.24(.20)	.32(.17)	12.79	.000
Intentional cohesion	.45(.32)	.61(.67)	.77(.48)	7.42	.001

According to Table 13, the differences among the grade levels were statistically significant for both the causal and intentional cohesion measures. Tukey HSD post hoc analyses also showed that all of the differences between the comparison groups (i.e., ME 1 vs. ME 2, ME 2 vs. ME 3, ME 1 vs. ME 3) were statistically significant for these situation models ( $p < .05$ ). That is, the causal and intentional cohesion of the reading materials in the middle school English textbooks systematically increased as the learners' proficiency level increases. The current findings imply that the L2 learners are progressively exposed to more causally and intentionally cohesive writings as their proficiency goes up. Thus, the text difficulty on the L2 English textbooks of the current study was systematically controlled for the two situation model indices.

## Discussion and Conclusion

The present study aimed to determine if a series of textbooks newly developed by the 2015 National Curriculum are based on a principled approach in which text difficulty linearly increases with grade levels. For this purpose, L2 English textbooks of first, second, and third graders in junior high school were analyzed by a total of 31 critical indices which enable an examination of a much wider range of features than used in previous studies. The analyses were conducted on word and sentence features, syntactic features, pronouns, reference and semantic cohesion, connectives, and situation model features.

The major findings are summed up as follows. The findings on word features demonstrated that they were partially controlled to increase text difficulty with grade levels. More specifically, the number of words and the average word length increased with grade levels. Namely, the age of acquisition of words progressively increased, whereas imageability, concreteness, familiarity, and meaningfulness measures decreased. Systematicity in text difficulty, however, was not observed in other word features such as word frequency (for both content words and all words) and the type-token ratio of content words. These findings suggest that textbook development need to be driven by more concrete guidelines so that infrequent words and diverse words are gradually presented in textbooks as target readers' proficiency increases.

Second, sentence and syntactic measures generally display the gradual progression of difficulty with grade levels. Specifically, the total number of sentences and the average sentence length increased as the grade level increased, and readability indices (FRE, FKGL, L2 reading ease score) indicated increased difficulty. The number of words before main verbs also increased, whereas syntactic similarity measures (both syntactic similarity measures) decreased with grade levels, thereby advancing the overall text difficulty. A potential account of the decrease in syntactic similarity across grade levels is that more diverse syntactic structures tend to be used in the textbooks of the higher grades. Despite this general pattern of a systematic increase in text difficulty, no significant differences were detected for the NP density measure. While syntactic complexity was properly controlled in the sentence level (as indicated in the number of words before main verbs), a smaller unit such as noun phrases was inclined to be ignored in determining the text difficulty in textbook development (Jeon, 2011).

Third, the analyses of text difficulty for personal pronouns indicated a greater use of first person pronouns, but no significant differences in the use of second and third pronouns as grade levels increased. An increase in pronoun use is purported to present a greater cognitive load as the processing burden of handling the coreference resolution caused by pronouns is higher (Cirilo, 1981). This is particularly the case for third person pronouns which refer to a larger range of entities compared to the first and second pronouns. On the other hand, first person pronouns only refer to the entity that is currently being in focus;

thus, less use of first person pronouns tend to be accompanied by an increase in text difficulty (Jeon, 2014). Another consideration in interpreting the results is that the text materials in the middle school English 1 contain a higher percentage of narratives in which the first person pronouns are used frequently. In contrast, texts in the middle school English 2 and 3 textbooks include more argumentative and expository texts in which third person pronouns are likely to be used more frequently than first person pronouns given their genre characteristics focus on objectivity. This indicates that first person pronouns were well controlled to manage text difficulty across grade levels, whereas a more careful matching of text difficulty are required for second and third person pronouns.

Fourth, referential and semantic cohesion measures (for both adjacent and all sentences) were overall incongruent with general expectations for the manipulation of text difficulty. All estimates displayed a rise in cohesion, which in turn led to decreased difficulty across grade levels. This finding implies that referential and semantic cohesion was not carefully considered by textbook developers, thereby indicating that cohesion features need to be controlled along with others for the balanced development of crucial reading components for L2 learners.

Finally, connectives and situation model measures showed a trend of increasing difficulty across expected proficiency levels. Connectives, in general, tend to create greater cohesion that facilitates comprehension and lowers text difficulty (Millis & Just, 1994; Murray, 1997). For temporal connectives, a progressive decrease in the number of connectives was observable in the textbooks. This is presumably due to the higher portion of narrative texts in the 1st grade textbook in which temporal connectives are likely to be used more frequently than other genres (e.g., argumentative and expository texts); note that tracking temporal progression is one of the key elements in comprehending narratives (Zwaan, Langston, & Graesser, 1995). For causal connectives, a different pattern was obtained in which the use of causal connectives increased with grade levels. Given that the use of many causal connectives is a genre characteristic (i.e., argumentative or expository texts) that requires more cognitive processing (Murray, 1997), the frequent occurrence of causal connectives, in fact, reflect increased difficulty. A similar logic is also applied to causal and intentional cohesion; they are likely to be higher in more challenging texts (expository and argumentative texts). Overall, the casual connectives and situation model measures indicate a systematic increase in text difficulty.

The current findings are comparable to the previous study conducted by Jeon (2015) in which he examined the text difficulty of L2 middle school English textbooks developed by the 2009 National Curriculum using a total of 22 indices. The analyses revealed that word features (the number of words, imageability, concreteness, and age of acquisition) generally increased with grade levels, whereas lexical diversity did not increase in difficulty accordingly; these results are consistent with the current study. Word frequency, however, was more systematically controlled in textbooks by the 2009 National Curriculum than by the 2015 National Curriculum. Namely, it was not statistically different in difficulty across grades in textbooks by the 2015 National Curriculum. This means that a lack of systematic arrangement (i.e., a gradual transition from more frequent words in lower grade to more infrequent words in higher grade) leads to inefficient learning for L2 learners with the recent curriculum than the previous one.

In addition to word features, sentence features (the number of sentences, the average sentence length, readability scores of FRE and FKGL), syntactic features (the number of words before main verbs), pronouns (first person pronouns), and casual connectives indicated a consistent increase in text difficulty in Jeon (2015) as in the current study. Second and third pronouns were also not significantly different across grades in Jeon (2015), which was also found in the current study. Contrary to the increase in difficulty for other aspects, referential and semantic cohesion measures in Jeon (2015) displayed a decrease in difficulty as in the present research. Along with the similar findings between two studies, however, a difference between the textbooks by the 2009 and the 2015 National Curriculum was identified: the difficulty by NP density was properly controlled in the previous curriculum, while it was not the case in the new curriculum.

In sum, the overall text difficulty of L2 English textbooks that are currently used in middle schools

gradually increased with grade levels for most measures; nevertheless, key elements in reading comprehension (e.g., word frequency, lexical diversity, reference cohesion) need to be more carefully manipulated in future textbook development in order to promote a balanced development of L2 reading proficiency. The current study advances the existing research base to the extent that it reveals how the newly developed textbooks adjust text difficulty as readers' grade levels increase with a broader range of indices. The present findings are in contrast to the expectations that more indices are controlled well in the textbooks newly developed by the 2015 National Curriculum than by the 2009 National Curriculum. Many elements that are presumably not reflected in textbook development are difficult to manipulate by visual inspection; as the current study demonstrates, future textbook development can be guided by evidence-based automatic text analysis tools such as Coh-Metrix. Furthermore, this study has implications for material development in general. Developers of materials, particularly linguistic materials, should follow the principle in which the text difficulty gradually increases with the expected proficiency level of the target population. Namely, this principle should be applied to as many features of the text as possible. Then, microscopic analyses on the text need to be conducted for surface level features (e.g., the average word length) and deeper level features (e.g., cohesion-related measures). Some of these features are, as mentioned above, difficult to identify with the naked eye. These, however, can be easily detected by computerized automatic text analysis tools such Coh-Metrix. Overall, this study shows that the Coh-Metrix system is useful for analyzing various types of English textbooks.

### The Authors

*Jisu Ryu* (first author) is a visiting scholar in the Konkuk Research Institute for Multilingualism and Multiculturalism at Konkuk University. Her main interests include L2 reading, corpus linguistics, psycholinguistics, eyetracking, EEG, cognitive linguistics, and language analysis tools.

Konkuk Research Institute for Multilingualism and Multiculturalism  
Konkuk University  
Email: jsryu0508@konkuk.ac.kr

*Moongee Jeon* (corresponding author) is a professor in the Department of English Language & Literature at Konkuk University. His main research interests include corpus analysis, language psychology, English education, artificial intelligence, cognitive science, and software development.

Department of English Language & Literature  
Konkuk University  
Email: mjeon1@konkuk.ac.kr

### References

- Ahn, M. (2016). A comparative study of vocabulary in high school English textbooks in South Korea and China: A corpus-based analysis of lexicon. *The Journal of Foreign Studies*, 36, 55-82.
- Almujaiwel, S. (2018). Culture and interculture in Saudi EFL textbooks: A corpus-based analysis. *The Journal of Asia TEFL*, 15(2), 414-428.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Bae, J. (2019). A continuity analysis of the reading passages in elementary and middle school English textbooks with 2015 Revised National Curriculum. *Modern Studies in English Language & Literature*, 63(2), 81-107.
- Berendes, K., Vajjala, S., Meurers, D., Bryant, D., Wagner, W., Chinkina, M., & Trautwein, U. (2018).

- Reading demands in secondary school: Does the linguistic complexity of textbooks increase with grade level and the academic orientation of the school track? *Journal of Educational Psychology*, 110(4), 518-543.
- Brill, E. (1995). Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational linguistics*, 21(4), 543-565.
- Chao, T. C. (2011). The hidden curriculum of cultural content in internationally published ELT textbooks: A closer look at new American inside out. *The Journal of Asia TEFL*, 8(2), 189-210.
- Chen, A. C. (2016). A critical evaluation of text difficulty development in ELT textbook series: A corpus-based approach using variability neighbor clustering. *System*, 58, 64-81.
- Choi, J. (2015). Cultural aspects on the Korean and Japanese middle school English textbooks: Based on the critical pedagogy. *Journal of Language Sciences*, 22(3), 219-253.
- Cirilo, R. (1981). Referential coherence and text structure in story comprehension. *Journal of Verbal Learning and Verbal Behavior*, 20, 358-367.
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33, 497-505.
- Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a foreign language*, 23(1), 84-101.
- Crossley, S. A., Skalicky, S., Dascalu, M., McNamara, D. S., & Kyle, K. (2017). Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas. *Discourse Processes*, 54(5-6), 340-359.
- Dubin, F. (1995). The craft of materials writing. In P. Byrd (Ed.), *Material writer's guide* (pp. 13-22). New York, NY: Heinle & Heinle Publishers.
- Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, 20(3), 586-592.
- Goh, G., & Park, J. (2007). A corpus-based study of continuity in Korean middle school English textbooks. *English Language and Linguistics*, 24, 27-45.
- Graesser, A. C., Jeon, M., Yan, Y., & Cai, Z. (2007). Discourse cohesion in text and tutorial dialogue. *Information Design Journal*, 15(3), 199-213.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202.
- Gupta, R. (2013). The readability of English language texts in the primary grades. *Language in India*, 13(5), 534-544.
- Haberlandt, K., & Bingham, G. (1978). Verbs contribute to the coherence of brief narratives: Reading related and unrelated sentence triples. *Journal of Verbal Learning and Verbal Behavior*, 17(4), 419-425.
- Heatley, A., & Nation, I. S. P. (2006). *Range 32H* (computer software). Wellington, New Zealand: Victoria University of Wellington.
- Im, Y., Cho, Y., & Jong, Y. (2015). A continuity analysis of the reading passages of elementary school 6th grade and middle school 1st grade English textbooks. *Foreign Languages Education*, 22(1), 125-148.
- Jang, N., & Kim, J. (2018). An analysis of cultural contents in elementary school English textbooks for grades 3 and 4 based on the 2015 Revised Korean National Curriculum. *The Journal of the Korea Contents Association*, 18(8), 1-8.
- Jeon, M. (2011). A corpus-based analysis of the continuity of the reading materials in middle school English 1 and 2 textbooks with Coh-Metrix. *The Journal of Linguistic Science*, 56, 201-218.
- Jeon, M. (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, M. (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, M., & Lim, I. (2009). A corpus-based analysis of middle school English 1 textbooks with Coh-

- Metrix. *English Language Teaching*, 21(4), 265-292.
- Jeon, M., & Lim, I. (2010). A corpus-based linguistic analysis of high school English textbooks. *Foreign Languages Education*, 17(1) 209-233.
- Jeong, H., & Kim, H. (2017). An analysis of grammar activities in middle school English textbooks and learners' perspectives. *Journal of the Korea English Education Society*, 16(1), 129-146.
- Kim, J. (2014). Continuity problems of elementary and secondary English education: Sequence analysis of English textbooks. *Language Research*, 50(1), 161-184.
- Kim, J., & Chun, Y. (2008). A study on the continuity of elementary and middle school English textbooks through a corpus-based analysis: The 6th grade elementary school English textbook and the 1st grade middle school English textbook. *English Teaching*, 63(2), 361-394.
- Kim, J., & Jeon, M. (2013). A corpus-based analysis of the continuity of the listening materials in middle school English textbooks for Coh-Metrix measures. *Journal of the Korean Data Analysis Society*, 15(4), 1987-2000.
- Kim, J., & Yang, J. (2012). An analysis of the continuity of elementary and middle school English textbooks using Coh-Metrix. *English Teaching*, 67(2), 319-341.
- Kim, M., & Song, H. (2007). A study of reading materials and activities in English textbooks for secondary education. *Korean Journal of English Language and Linguistics*, 7(2), 211-238.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Addison-Wesley Longman Ltd.
- Landauer, T. K. (2007). LSA as a theory of meaning. In T. K. Landauer, D. S. McNamara, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 3-34). Mahwah, NJ: Erlbaum.
- Lee, J., & Kim, H. (2007). Learners' and teachers' responses on high school English conversation textbooks and its relationship with CSAT. *English Language Teaching*, 19(2), 277-305.
- Lee, S. (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.
- Lim, I., & Jeon, M. (2013). A corpus-based analysis of the listening materials in middle school English 1 textbooks. *Korean Journal of English Language and Linguistics*, 13(1), 105-134.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- Miller, G.A., Beckwith, R., Feldbaum, C., Gross, D., & Miller, K. (1990). *Five papers on WordNet* (Tech. Rep.). Princeton University, Cognitive Science Laboratory.
- Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, 33(1), 128-147.
- Morales, B. C. (2019). Readability and types of questions in Chilean EFL high school textbooks. *TESOL Journal*. Online version of record before inclusion in an issue. doi: <https://doi.org/10.1002/tesj.4984>.
- Murray, J. (1997). Connective and narrative text: The role of continuity. *Memory & Cognition*, 25(2), 227-236.
- Nicol, C. C., & Crespo, S. M. (2006). Learning to teach with mathematics textbooks: How preservice teachers interpret and use curriculum materials. *Educational studies in mathematics*, 62(3), 331-355.
- Paul, T. (2003). *Guided independent reading*. Madison, WI: School Renaissance Institute.
- Plakans, L., & Bilki, Z. (2016). Cohesion features in ESL reading: Comparing beginning, intermediate and advanced textbooks. *Reading in a Foreign Language*, 28(1), 79-100.
- Ryu, J., & Jeon, M. (2020). A Coh-Metrix analysis of the first grade middle school English textbooks revised by the 2015 National Curriculum. *Korean Journal of English Language and Linguistics*, 20, 89-121.
- Son, M., & Lee, J. (2005). An analysis of writing activity materials in the middle school English textbooks. *English Language & Literature Teaching*, 11(3), 139-164.
- Song, H. (2013). Exploring the connectedness of listening tasks according to grades in high school English textbooks. *English Language & Literature Teaching*, 19(4), 279-305.

- Spencer, M., & Wagner, R. (2016). The comprehension problems for second-language learners with poor reading comprehension despite adequate decoding: A meta-analysis. *Journal of Research in Reading, 40*(2), 199-217.
- Tai, T. Y. (2016). A corpus-based analysis of discourse markers in curriculum-based English textbooks and the English entrance exam in Taiwan. *The Journal of Asia TEFL, 13*(4), 262-279.
- To, V. (2018). Linguistic complexity analysis: A case study of commonly-used textbooks in Vietnam. *Sage Open, 8*(3), 1-13.
- Tomlinson, B. (2012). Materials development for language learning and teaching. *Language Teaching, 45*(2), 143-179.
- Tsagari, D., & Sifakis, N. C. (2014). EFL course book evaluation in Greek primary schools: Views from teachers and authors. *System, 45*, 211-226.
- Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Young, M. J., & Riegeluth, C. M. (1988). *Improving the textbook selection process*. Fastback 275. Publications, Phi Delta Kappa, Eighth and Union, PO Box 789, Bloomington, IN 47402.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*(5), 292-297.

(Received May 27, 2020; Revised August 30, 2020; Accepted September 11, 2020)