Comparing the Difficulty of Classroom English in Primary English Teacher-Guidebooks by Levels Using Coh-Metrix

Philip-Yoongoo Jung and Jeong-Ah Shin[*]

Jung, Philip-Yoongoo, & Shin, Jeong-Ah. (2021). Comparing the difficulty of classroom English in primary English teacher-guidebooks by levels using Coh-Metrix. *English Teaching, 76*(4), 75-100.

This study investigated if the difficulty of instructional classroom English in primary teacher-guidebooks of English is adequately manipulated by learner-proficiency. Corpora of classroom English were compiled from 20 guidebooks from five publishers, approved following the 2015 Revised National Curriculum of Korea. Extracted materials from grades three and four were compared with those of grades five and six to observe variations in the difficulty. Coh-Metrix, a software application that computes an extensive range of measures on cohesion and language, was used for analyses. With evidence-based assessments on (psycho)linguistic features and patterns of classroom English, we report results both congruent and incongruent with the prospect that the difficulty should increase as learners become more proficient. Overall, although partial difficulty manipulations between the two levels were noted, inconsistent results and invariances were also observed, disclosing much room for improvement of classroom English in the guidebooks. Some implications toward teacher-guidebook development, particularly in its classroom English, are suggested.

Key words: primary English teacher-guidebooks, classroom English, text difficulty, Coh-Metrix, corpus

*First Author: Philip-Yoongoo Jung, Doctoral Candidate, Department of English, Dongguk University
Corresponding Author: Jeong-Ah Shin, Professor, Department of English, Dongguk University; 30, pildong-ro 1-gil, Joong-gu, Seoul 04620, Korea; Email: jashin@dongguk.edu

1. INTRODUCTION

Any day-to-day social interactions can certainly serve as an authentic source of language input, potentially leading to language acquisition (Krashen, 1976). However, in an *English as a Foreign Language* (EFL) context where English is not the medium of communication in society, learners are seldom exposed to such sorts of "natural" input. Rather, the core input is conveyed solely in a controlled learning site. In this mechanism, it is presumed that, for most students, English teacher-talk during an EFL lesson may be the only consistent source of discourse input that aids as a reliable model of authentic target language use (No & Jung, 2016; Pinter, 2006; Prabhu, 1987). Then, to complement EFL-learner-input of authentic English, the Ministry of Education in Korea mandated the *Teach English in English* (TEE) practice in 2008. Since the implementation, a growing number of English teachers have used English as the primary language of instruction and numerous studies thus far have reported positive perceptions of the TEE policy by both teachers and students (Kim & Kim, 2012; Lee & Lee, 2011; Oh & Joh, 2012).

To accompany and reinforce TEE, teacher-guidebooks of English provide samples of situational speech corpora of instructional classroom English (e.g., instructions/rules to a classroom activity) in a text format for non-native English teachers to refer to. Thus, particularly for novice teachers, classroom English outlined in the teacher-guidebooks would be a valuable resource in planning/practicing TEE lessons as well as in improving their classroom communication competence. Despite the imperative rationale toward the inclusion of classroom English in the guides, studies have yet revealed much room for its improvement in terms of its quality (Kim, 2015; Kim, 2018; No & Jung, 2016), and the lacking quality of classroom English is reported to be one of the major reasons for in-service teachers' general dissatisfaction toward the guidebooks (Jung & Shin, 2021). Though textbooks go through a variety of assessment measures before their approval, guidebooks are not subject to any screening or authorization process by the government. Consequently, while many studies have analyzed the linguistic features and patterns of the textbook materials (e.g., reading and listening passages in the textbooks), little attention has been paid to the content of the teacher-guidebooks.

The significance of teacher-instruction has been researched by previous literature (Cullen, 1998; Nunan, 1991; Richards & Lockhart, 1994), mostly with a proposition that teacher-talk should be carefully tailored with attention to its amount, type, and effects in relation to the pedagogical objectives. Accordingly, teacher-instruction must serve as a quality input for learners, yielding purposeful and incidental learning of the target language. Extending on this notion, classroom English is to be used considering the proficiency level of students. Hence, it is evident that the quality of classroom English of the teacher-guidebooks, which serves as a valuable referral source for non-native teachers of English be thoroughly

analyzed, not only qualitatively but with quantitative, evidence-based measures.

We thus present the following research question: Is the difficulty of instructional classroom English of primary English teacher-guidebooks substantially controlled based on learner proficiency? As English is taught from grades three to six in primary schools, 3rd and 4th grades are generally considered as "lower-level" and 5th and 6th as "higher-level" in the *National Curriculum* with many works of literature classifying the two grades (i.e., 3rd and 4th; and 5th and 6th) to be comparable (e.g., Jang & Kim, 2018; S. Lee, 2013; Yim & Huh, 2020; among many others). Therefore, in this study, classroom English of 3rd and 4th grades is compared to that of 5th and 6th to observe the variances in the difficulty between the two major levels of primary English education. Chunks of instructional classroom English from every approved and currently available guidebook of primary English are extracted and analyzed with Coh-Metrix, a software application capable of measuring an extensive range of linguistic and psycholinguistic features of English (Graesser, Jeon, Yan, & Cai, 2007). Twenty-seven indices of the Coh-Metrix system are selected under four major categories of analyses: general (i.e., basic counts such as number/length of words/sentences), lexical (i.e., type-token ratio, word frequency, word features), syntactic (e.g., structural complexity), and discourse (e.g., semantic cohesion) aspects.

2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1. Teachers' Language as a "Progressive" Source of Comprehensible Input

Among many necessary elements appointed to language acquisition, input that is comprehensible is claimed to be the foremost important factor (Krashen, 1981, 1985), and studies to define the sources and dynamics of input have been conducted widely (e.g., Carroll, 1999; Ellis, 1981; Krashen, 1985; Smith, 1993). Although many previous works of literature have suggested the significant role of the textbooks as a focal source of input for learning (Cunningsworth, 1995; Harmer, 2008; Jeon & Lim, 2010; Richards, 1993), the language materials of the textbooks, on their own, may not always be "comprehensible" for learners (but of course, we also acknowledge that some may be deliberately designed to elicit difficulties in comprehension to some extent (e.g., reading comprehension tasks/prompts)). Another fundamental type of input in an EFL site is the teacher-talk, that must ideally, in every case, be comprehensible for learners; that is, to fulfil the core functions of the teacher's English use: for class organization and communication with learners (Willis, 1981). By no means of disputing the argued significance of the textbooks, note that, with appropriate teacher-talk to complement the textbook materials, textbooks' role as a reliable source of

input can assuredly be better justified (as their comprehensibility improves). Following principles of the *Natural Approach*, any authentic and natural language use by the teacher functions as a major source of comprehensible input, leading to language development (Ellis, 2005; VanPatten, 2004); its applications in an EFL setting are further signified by Prabhu (1987) in that particularly for beginner learners, teacher-talk serves as the leading effective model of the target language use for any potential subsequent imitation. Teacher-talk then must be practiced appropriately to meet the educational and pedagogical objectives in the classroom.

Positing the significant impact that classroom English has on learners, its linguistic features and patterns must be controlled relative to its comprehender-status; specifically, they should be deployed based on the proficiency level of the learners. No educationalist would dispute against the scheme that an EFL teacher's instructional classroom English for a 12-year-old should be more advanced than the one for a nine-year-old, under the assumption that the child's language competence has gradually improved with the age. This rationale is further motivated by some widely accepted principles: *Zone of Proximal Development* (Vygotsky, 1978) and *Input Hypothesis* (Krashen, 1985). These theories imply that the difficulty of input must be precisely measured, and that it should always be marginally higher than the learners' present level. Therefore, as teacher-talk is argued to be an important source of comprehensible input for learners, it must prudently be used, closely aligning its difficulty level relevant to the learner's competence level. Ideally, the difficulty of a teacher's English instructions should linearly increase as the grade level rises.

2.2. Issues of Instructional Classroom English in the Teacher-Guidebooks

Classroom English integrated in the teacher-guidebooks is intended to better institute TEE environments by providing situational instructional English (e.g., instructions/rules to a classroom activity) in a text format; non-native teachers of English are encouraged to refer to it as a guideline for authentic instructional English. Such a prominent purpose of classroom English in the guidebooks seems quite relevant, as many Korean teachers of English have previously reported of their reluctance in using English as the medium of instruction due to having low confidence and limited proficiency (Min, 2008). Thus, guidebooks are meant to serve as a cherished resource for teachers in acquiring competence toward instructional English. Despite the upright implications, little attention has been paid toward the analysis of the teacher-guidebooks, relatively compared to the widely studied student-textbooks. The few previous studies relevant to the matter being discussed in the present study have suggested much room for the improvement of classroom English in the teacher-guidebooks of English.

No and Jung's (2016) qualitative study of classroom English in the primary teacher-

guidebooks of English from four publishers revealed some common problems defined based on four indices: comprehensibility, linguistic appropriateness, authenticity, and richness of input. Their overall verdict toward classroom English in the guidebooks was "fairly good," but with some systematic problems observed; many samples of classroom English were long-winded, linguistically inappropriate, inauthentic, and oversimplified. Concerning the present study of comparing the difficulty of classroom English by levels, their claim of classroom English being oversimplified is worth a detailed look. Having a routinized conversation (e.g., a display question such as an EFL teacher, in the morning, asking about the weather in English, every day) is known to have some valuable instructional functions (McCormick & Donato, 2000). While No and Jung (2016) agreed with this aspect, their analyses revealed an extensive "copy and paste" tendency of fixed and repeated expressions and that they were invariant over the course of four grade levels (i.e., grades three to six). Thus, they concluded that the richness of input is uncontrolled throughout the curriculum. A similar concern has been raised in Kim's (2018) study in that textbook/guidebook writers posited the necessity of including more diverse English expressions in the primary teachers' guides.

A corpus-based study by Kim (2015) quantitatively analyzed classroom English from three publishers of primary teacher-guidebooks for grades three to six and transcribed spoken instructional English by Korean teachers of English and native English teachers, based on lexical diversity/usefulness/frequency and syntactic complexity. The study found that, in the teachers' guides, the STTR (Standardized Type-Token Ratio) gradually increased, indicating the use of enhanced diversity of lexical elements as the grade level rose. However, lexical frequency/usefulness and syntactic complexity were revealed to be inconsistent across the grades. Spoken instructions practiced by native English teachers have demonstrated a systematic increase in the difficulty during their delivery of instructions as the grade level increased, but Korean teachers of English exhibited irregularity in all measured areas. As Korean teachers of English are known to refer to the guidebooks for samples of classroom English, Kim (2015) argued the very need for a strict screening toward the difficulty control of classroom English in the teacher-guidebooks.

To acquire an insight on how primary teachers assess classroom English of the guidebooks, Jung and Shin (2021) conducted a study of teacher-perception toward it. The study analyzed questionnaire responses from 32 pre-service and 41 in-service teachers with a variety of English teaching experiences. Overall, in-service teachers displayed their dissatisfaction toward classroom English in the guidebooks in terms of its quality, while pre-services teachers were highly reliant on the guidebooks for referrals of classroom English. Notably, in-service teachers reported some problems of classroom English, mainly toward its suitability and lack of difficulty manipulations by grades. Among 22 in-service teachers who raised a concern toward classroom English in the guidebooks, 13 saw it as being too simple,

11 pointed out that it does not consider students' proficiency of English (i.e., too difficult or too easy), and 9 agreed that it does not properly reflect the grade levels (multiple selections allowed in the questionnaire). As teachers, who the guidebooks are developed for, explicitly defined some specific lacking features of classroom English, there justifies a robust need for detailed analyses of classroom English in terms of how difficulty is controlled by learner-proficiency.

Previous studies directly relevant to the present study however revealed some limitations. While No and Jung (2016) offered a handle in suggesting some constructive alternative English expressions throughout the guidebooks based on the problems reported, their analyses were purely on qualitative measures, thus unable to provide sufficient objective, evidence-based conclusions in terms of assessing the difficulty-variance across the grade levels. Kim's (2015) corpus-based study of classroom English provided some valuable data in its evaluation, extending the analyses to its actual practice. However, as noted by the author himself, measures were limited to lexical frequency/usefulness/diversity and syntactic complexity of classroom English; no measure toward its semantic, cohesive, or discourse aspect has been carried out.

2.3. The Coh-Metrix Analyses of Student-Textbooks

Teacher-guidebooks, in nature, are intended to complement the student-textbooks. They are always published as a set; yet peculiarly, while the textbooks undergo a series of demanding assessments before they are approved for use, guidebooks are typically not subject to any screening process by the government. Thus, with much more emphases put toward the textbooks, many studies have been conducted using the Coh-Metrix system to systematically examine and compare the text difficulty of the textbook materials, often across grades and publishers (see Ahn & Ma, 2015; Bae, 2019; Berendes et al., 2019; Chen, 2016; Gupta, 2013; Jeon, 2011, 2014, 2015; Kim, 2014; Kim & Jeon, 2013; Kim & Yang, 2012; S.-H. Lee, 2013; Plakans & Zeynep, 2016; Ryu & Jeon, 2020; Song, 2013; To, 2018). Note that all these studies share a commonly agreed prospect that the difficulty level would increase as the grade level rises. For instance, Ahn and Ma (2015), with Coh-Metrix, analyzed the lexical, syntactic, and cohesion features of the text segments in primary English textbooks for grades four to six. Their results revealed that the difficulty control was well established for the lexical elements (i.e., number of words, concreteness, imageability of content words). The development-guidelines for the textbooks indeed mandate the number of new words to be introduced by grades thus such an observation was expected to an extent. Nevertheless, inadequate difficulty manipulations were found for the other measured areas, demanding the need of systematic attention in the process of developing/evaluating the textbooks. Kim and Yang (2012) compared the sixth grade primary school English textbooks

(six publishers) to the first grade middle school English textbooks (six publishers) to inspect the continuity of the two back-to-back school levels. Difficulty of the written texts have amplified significantly as the grade level increased but through many other measured indices that turned out to disclose insignificant variances, the study claimed a lack of continuity between the two levels. Also, Ryu and Jeon (2020) compared the textbooks for the first grade of middle school English, written by different publishers. Reading materials extracted from seven English textbooks were compared with one another (by publishers) with 27 indices of the Coh-Metrix system. Overall, they reported that significant differences were found among the seven textbooks in majority of measures including basic counts, frequency/features/diversity of words, syntactic complexity, connectives, readability, co-reference cohesion, and semantic similarity. Ryu and Jeon (2020) argued that the significant variances in the difficulty between different textbooks written for the same grade potentially indicate that the national development standards for textbooks were not sufficiently adopted.

Through several previous studies, the Coh-Metrix system has proven its legitimacy to cater reliable evidence-based interpretations in text segment comparisons. As aforementioned, previous studies all establish a robust underlying rationale for the assertion of difficulty manipulation of the textbook materials by learner-proficiency. We agree, along with many previous works of literature, that the difficulty of the language materials in the student-textbooks should increase as the grade level rises. Accordingly, we argue that the teacher-guidebooks should also model this pattern. Specifically, the notion further supports the need of any relevant input (e.g., instructional language) materials of the teacher-guidebooks to be aptly manipulated by grades.

2.4. Coh-Metrix Indices for the Study

Coh-Metrix is an automated tool developed by the Institute for Intelligent Systems at the University of Memphis to analyze the linguistic and psycholinguistic features of English text with an extensive range of measures (with over 200 indices) on cohesion and language (Graesser et al., 2007). Coh-Metrix measures selected for this study include basic counts, lexical diversity, word features, word frequency, syntactic complexity, reference cohesion, semantic cohesion, situation model, and connectives (see Table 1 for details). This section introduces the Coh-Metrix indices selected for the present study and their implications.

TABLE 1

Coh-Metrix Indices Selected for the Present Study

Aspect	Analyzed Variable	Coh-Metrix Index	Label
General	Basic counts	Number of words	DESWC
		Number of sentences	DESSC
		Average sentence length	DESSL
		Average word length	DESWLlt
Lexical	Diversity	Type-Token Ratio (TTR) of content words	LDTTRc
		Type-Token Ratio (TTR) of all words	LDTTRa
	Feature	Age of acquisition	WRDAOAc
		Concreteness	WRDCNCc
		Familiarity	WRDFAMc
		Imageability	WRDIMGc
		Meaningfulness	WRDMEAc
	Frequency	Centre for Lexical Information (CELEX) word frequency of content words	WRDFRQmc
		Centre for Lexical Information (CELEX) word frequency of all words	WRDFRQa
Syntactic	Complexity	Noun phrase (NP) density	DRNP
		Verb phrase (VP) density	DRVP
		Number of words before main verbs	SYNLE
Discourse	Referential cohesion	Argument overlap between adjacent sentences	CRFAO1
		Argument overlap among all sentences	CRFAOa
	Semantic cohesion	Latent Semantic Analysis (LSA) for adjacent sentences	LSASS1
		Latent Semantic Analysis (LSA) for all sentences in a paragraph	LSASS1p
	Situational cohesion	Causal cohesion	SMCAUSr
		Intentional cohesion	SMINTEr
	Connectives	Causal connectives incidence	CNCCaus
		Logical connectives incidence	CNCLogic
		Adversative/contrastive connectives incidence	CNCADC
		Temporal connectives incidence	CNCTemp
		Additive connectives incidence	CNCAdd

2.4.1. Basic counts

Four indices of basic counts are selected: number of words, number of sentences, average sentence length (calculated by the number of words in a sentence), and word length (calculated by the number of letters in a word). These indices are suggested to be important linguistic factors influencing text difficulty (Graesser et al., 2007). Generally, more words/sentences and/or longer words/sentences are known to elicit more processing burden.

2.4.2. Lexical aspect

The Coh-Metrix system computes the type-token ratio (TTR) which is a widely known index to examine the lexical diversity in a text segment. TTR is calculated by dividing the number of unique words (*types*) by the total number of words (*tokens*) in a segment of language. Therefore, the maximum TTR output is 1; the closer the TTR output is to 1, the greater the lexical richness of the analyzed segment. Two indices of TTR are selected: for content words and all words. Content words (e.g., nouns, verbs, adjectives, adverbs) are generally rarer than other words (e.g., prepositions, determiners, pronouns). They are also known to be semantically richer thus the frequent occurrences of them are associated with richer bodies of world knowledge (Beck, McKeown, & Kucan, 2002; Haberlandt & Graesser, 1985; Perfetti, 2007) and in turn suggesting an increased in the difficulty.

In addition, the Coh-Metrix system computes word features using the Medical Research Council (MRC) Psycholinguistics Database (Coltheart, 1981). For this present study, indices *age of acquisition* (AOA)*, concreteness, imageability, familiarity,* and *meaningfulness* are selected, as they are suggested to be essential factors affecting the difficulty of texts (Graesser, McNamara, Louwerse, & Cai, 2004; Jeon, 2015; Ryu & Jeon, 2020). *Age of acquisition* refers to the age-of-acquisition norms (Gilhooly & Logie, 1980), accounting the assumption that some words appear in children's language earlier than others. Words such as *cortex*, *dogma*, and *matrix* (AOA = 700) compute higher scores than relatively easier words such as *milk*, *smile*, and *pony* (AOA = 202) (from McNamara, Graesser, McCarthy, & Cai, 2014). The concept of AOA is also applicable in second/foreign language acquisition; previous works of literature distinguish and cover early and late vocabulary learning (see Service & Craik, 1993; Yoshida, 1978). The other four indices are ratings on the 1-7 scale, but the computed output is multiplied by 100 and rounded to the nearest integer to be able to present all the rating as integers on a scale from 100 to 700 (McNamara et al., 2014). *Concreteness*, based on human ratings, denotes how concrete or nonabstract a word is; a higher score of this index means more concrete the words are (e.g., *pen* would score higher than *policy*) in the text segment. *Imageability* refers to a measure of how easy it is to evoke a clear mental image from a content word in a text; a higher score of this index signifies a better imageability (e.g., *book* compared to *reason*). *Familiarity* is a rating of how familiar a word seems; the higher the computed score, the more familiar the lexical items are in the given text segment. *Meaningfulness* points to how meaningful a word is based on a corpus developed by Toglia and Battig (1978). It is related to the degree to which the word is associated with other words; therefore, a higher score of this index indicates the presence of relatively easier lexical items (e.g., *people* (612) compared to *abbess* (218), from McNamara et al., 2014).

Word frequency in Coh-Metrix is computed using CELEX, database from the Dutch

Centre for Lexical Information (Baayen, Piepenbrock, & Gulikers, 1995). Word frequency indices indicate how often specific words arise in the given text segments and are suggested to be important measures to observe the text difficulty (Graesser et al., 2007; Graesser, McNamara, Louwerse, & Cai, 2004).

2.4.3. Syntactic aspect

Density measurements of phrase types are significant contributors to observe the syntactic complexity of a text segment. Coh-Metrix measures information on the incidence of noun phrases (NP) and verb phrases (VP) which are expected to affect processing difficulty of text (McNamara et al., 2014). For instance, higher incidences of such phrase types imply that the text is more informationally dense with complex syntax (McNamara et al., 2014). In addition, the syntax of a text segment tends to be easier to comprehend with shorter sentences, particularly fewer words before the main verb of the main clause (McNamara et al., 2014). Three specific indices (i.e., incidences of NP/VP density and mean number of words that appear before the main verb) are selected to compare the syntactic complexity of classroom English in the grades 3-4 level and the grades 5-6 level.

2.4.4. Discourse aspect

A text segment with high referential cohesion comprises words and ideas that are overt across sentences and the entire text, providing explicit grounds that connect the text (McNamara et al., 2014). Two indices are selected: argument overlap (sharing of the same nouns or pronouns) in adjacent sentences and among all sentences. A higher reference cohesion score computed by the Coh-Metrix system represents an increase in the comprehension of the text, suggesting the decrease of its difficulty (Graesser et al., 2007).

In Coh-Metrix, the computation of semantic cohesion is measured through the application of the Latent Semantic Analysis (LSA) that is a mathematical and statistical technique for representing world knowledge, based on a large corpus of texts. LSA computes the semantic/conceptual similarities between words, sentences, and paragraphs (Landauer & Dumais, 1997; Landauer, McNamara, Dennis, & Kintsch, 2013). The computed values range from zero to one, and the text difficulty is predicted to increase with decreases in LSA scores. Two LSA indices were selected: for adjacent sentences and for all sentences in a paragraph.

The Coh-Metrix system provides two types of situation model ratio indices: the ratio of causal particles (e.g., *because*) to causal verbs (e.g., *break*) and the ratio of intentional particles (e.g., *so that*) to intentional verbs (e.g., *contact*) (McNamara et al., 2014). Precisely, the ratios are computed to reveal the necessity of connectives which ultimately depends on

the number of events described in the text. Thus, highly causal and intentional text segments require more cognitive resources by language learners (McNamara et al., 2014).

Connectives are known to contribute to discourse cohesion by explicitly connecting ideas at the clausal and sentential level (Britton & Gülgöz, 1991; Halliday & Hasan, 1976; Louwerse, 2001; McNamara & Kintsch, 1996; Sanders & Noordman, 2000). Therefore, the general expectation is to observe more incidences of connectives in classroom English targeted for the lower graders (i.e., 3rd and 4th graders) for easier processing (except for the case of causal connectives as they are mostly argumentative or expository that require more cognitive processing (Murray, 1997)). Coh-Metrix provides indices to measure the incidences of the five general classes of connectives: causal (e.g., *because, so*), logical (e.g., *and, or*), adversative/contrastive (e.g., *although, whereas*), temporal (e.g., *first, until*), and additive (e.g., *moreover*).

3. METHOD

3.1. Materials

In this study, a complete list of instructional classroom English provided in primary English teacher-guidebooks was selected to examine the text difficulty between the grades 3-4 level and the grades 5-6 level. The guidebooks were written following the *2015 Revised National Curriculum* of Korea and are approved for use currently. A total of 20 guidebooks for grades three to six by five publishers were sampled (Table 2). Classroom English materials from these 20 guidebooks were constructed as a corpus to examine the linkage between the two major levels.

TABLE 2
Teacher-Guidebook Information

English Teacher-Guidebook List	Publisher (Authors)
Primary English teacher-guidebook 3, 4, 5, 6	Chunjae Education Inc. (Ham et al.)
Primary English teacher-guidebook 3, 4, 5, 6	Daekyo Co. (Lee et al.)
Primary English teacher-guidebook 3, 4, 5, 6	Dong-A Publishing Co. (Park et al.)
Primary English teacher-guidebook 3, 4, 5, 6	YBM Co. (Choi et al.)
Primary English teacher-guidebook 3, 4, 5, 6	YBM Co. (Kim et al.)

3.2. Corpus Construction

For corpus formation, every chunk of classroom English was extracted from every guidebook. Each chunk is a set of instructions to a particular classroom activity/task (see Table 3 for samples of extracted materials from each grade; all samples were taken from one of the five sets of guidebooks). Specifically, across the five publishers, there were 494 chunks of classroom English from grade three, 513 from grade four, 563 from grade five, and 566 from grade six. One text file (in Unicode format) was constructed with all the chunks from one unit of each guidebook. The entire corpus package consisted of 59 text files (from 59 units) for the 3rd grade, 59 for the 4th grade, 62 for the 5th grade, and 62 for the 6th grade. Then, the files obtained from the 3rd and 4th grades were combined and those from the 5th and 6th grades were so as well to construct a total of 118 text files for the grades 3-4 level group and 124 for the grades 5-6 level group.

TABLE 3
Samples of Instructional Classroom English from Teacher-guidebooks by Grades

Grade	Instructional Classroom English
3	Write your name on the name card in Korean and put it on your clothes. Walk around the classroom, meet and greet your friends by saying "Hello, I'm ...," "Nice to meet you," and "Nice to meet you, too." When you're finished, go back to your seat and write the names of friends who you greeted in your book.
4	Take out the name cards from page 127. Write your name, class, and student number on each name card. Walk around the classroom, meet friends, and ask their names. Then, exchange name cards with them. Exchange all four name cards. Then, go back to your seat.
5	Let's make groups of six and stand in a row with your group. The first student (S1) in each group comes to me and asks, "How's it going?" and I will whisper the answer. S1 goes back to his/her group and when I give the signal, he/she whispers the answer to the next student. The last member of each group comes to me and says the answer. The first group with the correct answer wins.
6	I will choose a grade from one to six and write it on a piece of paper without showing it. The grade I choose becomes the grade of the class. I will ask, "What grade are you in?" and students will raise their hand and answer, "I'm in the ... grade." If the answer is lower than the grade written on the paper, I will say, "Up," and if it is higher, I will say, "Down." If the answer is correct, I will show the paper to the class and the student who said the correct answer gets one point. Continue the game, and the student with the most points wins.

3.3. Data Analysis

In this study, Coh-Metrix 3 was used to obtain the computed data for the selected indices. An independent samples *t*-test was performed to analyze the text difficulty of classroom English in primary English teacher-guidebooks using SPSS 27.0 (Statistical Package for the

Social Sciences). The independent variable in the analysis was the level (i.e., Grades 3-4 level or Grades 5-6 level) and the dependent variable was each index of Coh-Metrix. The assumption of homogeneity of variance was tested with Levene's Test of Equality of Variances. If the results of Levene's Test were insignificant, Welch's *t*-test results are reported. Each analysis was performed at a significance level of 5% ($p = .05$).

4. RESULTS

4.1. General Aspect: Basic Counts

Four indices of basic counts were analyzed: number of words/sentences and average sentence/word length. The results of basic count analyses are summarized in Table 4. For the number of words, a statistically significant difference was detected ($t(192) = -1.99$, $p < .05$; $d = -.26$), indicating an inclusion of significantly greater number of words in the grades 5-6 level ($M = 811.13$, $SD = 305.27$) compared to the grades 3-4 level ($M = 748.36$, $SD = 166.56$). A significant difference ($t(240) = -5.91$, $p < .001$; $d = -.76$) was also found for the average word length (calculated by the number of letters in a word), showing that the average word length in the grades 5-6 level ($M = 4.05$, $SD = .14$) is longer than the average word length in the grade 3-4 level ($M = 3.94$, $SD = .13$). No statistically significant differences were detected for the number of sentences and the average length of sentences between the two levels. In sum, the number/length of words seem to have increased as the grade level increased, however, the number/length of sentences showed no difference between the two levels.

TABLE 4
Results of Basic Counts

Index	Grades 3-4 Level ($N = 118$)	Grades 5-6 Level ($N = 124$)	t	Cohen's d
Number of words	748.36 (166.56)	811.13 (305.27)	-1.99*	-.26
Average word length	3.94 (.13)	4.05 (.14)	-5.91***	-.76
Number of sentences	69.65 (13.74)	72.02 (23.63)	-.95	-.12
Average sentence length	10.85 (1.34)	11.21 (1.63)	-1.86	-.24

*$p < .05$, **$p < .01$, ***$p < .001$; *M* followed by *SD* in parentheses

4.2. Lexical Aspect

4.2.1. Lexical diversity

Table 5 shows the results of lexical diversity. Our analyses revealed no significant

difference between the two levels for neither TTR of content words nor TTR of all words. The results signify that the level of diversity toward the use of vocabulary did not change between the two levels. In other words, no sign of difficulty-control was observed comparing the two levels in terms of lexical diversity.

TABLE 5

Results of Lexical Diversity

Index	Grades 3-4 Level (N = 118)	Grades 5-6 Level (N = 124)	t	Cohen's d
Type-token ratio (content words)	.444 (.054)	.438 (.071)	.67	.086
Type-token ratio (all words)	.284 (.035)	.282 (.059)	.45	.057

$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$; M followed by SD in parentheses

4.2.2. Word features

As reported in Table 6, statistically significant differences between the two levels were found for two indices of word features: *familiarity* ($t(240) = 4.08$, $p < .001$; $d = .53$) and *meaningfulness* ($t(222) = -2.75$, $p < .01$; $d = -.36$). For *familiarity*, an increase in the difficulty of lexical elements was observed from the grades 3-4 level ($M = 579.47$, $SD = 2.98$) to the grades 5-6 level ($M = 577.69$, $SD = 3.72$). However, the measure of *meaningfulness* revealed a decrease in the difficulty of lexical elements from the grades 3-4 levels ($M = 430.33$, $SD = 8.18$) to the grades 5-6 level ($M = 432.94$, $SD = 6.41$). For the other three indices analyzed (i.e., *imageability*, *concreteness*, *age of acquisition*), no statistically significant differences were detected. The inconsistent results and invariant differences suggest that the level of difficulty between the two levels is uncontrolled, particularly toward the features of words encompassed.

TABLE 6

Results of Word Features

Index	Grades 3-4 Level (N = 118)	Grades 5-6 Level (N = 124)	t	Cohen's d
Age of acquisition	299.70 (15.50)	302.99 (17.95)	-1.53	-.2
Concreteness	412.89 (10.76)	412.27 (9.85)	.47	.06
Familiarity	579.47 (2.98)	577.69 (3.72)	4.08***	.53
Imageability	438.95 (10.65)	439.58 (8.91)	-.502	-.06
Meaningfulness	430.33 (8.18)	432.94 (6.41)	-2.75**	-.36

$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$; M followed by SD in parentheses

4.2.3. Word frequency

As presented in Table 7, no statistically significant differences were found in measuring

the log frequencies of all words and content words between the grades 3-4 level and the grades 5-6 level. The finding suggests that the two levels, in terms of word frequency, did not demonstrate significant changes; no difficulty-manipulation was observed.

TABLE 7

Results of Word Frequency

Index	Grades 3-4 Level ($N = 118$)	Grades 5-6 Level ($N = 124$)	t	Cohen's d
Word frequency (content words)	1.58 (.12)	1.6 (.12)	-1.53	-.2
Word frequency (all words)	3.08 (.04)	3.08 (.05)	-.11	-.02

$^*p < .05, ^{**}p < .01, ^{***}p < .001$; M followed by SD in parentheses

4.3. Syntactic Aspect Syntactic Complexity

Table 8 reports the results of syntactic complexity. The analyses revealed no statistically significant difference in any of the three indices analyzed, indicating no presence of a sufficient difficulty-control between the two levels in terms of syntactic complexity.

TABLE 8

Results of Syntactic Complexity

Index	Grades 3-4 Level ($N = 118$)	Grades 5-6 Level ($N = 124$)	t	Cohen's d
NP density	368.06 (17.17)	371.47 (13.97)	-1.69	-.22
VP density	229.02 (17.30)	224.97 (17.78)	1.79	.23
Words before main verbs	3.21 (.53)	3.15 (.58)	.91	.12

$^*p < .05, ^{**}p < .01, ^{***}p < .001$; M followed by SD in parentheses

4.4. Discourse Aspect

4.4.1. Reference cohesion

Table 9 shows the results of referential cohesion. A statistically significant difference was observed between the two levels for the argument overlap score for adjacent sentences (t(240) = 2.12, $p < .05$; $d = .27$), suggesting an increase in the difficulty of classroom English for the grades 5-6 level ($M = .501$, $SD = .106$) compared to its counterpart grades 3-4 level ($M = .539$, $SD = .096$). The result suggests that classroom English consists of fewer overlaps between adjacent sentences in the grades 5-6 level compared to the grades 3-4 level. No difference was found for the argument overlap for all sentences.

TABLE 9

Results of Referential Cohesion

Index	Grades 3-4 Level ($N = 118$)	Grades 5-6 Level ($N = 124$)	t	Cohen's d
Argument overlap (adj. sentences)	.539 (.096)	.501 (.106)	2.12*	.27
Argument overlap (all sentences)	.367 (.067)	.358 (.071)	.99	.13

*$p < .05$, **$p < .01$, ***$p < .001$; M followed by SD in parentheses

4.4.2. Semantic/situational cohesion

Results of semantic cohesion and situation model are reported in Table 10. Independent samples t-test analyses indicated no statistically significant difference between the two levels for LSA for adjacent sentences and LSA for all sentences in each paragraph. In addition, the two types of situational ratio indices (i.e., causal and intentional cohesion) also resulted no statistically significant differences. The findings suggest the absence of difficulty manipulation in classroom English between the two levels in terms of semantic/situational cohesion.

TABLE 10

Results of Semantic Cohesion/ Situation Model

Index	Grades 3-4 Level ($N = 118$)	Grades 5-6 Level ($N = 124$)	t	Cohen's d
LSA (adj. sentences)	.27 (.04)	.271 (.041)	-.24	-.03
LSA (all sentences in a paragraph)	.224 (.057)	.216 (.052)	1.05	.14
Causal cohesion	.164 (.054)	.166 (.08)	-.27	-.03
Intentional cohesion	.385 (.167)	.417 (.179)	-1.43	-.18

*$p < .05$, **$p < .01$, ***$p < .001$; M followed by SD in parentheses

4.4.3. Connectives

As presented in Table 11, all indices except for the incidences of additive connectives demonstrated statistically significant differences between the two levels: causal ($t(240) = -2.15$, $p < .05$; $d = -.28$), logical ($t(240) = 3.59$, $p < .001$; $d = .46$), adversative/contrastive ($t(209) = 5.21$, $p < .001$; $d = .68$), and temporal connectives ($t(240) = 4.22$, $p < .001$; $d = .54$). Incidences of causal connectives were greater in the grades 5-6 level ($M = 19.13$, $SD = 7.75$) than the grades 3-4 level ($M = 17.1$, $SD = 6.86$). Incidences of logical connectives were fewer in the grades 5-6 level ($M = 35.33$, $SD = 8.8$) than the grades 3-4 level ($M = 39.44$, $SD = 9.06$). Incidences of adversative/contrastive connectives were fewer in the grades 5-6 level ($M = 3.86$, $SD = 2.86$) than the grades 3-4 level ($M = 6.23$, $SD = 4.06$). Incidences of temporal connectives were fewer in the grades 5-6 level ($M = 22.49$, $SD = 6.06$) than the grades 3-4 level ($M = 25.84$, $SD = 6.29$). Incidences of additive connectives did not show any significant

difference between the two levels. Frequent incidences of connectives are indications of easier processing (except for the causal connectives which indeed shows an increase in the difficulty, see more in Discussion). Thus, based on our analyses, the use of the four out of five types of connectives seems to have been controlled to demonstrate an increase the text difficulty in the grades 5-6 level compared to the grades 3-4 level.

TABLE 11
Results of Incidences of Connectives

Index	Grades 3-4 Level ($N = 118$)	Grades 5-6 Level ($N = 124$)	t	Cohen's d
Causal connectives	17.1 (6.86)	19.13 (7.75)	-2.15[*]	-.28
Logical connectives	39.44 (9.06)	35.33 (8.8)	3.59[***]	.46
Adversative/contrastive connectives	6.23 (4.06)	3.86 (2.86)	5.21[***]	.68
Temporal connectives	25.84 (6.29)	22.49 (6.06)	4.22[***]	.54
Additive connectives	53.36 (8.30)	52.91 (8.18)	0.43	.05

[*]$p < .05$, [**]$p < .01$, [***]$p < .001$; *M* followed by *SD* in parentheses

5. DISCUSSION AND CONCLUSION

The present study investigated whether the difficulty of instructional classroom English provided in primary teacher-guidebooks of English linearly increase as the students' grade level rises. The major findings and their implications are as follows.

First, analyses of basic counts revealed a partial control of difficulty between the two levels. Results showed that the number of words and the average word length in classroom English increased as the grade level escalated, potentially indicating that the level of difficulty is suitably controlled between the two levels. However, we also acknowledge that such findings may be due to the higher grades (five and six) likely comprising relatively more demanding prompts and complex classroom activities/tasks, which require more language (e.g., more and/or longer words) to instruct. In addition, by the curriculum, newer words that are more difficult (and potentially longer in length) than the words taught in the prior grades are introduced as the grade level rises. As classroom English is meant to directly complement the learning content outlined in the textbooks, we suspect that a similar phenomenon has been reflected in the analyses of classroom English. Further note that the length and number of sentences, suggested to be important linguistic factors influencing text difficulty (Graesser et al., 2007), did not differ between the two levels, revealing no difficulty manipulation taken place at the surface. Nevertheless, regarding the number/length of sentences, the conclusive interpretations of the output data seem to be arguable. Considering that the target comprehenders of classroom English are non-native children (mostly aged

nine to 12), the basic count analyses of sentences of classroom English may not exactly be in keeping with other analyses of complexity/readability on adults' spoken/written corpora. For instance, an EFL teacher may irregularly, depending on her instructional purposes (e.g., different types of classroom activities or tasks), choose or be required to use longer/shorter and more/fewer sentences regardless of the grade level (e.g., *okay?* vs. *did you all understand the rules?*). Such flexibilities are practical as teachers are known to continuously make modifications (e.g., simplification, rephrasing, prompting, speech rate, etc.) to their instructional English to facilitate learners' comprehension (Chaudron, 1988; Richards & Lockhart, 1994; Wong-Fillmore, 1985). Thus, spoken L2 instructional sentences are expected to be moderately flexible in terms of their length and amount; they are often repeated/rephrased and/or prompted. To pinpoint and precisely explain the correlations between the difficulty analyses of classroom English and the number/length of sentences, clearly, further research seems needed.

Second, analyses of the lexical aspect revealed a lacking control of difficulty between the two levels compared. Word frequency measures of all words and content words in the text segments revealed no differences between the two levels. Word frequency refers to how often specific words arise in the given text segment. The invariances indicate that classroom English demonstrated no difficulty-control between the two levels. Analyses of word features such as *familiarity* and *meaningfulness* produced statistically significant differences but with dissimilar interpretations; while an appropriate difficulty control was observed with *familiarity*, the computed results of *meaningfulness* indicated that the difficulty level indeed decreased as the grade level increased. Further, no significant differences in *age of acquisition*, *concreteness*, and *imageability* were observed. To observe the diversity of the lexical items, type-token ratio (TTR) of content words and all words were analyzed. TTR results computed by Coh-Metrix also implied that the difference of lexical diversity of content words and all words between the two levels is insignificant. The inconsistent and invariant results obtained from the analyses suggest that the difficulty of classroom English of the teacher-guidebooks in terms of the lexical aspect, is not suitably controlled. Surely, systematic and tangible guidelines in the process of teacher-guidebook development are needed, based on the principle that child-language-learners, as they become older, must receive more meaningful lexical input that are vividly more diverse and less frequent (e.g., *Age of Acquisition* from Gilhooly & Logie, 1980). We thus suggest that regular quantitative assessment measures (e.g., with Coh-Metrix) need to be programmed in the process of teacher-guidebook development, that is to effectively monitor and reflect the variations of lexical input across grades. Moreover, constructive lists of suggestions on lexical usage can be developed, which the authors of the guides can refer to. For instance, for a commonly used (and repeated) instruction, *make groups of five*, several alternative main verbs can be outlined (e.g., *form*, *arrange*, *organize*, etc.) to elicit the use of diverse words. With these

guidelines, authors can better control the difficulty of lexical elements in classroom English, preferably by learner-proficiency.

Third, for comparing the syntactic complexity between the two levels, NP/VP density incidences and the mean number of words before the main verbs were analyzed. Frequent incidences of NP/VP suggest that the text segment is more informationally dense with complex syntax (McNamara et al., 2014), potentially causing more processing burden. Similar logic is applied to the number of words before the main verbs as more words likely expose more complex syntax. For the three indices measured, no sign of difficulty differences between the two levels was observed. Hence, classroom English does not appear to comprise more sufficiently diverse sentence structures as the grade level increases. The findings are in accordance with Kim's (2015) corpus-based analyses of classroom English of primary teacher-guidebooks which noted that syntactic complexity was revealed to be irregular across grades. In addition, Kim (2018) reported a suggestion by the textbook/guidebook writers that having more variant English expressions in primary teacher-guidebooks is needed. Assuming that the teacher's instructional English is likely the most authentic discourse input for learners, it seems reasonable to suggest that the guidebook development programs adopt strategies designated to expose learners with variety of sentence structures. For instance, pilot samples of assorted sentence structures (that are quantitatively described/supported) can be developed for the textbook/guidebook writers to refer to.

Fourth, the statistical results showed that cohesion measures were irregular. A text segment with high referential cohesion consists of words and ideas across sentences and the entire text, as they provide explicit grounds that connect the text segments (McNamara et al., 2014). An increase in the difficulty from the grades 3-4 level to the grades 5-6 level was observed in analyzing the argument overlap between adjacent sentences but not for all sentences. While classroom English had fewer argument overlaps between adjacent sentences within a set of instructions in the grades 5-6 level compared to the grades 3-4 level, the overlaps within the entire text was found to be indifferent between the two levels. The findings indicate that classroom English in the lower grades comprised more repetitions of words and ideas, understandably for easier processing of back-to-back instructions (e.g., *Make groups of four. That means four students will be in a group*.). Semantic cohesion analyses (i.e., LSA), which compute semantic similarities between words, sentences, and paragraphs (Landauer & Dumais, 1997; Landauer et al., 2013;), revealed no difference between the two levels either for adjacent or all sentences in each paragraph. Researchers of discourse processing and cognitive science have been using the term *situation model* to denote the level of mental representation for a text that implicates how the words explicitly inferred their meaning. (Graesser & McNamara, 2011; Graesser, Singer, & Trabasso, 1994; Kintsch, 1998; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). Particularly, content

words and connectives are associated with the inferred meaning representations, but the explicit words alone cannot specify the deep meaning (McNamara et al., 2014). Situation model analyses (i.e., causal and intentional cohesion) resulted no difficulty variation between the two levels. It is generally assumed that discourse, highly associated with pragmatical knowledge is something we acquire along the acquisition of the first language. In contrast, the acquisition of L2 pragmatic fluency is suggested to require concentrated instructions (see Bouton, 1994; Cohen, 1996; House, 1996). Therefore, troubles by non-native guidebook authors in constructing difficulty-manipulated discourse corpora are expected to a notch. More intensive involvement of native English speakers in the guidebook development process is proposed to screen and apply precisely tailored difficulty manipulations in terms of the discourse aspect across grades. For quantitative assessments, guidebook developers are also suggested to utilize tools such as Coh-Metrix to control and assess the implementation of cohesion/discourse features within classroom English.

Lastly, measures on connectives showed a noticeable control of comprehension difficulty between the two levels. Connectives contribute to discourse cohesion by overtly linking ideas of text segments (Britton & Gülgöz, 1991; Halliday & Hasan, 1976; Louwerse, 2001; McNamara & Kintsch, 1996; Sanders & Noordman, 2000). Thus, it is generally expected to observe higher incidences of connectives in texts/discourse targeted for less proficient learners for easier processing. However, for causal connectives (e.g., *because*), frequent occurrences in fact reflect otherwise, that is, increased difficulty (Jeon, 2020) due to the use of causal connectives being mostly argumentative or expository that require more cognitive processing (Murray, 1997). Among the five types of connectives analyzed (causal, logical, adversative/contrastive, temporal, and additive), except for the additive connectives, all others signified an increase in the difficulty in the grades 5-6 level compared to the grades 3-4 level, congruent to the general expectation. For the case of additive connectives (e.g., *moreover*), no apparent variation between the two levels was observed. We presume that this may be due the nature of classroom English being often broken down (e.g., with excessive use of temporal additives such as *first*, *second*, *lastly*, *until* when delivering instructions of a classroom activity/task), countering the need for additive connectives in general. Overall, the use of connectives seems to be well established, exhibiting a noticeable rise in the difficulty of classroom English in the grades 5-6 level compared to the grades 3-4 level.

In sum, analyses of instructional classroom English provided in primary teacher-guidebooks of English confirmed appropriate difficulty control between the two levels in some areas. However, noticeable inconsistent results or no significant differences that are incongruent to the general expectation were found, disclosing much room for improvement of classroom English in many key areas of discourse science. Teacher-talk is highly purposeful under the premise that it functions as a cherished learning resource for language learners (Nunan, 1991), and classroom English in the guidebooks exists to assist this

rationale. In a similar vein with the findings of the present study, previous studies also pointed that one significant lacking feature of classroom English of teacher-guidebooks is its lack of difficulty manipulations by grades (Jung & Shin, 2021; Kim, 2015; No & Jung, 2016). Having authentically developed spoken corpus packages—that are carefully level-differentiated by grades—in the guidebooks can be one crucial rudiment in promoting a successful TEE environment. In addition, it is meaningful to note that, traditionally, majority of corpus analyses were vastly lexical-based, but it seems evident that the automated language analysis tool such as Coh-Metrix, with its wide range of indices, can offer a handle in expanding the capabilities of corpus linguistics by providing evidence-based suggestions. Certainly, future guidebook developers can amply benefit from it.

Finally, some limitations of the present study and suggestions for future research are addressed. First, during a delivery of instructional English to non-native child-learners, a teacher frequently, uses a variety of teaching-resources (e.g., visual aids, *realia*) to supplement learner-comprehension along with her spoken instructions, which this study obviously did not ponder upon. Such supplements can critically influence how teacher-talk is used and perceived in practice and how learners comprehend it. It is projected that Coh-Metrix analysis can also extend to such studies of progressive language outcomes, for example, by examining the linkage between the lively delivered spoken instructions (e.g., transcribed classroom English) and the pedagogical objectives (e.g., transcribed/written student-responses) accordingly. Second, we analyzed classroom English from four grade levels as two independent groups (i.e., grades 3 and 4 vs. grades 5 and 6) as it is generally divided so in the *National Curriculum*. For drawing more comprehensive proposals toward the improvement of classroom English, extensive comparisons among all four grade levels seem needed. Lastly, the general suitability of classroom English provided for the target learner groups is not defined or measured in this study. Namely, this study compared the difficulty of classroom English for the two major levels with a wide range of (psycho)linguistic features and patterns; however, it is unclear to determine if any classroom English outlined in each grade, initially, is appropriate for the target learners in that grade. Detailed qualitative analyses of classroom English in terms of its linguistic appropriateness certainly would complement the findings of this study.

Applicable level: Elementary

REFERENCES

Ahn, B.-K., & Ma, Y. (2015). A Coh-Metrix analysis of elementary school English textbooks. *English21, 28*(3), 435-460.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (CD-ROM). Philadelphia: 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.

Beck, I., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary development*. New York: Guilford.

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.

Bouton, L. F. (1994). Conversational implicature in the second language: Learned slowly when not deliberately taught. *Journal of Pragmatics, 22*(2), 157-167.

Britton, B. K., & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, *83*(3), 329-404.

Carroll, S. E. (1999). Putting 'input' in its proper place. *Second Language Research*, *15*(4), 337-388.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. New York: Cambridge University Press.

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.

Cohen, A. (1996). Developing the ability to perform speech acts. *Studies in Second Language Acquisition, 18*(2), 253-267.

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

Cullen, R. (1998). Teacher talk and the classroom context. *ELT Journal*, *52*(3), 179-187.

Cunningsworth, A. (1995). *Choosing your coursebook*. Oxford, England: Heinemann.

Ellis, R. (1981). The role of input in language acquisition: Some implications for second language teaching. *Applied Linguistics*, *2*(1), 70-82.

Ellis, R. (2005). Principles of instructed language learning. *System*, *33*(2), 209-224.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, *12*(4), 395-427.

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. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*(2), 371-398.

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.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371-395.

Gupta, R. (2013). The readability of English language texts in the primary grades. *Language in India, 13*(5), 534-544.

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, *114*(3), 357-374.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Harmer, J. (2008). How to teach English. *ELT Journal*, *62*(3), 313-316.

House, J. (1996). Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness. *Studies in Second Language Acquisition, 18*(2), 225-252.

Jang, N.-R., & Kim, J.-R. (2018). An analysis of cultural contents in elementary school English textbooks for grades 3 and 4 based on the 2015 Revised Korean National Curriculum. *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*(1), 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. (2010). A corpus-based linguistic analysis of high school English textbooks. *Foreign Languages Education, 17*(1) 209-233.

Jung P. Y., & Shin, J.-A. (2021). Primary teachers' perception toward classroom English in teacher-guidebooks. *Primary English Education, 27*(1), 99-120.

Kim, H. (2018). A case study for the development of teachers' guides in English textbooks. *English21, 31*(4), 205-223.

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.-R. (2014). Continuity problems of elementary and secondary English education: Sequence analysis of English textbooks. *Language Research, 50*(1), 161-184.

Kim, J.-R., & 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, S. (2015). A Study on the improvement of the teachers' guide classroom English through its comparison with real classes. *The Journal of Modern British & American Language & Literature, 33*(1), 267-292.

Kim, S.-Y., & Kim, T.-Y. (2012). The influence of elementary school students' interest, language anxiety, and English learning motivation about TEE class on the students' perceived English learning achievement. *Primary English Education*, *18*(1), 283-307.

Krashen, S. D. (1976). Formal and informal linguistic environments in language acquisition and language learning. *TESOL Quarterly, 10*(2), 157-168.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford, England: Pergamon.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London: Longman.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211-240.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2013). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum.

Lee, H.-Y., & Lee, K. (2011). Pre-service teachers' perceptions of teaching English in English. *Studies in Foreign Language Education*, *25*(2), 25-53.

Lee, S. (2013). Analysis of reading and writing activities in elementary English textbooks for 3rd and 4th graders. *Studies in Linguistics, 27*, 217-239.

Lee, S.-H. (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.

Louwerse, M. (2001). An analytic and cognitive parametrization of coherence relations. *Cognitive Linguistics*, *12*(3), 291-316.

McCormick, D. E., & Donato, R. (2000). Teacher questions as scaffolded assistance. In J. K. Hall & L. S. Verplaetse (Eds.), *Second and foreign language learning through classroom interaction* (pp. 183-201). Mahwah, NJ: Lawrence Erlbaum.

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

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247-288.

Min, C. K. (2008). A survey of the English teachers' perception on Teaching English in English (TEE). *English Language Teaching, 20*(2), 167-190.

Murray, J. (1997). Connective and narrative text: The role of continuity. *Memory & Cognition, 25*(2), 227-236.

No, G. H., & Jung, P. Y. (2016). Constructive suggestions toward communicative classroom English in the teacher guidebooks of primary English textbooks. *Primary English Education, 22*(3), 57-82.

Nunan, D. (1991). *Language teaching methodology*. New York: Prentice Hall.

Oh, M., & Joh, E. S. (2012). Primary school parent-teachers' perceptions of primary English education and of their own children's English learning. *Studies in English Education, 17*(1), 177-207.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357-383.

Pinter, A. (2006). Verbal evidence of task related strategies: Child versus adult interactions. *System*, *34*(4), 615-630.

Plakans, L., & Zeynep, B. (2016). Cohesion features in ESL reading: Comparing beginning, intermediate and advanced textbooks. *Reading in a Foreign Language, 28*(1), 79-100.

Prabhu, N. S. (1987). *Second language pedagogy*. Oxford, England: Oxford University Press.

Richards, J. C. (1993). Beyond the text book: The role of commercial materials in language teaching. *RELC journal*, *24*(1), 1-14.

Richards, J. C., & Lockhart, C. (1994). *Reflective teaching in second language classrooms*. Cambridge, England: Cambridge University Press.

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.

Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse processes*, *29*(1), 37-60.

Service, E., & Craik, F. I. M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language*, *32*(5), 608-623.

Smith, M. S. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, *15*(2), 165-179.

Song, H. S. (2013). Exploring the connectedness of listening tasks according to grades in high school English textbooks. *English Language & Literature Teaching, 19*(4), 279-305.

To, V. (2018). Linguistic complexity analysis: A case study of commonly-used textbooks in Vietnam. *Sage Open, 8*(3), 1-13.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.

VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5-31). Mahwah, NJ: Lawrence Erlbaum.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Willis, J. (1981). *Teaching English through English*. London: Longman.

Wong-Fillmore, L. (1985). When does teacher talk work as input. In S. M. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 17-50). Rowley, MA: Newbury House.

Yim, B.-H., & Huh, K. (2020). A study on the analysis of culture contents in English textbook for grade 5-6: Comparison of 2009 and 2015 Revised National Curriculum. *Journal of the Korea Contents Association, 20*(1), 99-106.

Yoshida, M. (1978). The acquisition of English vocabulary by a Japanese-speaking child. In E. M. Hatch (Ed.) *Second language acquisition: A book of reading* (pp. 91-100). Rowley, MA: Newbury House.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185.