



A Corpus-Based Analysis of the Inter-Grade Continuity of the Reading Passages of High School Mock CSAT English Tests*

Jisu Ryu (Konkuk University) · Hyejin Lee (Wonkwang University) · Moongee Jeon (Konkuk University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: March 31, 2024

Revised: April 18, 2024

Accepted: May 2, 2024

Ryu, Jisu (first author)

Konkuk University

The Konkuk Research Institute for Multilingualism and Multiculturalism
120 Neungdong-ro, Gwangjin-gu,
Seoul, Korea

E-mail: jsryu0508@konkuk.ac.kr

Lee, Hyejin

Wonkwang University

Department of English Education
460 Iksan Daero, Iksan,
Jeonbuk, Korea

Email: lee.hyejin0419@gmail.com

Jeon, Moongee (corresponding author)

Konkuk University

Department of English Language
& Literature
120 Neungdong-ro, Gwangjin-gu,
Seoul, Korea

E-mail: mjeon1@konkuk.ac.kr

*This paper was supported by Konkuk University in 2022.

ABSTRACT

Ryu, Jisu, Hyejin Lee and Moongee Jeon. 2024. A corpus-based analysis of the inter-grade continuity of the reading passages of high school mock CSAT English tests. *Korean Journal of English Language and Linguistics* 24, 616-631.

The aim of this study was to investigate the inter-grade continuity of reading passages of the high school mock College Scholastic Ability Test (CAST) English exams by using Coh-Metrix, a sophisticated text analysis tool. In the current study, a corpus consisting of 75 reading passage files, with 25 passage files for each grade level, was built based on the 2023 mock CSAT English exam administered by the Seoul Metropolitan Office of Education, which was then subjected to a wide range of Coh-Metrix indices covering various linguistic dimensions. These included basic counts, word frequencies, word features (imageability, concreteness, age of acquisition, familiarity), lexical diversity, personal pronouns, connectives, standard readability scores, syntactic complexity, co-reference cohesion, and semantic cohesion scores. The main findings from the broader literature on language learning and assessment indicated robust inter-grade continuity in the high school mock CSAT English test passages, particularly in terms of word count, average word and sentence length, word frequency, word concreteness, additive connectives, and standard readability indices. These findings suggest implications for improving the construction of reading passages in the high school mock CSAT English tests in terms of other less controlled features.

KEYWORDS

mock CAST English tests, continuity, corpus analysis, Coh-Metrix

1. Introduction

High school seniors in South Korea take the College Scholastic Ability Test (CSAT) every year to enter university. The CSAT holds tremendous significance as it is essential for admission to universities. The CSAT, introduced in 1994 to complement the university entrance examination system implemented in the early 1990s, has been conducted ever since and is still ongoing (Han and Lee 2022, Shin 2020).

The significance of the CSAT is also being greatly emphasized in the field of English education because the English section within the CSAT exam is relatively important (Kim and Cha, 2021). For these reasons, the proportion of private education across all aspects of English education has increased, posing a threat to public education (Oh and Shin 2020). To mitigate the negative effects of private education and promote the normalization of public education, the Education Broadcasting System (EBS)-CSAT linkage policy began to be implemented in the English section of the CSAT exam starting from 2011 (Koh and Shin 2017). Furthermore, the criterion-referenced evaluation system was introduced to the CSAT English exam in 2018, prompting research into its effectiveness (Chang 2022, Kim 2021). Unlike the norm-referenced evaluation system where academic grades are determined based on the distribution of CSAT scores for that year's examinees, the criterion-referenced evaluation system determines final grades based on the absolute scores of the exam, or in other words, based on achievement standards. Therefore, it is crucial to effectively adjust the difficulty level of the exam under the criterion-referenced evaluation system (Doo and Choi 2020, Shin 2019).

The new policies and evaluation methods implemented in the CSAT English exam have raised various questions regarding the validity, discriminability, and reliability of the exam. This has led to many studies being conducted from various perspectives to address these concerns. Those studies include research analyzing the complexity of reading passages in CSAT English tests under norm-referenced and criterion-referenced evaluation systems based on lexical density and vocabulary variation (Chang 2019), research comparing the vocabulary diversity and difficulty reflected in the listening scripts of CSAT English tests before the criterion-referenced evaluation and after the criterion-referenced evaluation (Kim and Cha 2021), research comparing the difficulty of reading passages in CSAT English tests with that of high school English textbooks and EBS CSAT Special Lecture (Ahn and Bae 2021, Yang and Lee 2019), and research comparing linguistic differences in the reading passages of CSAT English tests with those of EBS CSAT Special Lecture based on vocabulary sophistication and diversity (Lee 2020).

Prior studies analyzing the reading and listening passages of the CSAT English exam based on the level of comprehension and vocabulary characteristics have greatly contributed to evaluating the exam's validity, discriminability, and reliability. However, there is a relative lack of research on the mock CSAT English exam which aims to provide students with opportunities to prepare for the CSAT English exam and enable CSAT examiners to gauge the English proficiency levels of examinees for adjusting the difficulty of the exam (Lee 2017, Lee 2018). Since the introduction of criterion-referenced evaluation in the English CSAT exam, where students' academic grades are determined by their achievement level, it is crucial to carefully regulate the difficulty level of the CSAT exam (Chang 2019). This requires careful regulation to maintain a gradual progression of difficulty from the first to the third year of high school, facilitating a seamless academic transition for students and accurately reflecting their evolving English proficiency. Therefore, the necessity for research analyzing the validity and reliability of mock English exams, which play a decisive role in preparing for the CSAT English exam and regulating the level of difficulty of the exam, has emerged (Lee and Kang 2015, Song and Kim 2021, Yun et al. 2012). This study utilized the Coh-Metrix system to verify the validity and reliability of the high school mock CSAT English exam by analyzing the inter-grade continuity of the reading passages of high school mock CSAT

English tests conducted in 2023.

In this regard, this study aims to analyze the inter-grade continuity of the reading passages of high school mock CSAT English tests using the Coh-Metrix system, a multilevel language analysis tool (Graesser et al. 2004, Graesser et al. 2007, Graesser et al. 2008, Ryu and Jeon 2020). Specifically, the objective of this study is to construct a corpus of reading passages from the most recent high school first-, second-, and third-year mock CSAT English tests conducted in 2023 and then perform the analysis on the continuity among grades based on various Coh-Metrix indices. By constructing and analyzing a corpus from the 2023 mock exams across first to third-year high school levels, this study seeks to reveal the linguistic continuity that these exams exhibit. Such an analysis is crucial not only for reflecting the expected rigor of the actual high-stakes CSAT, but also for using these findings to promote educational equity. In doing so, it aims to fine-tune the difficulty of the exams, aligning them closely with the overarching goals of educational equity and effectiveness, while taking into account the washback effect on teaching methods and learning outcomes.

2. Research Background

2.1 Research on the CSAT English Tests

Changes in the CSAT system, such as the EBS-CSAT linkage policy and the implementation of criterion-referenced evaluation in English exams, have led to research on the analysis of CSAT English exams from various angles (Chang 2019, Han and Lee 2022, Kim 2021, Kim and Cha 2021, Kim and Choi 2015, Koh and Shin 2017, Lee 2020, Oh and Shin 2020, Shin 2019, Yang and Lee 2019).

For example, Koh and Shin (2017) compared the difficulty level of the English reading section of the CSAT between the periods before and after the EBS-CSAT linkage policy (i.e., 2001 to 2017) with Coh-Metrix program (Graesser et al. 2004, Graesser et al. 2007, Jeon 2015, Ryu and Jeon 2021). Specifically, they analyzed word count, sentence length, word frequency for content and all words, noun density, and two standard indices, FRE, and FKGL. Their findings showed that the difficulty of the CSAT exam increased after the implementation of the EBS-CSAT linkage policy compared to before the implementation of the EBS-CSAT linkage policy. These results indicated that students encountered more challenging test content after the EBS-CSAT linkage policy implementation, resulting in heightened academic pressure.

Oh and Shin (2020) constructed a corpus of reading passages consisting of a total of 11,424 words extracted from CSAT English tests conducted from 2017 to 2019 to evaluate the difficulty of the reading passages of them. They compared the constructed corpus with High School English II textbooks and reading passages from the Scholastic Aptitude Test (SAT), an entrance exam for American universities. Specifically, they analyzed the difficulty of the reading passages of the CSAT English exams at the vocabulary and sentence levels. VocabProfile program (Nation 2006) was used for vocabulary analysis, and L2 Syntactic Complexity Analyzer program (Lu 2010) was utilized for sentence complexity analysis. The research findings indicated that more vocabulary groups were used in the reading passages of the CSAT English exams compared to the High School English II textbooks. Additionally, the sentence complexity of the reading passages of the CSAT English exams was almost similar to that of the SAT level and higher than that of the High School English II textbooks. These results suggest that the examinees' burden for the CSAT English exam remains high even after the implementation of criterion-referenced evaluation.

Kim (2021) investigated the alignment between the difficulty level of the English section of the CSAT under

the criterion-referenced evaluation system and the National Curriculum. To accomplish this objective, a corpus-based analysis was conducted comparing the linguistic complexity of High School English textbooks with CSAT reading passages collected from 2015 to 2020. Results indicated that the difficulty level of CSAT reading passages under criterion-referenced evaluation was significantly higher than High School English textbooks. Moreover, there was minimal disparity in linguistic difficulty between the CSAT English sections under criterion-referenced versus norm-referenced evaluation. These results suggest implications for the effectiveness of the criterion-referenced evaluation system introduced to normalize public education.

Kim and Cha (2021) analyzed the characteristics of vocabulary reflected in listening scripts before and after the introduction of criterion-referenced evaluation. They constructed a corpus of CSAT English listening scripts for two years after the introduction of the criterion-referenced evaluation system in 2018 (i.e., 2018, 2019, 2020). The corpus of listening scripts for norm-referenced evaluation consisted of the listening scripts of CSAT English exams conducted between 2015 and 2017. To analyze the vocabulary characteristics of these criterion-referenced evaluation versus norm-referenced evaluation listening script corpora, measures such as vocabulary token count, vocabulary type count, vocabulary diversity, and vocabulary ease were analyzed. The findings indicated that the vocabulary token count, vocabulary type count, and vocabulary diversity values of criterion-referenced evaluation listening scripts were lower than those of norm-referenced evaluation listening scripts. Moreover, the analysis of vocabulary ease suggested that relatively easier vocabulary was used in criterion-referenced evaluation listening scripts. These results indicate that the difficulty of CSAT listening scripts has decreased somewhat since the introduction of criterion-referenced evaluation.

Chang (2019) analyzed the vocabulary characteristics reflected in the reading passages of CSAT English exams after the introduction of criterion-referenced evaluation. For this study, English reading passages were extracted from CSAT English exams conducted from 2016 to 2019, forming a corpus consisting of a total of 16,138 vocabulary items. To analyze the vocabulary characteristics, LCA program (Lu 2012) was utilized in this research. The results showed that the vocabulary complexity (vocabulary density, vocabulary level, vocabulary variability) of the reading passages based on norm-referenced evaluation (CSAT English exams in 2016 and 2017) did not statistically differ significantly from those based on criterion-referenced evaluation. In other words, the findings suggested that the level of vocabulary complexity reflected in CSAT English exams did not undergo significant changes even after the introduction of the English criterion-referenced evaluation system.

The previous studies related to CSAT English have significantly contributed to evaluating the impact of the introduction of EBS-CSAT linkage policies and the English criterion-referenced evaluation system on the discriminative ability, difficulty level, and other aspects of CSAT English exams. However, relatively few studies have focused on mock English exams, which play an important role in preparing for CSAT English exams and adjusting their difficulty level (Lee 2017).

2.2 Research on the Mock CSAT English Tests

Several studies have been conducted related to mock CSAT English exams. These include research analyzing the characteristics of listening and reading passages based on vocabulary complexity, syntactic complexity, achievement criteria, using a corpus consisting of CSAT English exams and mock CSAT exams conducted from 2016 to 2020 (Doo and Choi 2020), a study examining whether learners' English academic achievement varies depending on whether the language used in the questions, instructions, and answer choices is English or Korean (Yoon, Lee and Park 2012), a study investigating the effectiveness of the EBS-linked education policy introduced in 2011 based on a corpus consisting of CSAT English exams and mock CSAT exams conducted in 2014 and 2015

(Lee and Kang 2015), and a study analyzing significant differences in vocabulary diversity, vocabulary sophistication, and vocabulary density based on a corpus consisting of CSAT English exams and mock CSAT exams conducted in 2018 (Lee 2018).

For instance, Song and Kim (2012) examined the complexity of reading passages found in English textbooks including middle school English 3 and high school English 1 with those in mock CSAT English tests taken by first-grade high school students over three consecutive years (i.e., 2018, 2019, 2020). To achieve this, they utilized L2 Syntactic Complexity Analyzer (L2SCA) along with two different readability indices, the Flesch-Kincaid Grade Level (FKGL) and Flesch Reading Ease (FRE). The findings indicated that the reading passages in the mock tests exhibited the highest syntactic complexity. Furthermore, there was a more pronounced difference in readability between Middle School English 3 and the mock CSAT tests compared to the difference between Middle School English 3 versus High School English 1. These results suggest the need for adjustments in the difficulty levels of textbooks and emphasize the importance of implementing mock tests to alleviate students' learning burdens.

Lee (2018) examined the differences in vocabulary characteristics between the 2018 CSAT English exam and mock CSAT exams (i.e., mock exams conducted in June and September). Specifically, he extracted listening and reading passages from both exam types to build a corpus and then analyzed the diversity of vocabulary together with the frequency of function and content words reflected in the exams. The analysis revealed that the mock exam in June had the highest vocabulary diversity score, while the mock exam in September had the lowest. The vocabulary diversity score of the CSAT English exam fell between the scores of the two mock exams. However, the frequency scores of content and function words in the CSAT English exam were relatively higher. Moreover, the analysis on the frequency scores of the two mock exams showed that the mock exam in September had a higher score than the mock exam in June. These results indicate that the vocabulary difficulty of the listening and reading passages in the CSAT English exam differs from that of the mock exams.

Doo and Choi (2020) constructed corpora for the listening and reading sections of the CSAT English exams and analyzed the differences between the two areas. Specifically, they built corpora for the listening and reading materials based on CSAT English exams and mock exams conducted between 2016 and 2020, respectively. They analyzed the differences between the two sections in terms of vocabulary complexity, syntactic complexity, and achievement criteria. Vocabulary complexity was analyzed using LCA program (Lu 2012), while syntactic complexity analysis utilized the Coh-Metrix system (Graesser et al. 2004, Graesser et al. 2007, Jeon 2015, Ryu and Jeon 2020, Ryu and Jeon 2021). The results revealed that the vocabulary and sentences used in the reading passages were relatively more complex compared to those in the listening passages. In terms of achievement criteria analysis, a greater variety of achievement criteria items were applied to the reading section, and the allocation of achievement criteria points to the reading section was relatively higher than that of the listening section. These results indicate an imbalance between the reading and listening sections in terms of achievement criteria and difficulty.

Previous studies on mock CSAT English exams have contributed to examining the validity and reliability of the CSAT English exam by comparing the passages of mock exam English tests with those of CSAT English exams, English textbooks, and EBS English textbooks. However, research on the continuity of mock CSAT English exams across different grade levels, which can serve as a guide for students in preparing for the CSAT English exam step by step, is lacking. In this regard, the main purpose of this study is to examine the inter-grade continuity of the reading passages of high school mock CSAT English exams based on a wide range of Coh-Metrix indices (Graesser et al. 2007, Ryu and Jeon 2020) after constructing a corpus of the reading passages of mock CSAT English exams for high school first, second, and third graders conducted in 2023.

3. Research Method

3.1 Mock CSAT English Exam Corpus for the Current Study

The list of high school mock CSAT test corpora analyzed in this study was presented in Table 1. As indicated in Table 1, this study selected the mock CSAT test administered in 2023 targeting high school first, second, and third graders for continuity analysis. Specifically, only the mock test conducted in March was chosen as the subject of analysis. Mock CSAT tests are typically conducted four times a year for first and second graders, each organized by the Seoul Metropolitan Office of Education (March), the Busan Metropolitan City Office of Education (June), the Incheon Metropolitan Office of Education (September), and the Gyeonggi Province Office of Education (November). On the other hand, two additional tests are conducted for third graders, organized by the Korea Institute of Curriculum and Evaluation (June and September) as well as the four times conducted by different Offices of Education: the Seoul Metropolitan Office of Education (March and October), the Gyeonggi Province Office of Education (April), the Incheon Metropolitan Office of Education (July). To control for the effects of these different administering organizations and months of the tests conducted, this study focused on the mock CSAT English tests conducted in March, administered under the Seoul Metropolitan Office of Education. These tests were the only ones administered in both the same month and under the same organization across different grades.

Table 1. List of Mock CSAT English Exams for the Current Study

Grade	Year (Month)	Administering Organization
Mock CSAT corpus for first grader	2023 (March)	The Seoul Metropolitan Office of Education
Mock CSAT corpus for second grader	2023 (March)	The Seoul Metropolitan Office of Education
Mock CSAT corpus for third grader	2023 (March)	The Seoul Metropolitan Office of Education

3.2 Corpus Construction and Analysis Tool

For the analysis of the English mock test corpus, this study utilized the Coh-Metrix system created by the Institute for Intelligent Systems (IIS) at the University of Memphis (Graesser et al. 2004, Graesser et al. 2007). The reliability and validity of the Coh-Metrix system have been validated, for example, in a study analyzing the continuity of reading materials of middle school English textbooks (Jeon 2015), a study investigating the cohesion of tutorial discourse (Graesser et al. 2007), and a study examining the continuity of the listening passages of middle school English textbooks (Ryu and Jeon 2020). Specifically, the desktop version of the Coh-Metrix system was employed to analyze the continuity among different grades. Since this system only processes text files in Unicode format (i.e., txt files), all corpora constructed in this study were saved as txt files. For the continuity analysis, reading passages were extracted from the mock English tests conducted for high school first, second, and third graders in 2023.

When constructing the corpus of reading passages, a separate txt file was created for each item included in the reading passages. The corpora analyzed in this study for high school first, second, and third graders consisted of 25 files per grade. Therefore, this study analyzed the continuity among the grades using the Coh-Metrix system based on the mock English reading passage corpus comprising a total of 75 files. The Coh-Metrix indices used in the continuity analysis of this study consisted of basic counts (the mean number of words, the mean number of

sentences, average word length, average sentence length), word frequency (log word frequency for content words), word attributes (familiarity, imageability, concreteness, age of acquisition), lexical diversity (type-token ratio for content words), personal pronouns (first-person, second-person, and third-person pronouns), connectives (causal, additive, and temporal connectives), readability indices (Flesch Reading Ease score, Flesch-Kincaid Grade Level), syntactic complexity (noun density, subject density), co-reference cohesion (argument overlap for adjacent and all sentences), and semantic cohesion measures (LSA cosine for adjacent and all sentences). The selection of these measures is based on their widespread use in existing Coh-Metrix research (Graesser et al. 2008, Jeon 2015, Ryu and Jeon 2020), which highlights their effectiveness in dissecting the detailed layers of text complexity and continuity across grade levels in mock CSAT English exams. These measures allow for understanding not only the surface changes in text complexity, but also the underlying linguistic structures that influence learners' cognitive engagement with texts, thus providing a holistic view of the educational trajectory embodied in mock CSAT English exams.

In this study, IBM SPSS statistics 21 package was utilized to analyze the continuity of the reading passages of the mock CSAT English tests for high school first, second, and third graders conducted in 2023. Specifically, one-way Analysis of Variance (ANOVA) was conducted, with the independent variable being the grade level and the dependent variable being the Coh-Metrix measures. Each analysis was performed at a significance level of 5% ($p = .05$).

4. Results

4.1 Results of Basic Counts

Table 2 summarizes the results of basic counts, demonstrating a well-established continuity among the reading passages of high school mock CSAT English tests in terms of the number of words, average word length, and average sentence length across different grades. Regarding the size of each reading passage, the average number of words for the reading passages for first-year high school students ranged from 110 to 149. The average number of words for the reading passages for second-year high school students ranged from 139 to 178. The average number of words for the reading passages for third-year high school students ranged from 150 to 190. Specifically, as the grade level increased, there was an increase in the average number of words used in the mock test passages, as well as in the average word and sentence length. However, despite the increase in grade level, there was no significant difference in the average number of sentences. This can be understood as an attempt to increase the length of sentences rather than the number of sentences as the grade level increases.

Table 2. Means (Standard deviations) for Basic Counts

Coh-Metrix Indices	High School 1 ($n = 25$)	High School 2 ($n = 25$)	High School 3 ($n = 25$)	F	p
Number of words	130 (35.6)	159 (52.8)	170 (56.7)	4.507	.014*
Number of sentences	7.92 (3.24)	8.76 (5.87)	8.40 (4.90)	.193	.825
Average word length	1.47 (.11)	1.53 (.15)	1.64 (.17)	8.666	.000*
Average sentence length	17.2 (3.96)	20.3 (5.71)	22.1 (5.87)	5.623	.005*

* $p < .05$

Post hoc analysis using Tukey's post hoc test was conducted for the number of words, mean sentence length, and mean word length measures that showed significant differences in the one-way ANOVA. Since the sample size was consistent at 25 for each condition, Tukey's post hoc test was performed in this analysis. The analysis of the number of sentences and the average sentence length showed an increasing trend between high school second and third grades, although it was not statistically significant. However, it was statistically significantly higher compared to the mock test reading passages for first-year high school students. Regarding the mean word length score, there was no statistically significant difference between the mock test reading passages for high school first and second years, but it was lower compared to the reading passages for third-year high school students.

4.2 Results of Word Frequency

The result of word frequency analysis is presented in Table 3. Word frequency is a critical factor influencing the difficulty of a text (Graesser et al. 2004). Therefore, it is important to control the word frequency used in mock test reading passages for gradual manipulation of linguistic difficulty according to grade level when developing mock test items.

Table 3. Means (Standard deviations) for Word Frequency

Coh-Matrix Indices	High School 1 (<i>n</i> = 25)	High School 2 (<i>n</i> = 25)	High School 3 (<i>n</i> = 25)	<i>F</i>	<i>p</i>
Word frequency (Content words)	2.29 (.13)	2.23 (.23)	2.20 (.14)	1.542	.221

* $p < .05$

The Coh-Matrix system calculates word frequency measures using the CELEX lexical database (Graesser et al. 2004, Graesser et al. 2008). In this study, logarithmic transformation was applied to the word frequency measures for the analysis of grade-level continuity in mock test reading passages. Such logarithmic transformation is commonly employed to approximate the distribution of word frequency scores to a normal distribution, typically used for frequency analysis (Graesser et al. 2004).

As shown in Table 3, although there is a tendency for low-frequency words to increase as the grade level increases, there was no significant difference in word frequency among the three grades. These findings imply the need for more careful consideration when adjusting the frequency of words in the development of future mock test items.

4.3 Results of Word Features

The word feature measures provided by the Coh-Matrix system consist of imageability, concreteness, age of acquisition, and familiarity. Imageability indicates the degree to which a word can be visualized in the mind. Concreteness refers to the degree of specificity of the word. Finally, age of acquisition indicates the age at which the word was acquired. These measures are important factors that significantly influence the comprehension and difficulty of texts (Graesser et al. 2004). They are calculated based on the vocabulary information included in the MRC Psycholinguistic word database (Coltheart 1981). These measures are displayed as values between 100 and 700.

As shown in Table 4, statistically significant differences were observed in concreteness measures among the mock exam English reading passages for high school first, second, and third graders ($p < .05$). However,

statistically significant differences were not found among grades for imageability, age of acquisition, and familiarity measurements (all $ps > .05$). To examine the differences among grades, a Tukey post hoc test was conducted for the concreteness measure. As a result, the concreteness score for the first-grade high school passages was statistically significantly higher compared to the third-grade passages ($p < .05$). However, the differences between the first-grade and second-grade passages, and between the second-grade and third-grade passages were not statistically significant (all $ps > .05$). These results indicate that the continuity among grades in mock CSAT English exams was not well-established for the remaining word feature measures, excluding the concreteness measure.

Table 4. Means (Standard deviations) for Word Features

Coh-Matrix Indices	High School 1 ($n = 25$)	High School 2 ($n = 25$)	High School 3 ($n = 25$)	F	p
Imageability	413 (26.9)	402 (30.0)	395 (28.3)	2.541	.086
Concreteness	385 (25.2)	369 (29.4)	363 (32.8)	3.966	.023*
Age of acquisition	351 (43.4)	354 (46.4)	379 (43.7)	2.981	.057
Familiarity	573 (7.9)	570 (11.9)	567 (10.0)	2.334	.104

* $p < .05$

4.4 Results of Lexical Diversity

Lexical diversity indicates the degree to which different words are used in the text. The lexical diversity index provided by the Coh-Matrix system refers to the type-token ratio (TTR) of content words. Type represents the frequency of occurrence of individual vocabulary items, while token indicates the frequency of occurrence of specific vocabulary items (i.e., types). A higher TTR suggests a greater likelihood of diverse vocabulary usage in the passage, leading to higher vocabulary diversity in the text (Jeon 2015). Passages with higher lexical diversity tend to increase in difficulty due to the processing burden associated with handling various vocabulary items (Jeon 2015, Ryu and Jeon 2020, Ryu and Jeon 2021).

The summary of the lexical diversity analysis results is presented in Table 5. The analysis revealed that the lexical diversity observed in high school mock exam passages did not show statistically significant differences among grades ($p > .05$).

Table 5. Means (Standard deviations) for Lexical Diversity

Coh-Matrix Indices	High School 1 ($n = 25$)	High School 2 ($n = 25$)	High School 3 ($n = 25$)	F	p
Type-token ratio (Content words)	.81 (.11)	.79 (.08)	.79 (.10)	.438	.647

* $p < .05$

These results indicated that systematic continuity in terms of vocabulary diversity across grades was not reflected in mock English exams. Many research findings suggest that gradually increasing lexical diversity according to grade levels in passages is essential for effective English learning (Ryu and Jeon 2020). Therefore, these results suggest that when developing mock exam passages, it is important to incorporate a variety of vocabulary as grades increase to facilitate more effective English learning.

4.5 Results of Personal Pronouns

The Coh-Metrix system provides measures for pronouns. It is known that an increase in the use of pronouns in a text correlates with increased text complexity (Graesser et al. 2007). This is because when pronouns are used, readers need to search for their referents, increasing cognitive processing load for them and consequently raising the difficulty level of text comprehension. Particularly, third-person pronouns (e.g., *he, she, it* etc.) have a broader referential scope compared to first-person (e.g., *I, we*) and second-person pronouns (e.g., *you*), making texts with frequent use of third-person pronouns relatively more challenging to process (Ryu and Jeon 2020, Ryu and Jeon 2021). Therefore, effectively regulating the ratio of personal pronouns used in English mock exam passages is considered crucial for promoting effective English learning.

Table 6 presents the results of the usage of personal pronouns in the reading passages of high school mock English exams. The analysis revealed that statistically significant differences among grades were not found for all pronoun measures. Despite the importance of pronouns in text comprehension, these results showed that the effective adjustment of pronoun usage for controlling the difficulty level of exams across grades was not adequately reflected. In other words, the findings of this study indicated that linguistic features such as personal pronouns were not actively considered during the development of mock English exam passages. These results suggest that in the future, adjustments for difficulty based on pronoun usage should be more actively considered in the process of developing mock exam items.

Table 6. Means (Standard deviations) for Personal Pronouns

Coh-Metrix Indices	High School 1 (<i>n</i> = 25)	High School 2 (<i>n</i> = 25)	High School 3 (<i>n</i> = 25)	<i>F</i>	<i>p</i>
First-person pronouns	8.06 (11.1)	7.80 (13.9)	13.17 (16.5)	1.169	.317
Second-person pronouns	8.72 (12.8)	9.06 (15.2)	3.30 (6.4)	1.803	.172
Third-person pronouns	12.5 (12.2)	10.7 (15.0)	10.5 (15.0)	.151	.860

* $p < .05$

4.6 Results of Connectives

This study analyzed the patterns of connective usage in English mock exam passages, which directly or indirectly affect text complexity. The current study utilized Coh-Metrix measures such as causal connectives (e.g., *therefore, because* etc.), additive connectives (e.g., *moreover, and* etc.), and temporal connectives (e.g., *until* etc.) (Graesser et al. 2004). Specifically, causal and temporal connectives play a role in clearly indicating causal and temporal relationships between sentences, thus facilitating semantic and temporal cohesion between sentences (Jeon 2015). In other words, causal connectives tend to facilitate the degree of connection between causally related sentences, while temporal connectives contribute to establishing temporal cohesion by clearly indicating temporal relationships within the text. Additive conjunctions, on the other hand, may increase the syntactic complexity of sentences simply by connecting them. Therefore, connectives can contribute to adjusting the difficulty of text (Jeon 2015, Millis and Just 1994, Murray 1997).

As presented in Table 7, the analysis revealed no significant differences among grades in terms of causal and temporal connectives, which directly influence text comprehension and difficulty. However, significant differences were found in terms of additive connectives among grades ($p < .05$). A Tukey post-hoc test showed that more additive connectives were used in mock exams targeting second- and third-grade high school students

compared to first-grade students. There was no significant difference between second- and third-grade mock exams. These results suggest that when manipulating the difficulty of mock exams, linguistic elements such as causal and temporal connectives should be more actively considered.

Table 7. Means (Standard deviations) for Connectives

Coh-Metrix Indices	High School 1 (<i>n</i> = 25)	High School 2 (<i>n</i> = 25)	High School 3 (<i>n</i> = 25)	<i>F</i>	<i>p</i>
Causal connectives	26.4 (16.2)	26.5 (14.3)	26.5 (14.3)	.000	1.000
Additive connectives	7.03 (6.5)	14.8 (11.5)	13.3 (11.1)	4.316	.017*
Temporal connectives	17.5 (14.8)	19.3 (17.6)	16.9 (14.1)	.151	.860

* $p < .05$

4.7 Results of Standard Readability

The Coh-Metrix system provides useful measures for measuring the readability of texts, namely the Flesch Reading Ease (FRE) score and the Flesch-Kincaid Grade Level (FKGL) score. The FRE score indicates the ease of comprehension of a text on a scale from 0 to 100, where higher scores represent easier texts. The FKGL score indicates the grade level of a text on a scale from 1 to 12, with lower scores representing easier texts. These two indices are primarily based on word length and sentence length (Graesser et al. 2004). Therefore, if long words and sentences are used in a text, the FRE score tends to increase, while the FKGL score tends to decrease. As word length and sentence length can directly impact text comprehension, these measures can be useful for measuring text difficulty (Graesser et al. 2007).

As shown in Table 8, significant differences were found among grades in both FRE and FKGL scores for high school first-, second-, and third-grade mock English exam reading passages ($ps < .05$). Post-hoc Tukey tests for these two measures revealed that for the FKGL score, while there was no statistically significant difference between second- and third-grade passages, significant differences were found between first- and second-grade passages and between first- and third-grade passages (all $ps < .05$). In other words, compared to first-grade passages, second- and third-grade passages contained more linguistically complex texts. These results indicate the need for a clearer adjustment in the difficulty gap between the mock exam passages for the second and third grades.

Table 8. Means (Standard deviations) for Standard Readability

Coh-Metrix Indices	High School 1 (<i>n</i> = 25)	High School 2 (<i>n</i> = 25)	High School 3 (<i>n</i> = 25)	<i>F</i>	<i>p</i>
FRE	64.6 (11.6)	56.9 (14.8)	45.3 (17.7)	10.677	.000*
FKGL	8.5 (2.3)	9.8 (2.3)	10.8 (2.3)	6.817	.002*

* $p < .05$

On the other hand, the FRE score analysis revealed significant differences among all grades. In other words, unlike FKGL scores, significant differences were found between second- and third-grade passages. However, the difference between these two grades was not substantial. Since both FRE and FKGL scores are influenced by sentence length and word length, the analysis results of these two measures are similar to the results of basic counts presented in Table 2. Overall, the analysis of readability measures suggests the need for finer control over the length of words and sentences used in second- and third-grade passages to align them more closely with the grade levels.

4.8 Results of Syntactic Complexity

The syntactic complexity measures provided by Coh-Metrix consist of noun density and subject density. Noun density is calculated by dividing the number of modifiers modifying the head noun in a noun phrase by the total number of words of the noun phrase. Subject density represents the number of words preceding the main verbs. Higher values of these measures tend to indicate an increase in the difficulty of the text (Graesser et al. 2004, Graesser et al. 2007). Thus, syntactic complexity measures can play a direct role in assessing the difficulty of the text.

The results of syntactic complexity are presented in Table 9. As shown in Table 9, no significant differences were found in both measures across three grades. This suggests that, contrary to the expectation of different syntactic complexity in the sentences used in the three exam passages, there is a similarity in syntactic complexity among the passages. The analysis of syntactic complexity indicates the need to carefully consider the syntactic complexity of sentences when adjusting the difficulty in future mock test preparations.

Table 9. Means (Standard deviations) for Syntactic Complexity

Coh-Metrix Indices	High School 1 (<i>n</i> = 25)	High School 2 (<i>n</i> = 25)	High School 3 (<i>n</i> = 25)	<i>F</i>	<i>p</i>
Noun density	.87 (.12)	.88 (.27)	.89 (.21)	.070	.933
Subject density	4.7 (1.89)	4.9 (2.56)	5.7 (2.93)	1.238	.296

* *p* < .05

4.9 Results of Co-reference Cohesion

Co-reference cohesion, an important linguistic factor influencing text comprehension, is formed when two sentences share common referents (nouns, pronouns, noun phrases) within the text. Higher cohesion between sentences leads to better text comprehension, thus reducing the difficulty of the text (Cirilo 1981, Graesser et al. 2007, Graesser et al. 2008). The analysis of co-reference cohesion in reading passages of mock CSAT English exams for high school first, second, and third graders was presented in Table 10.

As presented in Table 10, the analysis revealed no statistically significant differences across grades for two argument overlap scores for adjacent and overall sentences (all *ps* > .05). These results indicated that co-reference cohesion was not systematically adjusted across grades. Despite its importance, co-reference cohesion is not visibly captured within the text, highlighting the need for tools like Coh-Metrix in mock test item creation to consider co-reference cohesion measures to establish continuity across grades.

Table 10. Means (Standard deviations) for Co-reference Cohesion

Coh-Metrix Indices	High School 1 (<i>n</i> = 25)	High School 2 (<i>n</i> = 25)	High School 3 (<i>n</i> = 25)	<i>F</i>	<i>p</i>
Argument overlap (Adjacent sentences)	.55 (.23)	.70 (.23)	.65 (.23)	2.769	.069
Argument overlap (All sentences)	.51 (.20)	.60 (.25)	.55 (.21)	1.008	.370

* *p* < .05

4.10 Results of Semantic Cohesion

Semantic cohesion indicates the extent to which sentences in a text are semantically connected to each other (Jeon 2015). Higher semantic cohesion of a text leads to clearer conveyance of meaning, resulting in lower text complexity. The Coh-Metrix system analyzes semantic cohesion using Latent Semantic Analysis (LSA) cosine measures, which are derived from the LSA mathematical model (Landauer 2007). LSA values are expressed as cosine values. To establish continuity across grades systematically, semantic cohesion reflected in exam passages should gradually decrease as grade levels increase.

Table 11 presents the results of the semantic cohesion analysis for reading passages in high school mock English exams. The analysis revealed no statistically significant differences across grades for LSA cosine measures for adjacent and overall sentences (all $ps > .05$), contrary to expectations. This indicated that, similar to co-reference cohesion measures, there was no systematic adjustment in semantic cohesion across grades. Therefore, as implied by the findings of the co-reference cohesion analysis, these results indicate the necessity of employing useful tools like Coh-Metrix to control semantic cohesion when creating mock test items.

Table 11. Means (Standard deviations) for Semantic Cohesion

Coh-Metrix Indices	High School 1 ($n = 25$)	High School 2 ($n = 25$)	High School 3 ($n = 25$)	F	p
LSA cosine (Adjacent sentences)	.24 (.15)	.25 (.09)	.26 (.14)	.070	.932
LSA cosine (All sentences)	.22 (.15)	.21 (.08)	.23 (.14)	.174	.840

* $P < .05$

5. Conclusion and Suggestions

In this study, the Coh-Metrix system, a corpus analysis tool developed by the Institute for Intelligent Systems at the University of Memphis, USA, was utilized to analyze the inter-grade continuity of mock CSAT English exams for first-, second-, and third-year high school students in 2023 based on multidimensional language indices provided by Coh-Metrix.

The main findings of this study were summarized as follows. Firstly, in the analysis of basic counts, the inter-grade continuity of high school English mock exams was well established in terms of average word count, average sentence length, and average word length. In other words, there was a tendency for word and sentence length to increase as the grade level advanced. These results indicated that superficial measures such as word and sentence length, which are visually apparent, are well reflected in the construction of reading passages in mock exams. However, contrary to expectations, statistically significant differences were not found in word frequency among grades. Word frequency is an important indicator as it directly measures the difficulty of word, thus it should be taken into consideration when composing questions to ensure a gradual increase in difficulty across grades. It is suggested that vocabulary frequency be well controlled when creating future mock exam questions.

Similarly, the analysis of word features revealed that inter-grade continuity was not statistically significant for measures such as imageability, age of acquisition, and familiarity, excluding concreteness. Furthermore, the analysis of lexical diversity also indicated that the inter-grade continuity of mock exam passages was not well established for this measure. Since the word attributes and diversity of vocabulary can significantly impact the

difficulty of passages (Jeon 2015), it is recommended that future English mock exam questions reflect the findings of this study. These findings about word features and lexical diversity require a strategic pedagogical shift towards a more deliberate and detailed approach to vocabulary selection in mock CSAT English exams. By more closely aligning exam content with students' incremental learning and cognitive development across grade levels, educators can better scaffold language acquisition and ensure a gradual and more natural progression of language proficiency.

The analysis of personal pronouns revealed that the inter-grade continuity of mock exam passages was not well established for first-person, second-person, and third-person pronouns. Additionally, the analysis of connective measures indicated that the continuity among grades was not well controlled for causal and temporal connectives. These results suggest that since personal pronouns and connectives used in passages significantly contribute to the formation of passage difficulty, such linguistic features should be well reflected when creating future English mock exam passages. The results of the pronoun and connective measures can also be usefully applied to English language teaching and learning. By explicitly teaching students about pronoun and connective usage, teachers can explore ways to improve their reading skills.

The analysis of standard readability indicated significant differences among grades for FRE and FKGL scores, suggesting that as grades increase, the overall difficulty of passages tends to rise. These readability measures are heavily influenced by word and sentence length, which aligns with the analysis results for word and sentence length. On the other hand, the analysis of syntactic complexity revealed that inter-grade continuity of mock exam passages was not well established. These results imply that when adjusting the difficulty of mock exams, not only word and sentence length but also syntactic structures such as noun phrase density and subject density should be considered.

Finally, the results of co-reference cohesion and semantic cohesion indicated that the inter-grade continuity of mock exam passages was not statistically significant. The reason for the lack of appropriate control over the inter-grade continuity of mock exam passages for co-reference cohesion and semantic cohesion measures is that these indices are not explicitly evident in the text (Jeon 2015, Ryu and Jeon 2020). In other words, it is challenging to visually control these measures. These results suggest that the Coh-Metrix system utilized in this study is beneficially employed in estimating the difficulty of English passages based on discourse-level measures. In addition, the results of co-reference and semantic cohesion can also be used in the classroom to improve students' reading skills through learning based on these cohesion measures, which affect the difficulty of the text.

Despite the implications and contributions of this study, there are limitations that should be briefly noted. Firstly, the analysis was limited to the reading passages of mock CSAT English exam conducted in 2023. Future research could benefit from analyzing a more extensive corpus, including mock exams conducted over recent years, to derive more generalizable results regarding inter-grade continuity. Additionally, this study only analyzed mock exam passages administered by the Seoul Metropolitan Office of Education. Future research can involve constructing a corpus using mock exam passages administered by educational authorities such as the Gyeonggi Provincial Office of Education, the Incheon Metropolitan Office of Education, and the Korea Institute of Curriculum and Evaluation, enabling comparisons with the results of this study. Furthermore, in future research, it will be possible to compare the inter-grade continuity of mock exam passages before and after the introduction of the criterion-referenced evaluation system, based on a corpus of mock exam reading passages.

The results of the current study suggest that the findings derived herein can be effectively utilized for analyzing mock CSAT English exam passages and further extend to the analysis of the relationship between mock English exams and EBS English textbooks, as well as English textbooks. This study anticipates that the Coh-Metrix

language analysis program used in the current study will be actively applied to control the difficulty of mock exam passages and various other exams in the future.

References

- Ahn, H. and J. Bae. 2021. A study about the analysis of linguistic difficulty among English textbooks with 2015 Revised National Curriculum, EBS-CSAT prep books, and College Scholastic Ability Test. *Secondary English Education* 14(3), 39-58.
- Chang, J. 2019. A comparison of lexical complexity in 2016-2019 CSAT English reading passages by the types of assessment (norm-referenced assessment vs. criterion-referenced assessment). *The SNU Journal of Education Research* 28(3), 85-110.
- Chang, J. 2022. A multi-dimensional comparison of CSAT English reading passages before and after the use of criterion-referenced assessment. *Language Research* 58(1), 61-89.
- 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.
- Doo, S. and I. Choi. 2020. A comparative analysis on difficulty of LC and RC of the KCSAT and the mock KCSAT English tests. *English Language Assessment* 15(2), 119-147.
- Graesser, A. C., M. Jeon, Z. Cai and D. S. McNamara. 2008. Automatic analyses of language, discourse, and situation models. In J. Auracher and W. van Peer, eds., *New beginnings in literary studies*, 72-88. Cambridge: Cambridge Scholars Publishing.
- Graesser, A. C., M. Jeon, Y. Yan and Z. Cai. 2007. Discourse cohesion in text and tutorial dialogue. *Information Design Journal* 15(3), 199-213.
- Graesser, A. C., D. S. McNamara, M. M. Louwerse and Z. Cai. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers* 36(2), 193-202.
- Han, J. and Y. Lee. 2022. A corpus-based analysis of the perceived difficulty increase in the 2022 CSAT English Test. *Korean Journal of English Language and Linguistics* 22, 336-354.
- 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.
- Kim, J. and I. Choi. 2015. A corpus-based comparative analysis of linguistic difficulty among high school English textbooks, EBS-CSAT prep books, and College Scholastic Ability test. *Multimedia-Assisted Language Learning* 18(1), 59-92.
- Kim, K. and K. Cha. 2021. An analysis on the vocabulary of recent CSAT's English listening assessment: Lexical scales, word difficulty, and high frequency content words. *English Language Teaching* 33(2), 23-36.
- Kim, S. 2021. A corpus-based analysis of linguistic difficulty of the CSAT English section under absolute grading. *English Language Assessment* 16(2), 205-227.
- Koh, N. and J. Shin. 2017. A comparison of the level of difficulty in the English reading part of the CSAT: Before and after the EBS-CSAT linkage policy. *Secondary English Education* 10(4), 3-24.
- Landauer, T. K. 2007. LSA as a theory of meaning. In T. K. Landauer, D. S. McNamara and W. Kintsch, eds., *Handbook of Latent Semantic Analysis*, 3-34. Mahwah, NJ: Erlbaum.
- Lee, D. and M. Kang. 2015. The analysis of linkage of the Korea Educational Broadcasting System (EBS) with

- the College Scholastic Ability Test (CSAT) & trial test for CSAT and the study on improvements for the linkage(in the English Part of the Test). *Secondary English Education* 8(4), 69-92.
- Lee, H. 2017. Analysis of the distribution of speech acts in the English conversations of the 2017 CSAT and CSAT simulations. *Secondary English Education* 10(3), 23-45.
- Lee, H. 2018. Vocabulary analysis of the English section of the 2017 CSAT and CSAT simulations. *International Language and Literature* 41, 195-215.
- Lee, H. 2020. Vocabulary analysis of CSAT English tests and CSAT-EBS preparation coursebooks, with reference to the reading tests. *Modern English Education* 21(3), 48-57.
- Lee, M. 2021. The prediction of the English preliminary CSAT by high school English performance test and regular English exam. *English Language Assessment* 16(2), 187-204.
- Lu, X. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15, 474-496.
- Lu, X. 2012. The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal* 96(2), 190-208.
- Millis, K. and M. Just. 1994. The influence of connectives on sentence comprehension. *Journal of Memory and Language* 33, 128-147.
- Murray, J. 1997. Connective and narrative text: The role of continuity. *Memory & Cognition* 25(2), 227-236.
- Nation, I. S. P. 2006. How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63, 59-82.
- Oh, J. and Y. Shin. 2020. A corpus-based analysis of the linguistic complexity levels of reading passages in the Korean college entrance examination. *English Language Teaching* 32(2), 109-126.
- Ryu, J. and M. Jeon. 2020. An analysis of the continuity of the listening scripts of middle school English textbooks with Coh-Metrix. *English Language & Literature Teaching* 26(3), 101-121.
- Ryu, J. and M. Jeon. 2021. An analysis of linguistic features in science textbooks across grade levels: Focus on text cohesion. *Journal of the Korean Association for Science Education* 41(2), 71-82.
- Shin, D. 2020. A validity study on the item type of 'identifying the emotional state or mood' in the English section of CSAT based on sentiment analysis. *Journal of the Korea English Education Society* 19(4), 159-179.
- Shin, Y. 2019. Analyzing CSAT reading passages by using Coh-Metrix and VocaProfile: Focusing on four years from 2016 to 2019. *Journal of Language Sciences* 26(4), 109-127.
- Song, J. and J. Kim. 2021. An analysis of the reading passages represented in the textbooks of middle/high school and national achievement English tests through syntactic complexity and readability. *Secondary English Education* 14(2), 48-66.
- Yang, S. and D. Lee. 2019. A corpus-based analysis of the topic distribution and vocabulary level of textbooks, EBS materials, and CSATs. *Journal of Learner-Centered Curriculum and Instruction* 19(4), 711-729.
- Yun, J., M. Lee and Y. Park. 2012. The influence of the language of directions, questions, and choices in practice CSAT on learners' test results. *Korean Journal of Applied Linguistics* 28(1), 59-85.

Examples in: English

Applicable Languages: English

Applicable Level: Tertiary