

Coh-Metrix를 활용한 고교 학년별 영어 영역 읽기 지문 난이도 분석

Analysis of the difficulty of reading passages between
grades in high school English tests by Coh-Metrix

2018년 6월

서울대학교 자유전공학부
김진영(2011-13272)

요약

본 연구에서는 고등학교 전국연합학력평가, 수능 모의평가 및 수능 영어 영역 읽기 지문의 난이도를 Coh-Metrix라는 웹 기반 언어 분석 도구를 활용하여 학년별로 분석하였다. 연구를 위해 최근 3개년도(2015~2017)에 대해 각 연도당 각 학년별로 2개씩 시험지 표본을 선정, 읽기 지문을 추출하여 각 시험지당 1개의 코퍼스로 구축하였다. 구축된 코퍼스를 Coh-Metrix로 분석하여 학년별로 결과치를 수집하고, 각 척도별 분석 결과에 대해 일원 분산분석을 실시하여 학년 간 평균치를 비교하였다. p-value가 유의미한 수준(<0.05)으로 나타난 척도에 대해서는 Tukey-HSD 사후 분석을 실시하여 학년 간 차이 유무를 확인하였다. 추가적으로 본 연구의 결과를 기존 연구 중 Coh-Metrix를 활용하여 고등학교 영어 교과서를 분석한 결과와 비교하여 교과서와 시험 간의 난이도를 비교하였다.

본 연구의 주요 결과는 다음과 같다. 우선, 영어 시험의 읽기 지문은 학년별로 총 단어 수, 이독성 지수, 단어 수준, 텍스트 난이도 지표에서 유의미한 차이를 보여 학년 간 연계성이 있는 것으로 나타났다. 학년이 올라감에 따라 지문의 총 단어 수는 증가하고 비일상적인 내용어의 출현 빈도가 증가하였으며, 텍스트의 서사성, 단어 구체성, 참조적 응집력은 감소하였다. 이독성 지수에서는 학년이 올라감에 따라 난이도가 증가하여 최종 학년의 경우 지문 난이도가 미국 학제 기준 10~11학년 수준으로 나타났다. 다만, 통사적 난이도는 학년 간 차이가 유의미하지 않아 상대적으로 어휘 수준 상의 차이가 두드러졌다.

기존 연구 결과를 토대로 시험 지문과 영어 교과서 읽기 지문의 수준을 비교한 결과, 고등학교 3학년 대상 교과서인 영어Ⅱ의 난이도가 고등학교 1학년 지문 수준에도 미치지 못해 교과서가 고교 영어 학습 준거로서의 기능을 상실한 것으로 나타났다. 이와 같은 상황에서 학년별 영어 시험 지문의 난이도를 비교분석한 본 연구 결과가 학습 자료 선정 및 현실적인 영어 학습 성취 수준 설정에 있어 기초 근거 자료로 활용될 수 있을 것으로 기대된다.

키워드 : Coh-Metrix(코메트릭스), 전국연합학력평가, 대학수학능력시험, 수능 영어 영역, 고등학교 영어, 독해 지문, 영어 교과서, 학년 간 난이도

Abstract

This study analyzed the difficulty of reading passages of English examinations of College Scholastic Ability Test(CSAT) and its mock tests in Korean high schools using Coh-Metrix, a web-based tool for text analysis. Tests administered in 2015 to 2017 were used, and two test sets were selected for research per grade each year. Reading passages from each test set composed a single corpus. Result data from Coh-Metrix were collected by grade, and one-way ANOVA was conducted to compare the differences. Tukey's range test was performed as a post hoc test to investigate significance between grades. Additionally, the result was compared with that of Coh-Metrix analysis of high school English textbooks from earlier studies.

The major results are as follows: Significant differences were observed among grade levels in the total number of words, readability indices, word familiarity, and text easability indices. The number of words escalated, and word familiarity, narrativity, word concreteness, and referential cohesion declined according to the grade levels. Increasing passage difficulty could also be observed through readability indices, and they showed that the level of the passages for the final(12th) grade students reached around 10~11th year in the US grade level. However, the differences of syntactic simplicity were not in the range of significance. This implied that vocabulary was the main source of the gaps among the levels.

When compared to the result of this study, passages from English II textbooks which are for the 12th year students were easier than those from 10th grade tests. This indicates that the textbooks cannot function as a norm of English study in high school, and explains the reality that the test papers are used as a de facto standard instead. As long as this phenomenon continues, it is expected that the result of this study can be of use when selecting proper reading materials and setting appropriate English-learning objectives.

Keywords: Coh-Metrix, College Scholastic Ability Test(CSAT), English test, mock test, high school, reading passage, English textbook, level of difficulty, grade level

< 목 차 >

요약	i
Abstract	ii
1. 서론	1
2. 선행 연구	3
1) 이독성 측정(readability tests) 관련 연구	3
2) 고교 영어 학습 자료 및 Coh-Metrix를 활용한 주요 연구	5
3. 연구 방법	7
1) 연구 과제	7
2) 연구 대상	7
3) 연구 도구	7
4) 연구 방법	10
4. 연구 결과 및 논의	12
1) 기술통계(descriptive) 지표 비교	12
2) 이독성 공식 지표 비교	13
3) 텍스트 난이도 지표 비교	13
4) 단어 관련 지표 비교	14
5) 기존 연구와의 비교	14
6) 논의	16
5. 결론	18
1) 결론	18
2) 한계 및 제언	19
참고문헌 목록	22

<표 목차>

[표 1] 코퍼스 구축	8
[표 2] 연구에 활용된 Coh-Metrix 지표	10
[표 3] 분석 결과 전체	12
[표 4] 기술통계 지표 비교	13
[표 5] 이독성 공식 지표 비교	13
[표 6] 텍스트 난이도 지표 비교	14
[표 7] 단어 관련 지표 비교	14
[표 8] 김재은, 최인철(2015)	15
[표 9] 이승환(2013)	15
[표 10] 텍스트 난이도 지표 비교	15
[표 11] 각 학년별 평균 읽기 속도(wpm: words per minute)	16
[표 12] [참고] z-score of Text Easability measures	17

1. 서론

초·중등교육법에서는 각급 학교에서 교육과정을 운영하여야 함을 못 박고 있지만, 입시가 차지하는 비중이 크다 보니 상당수의 고등학교에서는 정규 교육과정에 따라 편성된 교과서보다도 대학수학능력시험에 맞추어 수업이 진행되는 경우가 적지 않다. 바람직한 현상이 아님에도 불구하고 대학수학능력시험(이하 수능이라 칭함)이 고등학교 수업 내용의 사실상 준거 역할을 하고 있다는 점을 교육부에서조차 인정하고 있을 정도로 수능의 영향력은 막강하다. 수능이 고교 교육을 왜곡하는 일을 최소화하기 위해 수능 출제 원리에는 '고등학교 교육과정의 내용과 수준에 맞는 출제로 고등학교 학교교육의 정상화에 기여함'이 항상 포함되어 있지만, 해마다 시험 내용이 교육과정 수준과 동떨어져 있다는 크고 작은 논란이 있어 왔다. 특히 영어 과목 시험은 그 특성상 출제 기준이 되는 교과목(보통 고교 2, 3학년 학생을 대상으로 하는 영어 I, II 과목이 준거 과목으로 지정됨¹⁾)이 지정되어 있기는 하지만, 활용할 수 있는 지문의 범위에 대해 별도의 제한을 두고 있지 않다 보니 시험 내용이 정규 교육과정에 비해 너무 어렵다는 지적을 받아 왔다. 특히 2002년도 수능부터 9등급제가 시행된 이후, 지정된 비율대로 응시생의 등급을 구분하는 것이 시험 출제의 중요한 요소로 자리 잡음에 따라 상위권 학생의 등급 구분을 위해 문제의 난이도가 차츰 상승하는 경향이 나타났고, 모든 문제가 공개되는 수능의 특성상 이 현상이 해마다 반복적으로 나타나면서 점점 교육과정 수준과의 격차가 커지게 되었다. 실제로 수능 영어 시험의 대다수를 차지하는 읽기 영역의 경우, 1993년 첫 수능이 실시된 이래 지문의 길이가 길어지고 사용되는 어휘 수준과 글 자체의 난이도가 올라가고 있음이 지적되었다(김정렬, 2017).

수능 시험과 교육과정의 괴리는 수능 영어 독해 지문과 고교 영어 교과서 본문의 난이도를 비교함으로써 명확히 확인할 수 있는데, 2014년에 진행된 한 연구에 따르면 수능 영어 읽기 영역 지문과 고등학교 2, 3학년 학생을 대상으로 하는 영어 I, II 교과서 지문의 난이도를 GFI²⁾와 FKGL³⁾로 비교한 결과 두 지표 모두 수능 영어 지문의 수준이 영어 교과서에 비해 최소 3개 학년 이상 높은 것으로 나타났다.(김재은, 최인철, 2015). 이는 수능 대비라는 현실적인 과제에 있어서 교과서가 전혀 학습 준거로서 역할을 하지 못한다는 것을 의미한다.

교육과정에 따라 편성된 교과서의 경우, 엄격한 검수 절차를 거쳐 발간되기 때문에 학년이 올라감에 따라 학습 난이도도 점층적으로 상승하도록 설계되어 있다.

-
- 1) 2015학년도부터 매해 한국교육과정평가원이 발표하는 대학수학능력시험 시행기본계획에 명시됨
 - 2) Gunning Fog Index. 텍스트의 평균 문장 길이와 단어의 평균 음절 수를 바탕으로 한 이독성(readability) 측정 지표 중 하나. 결과값이 미국 학제의 학년과 일치되도록 설계되었다.
 - 3) Flesch-Kincaid Grade Level. GFI와 마찬가지로 문장당 평균 단어 수, 단어당 평균 음절 수를 토대로 텍스트의 이독성을 측정하는 공식이다. 이 또한 결과값이 미국 학제의 학년과 일치되도록 설계되어 있다.

영어 학습이 1년 만에 해결되는 일이 아님을 생각해 볼 때 학습자의 역량을 점진적으로 키워가는 것은 필수적이고, 이러한 점이 교과서 집필 과정에서도 충분히 반영되어 있는 것이다. 하지만 앞에서 언급했듯이 학습 준거로서 교과서의 기능이 적어도 고교 영어 교육에서는 상실된 상황에서, 학생과 교사 모두 어떤 기준으로 영어 학습을 진행하여야 하는지에 대해 뚜렷한 기준을 세우기가 쉽지 않다. 수능 지문을 고교 전체에서 학습 자료로 활용하기에는 학습 수준의 점진적 향상이라는 측면에서 바람직하지 않을뿐더러 자칫 저학년 학생들의 학습 부담이 너무 커져 학생들의 학습 의욕을 꺾을 가능성도 있다. 실제 학교 현장에서는 교과서의 부족분을 보충하기 위해 사설 출판사의 참고서나 문제집을 보조 교재로 활용하는 경우가 많지만, 이 또한 학년별 학습 준거로 활용하기에는 교재의 내용과 구성이 자주 바뀌는데다가 공교육에서 사설 교재를 기준으로 삼는다는 것 자체가 적절하지 않다는 지적도 존재한다.

이와 같은 현실에서 다수의 학교에서는 평균 3개월 단위로 실시되는 전국연합학력평가 영어 영역 문제를 학습 보조 수단으로 활용하고 있다. 전국연합학력평가(이하 학평으로 칭함)는 학년별로 문제가 출제되고 시도교육청이 평가를 주관하기 때문에 지문의 수준이 상대적으로 안정적으로 유지된다는 평을 받고 있다. 본 연구에서는 학평의 이러한 속성이 고등학교 영어 학습의 준거 역할을 하기에 비교적 적절하다는 판단 아래, 학평 영어 읽기 지문의 난이도를 전산언어학적으로 분석하여 학년별 계열성을 실증적으로 확인하고자 한다. 고교 3학년 학생을 대상으로 하는 수능 모의평가 및 수능 시험 지문은 그 동안 다양한 방법으로 비교분석이 시도되면서 실증적 데이터가 많이 축적되어 있지만, 고등학교 1, 2학년 학생들을 대상으로 하는 학평은 상대적으로 주목도가 떨어져 충분한 연구가 이루어져 있지 않다. 학습의 계열성·연속성이 학습 성취감 및 흥미 유지에 중요한 역할을 한다는 점에서 학평 지문을 코퍼스화하여 분석 대상에 포함시켜 고교 3년 과정에서 영어 학습의 전체적인 흐름을 찾는 것을 본 연구의 핵심 과제로 삼고자 한다. 이러한 과정을 통해 경험적으로 이루어져 오던 고등학교 영어 학습 자료 구성 및 선정에 언어학적으로 의미 있는 근거를 마련하고, 이를 토대로 학습자의 학년과 학습 성취 수준 등을 종합적으로 고려하여 보다 체계적이고 단계적인 학습 구성이 가능할 것으로 기대된다.

분석 시 활용할 코퍼스 분석 툴로는 기존 연구 조사 과정에서 영어 L2 학습에 활용되는 각종 읽기 자료를 연구하는데 가장 폭넓게 활용된 Coh-Metrix를 이용하고자 한다. 수능 및 고교 교과서 분석에 이미 다수 활용되어 성능이 검증된 툴을 채택하여 연구의 신뢰성을 제고하는 한편, 동일한 툴을 활용함으로써 과거 연구 결과와의 비교도 쉽게 가능할 것으로 기대된다.

2. 선행 연구

1) 이독성 측정(readability tests) 관련 연구

텍스트의 이독성을 측정하려는 시도는 과거로부터 다양한 목적에 의해 있어 왔다. 보통 이독성은 텍스트의 어휘, 문법, 글의 구조 등과 관련된 내용적 측면과 글꼴, 자간, 줄간격, 여백 등과 관련된 형식적 측면으로 나누어 연구되어 왔는데, 실제 피험자를 대상으로 측정하고자 하는 요소에 맞는 텍스트를 실험 자료로 한 이독성 조사(readability survey)가 전통적인 이독성 측정 방식으로 활용되어 왔다(William H. DuBay, 2004).

이독성에 대한 다양한 정의도 제시되어 왔다. Dale and Chall (1948)은 이독성을 '독자가 텍스트를 통해 얻을 수 있는 효용에 영향을 주는 텍스트의 모든 요소와 그 요소들 간의 상호작용의 합'으로 정의하였는데, 여기에서 언급된 효용은 텍스트에 대한 이해, 적절한 소요 시간 및 텍스트를 통해 얻을 수 있는 흥미를 모두 포함하는 용어로서 사용되었다. Bhagoliwal (1961)은 '이독성에서 가장 일반적으로 강조되는 요소로서 식별성(legibility), 흥미(interest), 읽고 이해함의 용이성(ease of reading and understanding)이 있으며, 이 세 요소는 상호 연관되어 있다'고 정의하였다. 구체적인 정의는 연구자마다 다르지만, 공통적으로 '읽고 이해하기 쉬운 정도'를 의미한다고 볼 수 있다.

'읽고 이해하기 쉬운 정도'는 독자에 따라 매우 주관적일 수 있는 요소이다. 독자의 교육 수준 및 성장 환경, 생활양식 등에 따라 같은 텍스트에 대해서도 이독성이 천차만별로 갈라질 수 있기에, 이독성의 이러한 주관적인 요소를 가급적 줄이기 위해서 보편적인 기준에 따라 이독성을 측정하고자 하는 노력이 있어 왔다. Sherman (1893)은 과거에 비해 영어 문장 당 단어 개수가 감소하고 있음을 지적하면서, 짧은 문장과 구체적인 용어가 독자로 하여금 글을 더 쉽게 이해할 수 있도록 한다고 주장했다. 이와 같은 연구가 축적됨에 따라 텍스트의 음절, 단어, 문장 개수를 기초로 한 이독성 공식(readability formulae)이 다양하게 제시되어 왔는데, 이러한 공식들이 문서나 문학 작품의 대략적인 이독성을 측정하는데 매우 효과적으로 활용되어 왔다.

단어와 문장 길이를 활용한 이독성 공식의 대표적인 예로 FRE, FKGL, GFI 등을 들 수 있다. Flesch (1948)는 문장당 평균 단어 개수와 단어당 평균 음절 개수를 바탕으로 보통 0에서 100 사이의 값을 갖는 Flesch Reading Ease(FRE) 공식을 제안했다. 이후 Kincaid (1975)는 FRE 공식의 가중치를 조정하여 미군에서 각종 기술 매뉴얼 작성 시 표준으로 활용하게 된 Flesch-Kincaid Grade Level(FKGL)을 개발하였다. FKGL은 이후 펜실베이니아 주를 비롯한 미국 다수의 주에서 법률 서류나 보험 약관 등을 작성할 때 이독성을 측정하는 표준 지표로

활용되고 있다(McClure, 1987).

단어와 문장 길이를 바탕으로 한 이독성 공식은 비교적 간단한 정보를 토대로 효율적으로 텍스트의 이독성을 측정하고 이를 학교 수업, 공문서 작성 등에 활용할 수 있었다. 하지만 본질적으로 '긴 단어와 긴 문장은 짧은 단어와 문장에 비해 이해하기 어렵다'는 가정을 바탕으로 만들어진 공식들인 만큼, 텍스트의 이독성을 좌우하는 다양한 요소들에 대한 고려가 충분히 이루어졌다고 보긴 어렵다(McNamara et al., 2012; Zakaluk & Samuels, 1988). 이러한 점을 보완하고자 텍스트의 구조적 측면이 이독성에 미치는 영향에 대해 연구가 진행되어 왔다. Kintch (1981)는 텍스트의 응집성(coherence)과 이독성의 관계성에 관해 연구하였고, Meyer (1982)는 텍스트 내용에 대한 전체적인 흐름을 제시하는 것이 독자들로 하여금 텍스트를 이해하고 평가하는데 큰 도움을 준다는 것을 밝혔다. Armbruster (1985)는 텍스트에 대한 이해를 좌우하는 가장 중요한 요소가 텍스트의 응집성임을 밝히고, 응집성을 광역(global) 응집성과 국소(local) 응집성으로 나누어 설명하였다. 여기서 광역 응집성은 텍스트 전체에서 공유되는 상위 차원의 아이디어를 모두 포괄한 것이고, 국소 응집성은 문장 내 또는 문장 간을 긴밀하게 연결하는 아이디어를 가리킨다.

이처럼 응집성에 대한 정의도 가독성만큼이나 다양하지만, 『Cohesion in English』로 텍스트 응집성 연구에 이정표를 세운 Halliday & Hasan (1976)은 담화의 한 부분에 대한 해석이 다른 부분의 해석에 의존하여 이루어질 때 응집력(cohesion)이 발생한다고 주장하였다. 하지만 텍스트가 최종적으로 응집성을 갖기 위해서는 단순히 텍스트 내부의 응집력만으로는 안 되고, 독자가 해당 텍스트에 대한 자신만의 심적 표상(mental representation)을 형성했을 때에 비로소 가능하다(Carrell, 1982; Xi, 2010). 즉, 응집력(cohesion)은 텍스트 내부에 존재하는 것으로 직접적인 측정이 가능하지만, 응집성은 독자의 정신활동의 범위 내에 존재하는 것이기에 간접적인 방법으로만 측정이 가능하다(McNamara et al., 2014).

응집성을 직접적으로 측정할 수는 없지만, 텍스트의 응집력이 낮을수록 독자는 텍스트의 내용 사이의 끊어지거나 약한 연결고리를 직접 이어야 하기 때문에 응집력이 높은 텍스트에 비해 상대적으로 부담을 느끼는, 다시 말하자면 읽기에 더 '어려운' 것으로 나타난다(McNamara et al., 2014). 이러한 사실은 텍스트의 응집력을 적절히 측정함으로써 텍스트의 난이도를 추정할 수 있다는 근거로 볼 수 있다.

텍스트의 응집력을 측정하고자 하는 시도는 지속적으로 이루어져 왔지만, 현재 응집력 측정 목적으로 Coh-Metrix라는 툴이 비교적 널리 활용되고 있다. Coh-Metrix는 2004년 미국 멤피스 대학교 지능형시스템연구소(Institute for

Intelligent Systems)에서 개발된 웹 기반 언어분석 시스템으로, 전산언어학 분야에서 광범위하게 활용되는 어휘집(lexicons), 코퍼스, 품사분석기(Brill, 1995), 통사구문분석기(Charniak, 2000), LSA(Latent Semantic Analysis, Landauer et al., 2007) 등으로 구성된다(전문기, 2011). Coh-Metrix의 가장 큰 특징은 위와 같은 언어학적 분석 항목을 자동적으로 분석하여 줌과 동시에 텍스트의 응집성에 영향을 미치는 응집력에 대한 분석도 함께 수행된다는 점이다(McNamara et al., 2010). Coh-Metrix에서 제공되는 여러 측정치 중 본 연구에서 활용된 값에 대한 구체적인 설명은 추후 연구 도구 설명부에서 상세히 서술하였다.

2) 고교 영어 학습 자료 및 Coh-Metrix를 활용한 주요 연구

고교 영어 교과서, 수학능력시험 지문 등을 언어학적으로 접근하여 분석하고자 하는 시도는 다양하게 이루어져 왔다. 김재은, 최인철 (2015)의 논문에서는 고등학교 영어 I, II 교과서, EBS 수능 연계 교재 및 대학수학능력시험 지문을 각각 코퍼스로 구축하여 난이도를 비교 분석하였는데, 전통적인 이독성 지수인 FKGL(Flesch-Kincaid Grade Level)을 기준으로 볼 때 고교 교과서(평균 7.5) < EBS 교재(평균 10.1) < 수학능력시험(평균 10.5) 순으로 난이도가 상승하는 것으로 나타났고, 특히 고교 교과서와 EBS 교재 간의 난이도 차이가 크게 나타났다. 이에 상응하는 결과로 참조적 응집력(Referential Cohesion: 지시하는 단어 간의 연결고리에 근거한 응집력)의 경우 수학능력시험 < EBS 교재 < 고교 교과서 순으로 높아지는 것으로 나타났는데, 이는 응집력이 텍스트 난이도에 영향을 미치는 사례로 볼 수 있다.

고나은, 신정아(2017)의 논문에서는 EBS-수능 연계 정책 전후로 수능 영어 영역 읽기 지문의 난이도가 유의미한 수준으로 변화하였는지를 Coh-Metrix를 통해 비교 분석하였는데, 단어 수와 문장 길이는 연계 정책 이전에 비해 이후 (2010년 실시 시험부터 해당)가 다소 증가한 것으로 나타났고, FKGL 지수도 약간 상승했지만, 응집력 지표에서는 뚜렷한 차이가 나타나지 않았다.

김정렬(2017) 및 문지현, 김해동(2017)의 논문에서는 2000년대 초반의 수능 영어 지문에 비해 2010년대의 지문의 난이도가 상승한 것으로 나타났다. 다만 김정렬(2017)에서는 EBS 연계 정책 이후 수능 지문의 난이도가 오히려 하락한 것으로 분석하여 고나은, 신정아(2017)의 분석과 맥을 달리 하였다.

성일호(2014)의 논문에서는 Coh-Metrix를 활용하여 고등학교 영어 교과서의 각 단원별 지문을 분석하였는데, 이독성 공식을 기준으로 볼 때 대체적으로 단원 배열이 난이도 순서와 상관없이 이루어진 것으로 나타났다.

이외에도 Coh-Metrix는 영어 독해 지문 및 영어 교과서를 비교 분석하는데 다양하게 활용되었는데, 국내의 대표적인 예로 수능 영어 문항 간 응집력과 어

휘 정보를 비교 분석한 연구(최민주, 김정렬, 2017), 초등학교에서 중학교로 넘어가면서 영어 교과서의 난이도가 급상승하여 학생들의 학습 의욕을 떨어뜨릴 수 있다는 연구(김정렬, 2014), 수능과 각종 공인 영어 시험(TEPS, TOEIC, EIKEN)의 독해 지문의 응집성을 비교 분석한 연구(Kim, S., 2016) 등이 있다. 유사한 목적으로 해외에서도 Coh-Metrix를 활용한 연구가 진행되었는데, 주로 영어를 L2로 사용하는 국가에서 많이 진행되었다. 그 사례로는 중국 전문대학(vocational college) 영어 교재 2종을 Coh-Metrix로 비교 분석한 연구(Zhang, R., 2016), 유럽언어기준 B2 수준의 서로 다른 두 영어 시험(Russian State Exam, Cambridge First Certificate)을 Coh-Metrix로 비교 분석한 연구(Solnyshkina, M.I. et al., 2014), 1999년부터 2009년까지 스웨덴 영어 쓰기 시험의 응시생 답안을 Coh-Metrix로 비교 분석한 연구(Törnqvist, J., 2015) 등이 있다.

3. 연구 방법

1) 연구 과제

- ① 전국연합학력평가 및 수능 모의평가, 수능 영어 과목의 읽기 지문의 수준이 각 학년별로 계열성과 연속성을 띄는가?
- ② 만약 계열성과 연속성이 있다면, 텍스트의 어떤 요소가 그러한 차이를 불러 일으켰는가?
- ③ 기존 연구 결과와 비교했을 때, 본 연구에서 활용된 지문의 수준이 기존 연구에서 활용된 영어 교과서의 수준과 어느 정도 차이를 보이는가?

2) 연구 대상

고등학교 각 학년별 전국연합학력평가 영어 영역 독해 지문을 대상으로 하되, 각 학기의 말에 시행되는 학평만을 대상으로 하여 시험 간 계열성을 확인하기 위한 충분한 시간 간격을 확보한다. 고등학교 3학년의 경우 전국연합학력평가 대신 수능 모의평가 및 수능 지문을 활용한다. 각 학년별 비교를 위한 최소한의 표본 확보를 위해 최근 3개년도 지문을 활용한다. 각 학평의 지문 중 문제의 유형에 따라 의도적으로 만들어진 빈칸이나 오류는 모두 보완하여 완전한 텍스트로 정제하고, 시험지의 발문과 선택지는 모두 제외하여 코퍼스를 구축한다. 코퍼스의 구체적인 목록은 [표 1]과 같다.

3) 연구 도구

본 연구에서는 각 학년별로 수집된 영어 영역 독해 지문을 다각적으로 분석하기 위해 Coh-Metrix(코메트릭스)를 활용하였다. Coh-Metrix는 미국 멤피스대학교 지능형시스템연구소(the Institute for Intelligent System)에서 개발한 웹 기반 언어분석 시스템으로(Graesser, McNamara, Louwerse, & Cai, 2004), 다양한 유형의 텍스트를 대상으로 광범위한 언어학적 측정치를 분석, 제공한다(Graesser, McNamara, & Kulikowich, 2011).

Coh-Metrix는 현재 전산언어학 분야에서 광범위하게 활용되는 어휘집(lexicons), 코퍼스, 품사분석기(Brill, 1995), 통사구문분석기(Charniak, 2000), LSA(Latent Semantic Analysis, Landauer, 2007) 등으로 구성되며(전문기, 2011), 제공되는 측정치는 Text Easability, Referential Cohesion, Lexical Diversity 등 총 11개 섹션으로 구분된다.

본 연구에서는 Coh-Metrix에서 제공되는 총 106개 측정치 중, 텍스트의 난이도와 밀접한 관련이 있고 시험 텍스트 분석에 유용하다고 판단되는 것만을 활용하였다. 연구에서 활용한 측정치에 대한 설명은 아래와 같다. 괄호는 해당 측

정치에 대해 Coh-Metrix에서 부여한 코드명이다. 본 연구에서 활용한 전체 지표에 대한 목록은 [표 2]와 같다.

[표 1] 코퍼스 구축

연도	코퍼스 번호	대상 학년	시행 월	시험 명칭
2015	15A1	1	6	전국연합학력평가
	15B1	1	11	전국연합학력평가
	15A2	2	6	전국연합학력평가
	15B2	2	11	전국연합학력평가
	15C1	3	6	수능 모의평가
	15C2	3	11	대학수학능력시험
2016	16A1	1	6	전국연합학력평가
	16B1	1	11	전국연합학력평가
	16A2	2	6	전국연합학력평가
	16B2	2	11	전국연합학력평가
	16C1	3	6	수능 모의평가
	16C2	3	11	대학수학능력시험
2017	17A1	1	6	전국연합학력평가
	17B1	1	11	전국연합학력평가
	17A2	2	6	전국연합학력평가
	17B2	2	11	전국연합학력평가
	17C1	3	6	수능 모의평가
	17C2	3	11	대학수학능력시험

① 기술통계(descriptive)

각 코퍼스의 총 문장 개수(DESSC), 총 단어 개수(DESWC), 단어 1개당 평균 글자 수(DESWLit)를 활용하였다. 문장 단위와 단어 단위는 각각 OpenNLP sentence splitter와 Charniak parser로 구분되었다.

② 이독성 지표(readability)

Flesch Reading Ease(RDFRE)와 Flesch-Kinkaid Grade Level(RDFKGL)을 활용하였다. 두 지표는 모두 주어진 텍스트의 전체 단어 수, 전체 문장 수, 전체 음절 수에 서로 다른 가중치를 부여하여 텍스트의 난이도를 정량화한다. FRE는 보통 0에서 100 사이의 값을 가지며, 값이 작을수록 텍스트의 난이도가 높아짐을 의미한다. FKGL은 보통 0에서 12 사이의 값을 가지며, 미국 학제의 학년과 그 수치가 연동되도록 설계되어 있다.

③ 텍스트 난이도 지표(Text Easability)

Coh-Metrix 연구진이 미국 학제 상 유치원부터 12학년에 걸쳐 제공될 수 있는 총 37,520개의 텍스트를 TASA 코퍼스(Touchstone Applied Science Associates Corpus)⁴⁾에서 추출하여 이를 대상으로 주요인분석(Principal component analysis)을 통해 확보한 8개 척도이다. Coh-Metrix에서는 입력 텍스트에 대해 각 척도별로 준거 코퍼스에 비교했을 때의 상대적인 위치를 z-점수와 백분위로 산출한다. z-점수와 백분위 모두 그 값이 클수록 텍스트의 난이도가 낮음을 의미하는데, 예를 들어 백분위 값이 80으로 나왔다면, 분석 대상 텍스트가 준거 텍스트의 80%보다 쉬움을 의미한다(McNamara et al., 2014). 각 척도에 대한 설명은 아래와 같다.

- 서사성(narrativity)(PCNAR) : 텍스트의 내용이 독자에게 친숙한 일상생활, 사건, 공간, 인물 등에 대한 서사적인 성격이 강한 정도를 나타낸다.
- 통사적 평이성(syntactic simplicity)(PCSYN) : 텍스트의 문장 구조가 단순하고 평이한 정도를 나타낸다.
- 단어 구체성(word concreteness)(PCCNC) : 텍스트의 단어 중 구체적인 심상을 쉽게 불러일으키는 단어의 포함 정도를 상대적으로 나타낸다.
- 참조적 응집력(referential cohesion)(PCREF) : 단어, 구절, 어간 등이 인접 문장 및 텍스트 전체적으로 얼마나 많이 중첩되는지를 상대적으로 나타낸다.
- 논리적 응집력(deep cohesion)(PCDC) : 텍스트에 인과 관계를 알리는 접속사 또는 접속어구가 얼마나 포함되어 있는지를 상대적으로 나타낸다.
- 동사 응집력(verb cohesion)(PCVERB) : 텍스트 내에서 동사의 중첩도를 나타낸다.
- 연결성(connectivity)(PCCONN) : 각종 접속·연결 어구가 얼마나 활용되었는지를 나타낸다.
- 시제성(temporality)(PCTEMP) : 시제나 상(aspect) 등이 얼마나 일관되게 유지되는지를 나타낸다.

본 연구에서는 기존 연구와의 비교를 위해 백분위 수치를 활용하였다.

④ 단어 관련 지표

타입-토큰 비율은, 텍스트의 타입-토큰 비율을 내용어(content words)에 한정하여 구한 값(LDTTRc)과 모든 단어에 대해 구한 값(LDTTRa)을 활용하였다. 단어의 난이도와 관련한 척도로는 CELEX Lexical Database(Baayen, Piepenbrock, & Gulikers, 1995)의 영어 부분 자료를 바탕으로, 주요 단어의 출현도를 분석한 측정치(WRDFRQc, WRDFRQa)와, 내용어의 친숙성(familiarity)

4) TASA Corpus : 미국 1~12학년 수준의 다양한 장르의 교육용 텍스트를 모은 코퍼스. 약 38000여 개의 문서로 구성되어 있다(Crossley et al, 2017).

을 측정한 지표(WRDFAMc, 매일 접하는 단어에는 7, 거의 접하지 못하는 단어는 1로 두고 단어별로 1~7점 사이의 점수를 부여하여 평균을 구한 뒤 100을 곱한 값), 의미성(meaningfulness – 수치가 낮을수록 추상적인 의미가 강하다)을 나타낸 지표(WRDMEAc)를 활용하였다.

[표 2] 연구에 활용된 Coh-Metrix 지표

구분	코드명	설명
기술통계 (descriptive)	DESSC	총 문장 개수
	DESWC	총 단어 개수
	DESWLit	1개 단어 당 문자 개수
이독성 (readability)	RDFRE	Flesch Reading Ease
	RDFKGL	Flesch-Kincaid Grade Level
텍스트 난이도 (text easability)	PCNARp	서사성(narrativity) 상대지표(z-점수/백분위)
	PCSYNp	통사적 평이성(syntactic simplicity) 상대지표
	PCCNCp	단어 구체성(word concreteness) 상대지표
	PCREFp	지시적 응집력(referential cohesion) 상대지표
	PCDCp	논리적 응집력(deep cohesion) 상대지표
	PCVERBp	동사 응집력(verb cohesion) 상대지표
	PCCONNp	연결성(connectivity) 상대지표
	PCTEMPp	시제성(temporality) 상대지표
단어 관련 지표	LDTTRc	타입-토큰 비율 (내용어 대상)
	LDTTRa	타입-토큰 비율 (모든 단어 대상)
	WRDFRQc	내용어 대상 CELEX 단어 출현 빈도
	WRDFRQa	모든 단어 대상 CELEX 단어 출현 빈도(로그)
	WRDFAM	내용어 친숙성(familiarity) 평균치
	WRDMEAc	내용어 의미성(meaningfulness) 평균치

4) 연구 방법

각 코퍼스별로 Coh-Metrix 분석을 실시하여 얻은 106개 지표의 결과치를 각 학년별로 수집한 후 평균치를 각각 산출하여 비교하였다. 일차적으로 탐색적 자료 분석(exploratory data analysis)을 실시하여 각 지표별 결과치의 전체적인 흐름과 특성을 파악한 후, 비교 과정에서 평균치의 차이가 유의미한지 검증하기 위해 일원 분산분석(one-way ANOVA)을 실시하였다. 분산분석 결과 p-value가 0.05 이하로 나타난 척도에 대해서는 Tukey-HSD 검정(Tukey's Honest Significant Difference test)을 사후 분석으로 실시하였다.

분산분석 및 사후 분석의 결과를 바탕으로 각 학년별 연계성을 검증하고, 각 척도별로 나타난 차이가 의미하는 바가 무엇인지를 다양한 기존 연구 성과를

바탕으로 해석하였다. 또한, 기존 연구에서 영어 교과서를 대상으로 한 Coh-Metrix 분석 결과를 본 연구와 비교하여 영어 교과서와 본 연구의 대상 텍스트 간의 차이점을 분석하였다.

4. 연구 결과 및 논의

각 지문 코퍼스별로 주요 척도를 분석한 후, 학년별로 묶어 일원 분산분석(one-way ANOVA)을 시행하여 각 학년별 차이의 유의성을 검증하였다. 분석 결과 표는 [표 3]과 같다. (평균치, 괄호 안은 표준편차)

[표 3] 분석 결과 전체

코드명	1학년	2학년	3학년	p-value
DESSC	218.8(11.79)	233.7(25.36)	247.2(7.03)	0.03
DESWC	3389(178.73)	3629(147.39)	3991.7(121.65)	0.00
DESWLIt	4.60(0.06)	4.65(0.07)	4.81(0.1)	0.00
LDTTRc	0.61(0.02)	0.62(0.02)	0.61(0.01)	0.65
LDTTRa	0.36(0.01)	0.37(0.01)	0.36(0.01)	0.22
PCNARp	46.6(2.94)	40.4(7.18)	33.3(4.8)	0.00
PCSYNp	64.1(6.45)	62.5(5.74)	66.6(3.17)	0.41
PCCNCp	57.7(7.63)	52.9(8.27)	43.3(6.85)	0.02
PCREFp	18.0(1.41)	14.9(3.17)	13.1(1.46)	0.01
PCDCp	78.8(7.68)	72.1(4.24)	72.3(7.44)	0.17
PCVERBp	48.3(10.91)	45.8(2.41)	39.9(6.24)	0.16
PCCONNp	1.68(1.44)	3.35(1.46)	1.32(0.9)	0.03
PCTEMPp	52.5(12.92)	48.6(5.98)	52.5(5.99)	0.69
WRDFRQc	2.26(0.06)	2.21(0.04)	2.16(0.04)	0.01
WRDFRQa	2.98(0.03)	2.95(0.01)	2.93(0.02)	0.00
WRDFAM	574.87(2.42)	572.38(1.86)	569.79(1.91)	0.00
WRDMEAc	436.5(2.43)	433.2(4.68)	429.8(3.33)	0.02
RDFRE	64.2(3.34)	62.2(3.5)	56.6(3.56)	0.01
RDFKGL	8.15(0.7)	8.47(0.8)	9.37(0.59)	0.02

1) 기술통계(descriptive) 지표 비교

전국연합학력평가, 수능 모의평가 및 수능은 모두 동일한 양식(문제 유형, 문제 배치 순서, 문재 개수가 모두 동일함)으로 출제되기 때문에, 주어진 텍스트의 길이가 난이도에 영향을 미친다. 각 학년별로 단위 시험 당 양적 비교를 한 결과는 [표 4]와 같다.

[표 4] 기술통계 지표 비교

코드명	1학년	2학년	3학년	p-value
DESSC	218.8(11.79)	233.7(25.36)	247.2(7.03)	0.03
증감률	-	▲ 6.8%	▲ 5.8%	-
DESWC	3389(178.73)	3629(147.39)	3991.7(121.65)	0.00
증감률	-	▲ 7.1%	▲ 10.0%	-
DESWLt	4.60(0.06)	4.65(0.07)	4.81(0.1)	0.00
증감률	-	▲ 1.1%	▲ 3.4%	-

각 학년별로 총 문장 수, 단어 수, 단어당 문자 수를 비교한 결과, 세 지표 모두 학년이 올라가면서 그 수치가 증가하였으며, 학년 간 차이가 유의미($p<0.05$)한 것으로 나타났다. 또한, 1학년과 2학년의 차이보다 2학년과 3학년의 차이가 더 큰 것으로 나타나 학년이 올라갈수록 텍스트의 절대량이 더 빠르게 증가하는 것을 알 수 있다.

2) 이독성 공식 지표 비교

연구 대상 코퍼스의 FRE(Flesch Reading Ease)와 FKGL(Flesch-Kincaid Grade Level) 결과값의 학년별 평균치는 [표 5]와 같다.

[표 5] 이독성 공식 지표 비교

코드명	1학년	2학년	3학년	p-value
RDFRE	64.2(3.34)	62.2(3.5)	56.6(3.56)	0.01
증감률		▼ 3.1%	▼ 9.0%	
RDFKGL	8.15(0.7)	8.47(0.8)	9.37(0.59)	0.02
증감률		▲ 3.9%	▲ 10.6%	

FRE는 그 값이 작을수록 난이도가 높음을 뜻하고, FKGL은 값이 클수록 난이도가 높음을 의미한다. 4-1)에서와 마찬가지로, 두 지표 모두 1학년과 2학년 사이보다 2학년과 3학년 사이의 간극이 크게 나타났다.

3) 텍스트 난이도 지표 비교

Coh-Metrix에서 제공하는 텍스트의 상대적 난이도 측정 지표의 학년별 분석 결과는 [표 6]과 같다.

8개 상대지표 중 학년 간 차이가 유의미하게 나타난 척도는 총 4개이다. 이 중 텍스트의 연결성을 나타내는 PCCONNp 척도의 경우, 그 차이가 극단적이고, 백분위 계산의 근거가 되는 z-점수 척도인 PCCONNz의 경우 $p=0.06$ 으로, 학년 간의 유의미한 차이가 없는 것으로 나타나, 실제 유의미한 차이를 보인 척도는 총 3개로 볼 수 있다.

[표 6] 텍스트 난이도 지표 비교

코드명	1학년	2학년	3학년	p-value
PCNARp	46.6(2.94)	40.4(7.18)	33.3(4.8)	0.00
증감률	-	▼ 13.3%	▼ 17.5%	-
PCCNCp	57.7(7.63)	52.9(8.27)	43.3(6.85)	0.02
증감률	-	▼ 8.3%	▼ 18.0%	-
PCREFp	18.0(1.41)	14.9(3.17)	13.1(1.46)	0.01
증감률	-	▼ 17.0%	▼ 12.0%	-
PCSYNp	64.1(6.45)	62.5(5.74)	66.6(3.17)	0.41
PCDCp	78.8(7.68)	72.1(4.24)	72.3(7.44)	0.17
PCVERBp	48.3(10.91)	45.8(2.41)	39.9(6.24)	0.16
PCCONNp	1.68(1.44)	3.35(1.46)	1.32(0.9)	0.03
PCTEMPp	52.5(12.92)	48.6(5.98)	52.5(5.99)	0.69

세 척도 모두 학년이 높아짐에 따라 백분위값이 평균 10% 이상의 비교적 큰 폭으로 하락하였다. 이에 비해 나머지 다섯 척도는 학년 간의 유의미한 차이가 나타나지 않아 고교 전체적으로 비슷한 수준을 유지하는 것으로 보인다.

4) 단어 관련 지표 비교

단어 관련 지표의 분석 결과는 [표 7]과 같다.

[표 7] 단어 관련 지표 비교

코드명	1학년	2학년	3학년	p-value
LDTTRc	0.61(0.02)	0.62(0.02)	0.61(0.01)	0.65
LDTTRa	0.36(0.01)	0.37(0.01)	0.36(0.01)	0.22
WRDFRQc	2.26(0.06)	2.21(0.04)	2.16(0.04)	0.01
WRDFRQa	2.98(0.03)	2.95(0.01)	2.93(0.02)	0.00
WRDFAM	574.87(2.42)	572.38(1.86)	569.79(1.91)	0.00
WRDMEAc	436.5(2.43)	433.2(4.68)	429.8(3.33)	0.02

우선, Type-token 비율은 내용어 대상과 전 단어 대상 모두 각 학년별로 유의미한 차이가 없는 것으로 나타났다. CELEX 데이터베이스를 바탕으로 한 주요 단어 출현 빈도 지표는 내용어와 전 단어 모두 학년이 올라감에 따라 그 값이 조금씩 하락하였고, 단어 친숙성 및 의미성 지표도 비슷한 양상으로 나타났다. 보다 명확한 비교를 위해 $p < 0.05$ 인 네 개 척도에 대해 Tukey post-hoc test를 실시했는데, 그 결과 모든 척도에서 1-2학년, 2-3학년 간의 차이는 무의미하지만 1-3학년 간의 차이는 유의미한 것으로 분석되었다.

5) 기존 연구와의 비교

Coh-Metrix를 활용한 연구 중 김재은, 최인철(2015)과 이승환(2013)에서 고등학교 영어 교과서(영어 I, 영어II)의 난이도 분석 결과치를 인용하여 본 연구의 학년별 시험 지문 분석 결과와 비교하였다.

- 이독성 지수 비교

[표 8] 김재은, 최인철(2015)

코드명	영어 I	영어II	평균	p-value
RDFRE	67.22	64.90	65.56	0.27
RDFKGL	7.27	7.69	7.50	0.41

[표 9] 이승환(2013)

코드명	영어 I	영어II	평균	p-value
RDFRE	69.28	65.84	67.56	0.06
RDFKGL	7.02	7.68	7.35	0.04

두 연구 모두 영어 I 교과서는 FRE 지수는 70을 약간 밑도는 수준, FKGL은 7 점대 초반으로 나타났고, 영어II 교과서는 각각 60대 중반, 7점대 후반으로 나타났다. 위 결과를 본 연구 결과와 비교해 보면, 영어 I 은 물론 영어II 교과서조차 두 이독성 지수에서 고등학교 1학년 수준만큼의 난이도(FRE 64.2, FKGL 8.15)도 되지 못함을 알 수 있다. 영어 I 과 II가 각각 고등학교 2학년과 3학년 학생들을 대상으로 하는 과목임을 고려했을 때, 시험을 통해 요구되는 영어 과목의 성취 수준에 비해 고등학교 영어 교과서가 난이도 측면에서 크게 뒤쳐져 있음이 뚜렷이 드러난다.

이승환(2013)의 연구에서는 이독성 지표 외에도 Coh-Metrix에서 제공하는 다른 척도의 결과치도 제시되어 있는데, 이 중 본 연구에서도 활용된 척도의 값은 [표 10]과 같다.

[표 10] 텍스트 난이도 지표 비교

코드명	이승환(2013)			본 연구
	영어1	영어2	p-value	1학년
PCNARp	53.44(21.84)	50.02(25.41)	0.42	46.6(2.94)
PCSYNp	67.79(14.18)	64.18(14.59)	0.17	64.1(6.45)
PCCNCp	62.62(24.76)	59.51(22.77)	0.47	57.7(7.63)
PCREFp	32.17(18.99)	27.20(17.39)	0.13	18.0(1.41)

위 척도는 백분위 값이 클수록 난이도가 쉽다는 뜻이기 때문에(예를 들어 80인 경우 분석 대상 텍스트가 준거 텍스트의 80%보다 더 쉽다는 의미), 비교 결과 통사적 평이성 척도를 제외한 나머지 척도의 값이 영어 I, II 모두 고등학교 1학년 지문 평균치보다 낮았다. 이는 단어 수, 문장 수, 음절 수라는 텍스트의

양적 측면에서뿐만 아니라 텍스트의 서사성, 응집성 등 질적인 측면에서도 교과서의 난이도보다 시험의 난이도가 더 높은 것으로 해석할 수 있다.

6) 논의

위 결과를 종합해 볼 때, 전국연합학력평가, 수능 모의평가 및 수능의 영어 읽기 영역 지문은 외형적으로는 지문의 길이(단어 수)가 길어지고, 내용적으로는 텍스트의 서사성과 단어 구체성이 감소하면서 상대적으로 추상적인 내용이 증가하는 것으로 추론해 볼 수 있다. 또한 참조적 응집력이 감소함에 따라 독자의 읽기 부담이 커지고, 독자의 배경지식의 활용도가 상대적으로 높아진다고 추론할 수 있다(McNamara et al., 2014).

연구 대상으로 활용된 시험은 문제 개수, 유형, 시험지상 배치 순서가 모두 동일하기 때문에, 텍스트의 길이가 난이도에 많은 영향을 끼친다. 응시자가 읽기 영역 문제에 할애할 수 있는 시간은 전체 시험 시간 70분 중 듣기 영역 방송 시간 20분을 제외한 50분 정도이다. 본 연구에서는 시험지의 모든 발문과 선택지를 제외한 순수 지문만을 대상으로 삼았기 때문에, 지문을 읽은 뒤 실제 문제를 풀고 답안지를 작성하는 시간을 고려한다면 적어도 40분 이내에 주어진 텍스트를 모두 읽어야 한다는 추산이 가능하다. 각 학년별 총 단어 개수 평균치를 40으로 나누어 요구되는 읽기 속도(reading rate)를 [표 11]과 같이 얻을 수 있다.

[표 11] 각 학년별 평균 읽기 속도(wpm: words per minute)

구분	1학년	2학년	3학년
wpm	84.73	90.73	99.79
조정된 wpm	108.93	116.65	128.30

기존 연구에 따르면 영어 L1 화자가 일반적인 수준으로 묵독(silent reading)을 할 때 평균적으로 250~300wpm 정도의 읽기 속도를 보이는 것으로 알려져 있다(Carver, 1990; Rayner, 1998). 또한, Nation(2005)에 따르면 텍스트를 읽은 후 그에 대한 이해도 측정(comprehension test) 과정에서 정답률이 70% 이하일 경우 읽기 속도를 줄일 필요가 있다는 지적을 고려해 볼 때 평균적인 읽기 속도를 유지하는 L1 화자는 약 70% 정도 텍스트의 내용을 올바르게 이해한다고 추정할 수 있다.

시험 지문이라는 특성상 텍스트의 내용을 70%만 이해하여서는 주어진 문제를 풀 수 있다는 보장을 할 수 없기 때문에, 90% 정도의 이해도를 기준으로 한다면 실제 요구되는 읽기 속도는 최초 구한 값에 9/7을 곱한 것으로 볼 수 있다. 이 결과를 바탕으로 볼 때, 고등학교 학생은 매해마다 평균 10wpm 정도 읽기 속도를 향상시켜야 함을 알 수 있다.

고등학교 3학년 학생에게 요구되는 최종적인 읽기 속도인 128wpm은, 일본 대학생의 평균적인 영어 읽기 속도가 65~90wpm이고(Robb & Susser, 1989; Taguchi et al, 2004) ESL 환경에서 중등학교(secondary school)와 대학교 학생들의 읽기 속도가 각각 120~150wpm, 200wpm 정도(Nuttall, 1982)라는 점을 고려해 볼 때 결코 느린 속도가 아님을 알 수 있다. 참고로 Nation(2005)은 읽기 유창성(reading fluency) 향상을 위해서는 최소 일주일에 3번 이상 연습을 해야 한다고 지적하고 있음을 언급해 둔다.

텍스트의 질적(내용적) 측면에서 특기할 점은, 서사성과 단어 구체성, 참조적 응집력은 학년이 올라감에 따라 감소하는 것으로 나타났지만, 통사적 평이성(syntactic simplicity)은 학년 간의 유의미한 차이가 나타나지 않았다는 점이다. 각종 단어 관련 지표가 학년이 올라감에 따라 조금씩이나마 값이 커지고 있음을 고려할 때, 고등학교 영어 읽기 지문의 난이도에 영향을 미치는 요소는 문장 구조보다도 어휘임을 추론할 수 있다. 서사성이 떨어진다는 것 자체가 일상적이지 않은 소재가 많이 등장한다는 의미인데, 이는 자연스레 어휘 수준의 상승을 불러일으키게 된다. 또한 통사적 평이성의 백분위 값이 학년과 관계없이 모두 60을 상회하는 것으로 나타나 연구 대상 텍스트의 문장 구조가 상대적으로 단순하다는 사실을 알 수 있는데, 이는 문법 위주의 영어 교육을 탈피하고자 하는 교육 당국의 의지가 일정 정도 반영된 것으로도 해석이 가능하다.

[표 12] [참고] z-score of Text Easability measures

코드명	1학년	2학년	3학년	p-value
PCNARz	-0.09(0.08)	-0.25(0.19)	-0.43(0.13)	0.00
PCSYNz	0.37(0.17)	0.33(0.16)	0.44(0.09)	0.43
PCCNCz	0.20(0.2)	0.07(0.22)	-0.17(0.18)	0.02
PCREFz	-0.92(0.05)	-1.05(0.13)	-1.12(0.07)	0.01
PCDCz	0.84(0.32)	0.59(0.13)	0.61(0.22)	0.17
PCVERBz	-0.04(0.28)	-0.11(0.06)	-0.26(0.16)	0.16
PCCONNz	-2.25(0.37)	-1.88(0.24)	-2.3(0.31)	0.06
PCTEMPz	0.07(0.34)	-0.03(0.16)	0.07(0.16)	0.69

마지막으로 학년 간 척도별 평균치를 비교하였을 때, 분산분석에서 $p < 0.05$ 로 나타난 척도 중 기술통계 및 이독성 지표 전체, 텍스트 난이도 지표 3개 중 2개에서 1학년과 2학년 사이의 차이보다 2학년과 3학년 간의 차이가 더 크게 나타났다. 이는 Tukey 사후 분석 결과에서도 나타났는데, $p < 0.05$ 로 나타난 32개 척도 중 1-2학년 및 2-3학년 간의 차이가 유의미하게 나타난 경우의 수가 각각 5개, 8개로 2-3학년 간의 차이가 더 두드러짐을 알 수 있었다.

5. 결론

본 연구는 대학수학능력시험 영어 영역의 난이도가 정규 교육과정 상의 영어 교과서에 비해 난이도가 높다는 기존 연구 성과를 토대로, 각 학년별로 실시되는 전국연합학력평가, 수능 모의평가 및 수능 영어 과목의 난이도가 학년에 따른 계열성과 연속성이 있는지를 읽기 영역 지문에 한정하여 살펴보았다. 각 학년별로 6개 시험 세트를 각각 코퍼스화하여 Coh-Metrix로 텍스트의 양적·질적 측면을 다각적으로 분석하고자 하였으며, 본 연구의 주요 결론과 제언은 다음과 같다.

1) 결론

전국연합학력평가, 수능 모의평가 및 수능의 영어 영역 읽기 지문은 학년 간의 난이도 차이가 뚜렷이 나타나 계열성이 있는 것으로 확인되었다. 난이도 차이가 발생한 주 원인으로는 텍스트 양적으로는 각 세트별 단어 개수의 증가가 두드러졌으며, 질적으로는 텍스트의 서사성과 참조적 응집력이 감소하고, 사용되는 단어의 추상성이 상대적으로 증가함을 찾을 수 있었다.

각 세트별 단어 개수의 증가 양상을 토대로 각 학년별로 요구되는 읽기 속도를 wpm(words per minute) 단위로 산출할 수 있었으며, 기존 연구 결과를 바탕으로 이해도 90% 수준의 읽기 속도를 추정한 결과 수능 시험을 안정적으로 치르기 위해서는 평균 130wpm 정도의 읽기 속도를 갖추어야 함을 알 수 있었다. 이는 ESL 환경의 중등학교(secondary school) 학생의 평균적인 읽기 속도인 120~150wpm의 범위에 해당되는 것(Nuttall, 1982)으로, EFL 환경에서 학습하는 한국 고등학생에게 상당히 부담이 갈 수 있는 수준으로 보인다.

텍스트의 양적 측면을 바탕으로 난이도를 측정하는 FRE와 FKGL 결과치 또한 학년이 올라감에 따라 난이도가 상승함을 나타내고 있으며, FRE는 평균 64~56 정도로 나타나 미국 학제 기준으로 대략 8~12학년 수준이고(Flesch, R., 2016), FKGL도 이와 유사하게 8~10학년 전후의 난이도임을 확인하였다.

텍스트의 질적 측면에서는 텍스트의 내용이 학년이 올라감에 따라 차츰 일상적인 것과 거리가 멀어지고, 구체적인 심상을 쉽게 떠올릴 수 있는 단어는 줄어들고 개념적이고 추상적인 단어의 출현빈도가 잦아짐을 확인할 수 있었다. 참조적 응집력이 감소함에 따라 독자의 배경지식이 텍스트 이해에 미치는 영향도 커지면서 전체적으로 난이도가 높아지는 것으로 나타났다. 다만, 문장 구조의 난이도는 고등학교 3학년 전체적으로 큰 차이가 나타나지 않아 문장보다는 단어 단위의 변화가 더 크다는 점도 확인하였다.

연구 결과를 종합적으로 보았을 때, 고등학교 3년 과정에서 실시되는 영어 시험에서 가장 강조되고 있는 부분은 어휘 능력임을 유추할 수 있다. 또한 각 척

도별 학년간 평균치 비교 및 Tukey 사후 분석을 통해, 1학년과 2학년의 수준 차이보다 2학년과 3학년의 수준 차이가 더 크게 나타남을 확인할 수 있었다.

기존 연구에서 영어 교과서를 분석한 자료와 본 연구 결과를 비교한 결과, 고등학교 3학년 학생을 대상으로 하는 영어II 교과서의 난이도가 이독성 공식 및 Coh-Metrix의 텍스트 난이도 척도 모두에서 고등학교 1학년 지문 수준에도 미치지 못하는 것으로 나타나 교과서가 고등학교 영어 학습에서 학습 준거로서의 기능을 상실한 것으로 분석되었다. 향후 수능 시험이나 영어 교육과정상의 대대적인 변동이 발생하지 않는 한, 고교 영어 학습에서 교과서의 위상은 쉽게 회복되기 어렵고, 학습자와 교수자 모두 실질적으로 요구되는 학습 성취 수준에 맞는 영어 읽기 자료를 확보하기 위한 노력도 지속되어야 할 것으로 보인다.

2) 한계 및 제언

본 연구에서는 고등학교 영어 영역 전국단위 시험의 읽기 지문을 분석하여 각 학년별 학습 성취 요구치를 간접적으로 측정하였다. 본 연구 과정에서 발견된 한계점은 다음과 같다.

첫째, 표본의 개수가 한정되어 있어 학년 간 충분한 비교가 이루어지지 않은 측면이 있다. 표본의 개수를 각 학년별 6개로 제한하였기 때문에 학년별 평균치가 뚜렷하게 차이가 나지 않은 척도가 많았다. 총 106개 척도 중 $p < 0.05$ 로 나타난 척도는 총 32개였는데, 표본의 개수를 고려했을 때 나머지 74개 척도에서 학년 간 차이가 없는 것으로 나타난 원인을 규명하기 위해서는 표본의 개수를 추가하여 연구할 필요가 있다. 실제 1, 2학년은 각 연도별로 4개의 표본을 얻을 수 있고, 3학년은 7개 표본을 얻을 수 있기 때문에 표본의 개수를 좀 더 많이 확보하였다면, 학년 간 차이의 유의미성이 달라지는 척도가 있을 가능성이 있다.

둘째, Tukey 사후 분석 결과를 해석하는 과정에 어려움이 있었다. 분산분석에서 $p < 0.05$ 로 나타난 32개 척도 중, 사후 분석에서 1학년과 3학년의 차이가 유의미하지 않은 것으로 나타난 척도가 6개나 되었고, 이 중 1개 척도는 모든 학년쌍의 차이가 유의미하지 않은 것으로 나타났다. 이는 학년 간 차이가 나타날 것이라는 일반적인 예상과 반하는 것인데, 본 연구에서는 이를 심도 있게 다루지 못하였다.

셋째, 시험 1세트 전체를 하나의 코퍼스로 둔 것이다. 현재 고등학교 영어 영역 시험은 마지막 5문항을 제외한 나머지 모든 읽기 문제에서 1지문 1문항 원칙을 유지하고 있다. 각 지문마다 제재와 주제가 다르기 때문에 코퍼스 전체를 대상으로 한 응집력은 낮게 나올 수밖에 없다. 보다 정확한 측정을 위해서는 각 지문을 분리하여 각각 분석해야 하지만, Coh-Metrix 프로그램에서는 입력 텍스트

트의 길이가 최소 200단어 이상이어야만 이독성 지표 등의 값이 신뢰할 수 있다(McNamara et al., 2014)고 명시하고 있으며, 각 지문의 평균 길이가 150단어 이하인 관계로 개별 분석을 진행하는 것이 적절하지 않다고 판단하였다. 실제 선행연구에서도 이 부분에 대한 고민이 드러나 있는데(고나은, 신정아, 2017), 어떤 방식을 채택할 것인지는 연구의 성격에 따라 달라져야 할 부분이다. 본 연구에서는 텍스트 길이 제한 이외에도 시험 전반적으로 비슷한 소재가 많이 등장할 경우 난이도가 낮아질 수 있다는 판단 아래 시험 1세트를 단일 코퍼스로 구성하였다.

넷째, 읽기 지문만을 연구 대상으로 하였다는 점이다. 본 연구에서 연구 대상으로 삼은 모든 영어 시험이 배점의 약 3분의 1 가량을 듣기 영역에 할애하고 있다. 학교 현장에서는 듣기 영역이 읽기 영역에 비해 쉽다는 의견이 우세한 것이 사실이기는 하지만, 일상생활에서 영어를 사용할 일이 없는 EFL 환경에서 영어를 학습하는 학생들에게 있어서는 읽기보다 듣기 영역이 더 부담스러울 수 있을 가능성이 있는 만큼, 듣기 영역에 대한 연구도 추후 이루어질 필요가 있다.

다섯째, 학년 간 차이를 규명하는 것 이외에 학습자와 교습자에게 필요한 실제적인 조언을 하지 못하였다. 단어 수준이 계속 상승한다는 측면을 밝히기는 했지만, 학습 자료 선정에 있어서 기존의 이독성 지수 이외에 즉각적으로 활용 가능할 만한 척도를 제공하지 못하였다. 이는 본 연구의 주 과제는 아니었지만, 연구의 효용성 측면에서 아쉬운 부분이다. 다만, 이독성 지수가 Coh-Metrix에서 제공하는 여러 다른 척도와 상당히 높은 상관관계를 갖는다는 연구 결과가 있는 만큼(McNamara et al., 2014), 본 연구에서 규명한 학년별 이독성 지수 수준을 고려하여 적절한 수준의 읽기 자료를 선정할 수는 있을 것으로 보인다.

영어 읽기 능력은 EFL 환경에서 거주하는 개개인의 학업적·직무적 역량을 크게 좌우할 수 있는 정보 습득력과 매우 밀접한 연관이 있다. 고등학교 영어 학습이 너무 독해 위주로 치우쳐져 있다는 비판이 꾸준히 있어 왔지만, 현실적인 여건상 필수 불가결한 측면이 있다고 생각된다. 다만, 지금과 같이 정규 교육과정과 실제 시행되는 시험 간의 난이도 격차가 매우 큰 현상이 지속된다면 이는 결코 바람직하다고 볼 수 없으며, 사교육을 경감하고자 하는 교육 당국의 정책 또한 실효성을 보기 어렵다. 시험에서 요구하는 읽기 수준이 EFL 환경에서는 상당히 도달하기 어렵고, 필요 독서량 또한 지금과 같은 교과서 체제로는 도저히 충족시킬 수 없는 수준이다.

교육과정과 시험에서 요구하는 학습 성취 수준에 큰 괴리가 있는 지금과 같은 상황은 결코 바람직하다고 보기 어렵다. 본 연구에서 밝힌 대로 단기적으로는 학력평가나 수능 시험이 학습의 기준으로 작용할 수는 있겠지만, 현실적으로

요구되는 수준으로 교과서의 난이도를 올리고 내용을 추가하거나 시험 난이도를 현 교육과정 수준으로 낮추는 등의 조치가 이루어지지 않는다면 앞으로도 학교 현장에서 영어 학습과 관련된 여러 혼란이 지속될 것임은 물론, 장기적으로는 영어 과목이 교과서 외의 학습 자료 및 공교육 외의 학습 기회가 상대적으로 부족한 지역 및 계층의 학생들에게 사회적 격차를 뛰어넘지 못하게 하는 유리천장과도 같은 역할을 할 우려가 있다. 이 점에 대하여 우리 사회의 깊이 있는 고민이 필요한 상황이다.

참고문헌 목록

- Armbruster, B.B., Anderson, T.H. (1985) Producing 'Considerate' Expository Text: or Easy Reading is Damned Hard Writing. *Journal of Curriculum Studies*, 17:3, 247-274
- Bahgoliwal, B.S. (1961). Readability formulae – their reliability, validity and applicability in Hindi. *Journal of Education and Psychology*, Vol XIX – No. 1, April, 13
- Brill, E. (1995). Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, 21, 543-566.
- Carrell, P.L. (1982). Cohesion Is Not Coherence. *TESOL quarterly*, 16:4, 479-488.
- Carver, R. (1990). *Reading Rate: A review of research and theory*. San Diego: Academic press.
- Chang, C-S. (2010). The effect of a timed reading activity on EFL learners: Speed, comprehension, and perceptions. *Reading in a Foreign Language*, 22(2), 284-303.
- Charniak, E. (2000). A maximum-entropy-inspired parser. *Proceedings of the 1st conference on North American chapter of the Association for Computational Linguistics*(132-139). San Francisco, California: Morgan Kaufmann Publishers.
- Crossley, SA., Dascalu, M., & McNamara, D. (2017). How important is size? An investigation of corpus size and meaning in both Latent Semantic Analysis and Latent Dirichlet Allocation. *FLAIRS 2017 - Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference*, 293-296.
- Dale, E. and Chall, J. S. (1948). A formula for predicting readability. *Educ. Res. Bull.*, 27, 11-20.
- Flesch, R. (1948). A new readable yardstick. *Journal of Applied Psychology* 32:221-33.
- Flesch, R. (1979). *How to write plain English: a book for lawyers and consumers*. Michigan: HarperCollins.
- Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. London: Longman Group Limited.
- Iwahori, Y. (2008). *Developing reading fluency: A study of extensive reading in*

- EFL. Reading in a Foreign Language, 20(1), 70-91.
- Jodai, H. (2011). Reading rate and comprehension. Non-journal.
- Kim, S. (2016), Analysing the cohesion of the reading passages on standardized English tests: CSAT, TEPS, EIKEN and TOEFL. 석사 학위 논문. 서울대학교 대학원.
- Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of new readability formulas (automated readability index, fog count, and Flesch Reading Ease formula) for navy enlisted personnel. Research Branch Report 8-75. Chief of Naval Technical Training: Naval Air Station Memphis.
- Kintsch, W. & Miller, J. R. (1981). Readability: A view from cognitive psychology. Teaching: Research reviews. Newark, DE: International Reading Assn.
- Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W. (2007). Handbook of Latent Semantic Analysis. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- McClure, G. (1987) Readability formulas: useful or useless. (an interview with J. Peter Kincaid). IEEE Transactions on Professional Communications, 30: 12-15.
- McNamara, D. S., Graesser, A., & Louwerse, M. (2012). Sources of text difficulty: Across genres and grades. Measuring up: Advances in how we assess reading ability.
- McNamara, D.S., Graesser, A.C., McCarthy, P.M., & Cai, Z. (2014). Automated evaluation of text and discourse with Coh-Metrix: Cambridge University Press.
- McNamara, D.S., Louwerse, M.M., & Graesser, A.C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. Discourse Processes, 47, 292-330.
- Meyer, B.J. (1982). Reading research and the teacher: The importance of plans. College composition and communication 33, no. 1
- Nation, P. (2005). Reading faster. PASAA, 36, 21-35.
- Nation, P. (2009). Reading faster. International Journal of English Studies, 9(2), 131-144.
- Nuttall, C. (1982). Teaching reading skills in a foreign language. London: Heinemann Education Books.
- Rayner, K. (1998). Eye movement in reading and information processing: 20 years of research. Psychological Bulletin, 124(3), 372-422.
- Robb, T.N., & Susser, B. (1989). Extensive reading vs skills building in an EFL

- context. *Reading in a Foreign Language*, 5(2), 239-251.
- Sherman, L. A. (1893). *Analytics of Literature: A manual for the objective study of English prose and poetry*. Boston: Ginn and Co.
- Solnyshkina, M.I., Harkova, E.V., & Kisel'nikov, A.S. (2014). Comparative Coh-Metrix Analysis of Reading Comprehension Texts: Unified(Russian) State Exam in English vs Cambridge First Certificate in English. *English Language Teaching*, 7(12).
- Taguchi, E., Takayasu-Maass, M., & Gorsuch, G.J. (2004). Developing reading fluency in EFL: How assisted repeated reading and extensive reading affect fluency development. *Reading in a Foreign Language*, 16, 1-23.
- Törnqvist, J. (2015). *Using Coh-Metrix to investigate changes in student texts*. Stockholm: Stockholms universitet.
- William H. DuBay (2004). *The principle of readability*. Costa Mesa: Impact Information.
- Xi, Y. (2010). Cohesion studies in the past 30 years: Development, application, and chaos. *Language Society and Culture*(31), 139-147.
- Zakaluk, B. L., & Samuels, S. J. (1988). *Readability: Its past, present, and future*. Newark: International Reading Association.
- Zhang, R., (2016). A Coh-Metrix Analysis of Two Textbooks: Successful English for Vocational Colleges and Vocational College English (An Integrated Skills Course). *US-China Foreign Language*, May 2016, 14(5), 351-356.
- 고나은, 신정아. (2017). Coh-Metrix를 이용한 수능 영어 읽기 영역 지문 난이도 비교: EBS-수능 연계 정책 전후. *Secondary English Education*, 10(4), 3-24.
- 김재은, 최인철. (2015). 고등학교 영어 교과서, EBS 수능 연계 교재, 대학수학능력시험의 코퍼스기반 난이도 비교 분석. *Multimedia-Assisted Language Learning*, 18(1), 59-92.
- 김정렬. (2014). 초중등 영어교육 연계성 문제: 영어교과서의 계열성을 중심으로. *Language Research*, 50(1). 161-184.
- 김정렬. (2017). 수능영어 읽기 지문에 대한 통시적 Coh-Metrix 분석. *초등교과교육연구* 2017, 27호, 63-78.
- 문지현, 김해동. (2017). 대학수학능력시험 영어 읽기 지문의 언어적 요소 분석. *Modern English Education*, 18(1), 193-211.
- 성일호. (2014). 이독성공식과 Coh-Metrix를 활용한 우리나라 고등학교 영어교과서 이독성 분석. *영어영문학연구*, 40(4), 299-320.
- 이승환. (2013). 코메트릭스(Coh-Metrix)를 이용한 고등학교 영어1과 영어2 교과서

- 문어 텍스트의 코퍼스 언어학적 난이도 분석. *Studies in Foreign Language Education*, 27(2), 131-148.
- 전문기. (2011). Coh-Metrix를 이용한 중학교 1학년과 2학년 개정 영어교과서 읽기 자료의 코퍼스 언어학적 연계성 분석. *언어과학연구*, 56, 210-218.
- 최민주, 김정렬. (2017). 수능 영어 문항 유형간 응집력과 어휘정보 분석. *한국콘텐츠학회논문지*, 17(12), 378-385.
- 한국교육과정평가원. (2014~2018). 대학수학능력시험 시행기본계획.
- 한문섭. (2010). 한국영어교육의 정책과 방향. 2010년 원어민 영어보조교사 교장단 직무연수. 서울: 국립국제교육원. 73-80.
- 황영민, 이경남. (2014). 고등학교 영어 교과서와 대학수학능력 영어시험의 어휘 비교. *Foreign Language Education*, 21(1), 191-214.