영어어문교육 *제30권* 2호 2024 여름
https://doi.org/10.35828/etak.2024.30.2.111

Coh-Metrix 시스템을 통한 고등학교 영어 모의시험 독해지문의 학년 내 연계성 분석: 고 3 전국연합학력평가를 중심으로

이혜진[*]
(원광대학교)
류지수[**]
(건국대학교)
전문기[***]
(건국대학교)

Lee, Hyejin, Ryu, Jisu, & Jeon, Moongee. (2024). Exploring the intra-grade continuity of high school English mock CSAT test reading passages via Coh-Metrix: Insights from the national academic achievement test for high school seniors. *English Language & Literature Teaching, 30(*2), 111-131.

This corpus-driven research aims to examine the intra-grade continuity of reading passages within high school English mock exams, specifically the national academic achievement tests for seniors in 2023. These tests were conducted quarterly by four regional educational offices: Seoul in March, Incheon in April, Gyeonggi Province in July, and Seoul in October. A corpus of 100 reading passages, 25 from each test, was analyzed via a range of Coh-Metrix measures, including word features, lexical diversity, connectives, readability, referential cohesion, syntactic complexity, semantic cohesion, and so on. These Coh-metrix results demonstrated that there was no compelling evidence to support strong intra-grade continuity, highlighting the necessity for meticulous calibration of text difficulty, which should incrementally escalate with successive test administrations. The analysis of basic metrics, lexical frequency and diversity, syntactic complexity, and readability revealed no consistent trends across test administrations. Additionally, lexical attributes such as concreteness, familiarity, imagery, and age of acquisition did not show significant continuity. Moreover, the analysis of syntactic complexity also showed no consistent pattern. The current inconsistency in text difficulty suggests that students may not be receiving the progressive challenge needed to develop their reading comprehension skills effectively. Drawing upon these findings, theoretical and pedagogical implications are elucidated.

---

[*] 제1저자 [**] 공동저자 [***] 교신저자

[Coh-Metrix/corpus/intra-grade continuity/national academic achievement test, 코메트릭스/코퍼스/학년 내 연계성/전국연합학력평가]

Ⅰ. 서론

대학수학능력시험(大學修學能力試驗, College Scholastic Ability Test, 이하 수능)은 문자 그대로 대학(大學, College)에서 수학(修學, Scholastic)할 수 있는 능력(能力, Ability)을 갖추었는지 검증하는 시험(試驗, Test)이다. 그러나 대학교육의 수학능력을 측정하는 수능의 본래 취지는 점차 퇴색되고, 수험생을 등급에 따라 분류하고 서열화하는 도구로 전락하고 있다(김용명, 2016; 김현미 외, 2024). 최근 주요대학의 정시비중이 확대되고, 의대정원 및 무전공 확대 등으로 이른바 N수생이 증가하면서 수능응시비율은 2020학년도 69.3%, 2021학년도 67.4%, 2022학년도 71.4%, 2023학년도 71.5%, 2024학년도 72.8%로 증가추세에 있다(이지희, 2024). 이렇듯 대학입시에서 학생선발의 결정적인 변수로 자리 잡은 수능은 그 결과가 수험생의 진학 및 향후진로에 결정적인 영향을 미치는 만큼, 사교육 시장에도 큰 영향을 주었고, 이는 학생과 학부모에게 상당한 압박으로 작용하게 되었다(김현미 외, 2024). 이와 같은 사교육 의존 현상이 심화되자, 교육부는 사교육 억제 및 공교육 정상화를 도모하고 학생간 불필요한 경쟁해소를 위하여 2011학년도 EBS-수능 연계정책, 킬러(초고난도)문항 출제배제, 2018학년도 수능영어 절대평가제도 시행 등 다양한 대책을 마련하였다(고나은, 신정아, 2017; 장지연, 2018, 2022).

특히, 수능영어 절대평가의 경우 정책이 집행되는 과정이 단기간에 진행됨에 따라 각 등급에 대한 명확한 기준 없이 10점 간격의 9 등급 체계로 적용되게 되었다(김정태, 2019; 이완기, 2015). 즉, 진정한 의미의 준거참조(criterion-reference) 절대평가체계가 아닌 '절대점수'로 등급이 정해지다보니 각 등급별 인원분포 비율이 매년 다르게 나타나게 되었다. 2023학년도에 시행된 모의수능 및 수능영어시험에서는 이러한 불안정성 문제가 특히 심각하게 나타났는데 2023학년도에 실시된 모의수능의 1등급(원점수 90점 이상) 비율은 3월 1.98%, 4월 4.97%, 7월 9.74%, 10월 15.31%로 나타났고, 11월에 실시된 2024학년도 수능에서는 킬러 문제를 제외하였으나 1등급 비율이 4.71%로 나타나 '불수능'을 넘어 '용암수능' 이었다는 평가를 받았다(중앙일보, 2023). 이는 등급 결정 기준에 일관성이 결여되어 있으며 시험의 난이도가 일정하지 못하다는 것을 의미한다.

시험 난이도에 영향을 미치는 요인은 다차원적으로, 수험자의 개인적 배경(성별, 나이, 사회경제적 상황, 준비 전략 등), 인지적 요소(시험 방식, 문제의 종류, 과제 유형, 배경 지식 등), 언어적 요소(문장의 화용론적, 조직적 특성 등), 그리고 기타 요인(오답의 매력도 등) 등이 포함된다(김재은, 최인철, 2015; Bachman & Palmer, 1996). 이 중 통제 가능한 시험 난이도 영향 요인으로는 텍스트의 내적

변인, 즉 언어적 정보가 있으며(김재은, 최인철, 2015; 황이수, 이제영, 2020), 선행연구에서 수능영어 텍스트 지문을 코퍼스로 구축하여 난이도와 관련있는 언어적 변인을 분석하고자 하였다. 이와 같은 기존연구들은 수능 난이도와 관련된 어휘적 특성, 통사적 복잡성 등의 결과를 다각도로 제시함으로써 시험의 타당도, 연계성 등을 분석하는데 크게 기여하였으나 대부분의 연구들이 한국교육과정평가원에서 주관하는 수능 혹은 수능모의평가로 한정되어 있다.

대부분 고3 재학생들은 수능을 준비하기 위해 한국교육과정평가원이 주관하는 6월과 9월의 수능 모의평가(이하 모평) 2회와 시도 교육청이 주관하는 전국연합학력평가 4회, 즉 총 6번의 모의시험을 치르게 된다(양성현, 2017; 염영희, 김해동, 2023). 전국연합학력평가(이하 학평)는 재학생들만 응시가 가능한 반면, 6월과 9월 수능모의평가에는 졸업생(재수생)들도 상당수 포함되어 응시인원 집단구성에 차이를 보인다(양성현, 2017). 즉, 학평은 단순히 수능시험의 예행연습이라는 측면을 넘어서 학생들에게 자신의 학력 수준과 위치를 전국단위를 기준으로 객관적으로 파악할 수 있는 기회를 제공한다. 이를 통해 교사와 학부모는 학생진로 및 진학상담에 필요한 주요정보를 얻을 수 있으며, 교사는 성취도 분석을 통해 향후 교수학습 방향을 설정하는 데 도움을 받을 수 있다. 다시 말하면 학평은 시험의 사회적 기능(선발기제, 사회통제, 서열위계화, 질서 정당화 등)뿐만 아니라 교육적 기능(학업성취 확인, 교육과정 결정, 향후학습 조언)을 수행할 수 있다는 것이다. 이러한 배경 하에 본 연구는 Coh-Metrix 프로그램을 사용하여 2023학년도에 실시된 총 4회 실시된 학평의 학년 내 연계성을 분석하고자 한다. 특히, 재수생을 포함하는 모평과 달리 재학생만 대상으로 하는 학평에서 2023학년도에 1등급 불안정성(3월: 1.98%, 4월: 4.97%, 7월: 9.74%, 10월: 15.31%)을 초래한 변인은 무엇이었으며 교육과정이 진행됨에 따라 특정 학년 내 시험의 난이도가 단계적으로 증가하는지 등의 변화양상을 살펴보고자 Coh-Metrix의 다양한 측정치를 통한 횡단적 연구를 수행하고자 한다. Coh-Metrix는 텍스트 이독성, 지시적 응집성, 상황 모델링 등 다양한 지표를 제공하여 텍스트 난이도와 관련된 여러 측면을 다각도로 분석할 수 있으며, 다양한 연구에서 널리 사용되고 검증된 도구로서 분석 결과의 신뢰성을 보장한다. 이러한 Coh-Metrix를 사용하여 학평의 시행 시기별 주관기관(3월: 서울특별시, 4월: 경기도, 7월: 인천광역시, 10월: 서울특별시)이 다르다는 것을 고려할 때 특정 학년도를 중심으로 독해 지문의 수평적 연계성이 잘 지켜졌는지에 대한 주요 시사점을 제공할 것으로 판단된다.

Ⅱ. 이론적 고찰 및 선행연구

1. 코메트릭스(Coh-Metrix)

Coh-Metrix는 멤피스대학 지능형시스템연구소에서 개발한 언어분석 프로그램(Graesser 외, 2004)으로, 11개의 주요 지표와 106개의 세부 측정치를 포함하고 있다(표 1). 11개 지표로는 어휘정보, 기초산출치, 상황모델, 텍스트 이독성, 지시적 응집성, 잠재 의미 분석, 어휘의 다양성, 연결어, 통사적 복잡성, 통사적 패턴 밀집성, 가독성로 구성되어 있다. 이처럼 텍스트의 표층적(예, 단어 수, 접속사 수, 대명사 수, 문장 내 단어 수 등) 측면뿐만 아니라 심층적(본동사 앞단어 수, LSA 측정치, 명사구 밀도), 어휘의 심리언어학 정보 등 포괄적인 측정치를 제공하여 텍스트의 종합적 분석이 가능하다(성일호, 2014; 장지연, 2018; McNamara 외, 2014).

표 1
Coh-Metrix 3.0의 11개 지표와 106개 세부 측정치

No.	측정 항목	수
1	Word Information (어휘 정보) e.g., WRDFAMc(Familiarity), WRDIMGc (Imageability), WRDCNCc (Concreteness), WRDAOAc (Age of acquisition)	22
2	Text Easability Principal Component Scores (텍스트 용이성 주성분 점수) e.g., PCNARp (Text Easability PC Narrativity, percentile), PCSYNz (Text Easability PC Syntactic simplicity, z score)	16
3	Descriptive Indices (기초 산출치) e.g., DESSC (Sentence count), DESWC (Word count) DESPC (Paragraph count), DESPL (Paragraph length)	11
4	Referential Cohesion (지시적 응집력) e.g., CRFCWO1 (Content word overlap, adjacent sentences, mean) CRFAO1 (Argument overlap, adjacent sentences, binary, mean)	10
5	Connectives (연결어) e.g., CNCCaus (Causal connectives), CNCLogic (Logical connectives) CNCTemp (Temporal connectives), CNCPos (Positive connectives)	9
6	Syntactic Pattern Density (통사적 양식 밀집도) e.g., SMCAUSvp (Causal verbs and causal particles incidence) SMINTEp (Intentional verbs incidence)	8
7	Situation Model (상황 모형) e.g., SMCAUSvp (Causal verbs and causal particles incidence) SMINTEp (Intentional verbs incidence)	8
8	Latent Semantic Analysis, LSA (잠재 의미 분석) e.g., LSASS1 (LSA overlap, adjacent sentences, mean), LSAGNd (LSA given/new, sentences, standard deviation)	8
9	Syntactic Complexity (통사적 복잡성) e.g., SYNNP (Number of modifiers per noun phrase, mean), SYNLE (Left embeddedness, words before main verb, mean)	7
10	Lexical Diversity (어휘적 다양성) e.g., LDTTRc (Lexical diversity, type-token ratio, content word) WRDFRQc (CELEX word frequency for content words, mean)	4

11	Readability (가독성) e.g., RDFRE (Flesch Reading Ease), RDFKGL (Flesch-Kincaid Grade level)	3
	Total (세부 측정치 총 합계)	106

국내에서 수행된 Coh-Metrix 관련 연구로는 전문기와 임인재(2009)의 중학교 1학년 영어교과서 코퍼스 분석연구를 필두로 다수의 수능지문(고나은, 신정아, 2017; 이주윤, 2022; 이혜진, 이제영, 2018; 장지연, 2018; 황이수, 이제영, 2020), 공무원 시험텍스트(신유선, 정영경, 2018), 중고등 영어교과서(김지은, 전문기, 2013; 류지수, 전문기, 2020; 성일호, 2014), 학습자 영작문(김소정, 전문기, 2016; 이지연, 2022) 등이 수행되었으며 분석의 외연을 확장하는데 기여하였다. 전술하였듯이 Coh-Metrix에서 제공되는 분석수치가 방대하다 보니 대부분 선행연구에서 목적에 필요한 주요측정치만 선별적으로 사용하는 경향이 있었는데 수능영어 지문 분석을 중점적으로 살펴보면 다음과 같다. 장지연(2018)은 2017-2018 수능독해 지문을 분석하기 위해 기초 산출치(문장 수, 단어 수 등), 가독성(Coh-Metrix 제2언어 이독성, FRE, FKGL), 텍스트 용이성(서사성, 단어구체성, 지시적 응결성 등)을 중점적으로 분석하여 절대평가 도입 전후를 비교하였다. 이주윤(2022)은 어휘 다양성(내용어 Type-Token 비율, 모든 단어의 CELEX로그 빈도평균 등), 응집성(인접한 문장 사이의 논항 중복 문장 평균 수, 인접 문장 간 의미적 유사성 등), 통사적 복잡성(본동사 앞 단어 평균 개수, 명사구 밀도 점수) 등의 측면에서 절대평가 이후의 수능영어 지문을 분석하였다. 이렇듯 대부분의 코퍼스 분석 연구는 고부담 속성(high-stakes nature)으로 인해 주로 수능 지문을 중심으로 진행되었으나 본 연구는 학평 독해지문 코퍼스를 분석할 예정이므로 류지수와 전문기(2023)의 고등학교 모의시험 연구에서 적용한 코퍼스 측정치를 주로 사용하고자 한다. 류지수와 전문기(2023)의 연구에 의하면 해당 측정치가 텍스트의 난이도 분석에 광범위하게 사용되었으며(Graesser 외, 2004; Graesser 외, 2007), 이러한 표층적 및 심층적 수준의 언어 측정치가 연계성 분석에 효과적으로 적용될 수 있다고 하였기 때문이다. 본 연구도 이를 통해 학평 지문의 언어적 특성과 구조적 복잡성을 면밀히 검토하여, 특정 시기에 시행된 학년 내 난이도 변화와 연계성을 분석함으로써, 학평의 단계적 난이도 설정과 교육과정의 조정에 대한 기초자료를 마련할 것으로 사료된다.

2. 전국연합학력평가

전국연합학력평가는 초.중등교육법 제9조와 전국 시도교육감협의회 합의 사항(2008. 10. 17)에 근거하여 사교육비 절감, 진로진학 자료 제공, 자기주도적 학습력 신장을 위해 전국 17개 시·도교육청의 합의에 따라 수능형식으로 시행되는 시험이다(서울특별시교육청, 2022). 2023년을 기준으로 고 3 대상의 학력평가는 3월(서울특별시교육청), 4월(경기도교육청), 7월(인천광역시교육청), 10월(서울

특별시교육청)에 총 4회가 실시되었으며 서울특별시 교육청 웹사이트에서 제공하는 응시인원, 평균, 표준편차, 1등급 인원 및 비율은 표 2와 같다.

표 2
2023학년도 고3 전국연합학력평가 영어영역 응시인원 및 결과

월	주관	응시인원 (비율)	평균	표준편차	1등급(90 이상)	1등급 비율
3	서울	308,404 (99.86%)	53.48	20.74	6,098	1.98%
4	경기	284,934 (99.81%)	56.16	21.86	14,171	4.97%
7	인천	268,401 (99.74%)	57.83	24.17	26,136	9.74%
10	서울	239,628 (99.69%)	63.26	24.38	36,695	15.31%

전국연합학력평가 분석과 관련하여 비교적 최근에 시행된 선행연구로는 다음과 같다. 류지수와 전문기(2023)는 2022학년도 고등 1, 2, 3학년 영어 모의시험의 연계성을 Coh-Metrix를 사용하여 분석하였으며 그 결과 기본측정치(평균어휘수, 문장길이, 어휘길이), 읽기 가독성(FRE, FKGL)의 학년간 연계성은 잘 성립되었으나 친숙성을 제외한 어휘 속성, 인칭대명사, 참조 정합성 등의 경우 학년 간 연계성이 잘 성립되지 않았음을 확인하였다. 이주윤(2023)은 2019학년도부터 2022학년도까지 6월, 9월, 11월에 실시된 고등학교 1학년 영어 모의시험 지문의 언어적 특성 및 난이도를 Coh-Metrix를 통해 분석하였다. 시기별(6월, 9월, 11월) 독해지문의 언어적 특성의 차이를 분석한 결과, 응집성, 어휘, 통사적 복잡성에서는 통계적으로 유의한 차이가 발견되지 않았으나, 문장 내 단어 수의 평균, 단어 수, 이독성에서 유의미한 차이가 확인되었다는 것을 확인하였다. 염영희와 김해동(2023)은 2018학년도부터 2023학년도에 시행된 고등학교 1, 2, 3 모의시험과 수능 간 텍스트 난이도 지표를 분석한 결과 내용어 빈도, 단어 수, 문장 길이, 명사구 수식어 수, 본동사 앞 단어 수, 가독성(FRE, FKGL) 지표는 유의한 차이를 보였으나, 문장 수, 타입-토큰 비율, 응집성 지표는 유의미한 차이가 없고 학년별 지표 평균값이 비슷하다는 결과를 확인하였다. 앞서 언급한 종단연구(류지수, 전문기, 2023; 염영희, 김해동, 2023)들은 학년 간 연계성을 확인하는 데 큰 기여를 했으나, 실제 수능 응시 여부를 결정하고 전국 단위에서 본인의 위치를 확인하려는 고 3을 대상으로 분석한 학년 내 연계성에 대한 횡단적 연구는 상대적으로 부족하다. 특히, 한국방송공사(2023)에 의하면 고 3을 기준으로 학평의 출제 범위는 3월에는 1, 2학년 전 범위를 시작으로 각 시행되는 달의 수준에 맞춰 점차 누적되어 출제된다는 것을 고려하였을 때 이상적으로 이러한 학평의 난이도는 시간이 지남에 따라 점진적으로 증가하여 10월 학평은 수능의 난이도에 근접하도록 설계되어야 할 것이다. 특히, 전국의 고등학교 재학생을 대상으로 2023학년도에 실시된 학평의 경우 3월(1.98%)과 10월(15.31%)의 1등급 비율 차이가 약 673.23%(13.33%p) 증가되는 등 1등급 비율의 변동 폭이 심하였다는 것을 고려할 때 본 연구는 2023학년도에 실시된 학평의 3월, 4월, 7월, 10월의 각 시험의 독해지문 난이도를 횡단적으로 분석하여, 학년 내 연계성과 난이

도 조정의 적절성을 평가하고자 한다. 이는 2024학년도 향후 학평 및 2025학년도 수능을 앞두고 있는 현 시점에서, 이러한 직전 년도를 중점적으로 분석한 학년 내 연계성 연구는 시의적인 함의를 제공할 것으로 사료된다.

III. 연구 방법

1. 2023학년도 고3 학평 코퍼스 구축

본 연구는 2023학년도에 실시된 고 3 학평 독해 지문의 학년 내 연계성 분석을 목적으로 한다. 전술하였듯이 고3 재학생을 대상으로 총 6회의 모의시험이 실시되는데 코퍼스 구축에 포함된 지문은 시도 교육청이 주관하는 전국연합학력평가 4회 (3월, 4월, 7월, 10월) 자료이며, 재수생이 참여할 수 있는 한국교육과정평가원에서 2회 (6월, 9월) 실시하는 모평은 제외하였다. 정확한 시행 시기는 시도교육청의 상황에 따라 달라질 수 있으며 2023학년도에 실시된 고3 학평의 경우 3월과 10월은 서울시 교육청, 4월은 경기도 교육청, 7월은 인천시 교육청 주관 하에 진행된다(한국교육방송공사, 2023). 코퍼스 구축을 위해 각 시도교육청 홈페이지에서 제공하는 기출문제 PDF 파일을 다운받아 Coh-Metrix에서 분석이 가능하도록 각 문항별 독해지문을 복사하여 유니코드 형식의 텍스트 (.txt) 파일 형식으로 저장하였으며, 본고의 코퍼스 구성은 아래 표3과 같다.

표 3
2023학년도 고3 전국연합학력평가 영어영역 응시인원 및 결과

월	시험일자	주관 교육청	지문 수	토큰	타입	구축형식 예 (.txt)
3	3월 23일(목)	서울특별시	25	4,801	1,335	H3_23_03_18
4	4월 12일(수)	경기도	25	4,853	1,429	H3_23_04_18
7	7월 11일(화)	인천광역시	25	4,760	1,445	H3_23_07_18
10	10월 12일(수)	서울특별시	25	4,730	1,415	H3_23_10_18

수능의 구조와 유사하게 학평도 총 45문항(듣기: 17문항, 독해: 28문항)으로 구성되며, 41번부터 42번, 43번부터 45번 문항은 하나의 장문독해 지문에서 파생된다. 이에 따라 각 25개의 지문이 4회(3월, 4월, 7월, 10월) 시행되어 총 100개의 학평 독해 지문이 코퍼스로 구축되었다. 지문을 텍스트 파일로 구축한 후 텍스트 정제과정을 수행하였는데 예를 들면, 본문과 관련 없는 불필요한 정보(예, 기호, 특수문자 제거)를 삭제하고 필요한 정보를 보강(예, 빈 칸 채우기 문제는 빈 칸을 채워 저장)하거나 수정(예, 내용의 순서를 묻는 지문은 정답에 참고하여 순서를 맞추어 저장)하는 등 전처리 작업을 진행하였다.

2. 2023학년도 고3 학평 코퍼스 분석

본 코퍼스 연구는 24개의 Coh-Metrix 측정치를 기반으로 학년 내 연계성을 분석하기 위해 IBM SPSS Statistics 21 통계프로그램을 사용하였다. 특히, 고3 학평의 난이도가 학년 내에서 점차적으로 증가하여야 한다는 점에 주목하여 통계적 유의성을 판단하기 위해 독립변인으로는 4회 시행된 각 시험이 사용되었고, 종속변인은 Coh-Metrix의 측정치 지표의 평균값으로 간주하여 유의수준(significance level) 5% ($p < .05$)에서 일원배치 분산분석(ANOVA)을 수행하였다. 각 조건 별로 표본 수가 25개로 동일하여 통계적으로 유의한 결과가 있는 경우 Tukey 사후검정을 실시하여, 시간의 증가에 따라 시험의 난도가 상승했는지 아니면 난도상승 체계이 위반되고 있는지를 파악하여 결과를 해석하였다.

Ⅳ. 연구결과 및 논의

1. 기본 측정치 분석결과

기본 측정치는 가장 기초적인 텍스트 분석 지표로 텍스트에 포함된 어휘 수 혹은 문장 길이가 증가할 수록 복잡성이 높아질 것으로 예측되기 때문에 난이도를 추정할 수 있다(Graesser 외, 2007). 2023학년도 고3 학평의 기본 측정치를 분석한 결과 전반적으로 학년 내 연계성이 잘 확립되지 않음을 확인할 수 있다. 분석된 네 가지 지표인 평균어휘수, 평균문장수, 평균문장길이, 평균어휘길이 모두 시간의 경과에 따라 소폭의 변동은 있었으나, 이러한 변화가 학기의 진행에 따라 일관된 상승 패턴을 보이지 않았으며, 통계적으로 유의미한 변화도 확인되지 않았다.

표 4
기본 측정치 분석 결과

기본측정치	3월 ($n = 25$)	4월 ($n = 25$)	7월 ($n = 25$)	10월 ($n = 25$)	F	p
평균어휘수	170(56.7)	171(49.6)	169(47.5)	167(50.7)	.030	.993
평균문장수	8.40(4.90)	9.28(4.16)	8.40(3.69)	8.68(3.38)	.260	.854
평균문장길이	1.64(.173)	1.66(.169)	1.62(.157)	1.60(.173)	.660	.578
평균어휘길이	22.1(5.87)	19.7(4.63)	21.5(6.12)	20.4(5.86)	.892	.448

* $p < .05$

표 4에 제시된 기본측정치 분석결과 평균어휘길이를 제외한 평균어휘수, 평균문장수, 평균문장길이의 세 가지 지표는 4월에 가장 높은 평균값을 기록하였음을 알 수 있다. 평균문장길이와 평균어휘길이는 유의미한 증가나 감소 경향을 보이지 않았으며 각 지표들의 변화가 통계적으로 유의미하지 않다는 점은,

시험회차진행에 따른 난이도의 체계적인 조절이 부재했음을 의미한다.

2. 어휘빈도수 및 어휘다양성 분석결과

어휘빈도수와 다양성은 텍스트의 난이도를 조절하는 데 중요한 요소이다. 텍스트 내에 저빈도 어휘가 많이 포함될 경우, 이를 해석하기 위해서 더 많은 인지적 노력이 요구되며 이는 텍스트 처리 속도 및 이해도에 영향을 미친다(Graesser 외, 2007; McNamara 외, 2014). 본 연구에서는 Coh-Metrix 도구를 활용하여 검증된 신뢰성을 바탕으로 포괄적인 어휘 데이터를 제공하는 네덜란드 어휘정보센터(Dutch Centre for Lexical Information)의 CELEX의 어휘 데이터베이스를 기반으로 광범위한 어휘 빈도 정보를 분석하였다. 본 연구에서는 어휘 빈도수를 원시 점수(raw scores)가 아닌 대수변환(logarithmic transformation)된 값으로 적용하여 분석하였다. 대수변환은 데이터의 분포를 정규분포(normal distribution)에 가깝게 조정하여 어휘 사용의 실제 패턴을 보다 정확하게 반영해주며 어휘의 실제 사용 빈도가 매우 낮거나 높은 경우에 발생할 수 있는 통계적 왜곡을 줄여주기 때문이다(류지수, 전문기 2020; Graesser 외, 2007). 표 5에 제시된 어휘빈도수 측정치 분석 결과, 시간이 지남에 따라 약간의 변동성은 있지만 전반적으로 학년 내 유의한 차이가 나타나지 않았다($p > .05$). 이는 시험 회차를 거듭하면서 어휘빈도와 관련된 난이도가 체계적으로 조정되지 않았음을 의미한다.

표 5
어휘빈도수 및 어휘다양성 측정치 분석 결과

Coh-Metrix 측정치	3월 ($n = 25$)	4월 ($n = 25$)	7월 ($n = 25$)	10월 ($n = 25$)	F	p
어휘빈도수(내용어)	2.20(.144)	2.17(.201)	2.19(.181)	2.24(.171)	.758	.521
타입-토큰비율(내용어)	.787(.097)	.808(.093)	.823(.064)	.818(.112)	.745	.528

* $p < .05$

어휘다양성은 특정 지문에 사용된 어휘의 다양성을 나타내는 지표로, Coh-Metrix 시스템은 타입-토큰 비율(type-token ratio)을 사용하여 이를 측정한다. 어휘다양성 지표에서 타입(type)은 텍스트 내에서 한 번 이상 나타나는 독립적인 어휘의 수를 의미하고, 토큰(token)은 텍스트 내에서 어휘가 총 몇 번 사용되었는지를 나타낸다. 예를 들어, 특정 지문에 '영어'라는 어휘가 총 4회 등장하면 이 어휘의 타입 값은 1이고 토큰은 4, 타입-토큰 비율은 0.25(1/4)인 것이다. 높은 타입-토큰 비율은 텍스트 내 어휘 반복이 적고, 어휘 사용의 다양성이 높아 텍스트의 복잡성을 증가시키며 독해에 더 많은 인지적 노력을 요구한다(류지수, 전문기 2020). 표5의 분석결과 고3 학평 독해 지문의 어휘 다양성은 학년 내에서 통계적으로 유의한 차이를 보이지 않았으며($p > .05$), 이는 학년 내 어휘의 빈도수 및 다양성 측면에서 연계성이 체계적으로 확립되지 않았음을 의미한다.

학습효과를 극대화하기 위해서는 시험 회차가 진행될수록 학생들이 접할 수 있는 지문 내 어휘 다양성을 점차 증가시켜야 할 것이다.

3. 어휘속성 분석결과

어휘 속성의 측정치는 심상성, 친숙성, 구체성, 습득나이가 포함되며 Coh-Metrix는 위와 같은 어휘정보를 MRC Psycholinguistic의 어휘 데이터베이스를 기반으로 100에서 700 사이의 값으로 산출한다(Coltheart, 1981). 심상성(imagability)은 어휘의 이미지 형상화 정도를 나타내는 수치이며 예를 들어 overtone는 268점, dogma는 327점, bracelet은 606점 등 측정치 점수가 높을수록 이미지 형상화가 용이하고 텍스트 난이도가 감소할 것으로 예상된다. 친숙성(familiarity)은 어휘의 친숙성 정도로 mother는 632점, milk는 588점, calix는 124점으로 낮은 점수는 일상에서 자주 접하기 어려운 생소한 어휘를 의미한다. 내용어의 구체성(concreteness)은 어휘가 얼마나 구체적인지를 나타내는 측정치로 difference은 270점, ball 은 615점 등 점수가 낮을 수록 추상적인 어휘이다. 습득나이(age of acquisition)는 어휘의 습득 연령대를 나타내며, cortex는 700점, milk는 202점 등 점수가 낮을수록 어린 연령대에 습득될 것으로 기대되는 어휘를 의미한다(최민주, 2018). 표 6의 어휘속성 분석 결과를 살펴보면, 3월부터 10월까지의 시험 회차에서 어휘의 심상성, 구체성, 습득나이, 친숙성 지표 값의 유의한 차이가 없다는 것을 알 수 있다(all *ps* >.05).

표 6
어휘속성 분석 결과

어휘속성	3월 (n = 25)	4월 (n = 25)	7월 (n = 25)	10월 (n = 25)	F	p
심상성	395(28.3)	408(28.1)	403(28.0)	405(30.6)	.835	.478
구체성	363(32.8)	381(30.7)	374(28.1)	370(36.9)	1.43	.239
습득나이	379(43.7)	370(37.4)	357(45.0)	366(44.0)	1.15	.334
친숙성	567(9.96)	568(10.6)	567(11.0)	569(8.96)	.371	.774

* $p < .05$

특히, 시험회차에 따라 어휘 속성의 지표값의 일관된 변화 패턴이 뚜렷하지 않음을 확인할 수 있는데, 심상성은 3월에 비해 4월에 소폭 상승하였으나 이후 큰 변화가 없었으며, 구체성 역시 4월에 다소 증가하였으나 이후 점차 감소하였다. 습득 나이의 경우 7월에 가장 낮은 점수를 기록하였고, 친숙성은 시험회차에 따라 큰 변화없이 비교적 일정하게 유지되었으며 통계적으로 유의한 차이도 없었다. 이러한 분석 결과는 어휘 속성의 변화가 학년도를 거치면서 일관된 패턴을 보이지 않음을 시사하며, 학생들의 학습수준이나 어휘 발달을 촉진할 수 있도록 체계적으로 조정할 필요가 있음을 의미한다.

4. 표준 텍스트 가독성 지표 분석결과

Coh-Metrix 시스템은 읽기 지문의 가독성을 측정하기 위해 Flesch Reading Ease (FRE, Flesch 읽기용이성)와 Flesch-Kincaid Grade Level (FKGL, Flesch-Kincaid 학년 수준) 점수를 활용한다. FRE 점수는 0에서 100 사이의 범위로, 점수가 높을수록 텍스트 이해가 용이함을 의미하며, FKGL 점수는 1에서 12까지 영어 원어민기준의 학년 수준(grade level)을 나타내며, 점수가 낮을수록 쉽고, 높을수록 복잡함을 의미한다(McNamara 외, 2014).

표 7
읽기 가독성 지표 분석 결과

Coh-Metrix 측정치	3월 (n = 25)	4월 (n = 25)	7월 (n = 25)	10월 (n = 25)	F	p
FRE 점수	45.3(17.7)	46.7(16.1)	48.0(15.3)	51.1(15.8)	.594	.621
FKGL 점수	10.8(2.28)	10.8(2.28)	10.8(1.99)	10.3(2.05)	.338	.798

* $p < .05$

표 7의 표준 읽기 가독성 지표 분석 결과에 따르면, FRE 점수는 시간이 지나면서 점차적으로 향상되어 텍스트의 난이도가 점진적으로 낮아지는 경향을 보이지만, 이 변화는 통계적으로 유의미하지 않다는 것을 알 수 있다($p > .05$). FKGL 점수 역시 모든 측정 시점에 걸쳐 비교적 일정하게 유지되며, FKGL 점수의 변화 또한 통계적으로 유의미하지 않았다($p > .05$). 시험회차에 따른 읽기 자료의 가독성이 시간이 지남에 따라 크게 변하지 않음을 나타내며, 수험자 관점에서는 텍스트의 난이도가 일정하게 유지되고 있음을 의미한다. 그러나 시험의 교육적 기능 관점에서 봤을 때 학습자의 독해 능력 향상을 위해서는 지문 난이도의 조정이 보다 세밀하게 이루어져야 할 필요성을 시사한다.

5. 통사적 복잡성 분석결과

Coh-Metrix는 텍스트의 통사적 복잡성을 측정하기 위해 명사구 밀도점수와 본동사 앞 어휘 수를 사용한다(Graesser 외. 2004). 명사구 밀도는 명사구에 포함된 수식어(형용사, 부사, 한정사 등)의 평균 수를 나타내며, 이 수치가 높을수록 문장 구조가 복잡해진다고 간주된다(Graesser 외. 2007). 본동사 앞의 단어 수는 문장의 중심 내용을 파악하기 위해 필요한 작업 기억(working memory)에 대한 인지적 부담을 나타내는 지표로 본동사 앞에 많은 단어가 있으면 중심내용을 파악하기 위하여 많은 정보를 기억해야 하므로 이 값이 크면 클수록 텍스트의 복잡도가 상승한다(성일호, 2014; Graesser 외, 2007).

표 8
통사구조 분석 결과

통사구조	3월 (n = 25)	4월 (n = 25)	7월 (n = 25)	10월 (n = 25)	F	p
명사구 밀도	.894(.207)	.857(.205)	.924(.189)	.879(.235)	.460	.711
본동사 앞 단어 수	5.71(2.93)	4.80(2.17)	4.77(2.28)	4.91(2.58)	.794	.500

* $p < .05$

표 8의 분석결과, Coh-Metrix를 사용한 통사적 복잡성 측정 지표인 명사구 밀도와 본동사 앞 단어 수 모두 시간에 따라 유의미한 변화가 없음을 보여주고 있다($p > .05$). 이는 각 시험 회차 간의 난이도에 큰 변화가 없음을 나타내며, 시간의 경과에 따라 시험의 구조적 복잡성이 일정하게 유지되었음을 의미한다. 그러나 시험의 교육적 관점에서 볼 때, 학생들의 언어적 능력과 통사적 이해를 증진시키기 위해 교육과정이 진행됨에 따라 통사적 복잡성을 조절하는 것이 필요할 것으로 사료된다.

6. 참조 정합성 및 의미적 유사성 분석 결과

참조 정합성과 의미적 유사성은 텍스트의 이해에 중요한 역할을 하는 언어학적 지표이다. 참조 정합성은 텍스트 내에서 문장들이 공유하는 명사, 대명사, 명사구 등의 주요 논항(argument)을 통해 측정되며, Coh-Metrix는 이를 인접 문장 및 전체 문장 간의 논항 중복을 통해 분석한다. 높은 참조 정합성은 문장 간에 논항이 반복되어 구조적 일관성을 제공하고 텍스트의 이해를 쉽게 하며, 따라서 난이도를 낮춘다(Graesser 외, 2004). 의미적 유사성은 텍스트 내 문장 간 의미적 연결의 정도를 나타내며, Coh-Metrix는 LSA(Latent Semantic Analysis) 기반의 수학적 모형을 통해 LSA 코사인 측정치로 어휘 및 문장 간 의미적 유사성을 측정한다. 높은 의미적 유사성은 텍스트 내용이 의미적으로 일관되게 연결되어 있음을 나타내며, 이는 인지적 부담을 줄이고 텍스트의 복잡성을 감소시킨다(Graesser 외, 2004). 표 9의 분석결과 참조 정합성과 의미적 유사성의 변화는 시간이 지남에 따라 일관된 패턴 없이 변동하고 있으며, 어떠한 통계적 유의성도 발견되지 않았다 (all $ps > .05$).

표 9
참조 정합성 및 의미적 유사성 분석 결과

Coh-Metrix 측정치	3월 (n = 25)	4월 (n = 25)	7월 (n = 25)	10월 (n = 25)	F	p
참조 정합성						
논항반복 (인접문장 간)	.653(.234)	.508(.253)	.587(.291)	.621(.266)	1.41	.246
논항반복 (전체문장 간)	.550(.209)	486(.224)	.539(.245)	.527(.256)	.360	.782
의미적 유사성						
LSA 코사인 (인접문장 간)	.255(.142)	.246(.130)	230(.102)	.238(.126)	.182	.908
LSA 코사인 (전체문장 간)	.231(.135)	.225(.130)	.209(.107)	.207(.119)	.222	.881

* $p < .05$

표 9의 분석결과를 살펴보면 참조 정합성의 경우, 인접 문장 간 논항 반복은 3월 .653(*SD* = .234)에서 4월 .508(*SD* = .253)로 감소하였으나, 이후 7월에는 .587(*SD* = .291)로 다시 증가하고 10월에는 .621(*SD* = .266)로 더욱 증가하여, 초기 측정치보다 높은 수치를 기록하고 있으며 통계적으로 유의미한 차이가 없었다($p > .05$). 의미적 유사성의 경우에도 인접 문장 간 LSA 코사인 측정치는 3월 .255(*SD* = .142)에서 시작해 점차 감소하는 경향을 보이며 10월에는 .238(*SD* = .126)로 나타났으며 통계적 유의성이 없는 것으로 나타났다($p > .05$). 이와 같은 결과는 시험 회차가 진행되는 동안 참조 정합성과 의미적 유사성이 일정한 방향성 없이 변동하고 있음을 나타내며 시험의 교육적 기능을 강화하기 위해서는 시험회차가 진행됨에 따라 지표값을 점진적으로 조정하여야 할 것이다.

7. 고3 학평의 어휘비율 및 가독성 비교분석

앞서 논한 학년 내 연계성 외에 2023학년도 고3 학평 시험에서 관측된 1등급 비율의 큰 변동 폭(3월: 1.98%, 4월: 4.97%, 7월: 9.74%, 10월: 15.31%)의 원인을 추측하기 위해 어휘비율 및 가독성을 비교 분석하였다. 어휘 비율은 Lextutor를 통해 학술어휘(Academic Word List, AWL)와 기본어휘(General Service List, GSL)의 사용 빈도를 측정하였으며, Range 프로그램을 활용하여 2022 개정 영어 교육과정에 따른 초등 권장어휘 800, 중고등 공통어휘 1200, 고등 선택어휘 1000의 분포비율을 각 시험 회차별로 분석하였다. 텍스트의 가독성은 Readability Analyzer를 통해 텍스트 내의 어렵고 생소한 어휘의 비율 분석을 포함하여 Gunning Fog Index와 SMOG Index의 가독성 지표를 살펴보았다. Gunning Fog Index는 문장 당 단어 수와 3음절 이상의 복잡한 단어 비율을 기반으로 계산되어 텍스트를 이해하는 데 필요한 교육 연한을 나타낸다. 반면, SMOG Index는 30문장 내의 3음절 이상 단어 수를 통해 산출되며, 주로 실용적인 텍스트의 난이도를 평가하는 데 사용된다. 또한, Readable 웹사이트를 통해 텍스트의 문장 길이와 단어 길이를

고려하여 자동화된 가독성 점수(Automated Readability Index)를 측정하여, 텍스트가 실제로 얼마나 이해하기 쉬운지에 대한 지표를 표 10과 같이 산출하였다.

표 10
고 3 학평의 어휘비율 및 가독성 분석

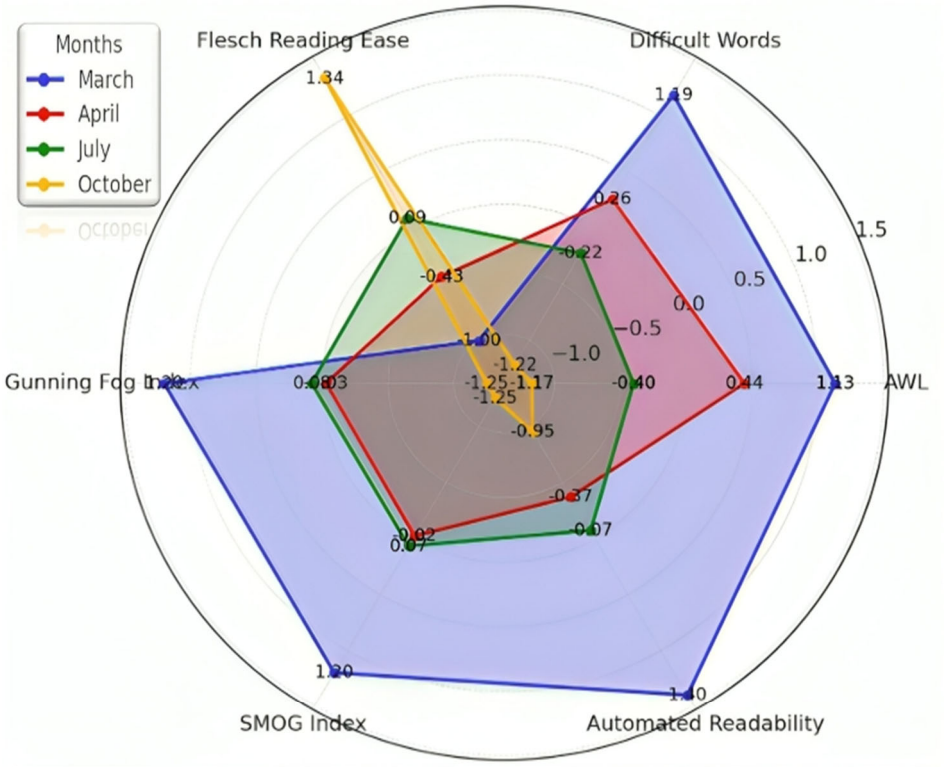
분석 지표 \ 월(1등급비율)		3월 (1.98%)	4월 (4.97%)	7월 (9.74%)	10월 (15.31%)
Lexical (Token /%)	학술어휘(AWL)	332(7.87%)	324(7.57%)	300(7.20%)	283(6.86%)
	기본어휘(GSL)	3,528(83.58%)	3,532(82.48%)	3,443(82.59%)	3,467(84.05%)
	초등권장어휘 800	2,861(67.91%)	2777(65.67%)	2807(67.23%)	2762(67.17%)
	중고등공통어휘 1200	814(19.32%)	809(19.13%)	749(17.94%)	760(18.48%)
	고등선택어휘 1000	233(5.53%)	270/(6.38%)	239(5.72%)	227(5.52%)
	어렵고 생소한 어휘	18.35%	17.44%	16.97%	16.00%
Textual	Flesch Reading Ease	45.3	46.7	48.0	51.1
	Gunning Fog Index	14.66	13.90	13.97	13.15
	SMOG Index	13.60	13.06	13.10	12.52
	Automated Readability	11.8	11.2	11.3	11.0

* $p < .05$

분석결과, 시험 회차가 진행될수록 학술어휘(AWL)의 사용 비율은 3월 7.87%에서 10월 6.86%로 점진적으로 감소하는 경향을 보인다. 2022 개정교육과정에서 권장하는 어휘분포를 살펴본 결과 고등 선택어휘의 반영비율은 4월에 일시적으로 증가한 후 다시 감소하는 경향을 보였다. Readability Analyzer 분석에 의하면 3월 학평지문에 어렵고 생소한 어휘의 비율(Zechmeister 외, 1995 참고)이 가장 높은 비율로 나타났으며 시험 회차가 진행될수록 이러한 어려운 어휘의 비율이 점차 감소하는 것으로 나타났다.

마찬가지로 가독성 지표 분석 결과 시험 회차가 진행될수록 독해지문이 점차 쉬워지고 있다는 것을 알 수 있는데 FRE의 경우 3월의 45.3에서 10월의 51.1로 계속적으로 상승하여, 문장이 더 직관적이고 이해하기 쉬운 구조로 사용되는 경향을 보여준다. Gunning Fog Index와 SMOG Index는 텍스트의 복잡성과 교육 수준을 나타내는데 사용되는데 이 두 지표 역시 Gunning Fog Index는 3월 14.66에서 10월 13.15로, SMOG Index는 13.60에서 12.52로 감소하였다. 이는 텍스트가 시간이 지남에 따라 구문적으로 간단해지고, 문장의 복잡성이 줄어들었음을 의미한다. 주로 텍스트 내의 단어 길이와 문장 길이를 기반으로 계산되는 Automated Readability 점수 역시 11.8에서 11.0으로 점진적으로 감소하여 시험회차가 진행될수록 전반적인 독해 난이도가 낮아졌음을 보여준다. 그림 1은 학술어휘(AWL), 어려운 단어 비율, FRE, Gunning Fog, SMOG, 자동 가독성 지표의 시계열 변화를 시험회차별로 비교하기 위해 Z-점수를 사용하여 표준화 하였으며 비교분석결과 다음과 같다.

그림 1
고3 학평의 어휘비율 및 가독성 비교 도식화



Metric	March Index	March z-score	April Index	April z-score	July Index	July z-score	October Index	October z-score
AWL	7.87%	1.13	7.57%	0.44	7.20%	-0.40	6.86%	-1.17
Difficult Words	18.35%	1.19	17.44%	0.26	16.97%	-0.22	16.00%	-1.22
FRE	45.30	-1.00	46.70	-0.43	48.00	0.09	51.10	1.34
Gunning Fog	14.66	1.20	13.90	-0.03	13.97	0.08	13.15	-1.25
SMOG Index	13.60	1.20	13.06	-0.02	13.10	0.07	12.52	-1.25
Automated Readability	11.80	1.40	11.20	-0.37	11.30	-0.07	11.00	-0.95

분석결과, 대부분의 지표에서 3월의 Z-점수가 가장 높게 나타났으며, 이는 3월 학평 독해지문 텍스트에 학술어휘 및 어려운 어휘가 다수 분포되어 있으며 텍스트의 복잡성과 난이도가 가장 높았음을 의미한다. Flesch 읽기 용이성 FRE 값의 경우, 높은 점수가 읽기 쉬운 텍스트를 의미하므로, 3월의 가장 낮은 Z-점수(-1.00) 역시 3월 독해지문 텍스트에 긴 문장과 어려운 단어의 사용이 많아 복잡성이 가장 높았음을 의미한다. 분석 결과에 따르면, 4월과 7월에는 대체로 Z-점수가 낮아지며, 텍스트의 복잡성과 난이도가 점차 감소하여 11월 대수능 시험에 근접해지는 10월에는 거의 모든 지표에서 Z-점수가 오히려 가장 낮게

나타나는 것을 알 수 있다. 실제로 2023년 11월 16일에 실시된 2024학년도 수능 영어의 경우, 절대평가 도입 이래 1등급 비율이 최저인 4.17%로, 역대급 불수능이라는 평가를 받았다. 이러한 결과에는 여러 가지 원인이 있을 수 있지만, 수능 시험에 대비해 점차 난이도 높은 텍스트에 익숙해져야 하는 시기에 학평과 수능의 난이도 간 괴리로 인해 학생들이 수능의 실제 난이도에 충분히 대비하지 못한 것이 원인 중 하나로 사료된다. 3월 학평에서 가장 높은 텍스트 난이도를 보인 후, 10월 학평에서는 텍스트의 복잡성과 난이도가 크게 감소한다는 것은 학생들이 수능 시험의 실제 난이도를 파악하지 못하게 하며, 학평 결과를 기반으로 잘못된 기대와 판단을 불러일으킬 수 있다. 이에 학생들의 학습전략 및 시험 대비를 효과적으로 지원하기 위해서는 난이도도 점차 높이는 방향으로 조절하여 학년 내 연계성을 강화해야 할 것이다.

V. 논의 및 결론

본 연구는 재학생만 응시 가능한 2023학년도 고3전국연합학력평가에서 관측된 1등급 비율의 큰 변동 폭과 시험 회차가 진행될수록 지문의 난도가 점차적으로 증가해야 고3 학생들이 이에 적응하고 본수능 시험에 효과적으로 대비할 수 있다는 가정 하에 학년 내 연계성을 분석하였다. 특히, 대다수의 선행연구에서 주목받아온 한국교육과정평가원에서 주관하는 수능 혹은 수능모의평가가 아닌 시행시기별로 서로 다른 시도교육청에서 실시하는 학평의 시험 회차에 따른 난이도 변화 및 학년 내 연계성을 분석하였다는데 의의가 있다. 2023학년도 고3 학평의 학년 내 연계성의 다차원적 언어 특성을 분석하기 위해 Coh-Metrix의 기본 측정치, 어휘빈도수 및 다양성, 어휘속성, 텍스트 가독성, 통사구조, 참조정합성, 의미적 유사성 측정치를 중심으로 분석하였으며, 본고의 주요 결과는 다음과 같다.

첫째, 기본측정치 분석결과, 평균어휘수, 문장수, 문장 및 어휘길이 모두 시험 회차에 따라 일관된 증가나 변화 추세를 보이지 않았으며, 이러한 변화들은 통계적으로 유의미한 차이를 나타내지 않아 학년 내 연계성이 성립되지 않음을 의미한다. 둘째, 어휘빈도수 및 다양성(타입-토큰 비율) 모두에서 통계적으로 유의미한 차이가 나타나지 않아 학년 내에서 이 두 지표가 일관되게 조정되지 않고 있다는 것을 확인하였다. 이는 시험의 교육적 기능 관점에서 볼 때 학습자들의 언어적 발달과 이해 능력 향상을 위해 보다 체계적인 접근이 필요함을 시사한다. 셋째, 어휘속성의 경우 구체성, 친숙성, 심상성, 습득나이 측정치에 대한 지문의 학년 내 유의미한 연계성을 보이지 않음이 확인되었다. 텍스트 이해에 있어 각 어휘의 이미지화 가능성, 일상적 친숙성, 구체적인 표현의 명확성, 그리고 습득 연령 등은 언어 처리 과정에 중대한 영향을 제공한다. 시험의 학습효과를 극대화하기 위해서는 이러한 측정치들의 체계적인 조정과 단계적인

발달을 고려하여 점진적으로 언어 능력을 개발할 수 있는 기회를 제공해야 할 것이다. 통사구조의 복잡성 측정에서도 명사구 밀도와 본동사 앞 단어 수 등이 일정한 변화 패턴을 보이지 않아, 텍스트의 구조적 난이도가 시험회차별로 체계적으로 조정되지 않은 것으로 나타났다. 학습자의 언어 처리 능력과 구문 분석 능력의 체계적인 향상을 도모하기 위해 시간의 경과에 따라 텍스트의 구조적 난이도를 점진적인 조정이 요구된다. FRE 읽기 가독성 측정 결과, 텍스트의 난이도가 시간이 지나면서 오히려 점진적으로 낮아지는 경향을 보였으며, 점진적으로 독해 능력을 향상시킬 수 있도록 지문의 복잡성을 단계적으로 증가시켜야 할 것이다. 마지막으로 인접 문장 간 및 전체 문장 간의 참조 정합성과 의미정합성 모두 시험 회차를 거듭하며 체계적으로 조정되지 않았으며, 통계적으로 유의미한 차이가 관찰되지 않았음을 확인하였다. 이와 같은 결과는 향후 지문 개발 시 학생들의 학습 발달 수준에 맞춰 텍스트의 언어적 요소와 구조적 복잡성을 체계적으로 조정할 필요성을 시사하며, 독해 지문을 단순한 정보 전달의 수단을 넘어서, 학습자가 정보를 효과적으로 처리하고, 의미를 통합하고 분석할 수 있는 시험의 교육적 도구로 활용해야 할 것이다. 마지막으로, 2023학년도 고3 학평의 어휘비율 및 가독성 비교 분석 결과, 3월 학평에서는 학술어휘(AWL) 사용 비율이 7.87%로 가장 높았고, 어려운 어휘 비율 역시 18.35%로 가장 높았으며, Flesch 읽기 용이성 지수(FRE)와 Gunning Fog Index, SMOG Index 모두 가장 복잡하고 난이도가 높은 텍스트를 나타냈다. 반면, 10월 학평에서는 대부분의 지표에서 Z-점수가 가장 낮아져 텍스트의 복잡성과 난이도가 크게 감소하였다. 이와 같은 경향은 효과적인 학습 전략 수립 및 시험 대비를 어렵게 하며, 학생들이 수능 시험의 실제 난이도를 파악하지 못하게 할 수 있다. 비록 시험 회차별로 서로 다른 교육청에서 주관하나, 학평의 교육적 기능을 강화하기 위해서는 학평의 체계적인 난이도 조절을 통한 학년 내 연계성을 높여 효과적인 학습전략을 수립할 수 있도록 지원해야 할 것이다.

본 연구는 상기의 결과들로부터 도출한 시사점에도 불구하고 다음과 같은 한계점을 가진다. 첫째, 특정 학년도 문항만을 분석 대상으로 하였으므로 후속 연구에서는 절대평가가 적용된 2018학년도 이후부터 실시된 학평이 출제기관 간 유의미한 차이가 있는지, 각 학년 내 연계성이 이뤄지는 지 등 연구 범위를 확장할 필요가 있다. 또한, 본 연구는 Coh-Metrix에서 제공하는 다양한 표층적 및 심층적 난이도 지표를 활용하여 텍스트 분석만을 수행하였다. 그러나 전술하였듯이 시험난이도에 영향을 미치는 요인은 다면적이며, 향후 연구에서는 수험자의 배경, 선택지의 매력도, 인지적 요인과 같은 다른 변수들을 통합적으로 분석한다면 보다 다각적인 접근이 가능할 것이다. 예를 들면, Coh-Metrix 도구를 사용하여 언어적 지표들을 분석함과 동시에, 문항 반응 이론(Item Response Theory, IRT)을 적용한다면 각 문항이 가지는 난이도, 추측 가능성, 차별화 정도 등을 수치화하여 개별 문항의 특성이 수험자에게 어떻게 작용하는지 파악할 수 있을 것이다. 이와 같은 확장적 연구는 시험의 전반적인 난이도 조정 및 타당성 제

고에 기여할 수 있을 것으로 판단된다.

물론, 수능과 마찬가지로 학평, 모평과 같은 대단위 고부담 표준화 시험의 특성상 고도의 보안 유지를 위해 집합식으로 문항이 출제되기 때문에 검수 시간이 부족하여 난이도 조절이 어려울 것으로 예상된다. 그러나 읽기 자료 자체가 갖는 언어적 특성을 객관적 지표를 통해 분석하고 이를 평가할 수 있는 기반을 마련해야 향후 문항 개발에 참고할 수 있으며, 문항 개선 방향에 대한 논의에 도움을 줄 수 있을 것으로 판단된다. 이러한 점에서 본 연구는 시험의 교육적 기능관점에서 텍스트 난이도 분석을 기초 자료를 제공하였다는 점에서 의의가 있다고 사료된다.

참고문헌

고나은, 신정아. (2017). Coh-Metrix를 이용한 수능 영어 읽기 영역 지문 난이도 비교: EBS-수능 연계 정책 전후. *중등영어교육, 10*(4), 3-24.

김소정, 전문기. (2016). Coh-Metrix를 활용한 초등학교 6학년 영어학습자의 영작문 분석 연구. *현대영어교육, 17(*3), 263-287. http://doi.org/10.18095/meeso.2016.17.3.11

김용명. (2016). 수능 전국연합평가 영어 점수와 고등학교 영어 내신 성적과의 상관관계 분석. *예술인문사회융합멀티미디어논문지, 6*(8), 187-195. http://doi.org/10.14257/AJMAHS.2016.08.26

김재은, 최인철. (2015). 고등학교 영어 교과서, EBS 수능 연계 교재, 대학수학능력시험의 코퍼스기반 난이도 비교 분석. *멀티미디어 언어교육, 18*(1), 59-92. https://doi.org/10.15702/mall.2015.18.1.59

김정태. (2019). 대학수학능력시험 영어영역 절대평가제의 타당성 검증과 성공적 정착을 위한 제언. *영어평가, 14*(1), 41-59. http://doi.org /10.37244/ela.2019.14.41

김지은, 전문기. (2013). 중학교 영어교과서 듣기자료의 Coh-Metrix 측정치에 대한 연계성 분석. *Journal of the Korean Data Analysis Society, 15*(4), 1987-2000.

김현미, 이수정, 최항섭, 변희현, 정은주, 주형미. (2024). 학령인구 감소 시대 미래 학교교육의 방향 탐색: 미래학교 시나리오를 바탕으로. *교육과정평가연구, 27(*1), 1-28. https://doi.org/10.29221/jce.2024.27.1.1

류지수, 전문기. (2020). Coh-Metrix 프로그램을 활용한 2015년 개정 교육과정 중학교 1학년 영어 교과서 비교 분석. *영어학 20,* 89-121. http://doi.org/10.15738/kjell.20..202003.89

류지수, 전문기. (2023). Coh-Metrix 시스템을 활용한 고등학교 영어 모의고사 읽기 지문의 학년 간 연계성 분석. *영어어문교육, 29*(2), 41-59. http://doi.org/10.35828/etak.2023.29.2.41

서울특별시교육청. (2022, 11). *2023학년도 고등학교 전국연합학력평가 시행 기본 계획*. 서울특별시교육청.

성일호. (2014). 이독성공식과 Coh-Metrix를 활용한 우리나라 고등학교 영어교과서 이독성 분석. *영어영문학연구, 40(*4), 299-320. https://doi.org/10.21559/aellk.2014.40.4.015

신유선, 정영경. (2018). 코메트릭스(Coh-Metrix)를 활용한 공무원 7급과 9급 영어 시험 지문 분석: 어휘 다양성, 통사적 복잡성, 응집성과 이독성 지수를 중심으로. *영미어문학, 129*, 143-161. http://doi.org/10.21297/ballak.2018.129.143

양성현. (2017). 전국연합학력평가와 수능 모의평가의 이질성이 학교 교육에 미치는 영향: 수학 영역을 중심으로. *한국학교수학회, 20*(1), 1-18. http://doi.org/10.30807/ksms.2017.20.1.001

염영희, 김해동. (2023). 전국연합학력평가와 대학수학능력시험의 영어 읽기 지문 언어적 비교 분석. *영어영문학21, 36*(3), 175-198. https://doi.org/10.35771/engdoi.2023.36.3.008

이완기. (2015). 수능영어 절대점수제의 건설적 발전 방향 모색. *영어평가, 10*(1), 9-24.

이주윤. (2022). 절대평가 실시 이후 수능 영어 영역 문항 유형별 언어적 특성 및 난이도 비교. *영어평가, 17*(1), 45-63. https://doi.org/10.37244/ela.2022.17.1.45

이주윤. (2023). 고등학교 1학년 전국연합학력평가 영어 지문의 언어적 특성 및 난이도 분석. *영어평가, 18*(1), 189-211. http://doi.org/10.37244/ela.2023.18.1.189

이지연. (2022). Coh-Metrix를 활용한 EFL 대학생들의 영작문 분석 연구. *융합영어영문학, 7*(1), 29-60. http://doi.org/10.55986/cell.2022.7.1.29

이지희. (2024. 1. 29.) 주요대 정시 확대…수능 응시하는 고3 2024학년도 최고치. *한국대학신문*. https://news.unn.net/news/articleView.html?idxno=558827

이혜진, 이제영. (2018). 한·중·일 대학 입학시험 영어 지문의 이독성 비교 분석. *인문사회 21, 9*(5), 441-453. https://doi.org/10.22143/HSS21.9.5.32

장지연. (2018). Coh-Metrix를 통한 2017-2018 수능 영어 독해 지문 비교: 기초 산출치, 이독성, 용이성을 중심으로. *Foreign languages education, 25*(4), 81-106. http://doi.org/10.15334/FLE.2018.25.4.81

장지연. (2022). 2013-2022학년도 수능 영어 읽기 지문의 텍스트 유형 비교. *현대영어교육, 23*(3), 10-23. https://doi.org/10.18095/meeso.2022.23.3.10

전문기, 임인재. (2009). 코메트릭스(Coh-metrix)를 이용한 중학교 1학년 개정 영어 교과서의 코퍼스 언어학적 비교 분석. *영어교육연구, 21*(4), 265-292. https://doi.org/10.17936/PKELT.2009.21.4.012

중앙일보. (2023. 12. 11.). 난이도 조절 실패 수능, 사교육 의존만 심해진다. https://www.joongang.co.kr/article/25213711#home

최민주. (2018). *코메트릭스(Coh-Metrix)를 활용한 수능 영어 읽기문항의 응집성과 어휘정보 분석*[미간행 석사학위논문]. 한국교원대학교.

한국교육방송공사. (2023). EBSi 모의고사 2023 시험일정. https://www.ebsi.co.kr/ebs/xip/xipa/retrieveExmSchedRng.ebs?targetCd=D300

황이수, 이제영. (2020). 수학능력시험 영어 영역 읽기 지문의 특성과 문항 난이도와의 상관관계: 어휘 다양성, 연결사, 상황 모델을 중심으로. *인문사회 21, 11*(4), 1871-1884. http://doi.org/10.22143/HSS21.11.4.132

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Graesser, A. C., Jeon, M., Yan, Y., & Cai, Z. (2007). Discourse cohesion in text and tutorial dialogue. *Information Design Journal, 15*(3), 199-213. https://doi.org/10.1075/idj.15.3.02gra

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36*(2), 193-202. https://doi.org/10.3758/BF03195564

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

Zechmeister, E. B., Chronis, A. M., Cull, W. L., D'Anna, C. A., & Healy, N. A. (1995). Growth of a functionally important lexicon. *Journal of Reading Behavior, 27*(2), 201−212. https://doi.org/10.1080/10862969509547878

예시언어(Examples in): English
적용가능 언어(Applicable Languages): English
적용가능 수준(Applicable Levels): Secondary

이혜진(교수)
원광대학교 사범대학 영어교육과
54538 전북 익산시 익산대로 460
Email: lee.hyejin0419@gmail.com

류지수(교수)
건국대학교 다언어다문화연구소
05029 서울시 광진구 능동로 120
Email: jsryu0508@konkuk.ac.kr

전문기(교수)
건국대학교 문과대학 영어영문학과
05029 서울시 광진구 능동로 120
Email: mjeon1@konkuk.ac.k

Received: May 5, 2024
Reviewed: June 2, 2024
Revised version received: June 20, 2024