

절대평가 도입 전후 수능 영어 읽기 지문의 다차원 비교

장지연[†]

명지대학교

A Multi-Dimensional Comparison of CSAT English Reading Passages before and after the Use of Criterion-Referenced Assessment

Ji-Yeon Chang[†]

Myongji University

ABSTRACT

This study aims to compare register variation in the reading passages of the CSAT (College Scholastic Ability Test) English section before and after the use of criterion-referenced assessment by applying an additive multi-dimensional analysis based on Biber (1988). To this end, it compiled a corpus of the CSAT English reading passages of the 2013-2022 academic years, using the computer software MAT (Multidimensional Analysis Tagger, Version 1.3.2) (Nini, 2019) for analysis. The median dimension scores indicate that the passages were informational, non-narrative, and abstract. The passages did not tend to persuade their readers, and the referents were specified in an explicit manner. However, the wide range and distribution of the dimension scores suggest that various registers were used in the reading passages. Finally, there was no difference between the two assessment groups in all dimensions except for Dimension 3, implying that the consistency of the CSAT English reading passages has been relatively well maintained regarding register variation.

Keywords: additive multi-dimensional analysis, MAT, register variation, dimensions, CSAT English reading passages

1. 서 론

2018학년도 대학수학능력시험 영어 영역(이하 수능 영어)에 절대평가가 도입되면서 난이도를 어떻게 하면 일정하게 유지할 수 있는지에 대한 관심이 더욱 높아졌다. 당해

[†] Corresponding author: jchang@mju.ac.kr



시험을 치른 수험생들의 성적을 기반으로 등급이 결정되는 상대평가와 달리 절대평가에서는 다른 수험생들의 성적과 상관없이 본인의 성취 수준에 따라 등급이 결정된다. 그런데 절대평가에서 수험생의 성취 수준은 당해 수능 시험의 난이도에 크게 영향을 받을 수밖에 없다. 따라서 수능 영어에 절대평가가 도입되면서 난이도를 안정화하는 작업이 더욱 중요해지고 있는 상황이다. 난이도에 영향을 끼치는 중요 변수들 중의 하나는 언어라 할 수 있다. 특히 언어라는 변수는 수험생이 체감하는 난이도에 직접적으로 영향을 끼치면서 사전에 미리 통제할 수 있다는 특징이 있다. 그러다 보니 Coh-Metrix, AntWord Profiler, VocabProfile, Lexical Complexity Analyzer, L2 Syntactic Complexity Analyzer 등을 사용하여 수능 영어의 읽기 지문이나 듣기 지문에서 사용된 언어를 분석하거나 비교하는 연구들이 꾸준히 있어 왔다(Chang, 2018, 2019a, 2019b; Kim, 2017; Kim & Cha, 2021; Koh & Shin, 2017; Moon & Kim, 2017; Shin, 2019). 다만 기존 연구들에서는 어휘나 구문과 같은 단편적인 언어 형태에 초점을 맞추는 경향이 있었다.

Biber(1988)는 다차원 분석 방법을 사용하여 텍스트에서 동시에 사용되는 언어 특성들(language features)의 집단을 파악하고, 각 집단이 텍스트에 어떤 기능을 부여하는지 분석하여 이를 차원(dimension)으로 규정했다. 그가 사용한 다차원 분석 방법은 학술지 논문, 교과서, 잡지, 대화문과 같은 특정 장르가 언어사용역 면에서 어떤 특징을 지니고 다른 장르와 얼마나 차이가 나는지 구체적인 수치로 보여줄 수 있다(Biber & Finegan, 2001; Conrad, 2001; Helt, 2001; Nini, 2019; Reppen, 2001; Rey, 2001; de Souza, 2014). 최근에는 영어교육 분야에서도 학습 자료를 비교하거나 학습자의 언어 수행을 비교하는 데 Biber의 다차원 분석 방법이 활용되고 있다(Al-Surmi, 2012; Park & Hwang, 2019; Park & Kim, 2016, 2019; Yoon, 2012). 기존 수능 영어의 읽기 지문과 듣기 지문을 분석한 연구들이 언어 기능이나 언어사용역(register) 측면을 간과했던 점을 고려하여 본 연구에서는 Biber(1988)의 다차원 분석을 사용하여 2013-2022 학년도 수능 영어 읽기 지문이 차원별로 어떤 특징을 지니고 절대평가 도입 전후 차이가 있었는지 파악하고자 한다.

2. 이론적 배경 및 선행 연구

2.1. Biber(1988)의 다차원 분석

Biber(1988)는 특정 어휘나 구문과 같은 개별 언어 형태를 사전에 지정하여 텍스트를 분석하기보다 텍스트에서 체계적으로 함께 사용되는 언어 특성들(language features)을 파악하여 그 언어 특성들의 집합인 차원으로 텍스트를 파악하고자 했다. Biber는 특히

텍스트를 분석함에 있어 거시적 분석과 미시적 분석을 모두 실시했다. 거시적 분석으로 Biber는 텍스트에서 체계적으로 함께 사용되는 언어 특성들의 집합인 요인을 파악하여, 이를 통해 텍스트의 차이를 설명하고자 했다. 미시적 분석에서 Biber는 요인을 구성하는 언어 특성들이 텍스트에 어떤 언어 기능을 부여하는지 면밀히 분석하여 이를 차원으로 해석했다.

우선 다차원 분석을 위해 Biber는 2개의 문어 코퍼스(The Lancaster-Oslo-Bergen (LOB) Corpus of British English, 개인 서한(personal letters) 및 전문 서한(professional letters)으로 구성된 서한 코퍼스)과 1개의 구어 코퍼스(The London-Lund Corpus of Spoken English(LLC))에서 다양성과 대표성을 고려하여 23개의 장르에 해당하는 481개의 텍스트(총 960,000단어)를 추출했다. 다음으로 Biber는 16개의 주요 문법 항목에 속한 67개의 대표 언어 특성을 선정했다. 다특성/다차원(multi-feature/multi-dimensional, MF/MD) 접근 방식으로 언어사용역 변이를 연구하기 위해 Biber는 다변량 통계 기법(multivariate statistical technique)인 요인 분석(factor analysis)을 실시했으며, 그 결과를 7개의 요인으로 결정했다. 다만 Biber는 0.35 미만의 적재값을 갖는 주요 언어 특성(major language feature)은 중요하지 않다고 간주하여 그 값이 비록 유의미하더라도 요인에서 제외했다. 따라서 요인 7은 적재값이 .40을 넘는 주요 언어 특성이 없는 관례로 요인에서 제외되었다.

주요 언어 특성이 갖는 적재값은 해당 요인 뿐만 아니라 해당 요인에 속한 다른 언어 특성과의 관련성이 얼마나 큰지를 나타낸다. 가령 특정 언어 특성의 적재값이 크면 해당 요인의 변이에 미치는 영향이 크다고 해석된다. 또한 적재값은 방향성(±)을 지니며, 요인 내에서 서로 상보적(complementary) 관계에 있다. 즉, 각각의 요인에서 양의 방향성을 지닌 언어 특성의 적재값이 클수록, 음의 방향성을 지닌 언어 특성의 적재값은 줄어든다. 이와는 반대로 음의 방향성을 지닌 언어 특성의 적재값이 클수록, 양의 방향성을 지닌 언어 특성의 적재값은 줄어든다. 각 요인에 적재된 주요 언어 특성과 방향성은 부록 1에 제시되어 있다.

Biber는 각 요인에 적재된 주요 언어 특성이 갖는 적재값과 방향성을 기반으로 연구에 사용된 23개 장르의 요인 점수를 계산했으며(부록 2), 일반 선형 모형(general linear models)을 실시하여 각 장르 간 요인 점수의 차이가 모두 통계적으로 유의미한지 확인했다. 이로써 Biber는 총 6개의 요인을 확정했으며, 각각의 요인을 구성하는 주요 언어 특성을 질적으로 분석하여 이를 차원으로 해석했다.

2.2. 언어사용역 변이의 6개 차원(Biber, 1988)

Biber(1988)는 각각의 요인에 적재된 주요 언어 특성이 텍스트에 어떤 기능을 부여하는지 면밀히 분석하여 이를 차원으로 해석했다(표 1). 첫 번째 차원(Dimension 1, D1)은 관여성-정보성(involved vs. informational production)으로 텍스트의 생성 목적

과 상황을 보여준다. 가령 첫 번째 차원의 요인 점수가 음수이면 양의 방향성을 지닌 주요 언어 특성보다 음의 방향성을 지닌 주요 언어 특성이 해당 텍스트에서 비교적 더 많이 사용되었음을 보여준다. 첫 번째 차원에서 음의 방향성을 지닌 언어 특성은 명사, 단어 길이, 전치사, 타입/토큰 비율, 한정 형용사이다. 이러한 언어 특성은 많은 정보를 집약해서 전달하고, 어휘 및 구문을 신중하게 선택하거나 고칠 수 있는 상황에서 나타난다. 따라서 첫 번째 차원의 요인 점수가 음의 방향으로 클수록 해당 텍스트는 정보성을 더 많이 지닌다고 해석할 수 있다. 반대로 첫 번째 차원의 요인 점수가 양의 방향으로 클수록 해당 텍스트에서는 양의 방향성을 지닌 언어 특성이 비교적 더 많이 사용된 것이다. 이러한 텍스트는 비교적 감정을 전달하거나 상호작용을 위한 목적으로 생성되었으며 일반적인 내용을 담고 있다. 또한 상대적으로 시간적 여유가 없는 상황에서 생성된 텍스트로 해석한다.

표 1. 언어사용역 변이의 6개 차원(Biber, 1988)

차원(요인)	차원명((+)(-))
1	관여성-정보성(Involved vs. informational production)
2	서사성-비서사성(Narrative vs. non-narrative concern)
3	명시성-상황 의존성(Explicit vs. situated-dependent reference)
4	설득성(Overt expression of persuasion)
5	추상성-비추상성(Abstract vs. non-abstract information)
6	실시간 정보 정교성(On-line informational elaboration)

두 번째 차원에서 양의 방향성을 지닌 언어 특성은 과거형 동사, 3인칭 대명사, 완료 동사 등이며, 이러한 언어 특성은 서사적 행위를 나타낼 때 주로 사용된다. 따라서 두 번째 차원(Dimension 2, D2)은 서사성-비서사성(narrative vs. non-narrative concern)으로 명명되었다. 가령 두 번째 차원 점수가 양의 방향으로 클수록 해당 텍스트에서는 과거형 동사, 3인칭 대명사, 완료 동사 등이 동시에 많이 사용된 것이며, 해당 텍스트는 서사적 목적으로 생성된 것으로 해석한다. 반면, 두 번째 차원 점수가 음의 방향으로 클수록 해당 텍스트는 설명을 하거나 묘사를 하는 등 비서사적 목적으로 생성된 것으로 해석한다.

세 번째 차원(Dimension 3, D3)은 명시성(상황 독립성)-상황 의존성(explicit vs. situated-dependent reference)으로 텍스트에서 지시 대상을 지칭하는 방식과 관련이 있다. 가령 텍스트의 차원 점수가 음의 방향으로 큰 경우, 해당 텍스트에서 음의 방향성을 지닌 시간 부사어구, 장소 부사어구, 부사와 같은 주요 언어 특성이 동시에 많이 사용됐기 때문이다. 이는 지시 대상을 외부 상황(물리적 상황, 시간적 상황)에 의존하

여 지칭하는 경향이 크다고 볼 수 있다. 반면, 텍스트의 차원 점수가 양의 방향으로 큰 경우, 해당 텍스트에서는 지시 대상을 맥락에 의존하지 않고 관계절과 명사화 등을 통해 명시적으로 기술하는 것으로 보인다.

네 번째 차원(Dimension 4, D4)은 설득성(Overt expression of persuasion)으로 설득 동사, 예측 법조동사, 필요 법조동사 등이 동시에 사용되는 정도를 나타낸다. 이러한 언어 특성은 화자/저자의 입장을 보여주거나, 어떤 사건의 가능성 혹은 타당성을 평가하는데 사용된다. 따라서 설득성은 청자/독자를 설득하려는 의도가 텍스트에서 어느 정도 분명하게 드러나는지를 보여준다. 따라서 설득성의 차원 점수가 클수록 해당 텍스트는 청자/독자를 설득하려는 정도가 크다고 해석된다.

다섯 번째 차원(Dimension 5, D5)은 추상성-비추상성(abstract vs. non-abstract information)으로 명명되었다. 다섯 번째 차원의 점수가 양의 방향으로 큰 경우, 해당 텍스트에 접속사, 수동태, 과거 분사절 등이 동시에 빈번하게 사용되었기 때문이다. 그리고 이를 통해 해당 텍스트가 정보를 전달하는 담화로서 추상적이거나, 기술적이거나, 형식적임을 알 수 있다. 반면, 음의 방향으로 차원 점수가 큰 텍스트는 이러한 특징을 보이지 않는 것으로 간주한다.

마지막 여섯 번째 차원(Dimension 6, D6)은 실시간 정보 정교성(on-line informational elaboration)이며, 세 가지 종속절(동사 보문으로서의 *that*절, 형용사 보문으로서의 *that*절, 목적어 위치의 *that*-관계절)과 지시사가 동시에 빈번하게 사용될수록 차원 점수가 양의 방향으로 높게 나타난다. 이러한 언어 특성은 실시간으로 정보를 전달할 때 주로 나타난다. 즉, 시간적 여유가 있는 상황이라면 화자/저자는 많은 정보를 보다 적은 구문으로 압축하여 제시할 수 있지만, 시간의 압박을 받는 상황에서는 종속절을 추가하는 방식으로 정보를 전달하게 된다. 따라서 여섯 번째 차원의 점수가 양의 방향으로 큰 경우, 해당 텍스트는 시간적 여유가 없는 상태에서 비교적 느슨하고 단편적인 방식으로 정보를 전달하는 경향이 크다고 해석한다.

제한된 어휘나 구문 중심으로 텍스트를 비교했던 기존 연구들과 달리 Biber(1988)의 다차원 분석은 양적 방법과 질적 방법이 모두 사용된 연구 방식으로 텍스트를 가장 체계적이면서도 정확하게 분석하고 비교할 수 있는 모형이라 할 수 있다. Biber는 자신의 다차원 분석을 사용하여 텍스트를 비교할 때 다음 두 가지 사항을 강조했다. 첫째, 일부 차원을 선택하기보다 모든 차원을 동원하여 텍스트를 비교해야 텍스트 간 관계를 적절하게 이해할 수 있다. 둘째, 각 차원은 양과 음의 이분법 척도가 아니라 연속 척도로 이루어져 있으므로, 텍스트를 각 차원의 연속선 상에 놓고 비교해야 한다.

2.3. Biber(1988)의 다차원 분석 기반 텍스트 분석 연구

Biber(1988)의 다차원 분석은 특정 텍스트 집단의 언어사용역이 기존 장르와 비교하여 어떤 특징을 가지는지 구체적으로 보여줌으로써 그 실체를 규명하는 데 사용될 수

있다. 그러다 보니 특정 텍스트 집단에서 사용된 언어 특성들을 대상으로 자체적으로 요인 분석을 실시하고 그 결과물인 요인을 구성하는 언어 특성들을 다시 질적으로 해석하여 새로운 차원을 발견하는 완전한 다차원 분석(a full multi-dimensional analysis, a full MDA) 대신 Biber(1988)의 연구 결과를 토대로 차원 점수를 계산하는 가산적 다차원 분석(an additive multi-dimensional analysis, an additive MDA)(Sardinha, Pinto, Mayer, Zuppari, & Kauffmann, 2019)이 여러 연구에서 사용되었다. 가령 Nini(2019)는 법정에서 법과학 증거(forensic evidence)로 채택되는 욕설, 협박, 비방을 포함하는 악의적 법과학 텍스트(malicious forensic texts, MFTs)를 모아 코퍼스도 구축한 후 Biber(1988)의 다차원 분석을 가산적으로 실시했다. 즉, Nini는 Biber의 연구에서 각 차원을 구성하는 주요 언어 특성들을 악의적 법과학 텍스트에서 찾은 다음 Biber의 분석 결과를 사용하여 각각의 차원 점수를 계산했다. Biber의 연구와 비교했을 때 악의적 법과학 텍스트는 전반적으로 개인 서한(personal letters)보다는 전문 서한(professional letters)에 근접했다. 가령 해당 언어사용역은 전문 서한과 마찬가지로 비서사적이고, 텍스트에서 지시 대상을 상황에 의존하지 않고 명시적으로 지칭했다. 다만 악의적 법과학 텍스트의 특이사항은 차원 4에서 발견되는데, 개인 서한이나 전문 서한과 비교하여 높은 설득성을 보여주었다. 즉, 악의적 법과학 텍스트는 두 서한 장르와 비교하여 독자를 설득하려는 의도가 비교적 더 분명하게 드러나는 언어사용역이라 할 수 있었다.

Al-Surmi(2012)는 Biber(1988)의 분석 결과를 토대로 차원 점수를 계산하는 가산적 다차원 분석이 영어교육 분야에서 어떻게 사용될 수 있는지 그 가능성을 보여주었다. Al-Surmi는 영어 수업에서 수업 자료로 흔히 사용되는 TV 프로그램인 시트콤(sitcom)과 연속극(soap opera) 중 어떤 장르가 더 실제 대화와 유사하고 더 높은 진정성(authenticity)을 갖는지 파악하고자 Biber의 분석 결과를 사용하여 실제 대화, 시트콤, 연속극의 차원 점수를 계산했다. 차원 점수를 비교한 결과, 시트콤은 차원 1(관여성), 차원 4(설득성), 차원 5(추상성)에서, 연속극은 차원 2(서사성)에서 실제 대화와 유사했다. 명시성을 보여주는 차원 3에서는 시트콤과 연속극 모두 실제 대화와 큰 차이를 보이지 않았다. Al-Surmi는 시트콤이 더 많은 차원에서 실제 대화와 유사하므로 시트콤이 영어회화 수업 자료로 더 낫다는 단순한 결론 대신 의사소통에서 어떤 부분을 지도할 것인지에 따라 수업 자료 선택이 달라져야 한다고 제안했다. 가령 수업에서 상호작용이 중요한 대화를 지도할 때에는 시트콤을, 서사성, 설득성, 추상성이 요구되는 대화를 지도할 때에는 연속극을 선택하는 것이 좋다는 점을 구체적인 차원 점수를 통해 보여주었다.

Yoon(2012)은 2009 개정 중학교 교과서에서 사용된 구어 장르가 원어민 구어 장르에 비해 얼마나 진정성이 있는지 파악하기 위해 Biber(1988)의 분석 결과를 사용하여 가산적 다차원 분석을 실시했다. 이를 위해 Yoon은 국내 2009 개정 중학교 교과서 총 45권(15권/학년 × 3개 학년)에 포함된 듣기 및 말하기 지문을 모두 추출한 후 5개 장

르(대화, 전화통화, 인터뷰, 방송, 연설)로 분류하여 차원 점수를 계산했다. Biber(1988)의 연구 결과와 비교했을 때 국내 중학교 교과서 구어 장르는 정보전달 중심이거나(예: 인터뷰, 연설), 서술적이거나(예: 대화, 인터뷰), 지시 대상을 명시적으로 지칭하거나(예: 방송, 인터뷰), 청자를 좀 더 설득하려고 하거나(예: 대화), 추상적이었다(예: 인터뷰, 연설). 이 결과를 토대로 Yoon은 국내 중학교 교과서 구어 장르가 원어민 구어 장르에 비해 보다 문어체의 형태를 지닌다고 보았으며, 학습자에게 진정성 있는 학습 자료를 제공하는 데 한계가 있다고 지적했다.

Park과 Kim(2016, 2019) 역시 Biber(1988)의 연구 결과를 기반으로 하는 가산적 다차원 분석을 통해 영어 프레젠테이션 장르에서 한국인 대학생 학습자와 원어민 간 차이가 있는지 살펴보고자 했다. 우선 2016년 연구에서는 64명의 대학생들이 작성한 94개의 영어 프레젠테이션 원고를 분석 대상으로 삼았다. Biber(1988)의 연구 결과와 비교해 보았을 때 한국인 대학생 학습자의 영어 프레젠테이션은 차원 2(서사성), 차원 4(설득성), 차원 5(추상성)에서 두드러진 차이가 나타났다. 이러한 결과는 영어 프레젠테이션을 할 때 한국인 대학생 학습자가 원어민에 비해 스토리텔링과 설득력이 부족하고, 프레젠테이션 원고가 수동태 구문이나 분사 구문 등이 많이 사용되는 추상적인 텍스트에 가깝다는 점을 보여주었다. 다만 이 연구에서 사용한 원고는 자기소개, 정보전달, 설득이라는 상반된 목적을 지니고 있어서 연구 결과를 일반화하기에 한계가 있었다. 이에 2019년에 실시된 후속 연구에서는 자기소개를 제외한 정보전달과 설득의 목적으로 작성된 128편의 프레젠테이션 원고를 분석대상으로 삼아, 차이가 있는지 파악하고자 했다. 2016년 연구와 달리 2019년 연구에서 학습자 코퍼스는 원어민 코퍼스와 비교하여 차원 2(서사성), 차원 3(명시성), 차원 5(추상성)에서 차이가 발생했다. 차원 4에서 원어민 코퍼스는 공문서에 가까운 설득성 수치를 지녔는데 반해, 학습자 코퍼스 중 설득 프레젠테이션 원고는 신문 사설에 가까운 정도로 높은 설득성 수치를 보여주었다. 이러한 결과를 토대로 Park과 Kim(2019)은 한국인 대학생 학습자의 영어 프레젠테이션은 원어민의 연설에 비해 보다 단조롭고, 보다 추상적이며, 보다 어렵게 들릴 수 있음을 시사했다.

Biber(1988)의 분석 결과를 토대로 Park과 Kim(2016, 2019)이 영어 프레젠테이션 원고를 대상으로 가산적 다차원 분석을 실시했다면, Park과 Hwang(2019)은 영어 논쟁적 에세이(argumentative essays)를 대상으로 가산적 다차원 분석을 실시했다. Park과 Hwang은 한국 예비 대학생이 작성한 에세이 881편과 LOCNESS 중에서 미국 국적의 영어 원어민이 작성한 에세이를 골라 차원 점수를 계산한 후 그 결과를 비교했다. 이 연구에서 학습자 코퍼스와 원어민 코퍼스 간 분명한 차이는 차원 1(관여성)과 차원 4(설득성)에서 발견되었다. 가령 원어민의 에세이에 비해 한국 예비 대학생들의 에세이에서는 관여성 정도가 즉흥 연설이나 인터뷰 정도로 높았던 반면, 설득성 정도는 소설이나 전화 대화 정도로 비교적 낮았다. 즉, 한국 예비 대학생들이 작성한 영어 논쟁적 에세이는 격식적인 문어 장르보다는 비격식적인 구어 장르에 근접했다고 볼 수 있

었다.

상기 연구들이 보여주듯이 Biber(1988)의 다차원 분석 결과를 사용하여 차원 점수를 계산하는 가산적 다차원 분석은 비록 새로운 차원을 제시하지는 않지만, 특정 텍스트 집단의 언어사용역이 기존 장르와 얼마나 유사하고 다른지 차원별 점수를 통해 다각도로 확인할 수 있다. 따라서 Biber의 연구를 기반으로 하는 가산적 다차원 분석은 다양한 텍스트의 언어사용역을 구체적으로 파악하고 비교하는 데 이용될 수 있다. 기존 수능 영어의 읽기 지문이나 듣기 지문을 분석한 연구들(Chang, 2018, 2019a, 2019b; Kim, 2017; Kim & Cha, 2021; Koh & Shin, 2017; Moon & Kim, 2017; Shin, 2019)은 아직까지 특정 어휘, 문장 구조, 응집성 등과 같은 단편적인 언어 형태에 의존했다는 점에서 한계가 있다. 따라서 본 연구에서는 절대평가 도입 전후 수능 영어 읽기 지문에 Biber의 다차원 분석을 가산적으로 실시하고, 그 결과를 언어사용역 측면에서 비교하고자 한다.

본 연구에서는 기존 연구들(Al-Surmi, 2012; Biber & Finegan, 2001; Park & Hwang, 2019; Park & Kim, 2016, 2019; Reppen, 2001)과 마찬가지로 마지막 여섯 번째 차원을 제외한 차원 1-5를 분석에 사용했다. 차원 6인 실시간 정보 정교성은 시간적 여유가 없는 상황에서 언어 사용자가 정보를 실시간으로 전달하는 경향이 얼마나 큰지를 보여주는 차원으로서 본 연구의 분석 대상인 읽기 지문에는 적합하지 않은 것으로 보였다. 따라서 본 연구의 연구 문제는 다음과 같다.

- (1) 수능 영어 읽기 지문은 차원 1에서 어떤 특징을 보이며, 절대평가 도입 전후 차이가 있는가?
- (2) 수능 영어 읽기 지문은 차원 2에서 어떤 특징을 보이며, 절대평가 도입 전후 차이가 있는가?
- (3) 수능 영어 읽기 지문은 차원 3에서 어떤 특징을 보이며, 절대평가 도입 전후 차이가 있는가?
- (4) 수능 영어 읽기 지문은 차원 4에서 어떤 특징을 보이며, 절대평가 도입 전후 차이가 있는가?
- (5) 수능 영어 읽기 지문은 차원 5에서 어떤 특징을 보이며, 절대평가 도입 전후 차이가 있는가?

3. 연구 방법

3.1. 코퍼스 구축

본 연구에서는 절대평가가 적용된 수능 영어 읽기 지문과 상대평가가 적용된 수능

영어 읽기 지문을 비교하기 위해 절대평가가 도입된 2018학년도(2017년 실시) 전후 각각 5년간의 수능 영어 읽기 지문을 추출하여 코퍼스로 구축했다. 따라서 본 연구에서 사용된 코퍼스는 2013학년도(2012년 실시)부터 2022학년도(2021년 실시)까지 총 10년간 실시된 수능 시험의 영어 읽기 지문을 포함한다. 2014학년도 수능 영어의 경우, 기존 시험보다 예외적으로 쉽게 출제된 A형과 기존 시험과 비슷한 난이도로 출제된 B형으로 나뉘어 시행되었으며(Ministry of Education, Science and Technology, 2012),¹⁾ 본 연구에서는 B형의 읽기 지문을 선택했다.

코퍼스를 구축하기 위해 우선 각 학년도 수능 영어 기출문제(출수형)와 정답지를 한국교육과정평가원 대학수학능력시험 홈페이지(<http://www.suneung.re.kr>)에서 다운로드 받은 후 읽기 지문을 골라 분석이 가능하도록 ANSI로 인코딩된 텍스트 파일(TXT)로 변환했다. 본 연구에서는 언어사용역 분석이 적절하게 이루어질 수 있도록 읽기 지문을 다음과 같이 변경했다.

- (1) 선택지는 포함하지 않았다.
- (2) 표와 그림은 포함하지 않았다.
- (3) 읽기 지문의 내용을 요약한 요약문은 포함하지 않았다(예: 2022학년도 40번 문제 요약문).
- (4) 번호(예: ①, ②, (a), (b) 등)와 불필요한 기호(예: ·, ※ 등)는 삭제했다.
- (5) 정답지를 바탕으로 전체 흐름과 관계없는 문장은 삭제했다.
- (6) 유니코드 포맷 문자는 ANSI로 인코딩된 문서로 저장될 수 있도록 유사한 기호로 교체했다.
- (7) 영어가 아닌 기타 외국어 알파벳(예: é)은 ANSI로 인코딩된 문서로 저장될 수 있도록 유사한 알파벳(예: e)으로 교체했다.
- (8) 정답지를 바탕으로 어법 상 틀린 것을 맞는 것으로 교체했다.
- (9) 정답지를 바탕으로 문맥상 낱말의 쓰임이 적절하지 않은 것을 적절한 것으로 교체했다.
- (10) 정답지를 바탕으로 주어진 문장이 들어가기에 가장 적절한 곳에 주어진 문장을 추가했다.
- (11) 정답지를 바탕으로 지문의 빈 칸에 가장 적절한 선택지를 추가했다.
- (12) 정답지를 바탕으로 읽기 지문의 내용을 순서에 맞게 배열했다.
- (13) 지문의 단락이 바뀌면 텍스트의 줄이 바뀌도록 설정했다.
- (14) 지문에서 동일 단락 안에서는 텍스트의 줄이 바뀌지 않도록 설정했다.

1) 2014학년도 수능 영어가 A형과 B형으로 나뉘게 된 계기는 영어과 교육과정 성취 기준의 달성 정도를 측정하는 것 외에 실용 영어 능력(A형)과 기초 학술 영어 능력(B형)을 각각 평가하기 위함이었으며, 평가 목표 뿐만 아니라 대상자, 평가 내용, 출제 범위 면에서도 분명한 차이가 있었다(Lee, 2013).

본 연구에서는 각각의 수능 영어 읽기 지문의 차원 점수를 파악할 수 있도록 지문별로 파일을 생성했다. 그 결과, 총 250개의 ANSI 텍스트 파일이 생성되었으며, 전체 토큰 수는 총 40,182개였다(표 2).

표 2. 수능 영어 읽기 지문 코퍼스(2013-2022학년도)

학년도 (유형)	평가	지문 수	단어 수(Number of tokens)					
			합계	평균	표준편차	중앙값	최소값	최대값
2013	상대	30	4,467	148.90	49.18	144.00	108	382
2014(B)	상대	20	3,530	176.50	57.93	162.50	126	395
2015	상대	25	3,801	152.04	46.03	148.00	93	308
2016	상대	25	3,989	159.56	48.34	155.00	105	351
2017	상대	25	4,066	162.64	51.50	159.00	94	363
2018	절대	25	4,163	166.52	60.22	158.00	105	415
2019	절대	25	3,997	159.88	50.89	154.00	83	347
2020	절대	25	4,077	163.08	51.82	163.00	87	363
2021	절대	25	4,027	161.08	49.61	162.00	87	347
2022	절대	25	4,065	162.60	49.14	161.00	110	349
전체		250	40,182	160.73	50.99	156.00	83	415

3.2. 코퍼스 분석

본 연구에서는 다차원 분석을 통해 텍스트 집단의 새로운 차원을 밝혀내는 것이 목적이 아니라, 기존 다차원 분석 결과를 토대로 텍스트 집단의 변이 정도를 파악하는 것이 목적이다. 따라서 본 연구에서는 자체적으로 요인 분석을 하는 것이 아니라, Biber(1988)의 다차원 분석 결과를 토대로 하는 가산적 다차원 분석(additive multi-dimensional analysis, additive MDA)(Sardinha, Pinto, Mayer, Zupparadi, & Kauffmann, 2019)을 실시했다. 이를 위해 본 연구에서는 MAT(Multidimensional Analysis Tagger, Version 1.3.2)(Nini, 2019) 컴퓨터 프로그램을 사용했다. 언어사용역 분석(register analysis) 프로그램인 MAT는 Biber(1988)의 다차원 분석 결과를 기반으로 입력된 텍스트를 자동으로 분석하며, 윈도우 운영체제 기반 무료 소프트웨어로서 누구나 손쉽게 이용이 가능하다. MAT는 특히 사이즈가 작거나 구조화되지 않은 코퍼스를 분석할 수 있다는 장점이 있다. 따라서 MAT는 수능 영어 읽기 지문을 분석하는 데 있어서 매우 적합한 도구라 할 수 있다.

MAT에서는 우선 the Stanford Tagger(Toutanova, Klein, Manning, & Singer, 2003)를 사용하여 입력된 텍스트의 모든 품사(parts of speech)를 태깅한다. 참고로 the Penn Treebank POS Tagset를 기반으로 하는 the Stanford Tagger는 여러 연구에서 폭넓게 사용되고 있으며(Lu, 2014), 97.24%의 정확도를 가진 것으로 알려졌다(Toutanova, Klein, Manning, & Singer, 2003). MAT에서는 태깅한 결과를 바탕으로 각각의 차원을 구성하는 주요 언어 특성(부록 1)의 빈도수를 100단어 기준의 빈도수로 정규화(normalization)한 후 표준 점수(standardized score)인 z -점수를 계산하는 데, 이때 Biber(1988, pp. 77-78)의 평균과 표준편차를 사용한다. 마지막으로 MAT에서는 각각의 차원을 구성하는 주요 언어 특성의 z -점수를 사용하여 차원 점수를 계산한다. 본 연구에서는 Biber(1988)의 다차원 분석을 사용했던 기존 연구들(Al-Surmi, 2012; Biber & Finegan, 2001; Park & Hwang, 2019; Park & Kim, 2016, 2019; Reppen, 2001)과 마찬가지로 마지막 차원 6(실시간 정보 정교성)은 분석에서 제외했다. 따라서 분석에 사용된 차원은 총 5개이다(차원 1-5).

최대한 Biber(1988)의 연구에 근접하기 위해 본 연구에서는 길이가 매우 짧은 텍스트에서 드물게 사용되는 언어 특성이 차원에 지나치게 큰 영향을 끼치지 않도록 하기 위해 5보다 큰 z -점수를 5로 고정하는 z -점수 보정(Z-score correction)은 실시하지 않았다. 또한 Biber(1988)의 연구에서 사용된 VASW 태그(VASW tags)만을 분석에 사용했다. 참고로 Biber(1988)의 연구에서는 1차원(관여성-정보성) 점수를 계산할 때 타입/토큰 비율(type/token ratio, TTR)이 포함되는 데, 전체 텍스트의 타입/토큰 비율이 아니라 텍스트에 등장하는 첫 400단어의 타입/토큰 비율이 사용되었다. 그리고 400단어보다 짧은 텍스트의 타입/토큰 비율은 차원 점수를 계산할 때 제외되었다. 따라서 본 연구에서도 역시 400단어 이상인 2018학년도 43-45번 문제 지문(415개)을 제외한 나머지 지문의 타입/토큰 비율은 차원 1(관여성-정보성)의 점수를 계산할 때 포함되지 않았다.

본 연구에서는 수능 영어 읽기 지문의 가산적 다차원 분석 결과가 절대평가 도입 이전(2013-2017학년도)과 이후(2018-2022학년도)에 차이가 있는지 알아보고자 했다. 이를 위해 분석 결과를 상대평가 집단(NRA, norm-referenced assessment)과 절대평가 집단(CRA, criterion-referenced assessment)으로 나누었다. 우선 10개의 모든 부분 집단(상대평가 집단의 차원 1-5 점수, 절대평가 집단의 차원 1-5 점수)에 Shapiro-Wilk 정규성 검정을 실시했는데, 두 부분 집단(절대평가 집단의 차원 1 점수와 차원 3 점수)을 제외한 나머지 부분 집단이 모두 정규분포를 따르지 않았다.²⁾ 이러한 점을 고려하여 본 연구에서는 평균 대신 중앙값을 사용하는 비모수검정인 Mann-Whitney U -검정을 실시했다($\alpha = .05$). 통계 분석에는 R(Version 4.1.2)이 사용되었으며, 각 차원에 따른 학년도별 차원 점수의 기술 통계 결과는 부록 3에 제시되어 있다.³⁾

2) 그림 1, 3, 5, 7, 9에서 각 집단의 평균은 x 로 표시되어 있다.

3) 지면 부족으로 인해 본 연구에서 사용된 R 코드 및 데이터는 저자의 이메일로 요청 시 제공될 예정이다.

4. 연구 결과 및 논의

4.1. 차원 1(관여성-정보성)의 차원 점수 비교

본 연구에서는 두 집단(NRA, CRA)의 차원 점수를 통해 차원별 특징을 살펴보고, 상대평가 집단(NRA)과 절대평가 집단(CRA) 간 차원 점수의 차이가 통계적으로 유의미한지 확인하고자 했다. 우선 차원 1은 관여성-정보성으로 텍스트의 생성 목적과 상황을 보여준다. 표 3과 그림 1에서 알 수 있듯이 NRA와 CRA의 최대값은 각각 24.81(2013학년도)과 21.39(2018학년도)였으며, 두 수치는 모두 Biber(1988)의 23개 장르 중 개인서한(PsL, 19.5)의 평균값보다는 높았으며, 면대면 대화(FC, 35.3)의 평균값보다는 낮았다(그림 2). 즉, 각 집단의 최대값을 갖는 읽기 지문은 두 장르의 중간 정도로 감정을 드러내거나 독자와의 상호작용을 하기 위해 작성된 텍스트임을 알 수 있다.

표 3. 절대평가 도입 전후 차원 1(관여성-정보성)의 차원 점수 비교

	상대평가(NRA)	절대평가(CRA)
지문 수	125	125
평균	-2.75	-3.96
표준편차	9.69	8.88
중앙값	-4.18	-5.88
최소값	-21.41	-22.98
최대값	24.81	21.39
범위	46.22	44.37
U	8,406.00	
Z	1.04	
p	0.30	
r	0.07	

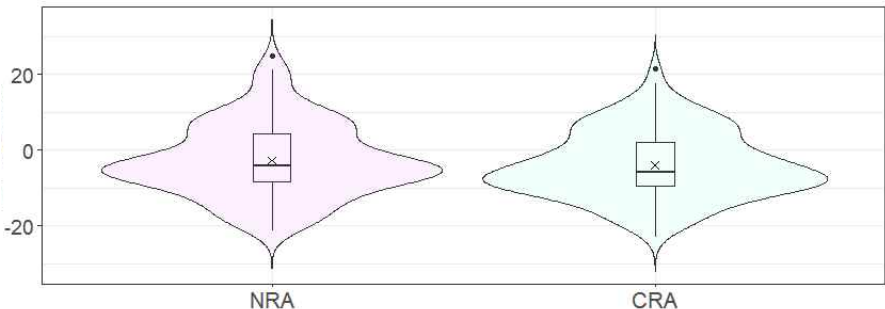
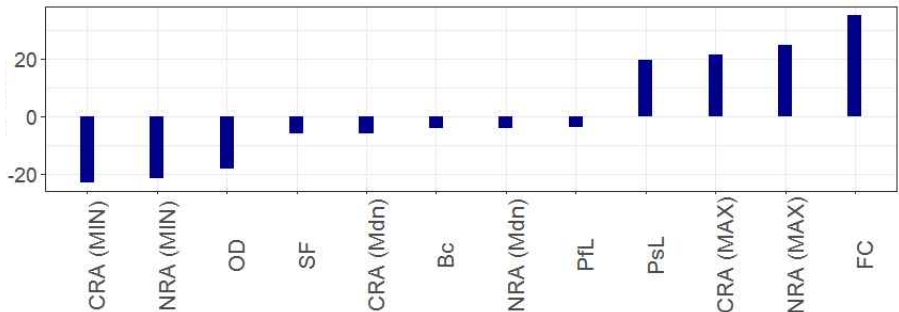


그림 1. 절대평가 도입 전후 차원 1(관여성-정보성)의 차원 점수 비교

NRA와 CRA의 최소값은 각각 -21.41(2016학년도)과 -22.98(2019학년도)이었으며, 두 수치 모두 공문서(OD, -18.1)의 평균값보다 낮게 나타났다. 이러한 결과는 두 읽기 지문이 공문서보다 어휘와 구문을 더 신중하게 선택하거나 고칠 수 있는 상황에서 많은 정보를 집약해서 전달하기 위해 작성된 텍스트임을 시사한다. 두 집단의 점수 범위와 분포를 보면(그림 1), 두 집단의 차원 1 점수가 음(-)의 방향으로 약간 치우쳐 있긴 하지만 양쪽 방향(+/-)에 고루 위치하고 있음을 확인할 수 있다. 이는 그만큼 다양한 수준의 관여성 텍스트와 정보성 텍스트가 각 집단에서 사용되었음을 의미한다.



OD: Official Documents, SF: Science Fiction, Bc: Broadcasts, Pfl: Professional Letters, PsL: Personal Letters. FC: Face-to-face Conversations.

그림 2. 장르에 따른 차원 1(관여성-정보성)의 차원 점수 비교

본 연구에서는 대부분의 부분 집단이 정규분포를 보이지 않으므로 차원 점수의 평균 대신 중앙값을 사용하여 절대평가 도입 전후 수능 영어 읽기 지문이 차이가 있는지 살펴해보았다. 우선 NRA와 CRA의 중앙값은 각각 -4.18과 -5.88로 관여성보다는 정보성을 지녔다. 상대평가 집단의 중앙값은 방송(Bc, -4.3)과 전문 서한(Pfl, -3.9) 사이에 놓여 있었으며, 절대평가 집단의 중앙값은 공상과학 소설(SF, -6.1)에 근접하는 정보성을 보여주었다. 근접한 장르는 달랐지만 Mann-Whitney *U*-검정을 실시한 결과, NRA와 CRA 간 중앙값의 차이는 통계적으로 유의미하지 않았다($U = 8,406.00$, $Z = 1.04$, $p = 0.30$, $r = 0.07$).

4.2. 차원 2(서사성-비서사성)의 차원 점수 비교

서사성-비서사성을 보여주는 차원 2에서 NRA와 CRA의 최대값은 각각 10.24(2015학년도)와 15.58(2019학년도)이었으며, CRA의 최대값이 비교적 매우 높게 나타났다

(표 4, 그림 3). 두 수치는 모두 소설 장르 중에서 서사성의 평균 수치가 가장 높은 로맨스 소설(RF, 7.2)을 훨씬 뛰어넘는 수준이었으며(그림 4), 그만큼 과거형 동사, 3인칭 대명사, 완료 동사 등을 동시에 많이 사용하여 과거에 일어난 사건을 이야기하는 읽기 지문이라는 점을 의미한다. NRA와 CRA의 최소값은 모두 -7.38이었으며, 이 수치는 Biber의 23개 장르 중 가장 서사성 수치가 낮은 방송(Bc, -3.3)보다도 낮았다. 이는 그만큼 과거형 동사, 3인칭 대명사, 완료 동사 등이 동시에 많이 사용되지 않았으며, 설명을 하거나 묘사를 하는 읽기 지문임을 보여준다. 최소값 수치를 가진 읽기 지문은 여러 학년도(2013, 2015, 2017, 2019, 2020, 2021)에서 발견되었다.

표 4. 절대평가 도입 전후 차원 2(서사성-비서사성)의 차원 점수 비교

	상대평가(NRA)	절대평가(CRA)
지문 수	125	125
평균	- 1.52	- 2.13
표준편차	4.10	4.44
중앙값	- 2.19	- 3.19
최소값	- 7.38	- 7.38
최대값	10.24	15.58
범위	17.62	22.96
<i>U</i>	8,717.00	
<i>Z</i>	1.58	
<i>p</i>	0.11	
<i>r</i>	0.10	

차원 2 점수의 범위와 분포(그림 3)를 보면 상당 부분이 비서사성으로 치우쳐 있긴 하지만 그림에도 양쪽 방향으로 퍼져 있었다. 즉, 각 집단에서 다양한 수준의 서사성 텍스트와 비서사성 텍스트가 사용되었음을 알 수 있다. 참고로 차원 2에서는 절대평가 집단에서만 극단적으로 양의 수치가 높은 이상치(outlier)가 몇 군데 발견되었다.⁴⁾ 이는 절대평가 집단에서 다른 지문과 비교했을 때 극단적으로 서사성이 높은 지문이 있었음을 의미한다.

4) 그림 1, 3, 5, 7, 9에서 각 집단의 이상치는 동그란 점으로 표시되어 있다.

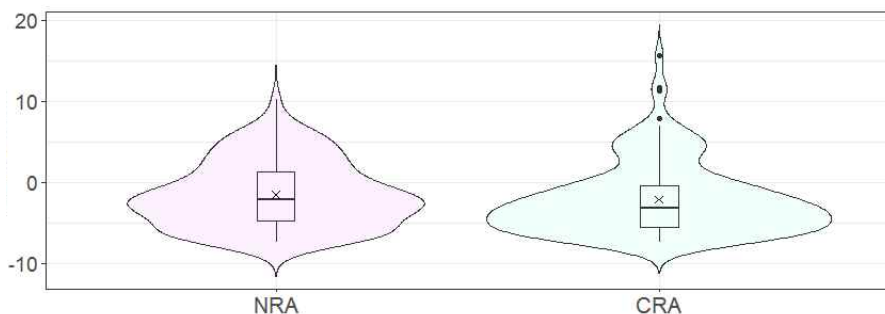
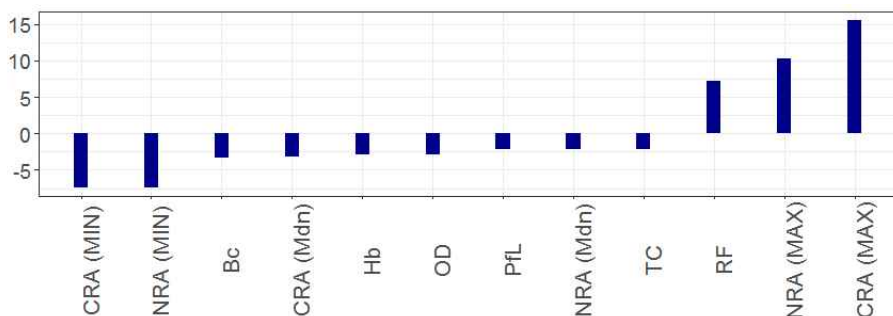


그림 3. 절대평가 도입 전후 차원 2(서사성-비서사성)의 차원 점수 비교

NRA와 CRA의 중앙값은 각각 -2.19, -3.19로서, 이는 각각 전문 서한(PfL, -2.2)과 방송(Bc, -3.3)의 비서사성에 근접한 수치였다. 비록 CRA의 최대값이 비교적 매우 높았고 NRA에는 없는 이상치가 CRA에서 발견되었지만 Mann-Whitney *U*-검정을 실시한 결과, 차원 1과 마찬가지로 차원 2 역시 NRA와 CRA 간 중앙값의 차이는 통계적으로 유의미하지 않았다($U = 8,717.00$, $Z = 1.58$, $p = 0.11$, $r = 0.10$). 즉, 상대평가 집단과 절대평가 집단이 서사성 차원에서 다르다고 볼 수 없었다.



Bc: Broadcasts, Hb: Hobbies, OD: Official Documents, PfL: Professional Letters, TC: Telephone Conversations, RF: Romantic Fiction

그림 4. 장르에 따른 차원 2(서사성-비서사성)의 차원 점수 비교

4.3. 차원 3(명시성-상황 의존성)의 차원 점수 비교

명시성(상황 독립성)-상황 의존성을 나타내는 차원 3에서 NRA와 CRA의 최대값은 각각 24.33(2016학년도)과 17.52(2022학년도)로, NRA의 최대값이 비교적 매우 높게

나타났다(표 5, 그림 5). 두 최대값은 모두 Biber의 23개 장르 중 공문서(OD, 7.3)를 훨씬 뛰어넘는 수치로서(그림 6) 지시 대상을 주변 맥락에 의존하지 않고 관계절과 명사화 등을 통해 명시적으로 기술함을 의미한다. 반면 NRA와 CRA의 최소값은 각각 -7.74(2016학년도)와 -8.05(2019학년도)로 전화 대화(TC, -5.2)와 방송(Bc, -9.0) 사이 정도의 수치였다. 즉, 해당 지문은 읽기 지문임에도 불구하고 구두 담화 장르와 마찬가지로 시간 부사어구, 장소 부사어구, 부사 등을 통해 맥락에 의존하여 지시 대상을 지칭하는 정도가 컸다고 볼 수 있다. 그림 5를 보면 두 집단의 차원 점수가 양쪽 방향으로 고루 분포되어 있지만, 절대평가 집단의 점수가 보다 범위가 좁았고, 보다 양의 방향(명시성)으로 치우쳐 있었다.

표 5. 절대평가 도입 전후 차원 3(명시성-상황 의존성)의 차원 점수 비교

	상대평가(NRA)	절대평가(CRA)
지문 수	125	125
평균	3.74	4.67
표준편차	5.64	4.80
중앙값	3.12	4.95
최소값	-7.74	-8.05
최대값	24.33	17.52
범위	32.07	25.57
<i>U</i>	6,608.00	
<i>Z</i>	-2.11	
<i>p</i>	0.04	
<i>r</i>	-0.13	

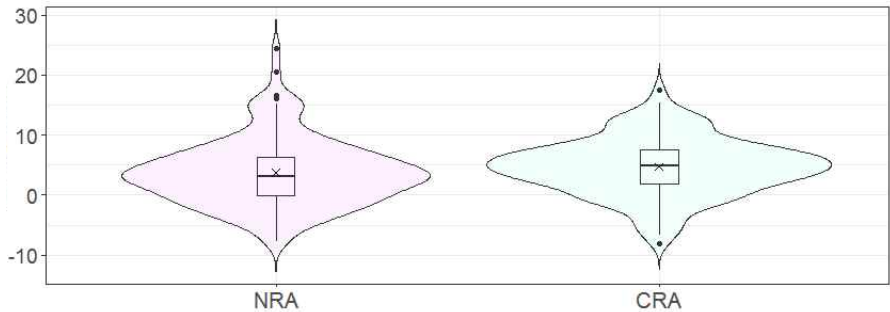
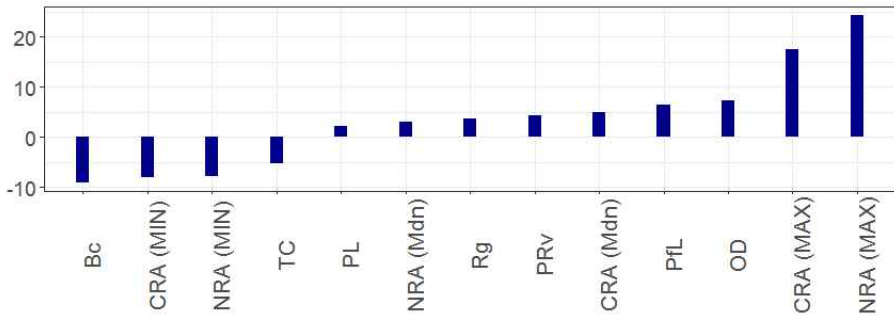


그림 5. 절대평가 도입 전후 차원 3(명시성-상황 의존성)의 차원 점수 비교



Bc: Broadcasts, TC: Telephone Conversations, PL: Popular Lore, Rg: Religion, PRv: Press Reviews, PfL: Professional Letters, OD: Official Documents

그림 6. 장르에 따른 차원 3(명시성-상황 의존성)의 차원 점수 비교

비록 최대값은 NRA가 CRA보다 매우 높았으나, 중앙값은 반대로 CRA가 NRA보다 높았다. 가령 NRA의 중앙값은 3.12로서, 이 수치는 대중 설화(PL, 2.3)보다는 높고 종교(Rg, 3.7)보다는 낮은 수준의 명시성이었다. 반면 CRA의 중앙값은 4.95로 언론 리뷰(PRv, 4.3)보다는 높고 전문 서한(PfL, 6.5)보다는 낮은 수준의 명시성이었다. Mann-Whitney *U*-검정을 실시한 결과, 다른 차원과 달리 NRA와 CRA 간 중앙값의 차이가 통계적으로 유의미했다($U = 6608.00$, $Z = -2.11$, $p = 0.04$, $r = -0.13$). 즉, 절대평가 집단의 명시성이 상대평가 집단의 명시성보다 더 높았다고 볼 수 있다.

4.4. 차원 4(설득성)의 차원 점수 비교

차원 4인 설득성의 경우, NRA와 CRA의 최대값은 각각 15.57(2015학년도)과 17.03(2019학년도)이었으며(표 6, 그림 7), 두 수치 모두 전문 서한(PfL, 3.5)을 훨씬 뛰어 넘는 매우 높은 수치였다(그림 8). 반면, NRA와 CRA의 최소값은 모두 -9.27로서 가급적 중립성을 지키려는 장르인 방송(Bc, -4.4)보다도 낮았다. -9.27의 최소값 수치를 지닌 읽기 지문은 2022학년도(최소값: -7.07)를 제외한 나머지 학년도에서 모두 발견되었다. NRA 및 CRA의 설득성 점수는 폭넓게 분포되어 있었으며(그림 7), 그만큼 독자를 설득하려는 텍스트와 그렇지 않은 텍스트가 수능 영어 읽기 지문에 모두 나타났음을 알 수 있다.

표 6. 절대평가 도입 전후 차원 4(설득성)의 차원 점수 비교

	상대평가(NRA)	절대평가(CRA)
지문 수	125	125
평균	-0.59	0.61
표준편차	5.56	5.60
중앙값	-1.08	-0.54
최소값	-9.27	-9.27
최대값	15.57	17.03
범위	24.84	26.30
U	6,770.50	
Z	-1.82	
p	0.07	
r	-0.12	

비록 양쪽 집단에서 모두 극단적으로 설득성 수치가 높은 이상치가 발견되긴 했지만 (그림 7), NRA와 CRA의 중앙값은 각각 -1.08, -0.54로 설득성을 보이지 않았다. 우선 NRA의 중앙값 -1.08은 모험 소설(AF, -1.2)에 근접했으며, CRA의 중앙값 -0.54는 학술 산문(AP, -0.5)에 상응한 수치였다. 이렇게 설득성 차원에서 근접한 장르를 다르더라도 Mann-Whitney U -검정을 실시한 결과, NRA와 CRA 간 중앙값의 차이는 통계적으로 유의미하지 않아($U = 6,770.50$, $Z = -1.82$, $p = 0.07$, $r = -0.12$) 상대평가 집단과 절대평가 집단이 설득성 차원에서 다르다고 볼 수는 없었다.

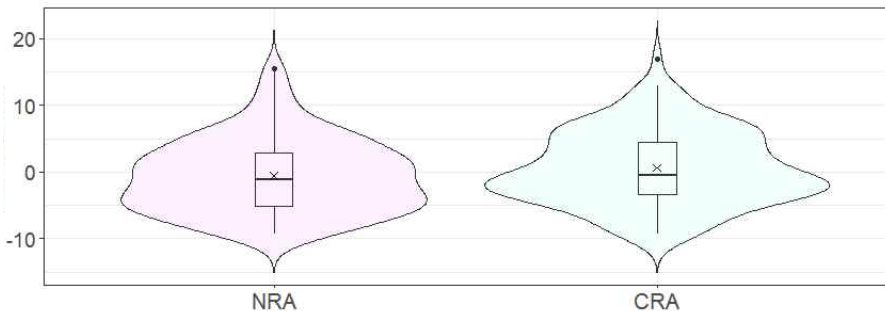
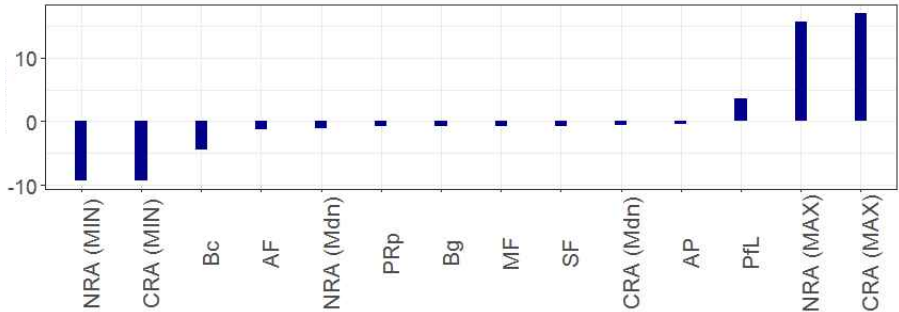


그림 7. 절대평가 도입 전후 차원 4(설득성)의 차원 점수 비교



Bc: Broadcasts, AF: Adventure Fiction, PRp: Press Reportage, Bg: Biographies, MF: Mystery Fiction, SF: Science Fiction, AP: Academic Prose, Pfl: Professional Letters

그림 8. 장르에 따른 차원 4(설득성)의 차원 점수 비교

4.5. 차원 5(추상성-비추상성)의 차원 점수 비교

마지막으로 차원 5는 추상성의 정도를 가리키며 NRA와 CRA의 최대값은 각각 26.50(2014B학년도)과 25.59(2022학년도)였다(표 7, 그림 9). 이들 수치는 Biber(1988)의 장르 중 가장 높은 추상성을 보이는 학술 산문(AP, 5.5) 장르보다도 훨씬 높은 값이었다(그림 10). 이러한 결과는 해당 읽기 지문에 접속사, 무행위자 수동태, 과거 분사 절, BY-수동태 등과 같이 추상성을 보여주는 언어 특성이 그만큼 동시에 많이 사용되었기 때문이다. NRA와 CRA의 최소값은 모두 -3.92였으며, 이 수치는 전화 대화(TC, -3.7)와 유사했다. 최소값 -3.92를 갖는 읽기 지문은 2022학년도(최소값: -3.03)를 제외한 나머지 학년도에서 모두 발견되었다. 차원 5 점수의 범위와 분포를 보면(그림 9), 다른 차원과 마찬가지로 다양한 수준의 추상성 혹은 비추상성을 갖는 텍스트가 두 집단에서 모두 사용되었음을 알 수 있다. 대부분의 차원 점수가 0에 근접하여 분포되어 있긴 하지만, 두 집단에서 모두 극단적으로 수치가 높은 이상치가 발견되었다. 이는 그만큼 다른 지문들에 비해 극단적으로 추상성이 높은 지문이 각 집단에 있었음을 의미한다.

표 7. 절대평가 도입 전후 차원 5(추상성-비추상성)의 차원 점수 비교

	상대평가(NRA)	절대평가(CRA)
지문 수	125	125
평균	3.00	3.91
표준편차	6.03	5.82
중앙값	1.64	2.67
최소값	- 3.92	- 3.92
최대값	26.50	25.59
범위	30.42	29.51
<i>U</i>	6,884.00	
<i>Z</i>	- 1.62	
<i>p</i>	0.10	
<i>r</i>	- 0.10	

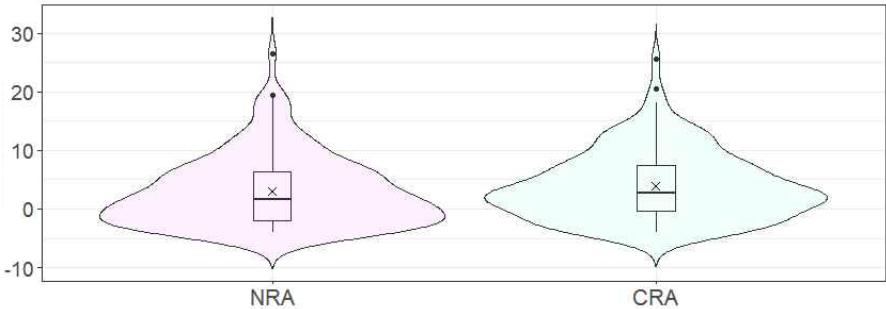
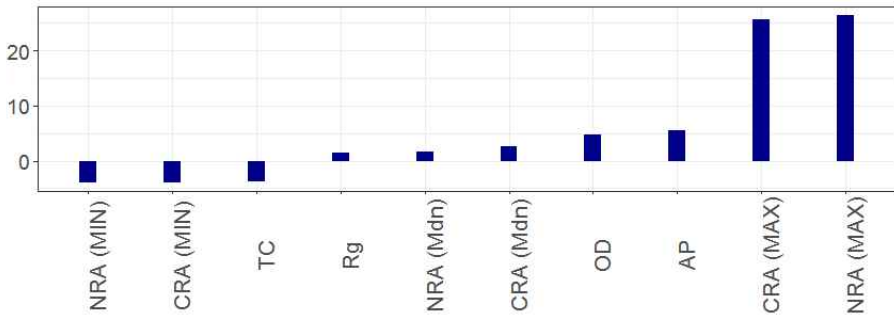


그림 9. 절대평가 도입 전후 차원 5(추상성-비추상성)의 차원 점수 비교

NRA와 CRA의 중앙값은 각각 1.64와 2.67로 종교(Rg, 1.4) 장르의 추상성에 근접했으며, 공문서(OD, 4.7) 장르의 추상성보다는 낮게 나타났다. 마지막으로 Mann-Whitney U-검정을 실시한 결과, NRA와 CRA 간 중앙값의 차이가 통계적으로 유의미하지 않아($U = 6,884.00$, $Z = -1.62$, $p = 0.10$, $r = -0.10$) 추상성 면에서 두 집단 간 차이가 존재한다고 볼 수 없었다.



TC: Telephone Conversations, Rg: Religion, OD: Official Documents, AP: Academic Prose

그림 10. 장르에 따른 차원 5(추상성-비추상성)의 차원 점수 비교

5. 결론 및 제언

기존 수능 영어 읽기 지문이나 듣기 지문을 분석한 연구들(Chang, 2018, 2019a, 2019b; Kim, 2017; Kim & Cha, 2021; Koh & Shin, 2017; Moon & Kim, 2017; Shin, 2019)이 특정 어휘나 구문과 같은 언어 형태의 차이에 주목했던 반면, 본 연구는 Biber(1988)의 다차원 분석을 사용하여 언어사용역 측면에서 수능 영어 읽기 지문을 여러 차원에서 분석하고, 그 결과를 Biber의 장르 분석 결과와 비교했다. 각 차원별 중앙값을 보았을 때 우선 차원 1(관여성-정보성)에서 수능 영어 읽기 지문은 독자와의 상호작용보다 전문 서한, 방송, 공상과학 소설 장르에 근접할 정도로 정보를 집약해서 전달하는 경향이 있었다. 차원 2(서사성-비서사성)에서 수능 영어 읽기 지문의 중앙값은 방송 및 전문 서한 장르와 유사했으며, 과거에 일어난 사건을 이야기하기보다 설명을 하거나 묘사를 하는 등 비서사적 성격을 보였다. 차원 4(설득성)에서 수능 영어 읽기 지문은 모험 소설 및 학술 산문 장르와 마찬가지로 독자를 설득하려는 목적을 지닌다고 볼 수 없었다. 차원 5(추상성-비추상성)에서 수능 영어 읽기 지문은 공문서 장르보다는 약하지만 종교 장르와 유사한 추상성을 보여주었다.

지시 대상을 지칭하는 방식(차원 3)에서 수능 영어 읽기 지문은 시간 부사어구, 장소 부사어구, 부사와 같은 언어 특성을 사용하여 주변 상황에 의존하기보다 관계절 및 명사화 등을 통해 대상을 텍스트 안에서 명시적으로 밝히려는 경향이 있었다. 다만 다른 차원과 달리 차원 3에서는 절대평가 도입 전후 중앙값에서 통계적으로 유의미한 차이가 있었으며 절대평가 시기가 상대평가 시기보다 좀 더 명시성이 높게 나타났다. 가령 상대평가 시기의 중앙값은 대중 설화와 종교 사이에, 절대평가 시기의 중앙값은 언론 리뷰와 전문 서한 사이에 위치했다. 각 차원별 점수 범위와 분포를 보았을 때 수능 영

어 읽기 지문은 모든 차원에 걸쳐 그 폭이 매우 넓게 나타났다. 특히 점수 범위와 분포 모두 한쪽 방향(+/-)으로 치우치지 않고 양과 음에 모두 걸쳐 있었다는 결과는 그만큼 수능 영어 읽기 지문이 전반적으로 언어사용역 측면에서 상당히 다양했다는 점을 보여준다.

본 연구에서는 절대평가 도입 전후 수능 영어 읽기 지문이 언어사용역 측면에서 차이가 나는지 차원별로 나누어 비교했으며, 지시 대상을 지칭하는 방식인 차원 3(명시성-상황 의존성)을 제외한 나머지 차원에서는 그 차이가 발견되지 않았다. 이러한 결과는 어휘나 구문과 같은 언어 형태 뿐만 아니라(Chang, 2018, 2019a, 2019b; Shin, 2019), 언어사용역 측면에서도 역시 수능 영어 읽기 지문이 평가 방식과 상관없이 대체로 일관성을 잘 유지하고 있다는 점을 의미한다. 후속 연구에서는 본 연구에 포함되지 않은 절대평가 도입 전후 수능 영어 듣기 지문을 언어사용역 측면에서 분석하고 비교해볼 수 있을 것이다. 또한 그 결과를 본 연구의 결과와 비교하여 수능 영어에서 나타나는 구두 언어와 문자 언어의 차이가 어느 정도인지 차원별로 파악해볼 수 있을 것이다.

마지막으로 본 연구의 분석 결과를 기반으로 수능 영어 읽기 지문이 영어로 된 교양 및 전공 대학교재 혹은 고등학교 영어 교과서와 언어사용역 측면에서 얼마나 연계성을 지니는지 파악할 수 있다. 대학수학능력시험의 영어 영역은 말 그대로 수험생이 영어 면에서 대학 수학 능력을 어느 정도 지녔는지 평가하는 시험이면서 동시에 영어과 교육과정 성취 기준을 어느 정도 달성했는지 측정하는 시험이다(Chang, 2019a, 2019b; Lee, 2013). 따라서 수능 영어가 각각의 역할을 얼마나 잘 수행하고 있는지 검증하기 위해서는 수능 영어 이전과 이후에 수험생이 경험하는 텍스트와의 비교가 필요하다. 가령 읽기 지문 측면에서 대학 수학 능력을 판단하기 위해서는 대학에서 수험생이 읽을 영어 원서 교재와의 비교가 필요할 것이며, 영어과 교육과정 이수 정도를 판단하기 위해서는 수능 영어의 출제 범위인 영어 I, 영어II(Korea Institute for Curriculum and Evaluation, 2021)와의 비교가 필요할 것이다. 이러한 연구는 수능 영어가 실제로 어떤 역할을 하는 시험인지 파악하는 데 도움이 되리라 본다.

Biber(1988)의 다차원 분석은 학술지 논문, 교과서, 잡지, 대화문과 같은 특정 텍스트 집단의 성격을 규명하거나 텍스트 집단 간 차이를 비교하기 위해 폭넓게 사용되어 왔다(Biber & Finegan, 2001; Conrad, 2001; Helt, 2001; Nini, 2019; Reppen, 2001; Rey, 2001; de Souza, 2014). 최근에는 영어 교육 현장에서 사용되는 학습 자료를 비교하거나 학습자의 구두 언어나 문자 언어를 비교하는 데에도 점차 그 활용 폭을 넓히고 있다(AI-Surmi, 2012; Park & Hwang, 2019; Park & Kim, 2016, 2019; Yoon, 2012). 본 연구에서 알 수 있듯이 Biber(1988)의 차원(관여성-정보성, 서사성-비서사성, 명시성-상황 의존성, 설득성, 추상성-비추상성, 실시간 정보 정교성)은 특정 텍스트 집단이 언어사용역 측면에서 어떤 특징을 지니고 어떤 장르와 얼마만큼 유사한지 구체적인 수치로 보여줄 수 있다. 비록 본 연구에서는 수능 영어 읽기 지문의 분석 결과를

Biber의 장르 분석 결과에 비추어 논의하는 데 그쳤으나, 학습자가 교실 안팎에서 경험하는 다양한 텍스트 집단에 대한 연구가 이루어질수록 더욱더 의미 있는 논의가 가능할 것이다. 이를 통해 좀 더 진정성 있는 교재 및 학습 자료를 선정하거나 개발할 수 있을 뿐만 아니라(Al-Surmi, 2012; Yoon, 2012), 학습자에게 요구되는 영어 사용 능력을 다각도로 측정할 수 있는 평가 지문 또한 적절하게 선정하거나 개발할 수 있을 것이다.

References

- Al-Surmi, M. (2012). Authenticity and TV shows: A multidimensional analysis perspective. *TESOL Quarterly*, 46(4), 671-694.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D., & Finegan, E. (2001). Intra-textual variation within medical research articles. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 108-123). New York: Routledge.
- Chang, J.-Y. (2018). A comparison of 2017-2018 CSAT reading passages via Coh-Metrix: Focusing on descriptive, readability, and easibility measures. *Foreign Languages Education*, 25(4), 81-106.
- Chang, J.-Y. (2019a). 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.-Y. (2019b). A comparison of syntactic complexity in CSAT reading passages before and after the introduction of criterion-referenced evaluation. *Journal of the Korea English Education Society*, 18(2), 161-188.
- Conrad, S. (2001). Variation among disciplinary texts: A comparison of textbooks and journal articles in biology and history. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 94-107). New York: Routledge.
- de Souza, R. C. (2014). Dimensions of variation in TIME magazine. In T. B. Sardinha & M. V. Pinto (Eds.), *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. 177-193). Philadelphia: John Benjamins.
- Helt, M. E. (2001). A multi-dimensional comparison of British and American spoken English. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 171-183). New York: Routledge.
- Kim, G.-H., & Cha, K.-W. (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, J.-r. (2017). A diachronic analysis of English KSAT reading passages. *Journal of the*

Elementary Education Society, 27, 63-78.

- Koh, N., & Shin, J.-A. (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.
- Korea Institute for Curriculum and Evaluation. (2021, March 16). *Announcement of the basic plan for the implementation of the 2022 College Scholastic Ability Test* [Press release].
- Lee, B. C. (2013). A study on the optimization of the CSAT categorization framework with a focus on item analysis. *Modern English Education*, 14(1), 267-296.
- Lu, X. (2014). *Computational methods for corpus annotation and analysis*. New York: Springer.
- Ministry of Education, Science and Technology. (2012, June 15). *Announcement of the basic plan for the 2014 College Scholastic Ability Test* [Press release].
- Moon, J.-H., & Kim, H.-D. (2017). An analysis of the linguistic elements of the text in the English reading section of the College Scholastic Ability Test. *Modern English Education*, 18(1), 193-211.
- Nini, A. (2019). The multi-dimensional analysis tagger. In T. B. Sardinha & M. V. Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 67-94). London: Bloomsbury Academic.
- Park, K., & Hwang, E. (2019). A multi-dimensional factor analysis of pre-university students' argumentative essays. *Modern English Education*, 20(3), 129-138.
- Park, K., & Kim, E.-J. (2016). A multi-dimensional analysis of Korean college English learners' speech transcript. *Studies in Foreign Language Education*, 30(3), 25-52.
- Park, K., & Kim, E.-J. (2019). A study of university English learners' oral presentation with multi-dimensional analysis: Focusing on informative and persuasive presentation. *Studies in Foreign Language Education*, 33(3), 1-26.
- R Core Team (2021). R: A language and environment for statistical computing (Version 4.1.2) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Reppen, R. (2001). Register variation in student and adult speech and writing. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 187-199). New York: Routledge.
- Rey, J. M. (2001). Changing gender roles in popular culture: Dialogue in Star Trek episodes from 1966 to 1993. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 138-156). New York: Routledge.
- Sardinha, T. B., Pinto, M. V., Mayer, C., Zuppari, M. C., & Kauffmann, C. H. (2019). Adding registers to a previous multi-dimensional analysis. In T. B. Sardinha & M. V. Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 165-186). London: Bloomsbury Academic.
- 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.
- Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech

tagging with a cyclic dependency network. In M. Hearst & M. Ostendorf (Eds.), *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology* (Vol. 1, pp. 173-180). Edmonton, Canada: Association for Computational Linguistics.

Yoon, H. (2012). A multi-dimensional analysis of oral registers in revised middle school English textbooks. *Korean Journal of English Language and Linguistics*, 12(4), 777-799.

장지연

부교수(전문)

미래융합대학 창의융합인재학부

명지대학교

03674 서울특별시 서대문구 거북골로 34

E-mail: jchang@mju.ac.kr

접수일자 : 2022. 2. 17

수정본 접수 : 2022. 4. 11

게재결정 : 2022. 4. 18

부록 1. 차원별 주요 언어 특성(Biber, 1988, pp. 89-90)

요인	차원	방향	주요 언어 특성
1	관여성 대 정보성	+	사적 동사, THAT 생략, 축약, 현재 시제 동사, 2인칭 대명사, 대동사 DO, 분석 부정(NOT), 지시 대명사, 일반 강조, 1인칭 대명사, 대명사 IT, 본동사 BE, 이유를 나타내는 종속 접속사, 담화 표지어, 부정 대명사, 일반 헤지, 증폭어, 관계절, WH-의문문, 가능성을 나타내는 법조동사, 절과 절을 연결하는 AND, WH-절, 문미 전치사
		-	명사, 단어 길이, 전치사, 타입/토큰 비율(TTR), 한정 형용사
2	서사성 대 비서사성	+	과거 시제 동사, 3인칭 대명사, 완료 동사, 공적 동사, 종합 부정(NO/NEITHER/NOR), 현재 분사절
		-	해당 사항 없음
3	명시성 대 상황 의존성	+	목적어 위치의 WH-관계절, 전치사 + 관계사, 주어 위치의 WH-관계절, 구와 구를 연결하는 AND, 명사화
		-	시간 부사어구, 장소 부사어구, 부사
4	설득성	+	부정사, 예측 법조동사, 설득 동사, 조건을 나타내는 종속 접속사, 필요 법조동사, 분리 조동사
		-	해당 사항 없음
5	추상성 대 비추상성	+	접속사, 무행위자 수동태, 과거 분사절, BY-수동태, WHIZ가 생략된 과거분사, 기타 종속 접속사
		-	해당 사항 없음
6	실시간 정보 정교성	+	동사 보문으로서의 THAT-절, 지시사, 목적어 위치의 THAT-관계절, 형용사 보문으로서의 THAT-절
		-	해당 사항 없음

부록 2. 장르별 평균 차원 점수(Biber, 1988, pp. 122-125)

장르	D1	D2	D3	D4	D5
언론 보도(Press Reportage, PRp)	-15.1	0.4	-0.3	-0.7	0.6
언론 사설(Press Editorials, PE)	-10.0	-0.8	1.9	3.1	0.3
언론 리뷰(Press Reviews, PRv)	-13.9	-1.6	4.3	-2.8	0.8
종교(Religion, Rg)	-7.0	-0.7	3.7	0.2	1.4
취미(Hobbies, Hb)	-10.1	-2.9	0.3	1.7	1.2
대중 설화(Popular Lore, PL)	-9.3	-0.1	2.3	-0.3	0.1
자서전(Biographies, Bg)	-12.4	2.1	1.7	-0.7	-0.5

부록 2. 계속

장르	D1	D2	D3	D4	D5
공문서(Official Documents, OD)	-18.1	-2.9	7.3	-0.2	4.7
학술 산문(Academic Prose, AP)	-14.9	-2.6	4.2	-0.5	5.5
일반 소설(General Fiction, GF)	-0.8	5.9	-3.1	0.9	-2.5
미스터리 소설(Mystery Fiction, MF)	-0.2	6.0	-3.6	-0.7	-2.8
공상과학 소설(Science Fiction, SF)	-6.1	5.9	-1.4	-0.7	-2.5
모험 소설(Adventure Fiction, AF)	0.0	5.5	-3.8	-1.2	-2.5
로맨스 소설(Romantic Fiction, RF)	4.3	7.2	-4.1	1.8	-3.1
유머(Humor, Hm)	-7.8	0.9	-0.8	-0.3	-0.4
개인 서한(Personal Letters, PsL)	19.5	0.3	-3.6	1.5	-2.8
전문 서한(Professional Letters, Pfl)	-3.9	-2.2	6.5	3.5	0.4
면대면 대화(Face-to-face Conversations, FC)	35.3	-0.6	-3.9	-0.3	-3.2
전화 대화(Telephone Conversations, TC)	37.2	-2.1	-5.2	0.6	-3.7
인터뷰(Interviews, Iv)	17.1	-1.1	-0.4	1.0	-2.0
방송(Broadcasts, Bc)	-4.3	-3.3	-9.0	-4.4	-1.7
즉흥 연설(Spontaneous Speeches, SS)	18.2	1.3	1.2	0.3	-2.6
준비된 연설(Prepared Speeches, PS)	2.2	0.7	0.3	0.4	-1.9

부록 3. 학년도별 차원 점수의 기술 통계 결과

차원	학년도(유형)	지문 수	평균	표준편차	중앙값	최소값	최대값	범위
D1	2013	30	- 3.44	9.70	- 5.22	- 18.44	24.81	43.25
	2014(B)	20	- 0.84	9.45	- 3.27	- 19.71	19.72	39.43
	2015	25	- 3.45	9.29	- 2.80	- 20.42	10.97	31.39
	2016	25	- 3.86	11.35	- 4.18	- 21.41	20.62	42.03
	2017	25	- 1.64	8.88	- 4.73	- 18.58	21.10	39.68
	2018	25	- 4.23	10.75	- 8.10	- 19.50	21.39	40.89
	2019	25	- 4.20	9.56	- 6.36	- 22.98	17.66	40.64
	2020	25	- 3.72	7.82	- 4.45	- 18.02	8.22	26.24
	2021	25	- 1.30	9.33	- 1.64	- 19.68	14.12	33.80
	2022	25	- 6.36	6.22	- 7.10	- 17.39	7.95	25.34

부록 3. 계속

차원	학년도(유형)	지문 수	평균	표준편차	증양값	최소값	최대값	범위
D2	2013	30	-0.27	4.00	-1.15	-7.38	5.92	13.30
	2014(B)	20	-2.00	2.94	-1.95	-7.14	6.24	13.38
	2015	25	-1.72	4.57	-3.07	-7.38	10.24	17.62
	2016	25	-2.08	4.49	-2.45	-7.18	7.84	15.02
	2017	25	-1.90	4.12	-2.93	-7.38	7.42	14.80
	2018	25	-2.75	4.11	-4.21	-7.00	7.86	14.86
	2019	25	-1.50	5.21	-3.19	-7.38	15.58	22.96
	2020	25	-2.59	4.04	-2.93	-7.38	11.31	18.69
	2021	25	-1.94	4.36	-2.09	-7.38	11.69	19.07
	2022	25	-1.84	4.64	-3.75	-7.02	6.34	13.36
D3	2013	30	2.20	5.83	2.23	-7.31	15.23	22.54
	2014(B)	20	3.77	3.98	3.26	-1.81	14.32	16.13
	2015	25	4.28	5.25	3.48	-4.03	16.11	20.14
	2016	25	3.83	6.77	2.33	-7.74	24.33	32.07
	2017	25	4.94	5.70	4.76	-6.70	16.54	23.24
	2018	25	3.78	5.02	4.46	-6.61	14.27	20.88
	2019	25	4.17	4.44	4.15	-8.05	11.87	19.92
	2020	25	4.95	4.13	5.32	-3.58	13.65	17.23
	2021	25	4.14	4.40	3.75	-5.14	12.11	17.25
	2022	25	6.30	5.80	6.46	-6.58	17.52	24.10
D4	2013	30	-0.73	5.17	0.53	-9.27	11.93	21.20
	2014(B)	20	-0.30	5.79	-1.31	-9.27	13.21	22.48
	2015	25	-0.66	6.00	-2.52	-9.27	15.57	24.84
	2016	25	-1.14	5.88	-1.91	-9.27	15.10	24.37
	2017	25	-0.05	5.43	0.38	-9.27	10.30	19.57
	2018	25	1.09	4.93	0.65	-9.27	10.88	20.15
	2019	25	-0.44	6.32	-0.93	-9.27	17.03	26.30
	2020	25	0.14	5.43	-1.85	-9.27	12.08	21.35
	2021	25	0.55	5.73	-0.01	-9.27	11.99	21.26
	2022	25	1.69	5.69	1.73	-7.07	12.92	19.99

부록 3. 계속

차원	학년도(유형)	지문 수	평균	표준편차	증양값	최소값	최대값	범위
D5	2013	30	3.61	7.17	1.22	-3.92	19.34	23.26
	2014(B)	20	2.65	6.77	1.31	-3.92	26.50	30.42
	2015	25	3.01	5.37	2.33	-3.92	13.92	17.84
	2016	25	3.51	5.84	2.39	-3.92	18.82	22.74
	2017	25	2.01	4.98	0.97	-3.92	16.88	20.80
	2018	25	2.62	5.71	1.27	-3.92	13.97	17.89
	2019	25	4.66	5.68	4.35	-3.92	20.54	24.46
	2020	25	4.29	5.32	3.07	-3.92	14.53	18.45
	2021	25	4.22	5.79	3.27	-3.92	18.10	22.02
	2022	25	3.79	6.76	1.75	-3.03	25.59	28.62