

수능 영어 문항 유형간 응집력과 어휘정보 분석

An Analysis of Cohesion and Word Information among English CSAT Question Types

최민주*, 김정렬**

한국교원대학교 영어교육과*, 한국교원대학교 초등교육과**

Minju Choi(minju9570@google.co.kr)*, Jeong-ryeol Kim(jrkim@knue.ac.kr)**

요약

본 연구는 대학수학능력시험 영어영역 읽기지문의 문항 유형별 응집력과 어휘정보를 파악하기 위해 코퍼스 기반 분석을 실시하였다. 본 연구에서는 한국교육과정평가원에서 제시하는 여섯 가지의 문항유형을 세 가지의 범주, 즉 맥락읽기, 세부읽기, 간접쓰기로 나누어 분석하였다. 이를 위해 처음 수능이 실시된 1994학년도부터 2017학년도까지의 수능 영어영역의 읽기지문의 코퍼스를 구축한 후, 코메트릭스 3.0을 활용하여 해당 코퍼스를 분석하여 각 문항들의 응집력과 어휘정보의 차이를 살펴보았다. 연구 결과 표층응집성 지표에서는 명사중복 측정치와 어간중복 측정치에서 통계적으로 유의미한 차이를 보였다. 연결사 지표에서는 역접의 연결사와 추가의 연결사에서 통계적으로 유의미한 차이를 보였다. 어휘정보 지표의 결과에서는 대명사의 발생정도, 습득나이 측정치, 내용어의 구체성 측정치, 심상성 측정치, 유의미성 측정치에서 통계적으로 유의미한 차이를 보였다. 이러한 문항유형간 응집력과 어휘정보 분석에 대한 정보는 교과서 집필 및 수능시험 출제에 활용될 수 있을 것이며, 학생들이 직접적으로 활용할 수 있는 읽기전략으로 활용 될 것으로 기대된다.

■ 중심어 : | 대학수학능력시험 | 코메트릭스 | 응집력 |

Abstract

The aim of this study was to analyze cohesion and word information among different types of questions in the English reading section of the College Scholastic Ability Tests (CSAT). The types of questions were divided into three categories: macro reading, micro reading, and indirect writing. Reading texts from 1994 to 2017 CSAT were analyzed by Coh-Metrix, an automated evaluation program of text and discourse. The findings of this study indicated that there were statistical differences among the three categories of questions for noun overlap, stem overlap, adversative and contrastive connective, additive connective, pronoun incidence, age of acquisition, concreteness for content word, imagability, and meaningfulness. The information of the findings bore pedagogic implications for developing textbooks, questions for CSAT, and reading strategies by students.

■ keyword : | CSAT | Coh-Metrix | Cohesion |

I. 서론

한국교육과정평가원에서는 2009개정 교육과정에 제시된[1] 성취기준에 상응하는 다양한 수능 문항을 제시하고 있다. 교육과정에서 제시하는 성취근거에 따른 수능문항의 유형은 대의파악, 문법·어휘, 세부사항, 상호작용, 간접쓰기, 복합으로 나누어 제시하고 있다[2]. 다양한 종류의 문항을 제한된 시간 내에 풀기 위해서는 무엇보다도 학습자들은 시험 문항 유형에 대해서 잘 이해하고 각 유형에 맞는 적절한 읽기 전략을 사용할 필요가 있다[3]. 이에 따라 한국교육과정평가원에서는 각 문항유형에 맞는 학습전략을 소개하고 있지만[4] 학습자들이 학습전략을 활용하여 직접 마주하는 각 문항의 유형에 대한 언어학적인 내용이 제시되어있지 않다. 언어학적인 내용의 부족은 수능문제 출제 및 교과서 집필에 있어 일관성 있는 출제를 어렵게 하는 요인이 될 수 있다. 더불어 영어뿐만 아니라 다양한 과목을 공부해야 하는 수험생 입장에서 이러한 부족함은 문제를 야기할 수 있다. 장경숙(2004)은 대학수학능력시험의 난이도가 해마다 급격하게 달라짐으로 인하여 사회적으로 큰 파장을 불러일으켜 왔다고 말하며[5] 수능의 일관성 있는 출제경향의 필요성에 대해 말하였다. 이러한 문제점을 고려해볼 때, 수능의 문항을 분석하여 각 문항이 가지는 특성을 파악할 수 있는 객관적인 자료를 추출해야할 필요가 있다.

이러한 객관적인 자료를 추출하기 위해 국내에서는 코퍼스를 활용하여 객관성과 정확성을 높이는 연구가 증가하였다. 하지만 코퍼스를 활용한 수능분석 연구들의 대다수가 어휘의 빈도수에 기초한 연구[6-9]가 대부분이다. 그러나 어휘의 빈도수와 같은 표층적인 측정치만으로는 수능 문항들이 가지는 언어학적인 특성을 종합적으로 파악하는데 한계가 있다. 이와 같은 다양한 측면의 언어학적인 분석을 가능하게 한 도구 중 하나가 코메트릭스(Coh-Metrix)이다[10]. 코메트릭스는 기존의 연구들에서 분석한 표층적 측면뿐만 아니라 문장의 응집력 및 어휘의 심리언어학적 정보까지 포괄적으로 분석할 수 있도록 측정치를 제공해준다.

따라서 본 연구에서는 각 문항 유형이 가지는 언어학

적인 특징을 코메트릭스에서 제공하는 응집력과 어휘정보의 차이를 분석하고자 한다. 본 연구의 연구문제는 다음과 같다.

- (1) 분류된 읽기문항의 유형별 응집력의 차이를 비교, 분석한다.
- (2) 분류된 읽기문항의 유형별 어휘정보에 대한 차이를 비교, 분석한다.

II. 이론적 배경

1. 코메트릭스(Coh-Metrix)

코메트릭스(Coh-Metrix)는 코퍼스 분석 도구 중 하나로서 미국 멤피스대학교(University of Memphis) 지능형시스템연구소(Institute for intelligent Systems)에서 개발한 웹기반 언어분석 시스템이다[11]. 이는 다양한 유형의 텍스트를 다양한 언어학적 측정치들에 의해 분석하는 자동화된 컴퓨터 도구이다[12]. 현재 웹버전 코메트릭스3.0은 11개의 지표와 11개의 지표를 106개의 세부 측정치로 나누어 제공하고 있다. 11개의 세부 지표는 기초산출치, 지문 이독성 요소 수치, 표층응집성, 잠재 의미 분석, 어휘의 다양성, 연결사, 상황모델, 통사적 복잡성, 통사적 패턴 밀집성, 어휘 정보, 이독성의 열한가지로 제시된다. 이처럼 다양한 언어학적인 측정치를 제공하는 코메트릭스를 통하여 여러 장르의 텍스트를 비교하여 언어학적 차이를 밝힐 수 있다.

2. 선행연구

1994학년도부터 시행된 수능 영어영역에 관한 연구가 활발하게 진행되어 왔다. 특히 최근 컴퓨터 기술이 발달하면서 코퍼스 기반 연구가 진행되고 있는데 국내에서도 코퍼스 기반 수능 영어영역의 연구가 여러 차례 진행되어 왔다. 김낙복(2008)은 수능에서의 언어적 쌍인 어휘적 뭉치말이 빈약하다는 연구결과를 보여주었다[6]. 김제우, 이동주(2016)은 교과서, 수능, EBS 교재 사이의 어휘수준을 비교하여 EBS교재가 수능과 교과서보다 높은 수준의 어휘를 요구하였고 교과서가 가장 적은 어휘 수준을 요구함을 보여주었다[9]. 이러한 대부

분의 코퍼스 기반 연구는 어휘에 대한 분석이 많이 실시되었다. 빈도수와 어휘수준에 대한 연구들은 코퍼스 연구가 주는 큰 장점이지만 여전히 어휘의 기초산출 레벨에 머물러 있다는 단점이 있다.

이에 최근 담화의 응집성까지 측정해줄 수 있는 웹프로그래머이자 코퍼스 도구인 코메트릭스를 활용한 연구가 진행되었다. 김재은, 최은철(2015)은 코메트릭스를 활용하여 EBS교재, 수능, 교과서의 차이를 비교하였다. 그 결과 EBS교재와 수능의 독해 난이도가 교과서에 비해 높게 나타났고 EBS교재와 수능 지문의 난이도가 미국의 New York Times 또는 Time Magazine의 평균 이독성 지수와 비슷함을 밝혀내었다. 또한 구문의 복잡도가 EBS교재와 수능이 교과서보다 높게 나타남을 밝혀내었다[13]. 문지현, 김해동(2017)은 1994학년도부터 2016학년도까지 시행된 수능 영어 읽기지문을 코메트릭스를 통해 분석한 뒤, 코메트릭스에서 제공하는 측정치 중에서 꾸준히 상승하거나 하락하는 13개의 언어적 요소와 2개의 이독성 지표를 분석하였다. 그 결과로 수능의 읽기 지문 난이도가 점차적으로 증가되었고 단어의 수가 지문의 어려운 정도와 연관성이 있는 변인으로 확인하였다. 또한 어휘습득나이와 문장의 통사적 복잡성이 증가하고 있음을 밝혀내었다[14]. 권경화, 이정원(2015)은 문항의 유형별 학생들이 사용하는 전략에 대해 연구하였다. 그 결과로 빈칸 추론 유형에서는 연역적 추론을 통한 핵심정보 파악하기, 글의 순서 유형에서는 문맥의 대명사 지칭어나 연결어 등을 활용하여 내용의 흐름파악하기, 주제 파악 유형에서는 핵심 부분만을 선별적으로 읽기와 같은 전략의 사용빈도가 높음을 연구결과로 보였다[15]. 이상과 같은 선행연구의 논문을 볼 때 수능 시험의 문항간의 언어학적인 요소를 밝혀낸 연구가 부족하며 수능문항유형을 모두 포함하지 않았음을 알 수 있다.

III. 분석방법

1. 분석대상

연구 대상이 되는 수능 영어영역 읽기지문은 수능이

시작된 1994학년도부터 2017학년도까지 출제된 읽기지문이다.

수능 문항의 유형은 한국교육과정평가원에서 제시한 대의과약, 문법·어휘, 세부사항, 상호작용, 간접쓰기, 복합으로 여섯 가지 문항유형이다. 하지만 위의 여섯 가지 문항유형은 다소 겹치는 부분이 있다. 이에 따라 본 연구에서는 조금 더 넓은 개념으로 문항유형을 묶은 김정렬(2016)의 분류[16]를 토대로 맥락읽기, 세부읽기, 간접쓰기로 나누어 분석하였다. 대의과약문항은 맥락읽기로 분류하였고 세부사항문항과 문법·어휘 문항을 세부사항문항에 포함시켰고 빈칸추론문항은 간접쓰기에 포함시켰다. 간접쓰기 문항 유형 중 문장요약문항은 전체적인 문맥의 내용을 파악한 후 요약하는 문항이기 때문에 맥락읽기에 포함시켰다. 또한 다양한 종류의 문항유형을 합친 장문독해는 연구대상에서 제외하였다. 상세한 문항유형은 [표 1]에 제시하였다.

표 1. 문항유형

범주	문항
맥락읽기	주제/제목 찾기
	요지/주장/목적 파악
	분위기/심경파악
	요약문완성
세부읽기	내용일치(실용자료/도표)
	문법/어휘
간접쓰기	빈칸/연결사 추론
	무관한 문장
	문장삽입
	글의 순서

2014학년도 대학수학능력시험은 난이도에 따라 A형과 B형으로 나뉘었기 때문에 코퍼스에서 제외하였다. 보다 정확한 분석을 위해 완전한 문장으로 이루어진 지문의 단락만을 코퍼스에 포함시키며 문제의 지시문, 읽기 지문의 제목 및 문단 제목, 삽화, 광고 등을 제거하여 코퍼스를 구축하였다. 최근 수능 읽기문항과 다른 유형인 시사점 찾기, 글의 성격, 앞 내용 추론하기와 같은 문항은 포함하지 않았다. 또한 코메트릭스의 특성인 응집력을 오류 없이 측정하기 위해 코퍼스를 구축하기 전에

수능에 출제된 지문을 수정하였다. 예를 들어, 빈칸 추론과 같은 지문의 경우에는 빈칸에 알맞은 내용에 해당하는 지문의 답안을 채워 넣어 코퍼스를 구축하였다. 이와 같이 수정한 문항유형은 문법, 어휘, 연결사 추론, 빈칸 추론, 무관한 문장 고르기, 문장 삽입, 글의 순서이며 해당하는 문항들은 각각 문법에 맞게 고치거나 무관한 문장을 골라내어 코퍼스로 구축하였다. 주어진 지문의 내용을 요약하는 요약문의 경우에는 요약문을 제외하여 코퍼스에 포함하였다.

2. 분석도구

1994학년도부터 2017학년도까지 시행된 수능시험을 분석하기 위해 텍스트의 다양한 언어학적 특징을 분석하는 코메트릭스를 활용하였다.

3. 분석항목별 수치 측정 방법

3.1 표층응집성(Referential Cohesion) 분석 항목

명사중복(noun overlap)은 인접한 문장 간의 명사의 중복을 나타내는 측정치로서 명사중복 측정치는 다른 중복 측정치에 비해 가장 엄격한 측정방법을 가지고 있다. 형태와 복수성이 모두 같은 모양을 지녀야 중복으로 측정된다. 예를 들어서 단수형태의 'cell'과 복수형태의 'cells'는 명사중복으로 측정되지 않는다. 측정치는 중복이 되는 경우의 0과 중복이 되는 경우의 1의 점수로 제공된다[17].

논항중복(argument overlap)은 명사중복과는 다르게 단수 형태와 복수 형태를 모두 중복으로 간주한다. 또한 대명사의 중복도 측정한다. 예를 들어서 단수형태의 'cell'과 복수형태의 'cells'를 중복으로 취급하고 인접한 문장 내에 같은 대명사인 'he'와 'he'를 중복으로 처리한다. 측정치는 0과 1의 점수로 제공된다[17].

어간중복(stem overlap)은 인접한 문장 내에 존재하는 명사와 일치하는 명사, 동사, 형용사, 부사와 같은 공통된 레마(lemma)를 모두 중복으로 처리한다. 예를 들어서 'tree'와 'treed'는 중복으로 처리하여 취급한다. 측정치는 0과 1의 점수로 제공된다[17].

내용어중복(content word overlap) 측정치는 인접한

문장 내에 존재하는 내용어의 중복 비율을 나타낸다. 예를 들어 적은 어휘를 지니고 있는 두 개의 문장 사이에 두 개의 내용어중복을 지닌다면 많은 어휘를 지닌 두 개의 문장의 비율보다 더 높은 비율 점수를 얻는다 [17].

3.2 연결사(Connectives) 분석 항목

연결사는 개념과 문장을 이어주는 응집력 장치를 만들어주는 중요한 역할을 하며 글의 구조에 대한 단서를 제공한다. 코메트릭스에서는 1,000 단어 당 연결사의 발생 정도를 분석하여 제공한다. 제공되는 측정치로는 모든 연결사에 대한 발생정도, 'because'나 'so'와 같은 인과 관계를 나타내는 연결사(casual)의 발생 정도, 'and'나 'or'와 같은 논리적 흐름을 나타내는 연결사(logical), 'although'나 'whereas'와 같은 역접(adversative and contrastive), 'first'또는 'until'과 같은 시간을 보여주는 연결사(temporal)와 더불어 'and'나 'moreover'와 같이 내용을 덧붙일 때 사용하는 연결사(additive)의 발생 정도에 대한 측정치를 제공한다[17].

3.3 어휘정보(Word Information) 분석 항목

코메트릭스에서 제공하는 어휘정보 중에서 품사정보는 자연어 처리 도구인 Charniak parser에 기초하여 텍스트에 포함된 모든 어휘에 대한 품사정보를 제공한다. 만약 'bank'와 같이 두 개 이상의 품사를 지닌 어휘가 나타날 경우에는 통사적 구조에 기초하여 알맞은 품사를 배정한다. 연결사와 마찬가지로 1,000단어 당 품사의 발생 정도에 대한 측정치를 제공한다. 코메트릭스에서 제공하는 품사정보는 명사, 동사, 형용사, 부사 그리고 대명사의 발생 정도가 있다[17].

또한 코메트릭스에서는 의미적이고 심리적인 차원에 기초한 추가적인 어휘정보를 제공한다. MRC Psycholinguistic Database에 기초한 MRC 코퍼스는 150,000개 이상의 단어로 구성되어 다양한 심리적 특징을 나타내는 측정치를 제공한다. 측정치는 7등급으로 나뉘며 제공되는 측정치는 100부터 700까지의 숫자로 나타낸다. 습득나이(age of acquisition)는 아이들에게 주로 나타나는 어휘에 대한 측정치이다. 친숙성

(familiarity)은 성인에게 친숙한 어휘를 나타내는 측정치이다. 내용어의 구체성(concreteness for content word)은 어휘가 얼마나 구체적인지를 나타내는 측정치이다. 심상성(imagability)은 얼마나 정신적 이미지를 구상하기 쉬운지를 나타내는 측정치이다. 유의미성(meaningfulness)은 다른 어휘와 의미적으로 얼마나 연결되는지에 대한 측정치이다[17].

4. 통계분석 방법

각 연도별로 산출된 문항은 맥락읽기, 세부읽기, 간접쓰기로 나누어 코딩을 실시하였다. 각 문항별 측정치에 대해 유의미한 차이가 있는지를 알아보기 위해 SPSS 18.0을 활용하여 일원배치 분산분석을 실시하였다. 또한 어느 문항유형 사이에 유의미한 차이가 있는지를 알아보기 위해 Tukey 사후분석을 실시하였다.

IV. 결과

1994학년도부터 2017학년도까지 시행된 수능 영어문항에 대한 각 문항별 코메트릭스 측정치 중에서 통계적으로 유의미한 차이를 보이는 결과는 다음과 같다.

1. 응집성

텍스트의 응집성을 분석하는데 표층응집성 지표와 연결사 지표가 사용되었다. 본 연구의 결과 분류된 문항 유형에 따라 표층응집성과 연결사 지표의 통계적 차이를 보였다. 자세히 들여다보면 본 연구에서 사용한 표층응집성 지표에서의 세부지표의 명사중복비율과 어간중복비율은 인접한 문장사이의 중복되는 비율을 측정하는 값으로서 문항의 유형 사이에 명사중복비율($f=5.776, p=0.003$), 어간중복비율($f=4.485, p=0.012$) 모두에서 통계적으로 유의미한 차이를 보이고 있다. 연결사 정보에서는 역접과 추가의 연결사 측정치가 사용되었고 역접($f=3.524, p=0.03$), 추가의 연결사($f=5.359, p=0.005$) 모두에서 통계적으로 유의미한 차이를 보이고 있다. 응집성의 상세 내용은 [표 2]에 제시하였다.

표 2. 응집성 평균(표준오차)

측정치	맥락읽기	세부읽기	간접쓰기	f값	p값
표층응집성 명사중복	0.28 (0.25)	0.36 (0.30)	0.36 (0.29)	5.776	.003*
어간중복	0.39 (0.28)	0.44 (0.33)	0.47 (0.30)	4.485	.012*
연결사 역접	18.56 (14.91)	16.21 (13.55)	20.08 (14.00)	3.524	.03*
추가의 연결사	45.72 (19.49)	40.68 (22.58)	47.46 (20.71)	5.359	0.005*

* $p < 0.05$

구체적으로 어느 문항 사이에 유의미한 차이를 보이는지 알아보기 위하여 실시한 Tukey 사후분석 결과를 살펴보면 표층응집성지표의 명사중복지표에서는 맥락읽기문항과 세부읽기문항($p=0.024$), 맥락읽기문항과 간접쓰기문항($p=0.006$)에서 유의미한 차이를 보이고 어간중복지표에서는 맥락읽기문항과 간접쓰기문항($p=0.010$)에서 유의미한 차이를 보인다. 연결사지표의 역접측정치에서는 세부읽기문항과 간접쓰기문항(0.023)에서 유의미한 차이를 보였고 추가의 연결사에서는 맥락읽기문항과 세부읽기문항($p=0.039$), 세부읽기문항과 간접쓰기문항($p=0.004$)에서 유의미한 차이를 보였다. 사후분석의 상세 내용은 [표 3]에 제시하였다.

표 3. 응집성의 사후분석

측정치	맥락-세부	세부-간접	맥락-간접
표층응집성 명사중복 어간중복	$p=.024^*$ $p=.157$	$p=.957$ $p=.641$	$p=.006^*$ $p=.010^*$
연결사 역접 추가의 연결사	$p=.219$ $p=.039^*$	$p=.023^*$ $p=.004^*$	$p=.494$ $p=.645$

* $p < 0.05$

2. 어휘정보

본 연구의 결과 분류된 문항 유형에 따라서 어휘정보에 대한 통계적인 차이를 보였다. 조금 더 자세히 들여다보면 어휘정보에는 텍스트에 나타난 어휘의 품사의 정보와 MRC database에 기초한 심리언어학적인 어휘의 정보에 대한 지표다. 어휘정보에서 제공하는 품사측정치에서 대명사발생률($f=4.641, p=0.01$)에서 유의미한 차이를 보였다. MRC에서 제공하는 어휘정보에서는 습

득나이($f=4.995$, $p=0.007$), 내용어의 구체성($f=6.290$, $p=0.002$), 심상성($f=7.811$, $p=0.000$), 유의미성($f=3.443$, $p=0.003$)에서 유의미한 차이를 보이고 있다. 어휘정보의 상세 내용은 [표 4]에 제시하였다.

표 4. 어휘정보 평균(표준오차)

측정치	맥락읽기	세부읽기	간접쓰기	f값	p값
품사정보 대명사발생	85.15 (52.96)	76.75 (52.51)	71.01 (43.52)	4.641	.01*
MRC 습득나이	343.88 (48.92)	330.86 (59.35)	347.79 (55.38)	4.995	.007*
내용어의 구체성	375.34 (40.14)	386.38 (37.35)	373.01 (38.62)	6.290	.002*
심상성	410.34 (38.30)	419.25 (33.22)	404.82 (34.45)	7.811	.000*
유의미성	434.23 (24.24)	436.06 (24.39)	429.84 (24.13)	3.443	.003*

* $p < 0.05$

구체적으로 어느 문항 사이에 유의미한 차이를 보이는지 분석하기 위해 실시한 Tukey 사후분석을 살펴보면 품사지표의 대명사발생정도 측정치에서는 맥락읽기와 간접쓰기($p=0.007$)에서 유의미한 차이를 보인다. MRC 심리언어학적 지표의 습득나이 측정치에서는 맥락읽기와 세부읽기($p=0.041$), 세부읽기와 간접쓰기($p=0.007$)에서 유의미한 차이를 보였다. 내용어의 구체성 측정치에서는 맥락읽기와 세부읽기($p=0.012$), 세부읽기와 간접쓰기($p=0.002$)에서 심상성에서는 맥락읽기와 세부읽기($p=0.032$), 세부읽기와 간접쓰기($p=0.00$)에서 유의미한 차이를 보였다. 마지막으로 유의미성 측정치에서는 세부읽기와 간접쓰기($p=0.034$)에서 유의미한 차이를 나타내고 있다. 사후분석의 상세 내용은 [표 5]에 제시하였다.

표 5. 어휘정보 사후분석

측정치	맥락-세부	세부-간접	맥락-간접
품사정보 대명사발생	$p=.207$	$p=.501$	$p=.007^*$
MRC 습득나이	$p=.041^*$	$p=.007^*$	$p=.722$
내용어의 구체성	$p=.012^*$	$p=.002^*$	$p=.800$
심상성	$p=.032^*$	$p=.000^*$	$p=.228$
유의미성	$p=.727$	$p=.034^*$	$p=.132$

* $p < 0.05$

V. 결론 및 제언

본 연구에서는 수능 영어영역 지문을 맥락읽기, 세부읽기, 간접쓰기로 나누어 분석하였다. 본 연구에서는 총 9개의 코메트릭스 측정치를 활용하여 분석하였다. 분석 결과 다음과 같은 결론을 얻었다.

첫째, 명사중복 측정치에 대한 결과 맥락읽기문항과 세부읽기문항사이와 맥락읽기문항과 간접쓰기 문항사이에 통계적으로 유의미한 차이가 발견되었다. 명사중복 측정치보다 조금 더 관대한 측정을 하는 어간중복 측정치에서는 맥락읽기문항과 간접쓰기문항에서 통계적으로 유의미한 차이를 보였으며 두 개의 측정치에서 모두 맥락읽기문항이 가장 저조한 중복을 보였다. 하지만 대명사발생빈도 측정치에서는 맥락읽기문항과 간접쓰기문항사이에 통계적으로 유의미한 차이를 보였고 간접쓰기문항이 가장 저조한 발생빈도를 보였다. 이를 통해 학습자들이 전체적인 맥락을 파악하는 맥락읽기 문항을 접할 때에 명사 또는 어간의 중복보다는 대명사를 통해 전체적인 맥락을 파악하며 읽어야 함을 암시한다.

둘째, 연결사 측정치에 대한 결과 역접의 연결사에서 세부읽기문항과 간접쓰기 문항사이에서 통계적으로 유의미한 차이를 보였다. 추가의 연결사에서는 맥락읽기문항과 세부읽기문항, 세부읽기문항과 간접쓰기문항 사이에서 통계적으로 유의미한 차이를 보였다. 두 개의 측정치에서 세부읽기문항이 가장 저조한 연결사 점수를 보인다. 이를 통해 글의 전체적인 맥락을 파악하는 맥락읽기와 논리적 글의 흐름을 파악하는 간접쓰기 문항에서 연결사를 신경 쓰며 글을 읽어야 함을 시사한다.

셋째 MRC 심리언어 지표의 결과 습득나이 측정치에서 맥락읽기문항과 세부읽기문항, 세부읽기문항과 간접쓰기문항에서 통계적으로 유의미한 차이를 보였고 세부읽기문항이 가장 낮은 점수를 나타내었다. 내용어의 구체성 측정치에서는 맥락읽기와 세부읽기, 세부읽기와 간접쓰기에서 통계적으로 유의미한 차이를 보였고 간접쓰기문항이 가장 저조한 점수를 보였다. 심상성 측정치에서는 맥락읽기와 세부읽기, 세부읽기와 간접쓰기에서 통계적으로 유의미한 차이를 보였고 간접쓰기문항이 가장 낮은 점수를 획득하였다. 유의미성 측정

치에서는 세부읽기와 간접쓰기문항에서 통계적으로 유의미한 차이를 보였고 간접쓰기문항에서 가장 낮은 점수를 나타내었다. 이를 통해 각 문항의 심리언어학적인 어휘의 수준이 측정치마다 다르게 나타났다. 각 측정치에 따르면 문항별 어휘의 수준이 통일되지 않았고 이는 교육과정의 목표를 측정하는 각기 다른 영역의 문항들이 서로 다른 어휘의 난이도를 지님을 나타낸다. 기존 수능 문항에 대한 선행 연구 중 다수가 어휘의 빈도수 [6][9]나 수능 전체의 난이도 및 난이도에 영향을 미치는 요소에 관한 연구[13][14], 또는 제한된 문항을 가지고 실행한 연구[15]에 비해 본 연구에서는 장문독해를 제외한 전체 수능 문항을 나누어 분석하여 차이점을 제시하였다는 점에서 그 특징이 있다.

전체적으로 문항유형별로 특징을 나타내는 언어학적인 요소가 잘 반영되어있지 않아 수능 영어영역 읽기 지문의 난이도에 영향을 미칠 수 있음이 시사되었다. 본 연구에서는 지문의 전체적인 의미를 파악하기위해 글의 전체를 훑어 읽어야 하는 맥락읽기의 특성상 응집력이 높을 거라 가설을 세웠지만 명사중복과 어간중복 모두에서 가장 낮은 점수를 획득하였다. 각 문항별로 MRC 심리언어학적인 어휘의 수준이 모두 다르게 나타나 문항 유형별로 서로 다른 어휘 난이도를 나타내고 있음이 시사되었다. 따라서 본 연구에서 분석된 응집력과 어휘정보 측정치들이 차후 교재 집필과 수능 출제에 반영되기를 기대한다. 또한 학생들이 수능문제에 접할 때에 이러한 응집력에 대한 차이를 알고 활용할 수 있도록 하는 전략 교수가 교육과정에 반영되기를 기대한다.

이후의 후속 연구는 다음과 같이 진행될 수 있다. 우선 각 문항의 코퍼스 크기에 대한 보완이 필요하다. 본 연구에서는 한국교육과정평가원에서 제공하는 수능문제를 사용하였지만 각 문항이 가지는 텍스트의 수가 서로 달라 이를 표준화할 필요가 있다. 다음으로는 연구를 통해 밝혀진 차이점을 실제 학생들이 활용할 수 있는지에 대한 연구가 필요하다.

참고 문헌

- [1] 교육과학기술부, “2009개정 교육과정에 따른 영어과 교육과정,” 교육과학기술부 고시 제 2011-361호, 2011.
- [2] 한국교육과정평가원, 2015학년도 통합형 수능 영어 영역 학습 안내 및 Q&A자료집, 2014.
- [3] J. W. Lee, “An exploratory study on reading comprehension test-taking process and strategies in the EFL context,” *English Teaching*, Vol.57, No.4, pp.177-195, 2002.
- [4] 한국교육과정평가원, 2018학년도 수능 영어 절대평가 학습안내, 2016.
- [5] 장경숙, “대학수학능력시험 외국어(영어)영역 읽기 난이도 예측 모형 개발,” *foreign language education*, 제11권, 제1호, pp.111-130, 2004.
- [6] 김낙복, “대학수학능력시험 외국어(영어)영역의 코퍼스 언어학적 어휘 비교 분석,” *영어어문교육*, 제14권, 제4호, pp.201-221, 2008.
- [7] 송선화, 정양수, “대학수학능력시험 외국어영역에 나타난 Make, Get, Take의 코퍼스 기반 언어 분석,” *인문학연구*, 제85권, pp.223-251, 2011.
- [8] 강지은, 이동주, “고등학교 영어 듣기 지문의 Lexcial Bundles 분석,” *학습자중심교과교육연구*, 제16권, 제12호, pp.271-294, 2016.
- [9] 김제우, 이동주, “코퍼스기반 고등학교 어휘 사용 연구,” *언어과학연구*, 제80권, pp.51-74, 2017.
- [10] A. Graesser, D. McNamara, and J. Kulikowich, “Coh-Matrix: Providing multilevel analyses of text characteristics,” *Educational Researcher*, Vol.40, No.5, pp.223-234, 2011.
- [11] 전문기, “고등학교 개정 영어교과서 읽기 자료의 연계성 분석,” *Journal of the Korean Data Analysis Society*, 제16권, 제2호, pp.925-938, 2014.
- [12] 전문기, 임인재, “고등학교 영어 교과서 읽기 자료의 코퍼스 언어학적 비교 분석,” *Foreign language education*, 제17권, 제1호, pp.209-233,

2010.

- [13] 김재은, 최은철, “고등학교 영어 교과서, EBS 수능 연계 교재, 대학수학능력시험의 코퍼스기반 난이도 비교 분석,” 멀티미디어 언어교육, 제18권, 제1호, pp.59-92, 2015.
- [14] 문지현, 김해동, “대학수학능력시험 영어 읽기 지문의 언어적 요소 분석,” 현대영어교육, 제18권, 제1호, pp.193-211, 2015.
- [15] 권정화, 이정원, “영어 읽기 시험의 문항 유형별 읽기 시험 수행 전략 사용 연구,” 인문학연구, 제101권, pp.79-100, 2015.
- [16] 김정렬, “말하기/쓰기 표현기능 강화 영어교육 방안,” 올바른 서울영어교육을 위한 실용영어 정책 활성화 마련 정책토론회, 2016.
- [17] D. McNamara, A. Graesser, P. McCarthy, and Z. Cai, Automated Evaluation of Text and Discourse with Coh-Metrix, Cambridge, 2014.

김정렬(Jeong-ryeol Kim)

정회원



- 1996년 ~ 현재 : 한국교원대학교 초등교육과 교수
 - 2012년 ~ 현재 : 한국영어다독학회 회장
 - 2011년 ~ 2012년 : (전) 초등영어교육학회 회장
 - 2013년 ~ 2014년 : (전) 외국어교육학회 회장
 - 2012년 ~ 현재 : Extensive Reading Foundation 이사
 - 2014년 ~ 현재 : Reading in Foreign Languages 저널 편집이사
 - 2015년 ~ 현재 : 교육부 교육과정심의위원
- <관심분야> : 컴퓨터활용 영어교육, 초등영어교육, 영어교수법

저자소개

최민주(Minju Choi)

준회원



- 2016년 2월 : 한밭대학교 영어영문학과(문학사)
- 2016년 3월 ~ 현재 : 한국교원대학교 영어교육과 석사과정

<관심분야> : 대학수학능력시험, 코퍼스 분석, CohMetrix