creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

교육학석사 학위논문

중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 텍스트 개발

2023년 8월

서울대학교 대학원

AI융합교육학과

이 은 주

중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 텍스트 개발

지도교수 김 선 희

이 논문을 교육학석사 학위논문으로 제출함
2023년 8월

서울대학교 대학원
AI융합교육학과
이 은 주

이은주의 석사 학위논문을 인준함
2023년 8월

위 원 장 _____ 안 현 기 _____ (인)

부위원장 _____ 이 병 민 _____ (인)

위 원 _____ 김 선 희 _____ (인)

국문 초록

본 연구는 AI 기술을 활용하여 학습에 어려움을 겪는 학생들의 영어 읽기 능력 향상을 돕기 위한 읽기 자료 제작을 목표로 한다. 2015년 개정 교육과정의 초등학교 6학년과 중학교 1학년의 영어 교과서 읽기 본문을 대상으로 하여 자연어 처리 기법과 생성형 AI 엔진을 활용하여 학습 부진 학생들의 소리 내어 글 읽기 활동에 활용할 수 있는 최소대립쌍과 문장 읽기 자료를 생성했다. 개발된 읽기 자료는 학교 현장에서 쉽게 사용할 수 있는 Reading Progress와 같은 AI기반 읽기 학습용 프로그램에 즉시 활용할 수 있는 형태로 제작하였다.

구체적인 읽기 자료 개발 단계는 다음과 같다. 첫째, 코퍼스 생성을 위해 연구 대상 텍스트 전체를 하나의 문서로 통합했다. 둘째, 자연어 처리 기술을 활용해 통합된 텍스트에서 주요 어휘 1,500개를 빈도 기반으로 추출하고, 실생활 빈도와 교육과정 내의 빈도, 음절 수, 읽기 난이도 등을 수치화하여 데이터프레임을 생성했다. 셋째, 수치화된 난이도 데이터를 바탕으로 생성형 AI를 활용하여 주요 단어의 최소대립쌍과 주요 단어를 활용한 짧은 문장을 생성하고, 읽기 난이도별로 문장을 분류하였다.

개발된 텍스트에 대한 타당성 검토를 위하여 읽기 자료 개발 과정, 기준, 결과물은 영어교육, AI융합 교육, 자연어 처리에 전문성이 있는 현장 교사 전문가 6인의 전문가 검토를 거쳤다. 총 두 차례 전문가의 검토 의견을 반영하여 최종 텍스트를 개발했다.

최종 개발된 최소대립쌍은 전체 텍스트에서 산출한 빈도가 높은 1,500개 단어 중 난이도 분류에서 아주 낮은 난이도 59개, 낮은 난이도 99개의 단어를 기준으로 생성되었다. 최종 개발된 최소대립쌍은 20개 음소에 대하여 총 288개이며, 최종 개발된 문장은 중복을 제외하고 총 112문장으로, 하 수준 44개, 중 수준 58개, 상 수준 10개로 이루어져 있다.

본 연구는 학습 부진 학생의 읽기 학습자료를 현행 교과서 읽기 지문에 포함된 단어들의 교과서 내 출현 빈도, 단어의 음절 수, 음절 결합 형태, 음소 구성에 근거하여 개발하였다는 점에서 의의를 가진다. 또한 결과물은 AI 기반 학습 프로그램을 활용하여 중학교 학습 부진 학생을 지도할 때 유용하게 활용할 수 있다는 점, 개발된 코드에 다른 코퍼스를 활용하여 유사한 결과물을 생성할 수 있다는 점에서 현장에 도움을 줄 수 있다.

그러나 문장 난이도 산출시 활용한 음소 기반 분석 도구는 문장 단위에서의 연음이나 억양 등을 고려하지 못한다는 점, 생성형 AI 엔진의 기술적 한계로 인하여 다양한 문장을 생성하지 못했다는 한계점을 가진다. 위와 같은 한계점에도 불구하고, 본 연구의 자연어 처리 기법, 생성형 AI를 활용한 학습자료 개발 시도는 현장 교사들에게 AI 활용 수업 자료 개발 과정을 참고할 수 있는 사례로 기능할 것으로 기대한다.

주요어 : 학습 부진, 기초 학력 지도, 읽기 유창성, AI 활용 교육, 자연어 처리
학 번 : 2021-22351

목 차

제 1 장 서 론 ····· 1
제 1 절 연구의 배경 및 필요성 ····· 1
제 2 절 연구의 목적 ····· 3

제 2 장 이론적 배경 ····· 4
제 1 절 영어 학습부진 학생 ····· 4
제 2 절 이독성과 학습 부진 ····· 6
제 3 절 작업 반복과 유창성 ····· 8
제 4 절 맞춤형 수업을 위한 AI 활용 ····· 9

제 3 장 연구 방법 ····· 13
제 1 절 연구 대상 및 도구 ····· 13
제 2 절 읽기 자료 개발 기준 ····· 14
제 3 절 읽기 자료 개발 절차 ····· 15
제 4 절 내적 타당화 ····· 19

제 4 장 연구 결과 ····· 21
제 1 절 초기 개발 읽기 자료 및 전문가 검토 결과 · 21
제 2 절 2차 개발 읽기 자료 및 전문가 검토 결과 ··· 34
제 3 절 최종 개발 읽기 자료 ····· 45

제 5 장 논의 및 결론 ····· 49
제 1 절 결론 및 제언 ····· 49

참고문헌 ····· 51
Abstract ····· 55
부록 ····· 58

표 목 차

[표 1] 영어 학습에 어려움을 겪기 시작하는 시점_중학교 교사(김성혜, 2019) ····· 4
[표 2] 영어 학습 부진 학생들이 어려워하는 영역과 학습해야 할 영역(김성혜, 2019) ····· 5
[표 3] 연구 대상 교과서 리스트 ····· 13
[표 4] 읽기 자료 개발 연구 절차 ····· 15
[표 5] 내적 타당화 참여 전문가 프로필 ····· 19
[표 6] 내적 타당화 전문가 검토지 문항 ····· 20
[표 7] 초기 최소대립쌍 개발 결과 ····· 21
[표 8] 초기 문장 개발 결과 ····· 25
[표 9] 1차 전문가 타당화 결과 ····· 29
[표 10] 1차 전문가 검토 개방형 응답 의견 및 반영 사항 30
[표 11] 1차 전문가 검토 결과를 반영한 최소대립쌍 읽기 자료(2차) ····· 35
[표 12] 1차 전문가 검토 결과를 반영한 문장 읽기 자료 (2차) 39
[표 13] 2차 전문가 타당화 결과 ····· 42
[표 14] 2차 전문가 검토 개방형 응답 의견 및 반영 사항 42
[표 15] 최종 최소대립쌍 음소별 분류 ····· 46
[표 16] 최종 문장 읽기 난이도별 분류 ····· 48

그 림 목 차

[그림 1] 2017년·2021년 국가수준 학업성취도 평가 1수준(기초학력 미달) 비율 추이(교육부, 2022) ····· 1
[그림 2] 영어 학습 부진 심화 패턴(김태은, 2020) ····· 2
[그림 3] Insights기능을 활용한 학생 수행 결과 자동 분석 (Microsoft, 2022) ····· 11
[그림 4] Reading Coach 기능(Microsoft, 2022) ····· 11
[그림 5] 학생의 오류 음소 패턴 자동 분석 및 개별화된 추가 과제 자동 생성 (Microsoft, 2022) ····· 12

부 록

[부록 1] 최종 개발 읽기 자료(최소대립쌍) ····· 58
[부록 2] 최종 개발 읽기 자료(문장) ····· 62
[부록 3] 1차 전문가 검토지 ····· 65
[부록 4] 2차 전문가 검토지 ····· 83
[부록 5] 최종 개발 코드 ····· 104

제 1 장 서 론

제 1 절 연구의 배경 및 필요성

해가 거듭될수록 학생 간 학습격차는 심화되어 기초학력이 부족한 학생들을 위한 정책의 필요성이 더욱 대두되고 있다. 교육부는 제1차 기초학력 보장 종합계획(2023~2027)을 발표하여 코로나로 인한 비대면 수업 기간 동안 심화된 학생들의 학습 결손과 정서 안정을 지원하는 국가 단위의 기초학력 보장 체계를 마련하고자 했다. [그림1]은 2017년과 2021년도의 국가수준 학업성취도 평가에서 기초학력 미달 학생들의 비율을 나타내는데, 코로나 이전에 비하여 기초학력 학생의 비율이 모든 과목에서 급격히 증가했음을 알 수 있다.

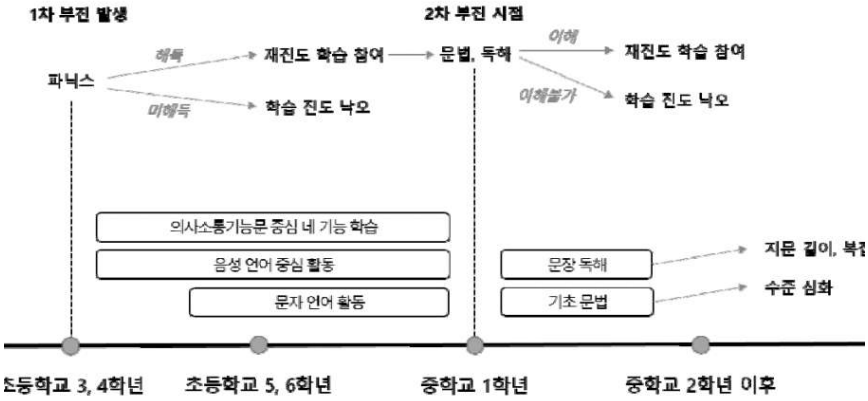
[그림 1] 2017년 · 2021년 국가수준 학업성취도 평가 1수준(기초학력 미달) 비율 추이(교육부, 2022)

교육부는 이번 종합계획에서 인공지능(AI) 기반 진단체계 구축을 강화하여 에듀테크 기반 학생 개별화 학습 지원을 강조했다. 이를 위해 국가 기초학력 지원센터를 운영하며, 2022 개정 교육과정에 따른 '최소한의 성취기준'을 개발하고, 국가 수준의 진단 도구 및 교수·학습 자료를 보급하고자 한다(교육부, 2022).

현재 서울특별시교육청은 기초학력 진단-보정 시스템(s-basic)홈페이지를 통해 학년 초 학생들의 교과 학습 능력(국어, 영어, 수학 중심) 진

단검사를 진행한다. 영어 진단평가의 경우 4기능(듣기, 말하기, 읽기, 쓰기)를 모두 포함하고 있으나, 말하기 영역은 학생의 발화를 직접 녹음하여 분석하는 형태가 아닌, 맥락에 맞는 적절한 응답 고르기 형태의 간접 평가로 진행된다. 이러한 방식은 학생의 단어 음소 인식 및 영어 말하기 유창성 진단에 한계가 있어 학습 부진을 예방하고 지원하는 데 한계가 있다.

김태은(2020)은 기초학력 진단-보정 시스템이 유일한 기초 영어 능력 검사 도구라는 점을 강조하며, [그림2]와 같이 초등학교에서 발생하는 1차 부진(파닉스) 영역에 대한 진단 도구의 부족을 지적했다. 파닉스 능력



[그림 2] 영어 학습 부진 심화 패턴(김태은, 2020)

진단과 선제적 지원을 제공 한다면 학습 부진이 발생하는 초기에 교사가 개입하여 학습 부진을 예방할 수 있다는 것이다.

중학교 영어 학습 부진 학생들을 지도하기 위해 서울특별시교육청이 교사들에게 제공하는 자료는 각 학년별 의사소통영역 중심으로 이루어진 영어 '늘품이' 교재와 음소별로 학습이 가능한 '15분 파닉스' 교재이다. '15분 파닉스' 교재는 온라인 단어 학습 사이트인 클래스 카드에 단어 세트를 업로드하여 교사들이 사용할 수 있도록 하고 있다. 하지만 중학교 기초학력 미달 학생들의 음성 발화를 분석하고 즉시 피드백을 줄 수 있

는 AI기반 프로그램과 연계하여 활용할 텍스트는 별도로 개발되어 있지 않다. 따라서 AI기반 프로그램을 활용하여 학생의 음소 인식 능력을 진단하거나, 학습 부진 학생을 지원을 위한 기초적인 수준의 읽기 자료를 개발할 필요성이 있다.

제 2 절 연구의 목적

본 연구는 기본적인 음철법은 이해했으나 단어나 문장을 소리 내어 읽는 것에 어려움을 겪는 중학교 영어 학습 부진 학생들을 대상으로 읽기 유창성 향상을 위한 학습을 진행할 때 교사들이 활용할 읽기 자료 세트를 제작하는 것을 목적으로 한다. 개발 결과물은 학습 부진 학생들을 지도하는 과정에서 AI프로그램을 사용할 때 즉시 활용할 수 있는 읽기 자료의 형태로 제작하며, 2015 개정 교육과정의 초등학교, 중학교 교과서의 읽기 지문을 코퍼스로 활용하여 교과서에 출현하는 빈도가 높은 단어들을 중심으로 최소대립쌍, 짧은 문장으로 이루어진 단계별 읽기 자료를 개발하고자 한다.

제 2 장 이론적 배경

제 1 절 영어 학습 부진 학생

1. 영어 학습 부진 학생 정의

기초학력 부진 학생들은 학습 속도가 느리고 어휘력, 언어 이해, 언어 표현이 부족하여 낮은 자존감과 낮은 학습 동기, 불안, 위축 등의 특성을 보인다(김태은 외, 2020). 본 연구에서는 학습 부진 학생을 기초학력 미달 또는 기초학습 부진학생과 구분하여 영어 교과 교육과정의 성취기준에 도달하지 못한 영어 교과 학습 부진 학생을 대상으로 한다. 김성혜(2019)의 연구에 따르면 중학교 영어 성취수준이 '하' 수준인 학생의 경우 20~50% 미만의 학생들이 수업 내용을 잘 이해하지 못하며, 교육과정 성취기준을 미충족한다는 것을 보여준다. 중학교 영어 학습 부진 학생들은 초등학교 시기부터 누적된 학습 부진이 심화되어 다른 학생과의 수준 차이가 더 심각해지며, 학교 수업만으로 학습 내용을 따라가기에 많은 어려움을 겪고 있다(김성혜, 2019).

단위: 빈도(%)

항목	초등학교	중학교 1학년	중학교 2학년	중학교 3학년
전체	124(37.3)	114(34.3)	87(36.3)	7(2.1)

[표 1] 영어 학습에 어려움을 겪기 시작하는 시점_중학교 교사(김성혜, 2019)

2. 영어 학습 부진 학생 대상 읽기 지도 필요성

초등학교의 영어 수업 시간은 읽기 학습에 충분하지 않아 일부 학생들은 단어 해독 능력이 있음에도 문장이나 문단 이해에 어려움을 겪는다. 따라서, 읽기 발달 초기 단계에 있는 학생들의 읽기 유창성을 향상 시키기 위한 노력이 필요하다(최수진, 이윤, 2015). 중학생 중에서 영어 학습

에 어려움을 겪는 학생들의 영어 읽기 능력 부족은 기본적인 음철법의 적용에서부터 동료들의 영어 독해 이해도 및 독해 전략의 활용에 이르기까지 그 범위가 광범위하며, 이는 초등학생들에 비해 더 큰 편차를 보인다. 이러한 중학생들의 영어 읽기 능력 부족 문제는 반드시 해결해야 하는 중요한 과제이다. 영어 학습의 어려움은 읽기 능력과 깊은 연관이 있으므로, 이러한 학습 어려움을 해결하기 위한 교육과정 개선은 문자의 노출 양과 시기, 방법뿐만 아니라 문자 언어 관련 성취기준의 연계성을 강화하는 것이 필요하다(한국교육과정평가원, 2019).

3. 영어 학습부진 학생 지도의 어려움

중학교 1학년부터 영어 교과는 점차 음성언어에서 문자 언어 중심으로 전환되며, 학습 어휘의 양이 급증하고 언어 형식이 강화되는 시기이며, 영어 교사들은 [표 2]와 같이 학습 부진 학생들이 '어휘'와 '읽기' 영역을 우선적으로 학습해야 한다는 입장을 보였다(김성혜, 2019).

단위: 빈도(%)

항목	듣기	말하기	읽기	쓰기	어휘	문법
어려운 영역	7(2.1)	29(8.7)	36(10.8)	112(33.7)	20(6.0)	128(38.6)
우선 학습 영역	25(7.5)	33(9.9)	93(28.0)	7(2.1)	167(50.3)	7(2.1)

[표 2] 영어 학습 부진 학생들이 어려워하는 영역과 학습해야 할 영역 (김성혜, 2019)

해당 연구에 참여한 교사들은 문자 언어와 관련된 교육과정의 학습 수준이나 학습량 적정화보다는, 기존의 어휘 및 언어 형식을 제시하는 방법의 위계성이 부족하다는 점과 주제별로 구성된 교과서에 나타난 어휘 체계의 설계 기준이 모호함을 지적하며, 교사들의 기본학력 학생 지도의 어려움을 호소했다.

김성혜(2019)는 2015 개정 교육과정 영어과의 성취기준, 어휘, 언어형

식이 학교급별로 구분되어 있어 학습 부진 학생 지도에 활용도가 떨어짐을 지적하며 일상생활에서 사용 빈도가 높은 낱말과 의사소통 기능 표현을 순차적으로 익힐 수 있도록 학습 자료가 세분화 되어야 한다는 것을 강조했다. 또한, 필수 성취기준과 필수 어휘, 의사소통기능, 언어형식을 기초군으로 분류하여 교육과정에 제시해야 하며, 초기 읽기 능력 습득이 학습 부진의 경계가 되는 지점이므로 고빈도어를 토대로 한 기초 어휘목록이 교육과정에서 제시되어야 한다고 강조했다.

제 2 절 이독성과 학습 부진

1. 이독성의 정의

이독성(Readability)은 주어진 읽기 자료 내의 모든 요소들에 대해 글을 읽는 사람의 이해도, 최적의 속도로 읽는 능력, 그리고 그 읽기 자료에 대한 흥미 정도에 대한 종합적인 개념이다(Dale & Chall, 1949). Graves와 Graves(2003)는 이독성에 영향을 미치는 두 가지 요인으로 '읽기 자료에 내재된 요인들(factors largely inherent in the text itself)'과 '독자 및 읽기 자료와 관계있는 요인들(factors involving the reader and the text)을 제시했다. 내재된 요인으로는 어휘, 문장 구조, 문장의 길이, 풍부한 설명(elaboration), 일관성(coherence), 텍스트 구조가 있으며, 외재적 요인으로는 내용 친숙도, 배경지식, 독자의 인지적 수준, 글의 특징, 독자의 흥미가 있다. 최근의 연구들이 제시하는 이독성 공식들은 단어의 음절이 길수록, 단어의 철자 수가 많을수록, 저빈도(low frequency)의 단어가 많을수록, 한 문장 내의 단어가 많을수록 그 읽기 자료는 이해하는 것이 어렵다고 가정한다(임영진, 조윤경, 정영경, 2015).

이윤(2016)은 읽기 교재가 어떤 어휘로 구성되는가에 따라 텍스트의 난이도가 달라지므로 학생들의 읽기 능력에 적합한 텍스트를 제공하는 것이 읽기 학습에서 중요한 역할을 한다고 했다. 그러나 음운 규칙 기반

어휘로만 텍스트를 구성하는 것은 어려운 일이며, 영어의 특성상 초기에 습득해야 하는 어휘들 중 다수가 음운 규칙이 적용되지 않는 의미 중심 어휘이므로, 읽기 학습 초기에 이 특성을 고려하여 난이도를 조정하는 것이 필요하다고 강조했다.

2. 영어 교과서의 이독성 수준

영어 교과서의 이독성 수준이 학습자의 이해 능력을 초과하는 경우 지식 전달 과정에서 공백이 생기게 되어 학습 의욕 저하를 유발하므로 학습자의 이해력에 맞는 교재를 활용해야 한다(차경환, 손석완 외, 2006). 특히 중학교로 올라가면서 학습 부진의 편차가 커지는 것은 읽기 영역의 학습 부진에서 기인한 것으로 문자와 관련된 영역의 교과 내용의 적정성을 제고할 필요가 있다.

우리나라 2015 개정 교육과정의 초등학교 교과서 읽기 지문의 가독성 지수를 분석한 결과, 5학년과 6학년 사이의 가독성은 미국 학제 기준으로 한 학기 정도의 차이가 났지만(황은경, 박광현, 2019), 초등학교 6학년과 중학교 1학년 읽기 지문 사이에는 미국 학제 기준으로 한 학년 이상의 가독성 수치 차이가 나타난다(배지영, 2019). 임영진, 조윤경, 정영경(2015)은 초등학교 6학년과 중학교 영어 교과서의 이독성과 어휘 난이도를 분석였는데, 이독성 지수는 학교급간에 1점 이상 증가하였으며, 중학교 교과서의 어휘 수는 초등학교 교과서 대비 세 배 이상 증가하였다. 무작위로 증감하는 이독성 수치는 학습자에게 어려움을 줄 수 있으므로, 학습 자료는 저난이도에서 고난이도, 적은 학습량에서 많은 학습량으로 조직화되어야 한다는 결론을 도출하였다.

이윤(2016)은 초등영어 4종의 교과서를 분석해 본 결과, 공통적으로 어휘 반복율, 어휘 난이도 지수 측면에서 모두 읽기 발달 초기 학습자의 특성을 고려하지 못하는 것을 발견했다. 따라서, 중학교 영어 학습 부진

학생은 교과서를 활용한 읽기 활동에 어려움을 겪을 수 있으며, 초등학교 교과서 텍스트를 활용한 읽기 학습을 하더라도, 여전히 학습부진 학생을 읽기 능력 발달을 위한 학습에 적절한 텍스트라고 보기 어렵다.

제 3 절 작업 반복과 유창성

1. 영어 학습자의 반복 읽기 효과성

여러 학자들은 연구를 통해 언어학습자들이 언어학습을 반복(task-repetition)할 때 언어 사용의 복잡성(complexity), 정확성(accuracy), 그리고 유창성(fluency) 등이 제2언어(L2) 학습 및 발달에 어떻게 영향을 미치는지를 탐구해왔다(Hawkes, 2012; Takimoto, 2012). Bygate (1996, 2001)는 Levelt의(1989) 연설 생산 모델을 이용해 작업 반복의 잠재적인 심리언어학적 이점을 설명하였다. 이 모델에 따르면 언어 사용은 개념화(conceptualization), 구현(formulation), 및 발성(articulation)의 세 단계를 거친다. 작업을 반복함으로써 학습자들은 이전의 처리 단계(즉, 개념화, 구현 및 발성)를 기억 저장소에 보관하고 다음번에 재사용할 수 있다. 최수진, 이윤(2015)은 국내 초등학생들의 영어 읽기 유창성 향상을 위해 반복적인 읽기 학습에 음소 지도를 도입하여 음소 인식을 포함한 반복 읽기 방법이 단어 인식의 정확성, 운율, 이해력 증진 등에서 더 큰 향상을 보였음을 확인하였고, 반복적인 읽기가 읽기 유창성과 독해력 향상에 효과적임을 발견했다.

2. 학습 부진 학생의 음절 기반 읽기 활동의 필요성

Metsala와 Walley(1998)은 "lexical restructuring hypothesis"를 제시하여 어휘 성장과 음운인식, 독서 발달 간의 밀접한 관계를 주장하였다. 이 연구에서는 어린이들의 어휘가 확장되고 음운적으로 유사해짐에 따라 음절적 표현이 발달되며, 이러한 음절적 표현이 음운인식의 발달에 기여

하고 독서와 쓰기 학습에 영향을 미친다고 언급했다. Ayres(1998)는 음운인식 프로그램이 영어를 학습하는 어린이들에게 어휘 학습을 촉진 시킨다는 것을 밝혔다. 이 연구에서는 음운인식 프로그램과 어휘 학습 활동을 통합하여 구성된 교육 프로그램이 어린이들의 단어 분절(segmentation) 능력에 긍정적인 영향을 미친다고 결과를 제시했다. 이러한 선행 연구들은 ESL 학생의 음절 연계 단어 학습의 중요성을 강조하며, 구두 언어 능력과 어휘 수준을 고려하여 교육 프로그램을 설계해야 한다는 것을 나타낸다.

엄월영, 김혜련(2014)은 영어 학습 부진아들이 반복 학습 및 효과적인 학습 방법 등을 통해 보다 높은 학습 수준에 도달할 수 있다고 강조하였다. 김은정과 이병민(2021)은 초등학교 6학년 학생들의 영어 읽기 능력을 세분화하여 분석하였는데, 영어는 문자와 파닉스가 일정하지 않음에도 음운법 규칙을 통해 어휘의 84%를 규칙성을 적용해 읽을 수 있고, 알지 못하는 어려운 단어를 유추하는 데에 유용하기 때문에 학생들에게 맞춤형 읽기 보충학습은 필수적이라고 주장했다. 이를 위해서는 단어와 문장 유창성과 관련된 다양한 영어 독서 자료와 함께 해독 능력을 향상하기 위한 교육이 제공되어야 하며, 학생들 간의 개인차는 정규 영어 수업 또는 방과후 교육 등을 통해 공교육 내에서 처치가 이루어져야 한다고 강조했다.

제 4 절 맞춤형 수업을 위한 AI 활용

1. 학습 부진 학생 지도에서의 AI프로그램 활용 필요성

교과 학습 부진 학생을 지도하기 위해서는 각 학생에 맞는 개별 보충지도를 제공해야 하므로 다인수 학급에서는 적절한 수업이 이루어지기 어렵다. 대상 학생 간에도 학습 부진의 정도에 있어 차이를 보이기 때문에 학생들의 영어 수준에 따라 다른 종류의 보충학습이 제공되어야 한다

(엄월영, 김혜련, 2014). 김태은(2020)의 연구에 따르면 학습 부진 학생들의 성장을 위해서는 개별 학생의 학습 과정과 어려움의 원인을 '들여다보기', 학습 단계와 자료를 세밀하게 '세분화하기', 반복적이고 꾸준하게 반복하는 '기초 근육 만들어주기', 꾸준히 성장 과정을 지켜보며 감정적인 지원을 제공하는 '견뎌주기' 과정이 필요하다. 김태은(2020)은 효과적인 학습 부진 지도를 위해서는 교사의 지도계획과 실행이 현실적으로 적정화 되어야 하며, 그 과정에서 AI기반 프로그램을 활용하는 것이 필요하다고 강조했다. 이는 개별지도 효과를 높이면서도 교사의 부담을 줄이고, 실수를 두려워하는 학생들이 정의적 여과(affective filter)가 낮은 상태에서 학습에 임할 수 있도록 하기 때문이다.

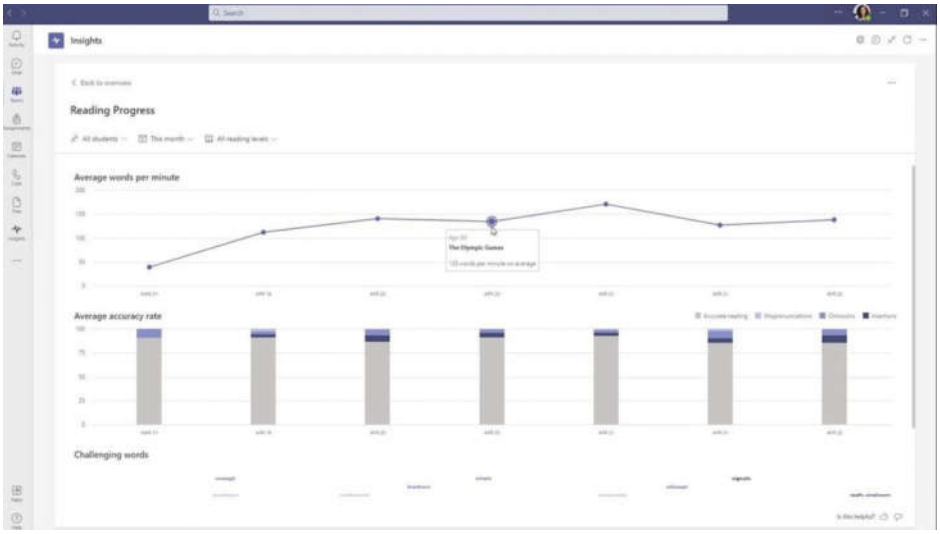
2. 소리 내어 읽기 활동을 위한 AI기반 프로그램

AI기술을 기반으로 한 학습용 애플리케이션 중, 학교 현장에서 가장 대중적으로 사용되는 Microsoft의 Reading Progress 프로그램을 이번 연구 텍스트의 활용 모델로 설정하였다. 해당 프로그램은 교사가 업로드한 텍스트를 학생이 소리 내어 읽으면 학생의 수행 결과를 자동으로 분석한 결과를 [그림 3]과 같이 보여준다.

분석 결과에는 텍스트를 읽는 학생의 영상이 좌측에 나타나고, 학생의 발화가 녹음되어 저장된다. 중앙에 있는 결과 리포트 상단에는 1분당 읽은 올바른 단어 수(correct word per minute)와 정확도 비율(accuracy rate)이 숫자로 표기되어 나타난다. 학생의 발화 오류는 자동으로 다섯 가지 영역(mispronunciation, omission, self-correction, insertion, repetition)으로 분석되어 텍스트에 각기 다른 색깔로 표기된다.

교사는 [그림 3]과 같이 Insights기능을 활용하여 학생의 과제별 수행 결과 그래프를 볼 수 있다. 그래프 하단에는 학생이 부정확하게 읽은 단어가 자동으로 분석되어 워드 클라우드의 형태로 제공된다. 기간을 설정

하여 그룹 평균 수행 대비 해당 학생의 수행 결과를 비교하여 볼 수도 있다.



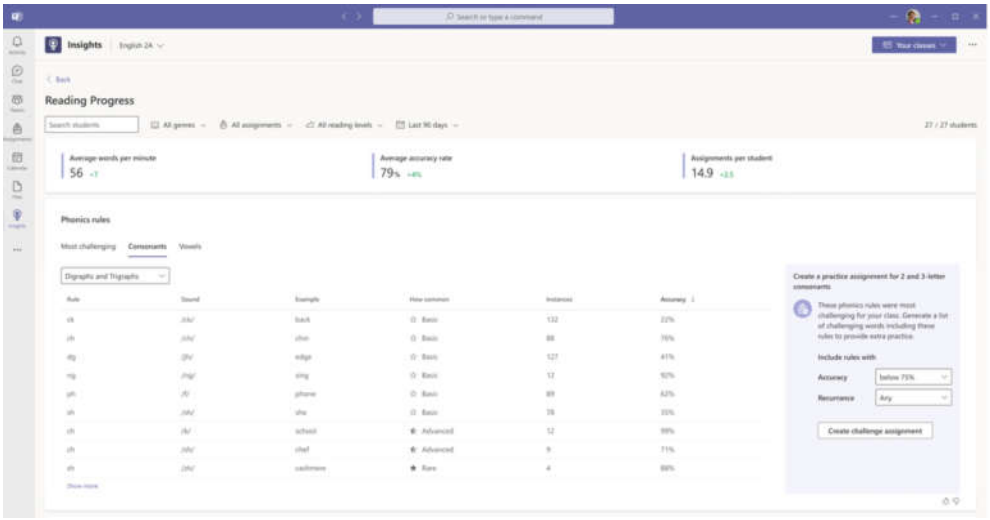
[그림 3] Insights기능을 활용한 학생 수행 결과 자동 분석(Microsoft, 2022)

과제 설정 시 교사는 읽기 후에 리딩 코치(reading coach) 기능을 추가할 수 있다. 리딩 코치 기능은 [그림 4]와 같이 학생이 글을 읽을 때

[그림 4] Reading Coach 기능(Microsoft, 2022)

부정확하게 발음한 단어를 자동으로 분석하여 제시하며, 올바른 발음을 음절별로 다시 들을 수 있도록 음원을 제공한다. 학생은 리딩 코치가 제공하는 발음을 듣고 다시 단어를 올바르게 읽도록 연습하며, 일정 수준 이상의 정확도에 도달하면 다음 단어로 넘어갈 수 있다.

모든 과제가 끝나면 학생이 정확히 읽지 못한 모든 단어가 자동으로 분석되며, 교사는 개별 학생의 이름을 클릭하여 해당 학생이 어떤 음소를 부정확하게 발음했는지를 [그림5]의 화면을 통해 파악할 수 있다. 오류 비율이 높은 음소를 교사가 선택하여 추가 도전 과제 만들기 버튼을 클릭하면 해당 학생에게만 개별화된 단어 읽기 과제가 부여된다.



[그림 5] 학생의 오류 음소 패턴 자동 분석 및 개별화된 추가 과제 자동 생성

제 3 장 연구 방법

본 연구는 학습부진 학습자를 위한 읽기 자료를 개발하는 것을 목표으로 하여 Richey와 Klein(2005)의 설계개발 연구 방법 중 유형 1 개발 연구 방법을 활용하였다. 유형 1 개발 연구는 특정 맥락에 활용되는 특정 산출물이나 도구의 개발 등을 포함한다.

제 1 절 연구 대상 및 도구

1. 연구 대상

텍스트 개발을 위한 연구 대상은 [표 3]에 나타난 2015 개정 교육과정 초등 6학년 5종, 중학교 1학년 9종 교과서의 읽기(reading)영역의 본문을 대상으로 한다.

학교급	출판사	저자
초등	동아	박기화 외
	대교	이재근 외
	천재	함순애 외
	YBM(김)	김혜리 외
	YBM((최)	최희경 외
중등	능률(김)	김성곤 외
	동아(윤)	윤정미 외
	동아(이)	이병민 외
	미래엔(최)	최연희 외
	비상(김)	김진완 외
	지학사	민찬규 외
	천재(이)	이재영 외
	천재(정)	정사열 외
	YBM(송)	송미정 외

[표 3] 연구 대상 교과서 리스트

제 2 절 읽기 자료 개발 기준

이윤(2016)은 교과서 텍스트는 영어 학습자의 학습 내용이자 도구이므로, 초기 읽기 단계에서 학생들이 알파벳 원리를 익혀서 단어 인지 과정을 학습하도록 교재를 체계적으로 구성해야 함을 강조했다. 현재 교과서 텍스트의 경우 내용만 고려하여 어휘 선정을 하는 경우, 등장 인물이나 장소의 이름이 발음하기 어렵거나 길이가 긴데, 이것은 읽기 난이도를 높이는 주 원인임을 지적했다. 따라서 낮은 난이도의 텍스트에서는 소리와 철자의 대응이 명확한 어휘와 고빈도어 중심으로 텍스트를 구성하거나 특정 모음 패턴을 반복적으로 제시하는 등의 방법을 통해 읽기 텍스트가 학생들의 읽기 학습을 돕도록 특별하게 체계화되어야 한다고 강조했다.

Klare(1954)은 이독성과 관련하여 자주 사용되는 두 변수인 '의미 측정'(어휘 난이도 등)과 '구문 측정'(평균 문장 길이 등)이 텍스트 난이도의 가장 효과적인 예측 요인임을 연구를 통해 입증했고, 이 두 변수를 통해 읽기 난이도를 추정하였다. 또한, Klare(1968)의 연구에 따르면, 사람들은 일부 단어를 다른 단어보다 훨씬 더 자주 사용하며, 이런 빈번하게 사용하는 단어들을 더 빨리 인식하고 선호하며, 더욱 쉽게 이해하고 배운다. 그래서 이 '단어의 사용 빈도'라는 변수가 텍스트의 읽기 쉬운 정도를 결정하는 데에 중요한 역할을 한다고 주장했다.

본 연구는 연구의 대상인 교과서 텍스트를 분석할 때 문서에 나타난 단어의 빈도, 단어 음절의 수, 음절의 구성(단모음, 단자음, 이중모음, 이중자음)을 기반으로 한 읽기 난이도를 수치화하여 개발된 최소대립쌍과 짧은 문장을 낮은 난이도부터 배열하였다.

제 3 절 읽기 자료 개발 절차

본 연구는 AI 기술을 활용해 학습 어려움을 겪는 학생들의 영어 읽기 능력 향상을 돕기 위한 읽기 자료를 개발하기 위해 [표4]과 같은 단계를 거쳐 연구를 진행하였다. 우선 초기 읽기 자료를 개발하기 위해 초등학교 6학년과 중학교 1학년의 교과서 읽기 본문 텍스트를 통합하고, Google Colab을 활용하여 연구 대상 텍스트를 분석하였다. 구체적인 텍스트 개발은 다음 4가지 단계로 진행하였다.

연구 방법	세부 내용	산출물
초기 읽기 자료 개발	• 연구 대상 텍스트 선정 및 코퍼스 수집 • 텍스트 전처리 및 고빈도 단어 추출 • 단어의 읽기 난이도 분석 • 최소대립쌍 및 문장 생성 • 최종 문장 확인 및 수정	초기 개발 코드 및 읽기 자료
1차 전문가 검토	• 1차 전문가 검토 • 영어교육, AI 활용 교육 전문가 6인 대상 전문가 검토 • 초기 읽기자료 코드 수정 및 보완	2차 코드 및 읽기 자료
2차 전문가 검토	• 1차 전문가 검토 • 영어교육, AI 활용 교육 전문가 6인 대상 전문가 검토 • 2차 읽기 자료 코드 수정 및 보완	3차 코드 및 읽기 자료
최종 읽기 자료 도출	• 3차 코드 및 읽기 자료 수정 및 보완	최종 코드 및 읽기 자료

[표 4] 읽기 자료 개발 연구 절차

1. 코퍼스 수집

텍스트 개발을 위한 코퍼스를 수집을 위해 본 연구에서는 2015 개정 교육과정의 초등학교 6학년 교과서 5종과 중학교 1학년 교과서 9종의 읽

기 지문 텍스트를 추출하여 하나의 통합된 파일로 작성했다. 이렇게 수집된 코퍼스는 본 연구의 주요 데이터로 활용하였다.

2. 텍스트 전처리 및 고빈도 단어 추출

Google Colab에서 통합된 텍스트에 나타난 단어를 NLTK(Natural Language Toolkit)를 활용하여 토큰화하는 과정을 거쳤다. 토큰화란 텍스트를 개별 단어나 표현으로 분리하는 과정이다. 이를 통해 불필요한 요소를 제거하고, 분석에 필요한 핵심 요소만을 추출했다. 특히, 문장부호, 숫자, 'is', 'the', 'and' 같이 문장 내에서 자주 등장하지만 실질적인 의미를 가지지 않는 불용어 처리를 위해 nltk의 'stopwords' 기능을 사용했다.

불용어를 제거한 후, 텍스트에서 빈도수가 높은 단어를 파악하기 위해 'collections' 모듈의 'counter'를 사용했고, 이를 통해 가장 빈도가 높은 상위 1,500개의 단어를 추출했다. 이후, sklearn의 'TfidfVectorizer'를 이용하여 역문서빈도(TF-IDF)를 계산했다. TF-IDF수치는 특정 문서에서의 단어의 중요성을 판단하는 수치로, 단어가 특정 문서에서 얼마나 자주 등장하며, 전체 문서에서 얼마나 희귀하게 등장하는지를 모두 고려한 값이다.

마지막으로, pandas 라이브러리를 활용해 데이터를 가공하고, TF-IDF 값과 빈도수를 포함한 데이터프레임을 생성했다. 이를 Google Drive 경로에 엑셀 파일로 저장해 분석 결과를 관리하고 추후에 활용할 수 있도록 했다. 이처럼 텍스트 분석 과정을 통해 특정 텍스트 내에서 중요하고 자주 등장하는 단어들을 추출하고 이를 기반으로 읽기 난이도 분석을 위한 기초 데이터를 생성하였다.

3. 단어의 읽기 난이도 분석

이후 각 단어의 읽기 난이도를 결정하기 위해 Pyphen 라이브러리를

활용하여 음절을 분리하여 단어별 음절 수를 산출한 뒤, 음절 구조에 따른 난이도를 매핑하였다. 난이도 설정은 자음('C')과 모음('V')의 조합으로 이루어진 각 패턴에 대해 특정 점수를 부여하는 방식으로 이루어졌다. 가장 단순한 구조인 'C', 'CV', 'VC'의 경우 낮은 난이도 점수인 0.1을, 'CVC', 'VCV'와 같은 구조는 중간 난이도 점수인 0.2를, 'CCV', 'VCC', 'CCVV', 'VVCC'와 같은 상대적으로 복잡한 구조는 0.3의 난이도 점수를 부여하였다. 더욱 복잡한 구조에는 'CCVC', 'CVCC', 'CCVCC', 'CCVVC', 'CVVCC', 'CCVVCC' 등에는 0.4, 그리고 'CCCV', 'CCCVC', 'CCCVV', 'CCCVVC', 'CCCVCCC', 'CCCVVCCC'와 같은 매우 복잡한 구조는 0.5의 난이도 점수를 부여했으며, 그 외 구조는 0.6을 부여했다. 이러한 난이도 점수는 단어가 얼마나 복잡한 음절 구조를 가지고 있는지를 대략적으로 나타낸다. 이렇게 단어의 음절 구조에 따른 난이도를 평가한 뒤, 모든 정보를 데이터프레임에 담아 tf-idf값, 빈도 값과 함께 단어의 음절수와 음절 난이도를 함께 고려하였다.

4. 최소대립쌍 및 문장 생성

최소대립쌍 생성은 난이도 분석을 통해 분류된 단어를 바탕으로 '아주 낮은 난이도'와 '낮은 난이도' 그룹으로 분류된 단어 총 1,500개를 활용했다. 그 후 대상 단어의 발음을 기준으로 음절을 분리하기 위하여 CMUDict 패키지를 활용하여 음절을 분석했다. CMUDict는 Carnegie Mellon University Pronouncing Dictionar의 약어로 134,000개 이상의 단어와 그 발음을 포함하는 발음 사전이다. 총 39개의 음소를 기준으로 하여 주어진 단어의 음절을 구분할 수 있다. 분석 결과를 바탕으로 기준 단어의 발음에서 하나의 음절만 다른 단어를 찾아서 최소대립쌍 세트로 반환하도록 코드를 작성했다.

다음으로 OpenAI의 API를 연결하여 GPT-3 엔진인 text-davinci-003

엔진을 기반으로 문장을 생성했다. 이 과정에서 추출한 1,500개의 단어를 활용하여 최소 2단어에서 5단어 사이의 길이의 짧은 문장을 생성하도록 했다. 문법적인 문장 생성이 가능하도록 기존 단어의 비율을 1.0부터 점차 조금씩 줄여가며 기능어의 비율을 조금씩 높이는 방식으로 반복적으로 문장 생성을 시도하였으며, 이를 통해 기존 추출 단어의 비율을 가장 높이 유지하면서도 문법적 문장 생성이 가능한 적정 수치를 찾아서 문장 생성 코드를 작성하였다. 기존에 추출한 코퍼스 주요 단어의 활용 비율을 최대 0.8로 지정하고, be동사, do동사, 전치사 등의 간단한 단어 리스트의 비율을 0.2로 설정하여 자연스러운 문장을 생성하도록 했다.

생성된 문장의 난이도 수치 산출은 문장을 구성하는 단어의 난이도를 통해 근사하게 문장 난이도를 산출하는 방식을 사용하였으며, 문장 길이와 함께 이전에 산출한 각 문장을 구성하는 단어의 음절 수, 음절 구조 난이도, 교과서 내 출현 빈도를 합산한 문장의 난이도 수치를 산출했다.

문장 길이는 문장에 포함된 단어의 수로 계산되었다. 먼저, 문장을 공백을 기준으로 분리하여 단어의 리스트를 만들었고, 이 리스트의 길이가 곧 문장의 길이가 되었다. 음절 수는 초기에는 문장에서 공백을 제외한 문자의 수로 계산하였으나, 후에 'textstat' 라이브러리의 'syllable_count' 함수를 사용하여 보정 하였다. 이 함수는 주어진 텍스트에 대해 음절 수를 정확하게 계산하는 기능을 제공한다.

마지막으로, 문장 길이, 음절 수, 문장 난이도를 기반으로 난이도 카테고리('하', '중', '상')를 부여하였다. 이를 위해 'categorize_difficulty' 함수를 활용하였으며, 이 함수는 각 기준에 따라 점수를 부여하고, 그 점수의 합에 따라 최종적인 난이도 카테고리를 결정한다. 총 3가지 기준인 문장의 길이, 음절 수, 그리고 단어 난이도를 기준으로 하여 각 요소마다 1점~3점의 난이도 점수를 부여했다. 문장 길이는 문장을 구성하는 단어의 수가 2 이하인 경우 1점, 2~3인 경우 2점, 그 이상인 경우 3점을 부여

하였다. 문장의 음절 수는 3 이하인 경우 1점, 4~6인 경우 2점, 7 이상인 경우 3점을 부여하였다. 단어 난이도는 0.4이하인 경우 1점, 0.4~0.8인 경우 2점, 0.8 이상인 경우 3점을 부여했다. 각 요소가 모두 낮을 경우 최소 점수인 3점을 받아 '하' 난이도로 분류되고, 반대로 모두 높을 경우 최대 점수인 9점을 받아 '상' 난이도로 분류되도록 하였다.

산출된 결과물을 난이도 카테고리별로 분리하여 데이터프레임을 생성한 후, 각각의 데이터프레임을 문장 길이를 기준으로 정렬하고 별도의 엑셀 시트에 저장하였다. 이러한 방식을 통해 문장의 대략적인 읽기 난이도를 정량적으로 측정하고 분류하였다.

제 4 절 내적 타당화

이때 개발된 초기 개발 코드 및 읽기 자료는 Richey와 Klein(2005)의 설계개발 연구 방법 중 유형 1 연구 방법에 따라 1, 2차 전문가 검토를 거쳐 최종 결과물을 산출하였다.

초기 개발 코드 및 읽기 자료의 내적 타당도를 검증하기 위하여 1, 2차에 걸쳐 전문가 타당화를 실시하였다. 내적 타당화에 참여한 전문가 6인은 영어교육, AI 융합 교육 관련 경험이 풍부한 현장 교사로 선정하였으며, 초등 교사 1명, 중등 교사 4명, 고등 교사 1명으로 구성하였다. 6명의 전문가는 모두 본 연구 개발 결과물의 예상 활용 모델인 Reading Progress 프로그램의 해당 프로그램을 학생 교육 목적으로 사용해 본 경험이 있다. 내적 타당화에 참여한 전문가의 6인의 프로필은 [표5]와 같다.

학교급	구분	직위	경력(년)	최종 학력	연구 분야
초등	A	교사	6년	석사	초등 영어 교육, 자연어 처리, 데이터 사이언스
중등	B	교사	10년	석사	영어교육, 교실 평가

	C	수석교사	20년	석사	영어교육, 교육학, 에듀테크 활용 수업
	D	교사	14년	박사수료	영어교육
	E	교사	5년	학사	영어 교육, 에듀테크 활용 수업
고등	F	교사	21년	석사	영어교육, AI 융합교육

[표 5] 내적 타당화 참여 전문가 프로필

내적 타당화에 활용한 전문가 검토지는 나일주, 정현미(200)가 제안한 타당성, 설명력, 유용성, 보편성, 이해도에 관한 문항을 활용하여 [표6]과 같이 작성하였다.

영역	문항
타당성	본 텍스트는 중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 수업을 실시하고자 할 때 활용할 수 있는 것으로 타당하다.
설명력	본 텍스트는 중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 수업을 실시하고자 할 때 활용 할 수 있는 내용들을 잘 설명하고 있다.
유용성	본 텍스트는 중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 수업을 실시하고자 할 때 유용하게 활용될 수 있다.
보편성	본 텍스트는 중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 수업을 실시하고자 할 때 보편적으로 적용할 수 있다.
이해도	본 텍스트는 중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 수업을 실시하고자 할 때 이해하기 쉽게 표현되었다.

[표 6] 내적 타당화 전문가 검토지 문항

각 영역에 대한 타당도를 4점 척도로 평가하도록 하였으며, 수집된 결과물의 내용 타당도 지수(CVI) 및 평가자간 일치도 지수(IRA)를 분석하였다. CVI산출은 3점과 4점을 받은 문항은 1점으로 환산하고, 2점과 1점을 받은 문항은 0점으로 환산하여 합산한 후, 평가에 참여한 전문가의

수로 그 값을 나누었다. 5~9명의 전문가가 참여할 때, CVI가 0.80 이상이면 타당하다고 해석한다(Grant&Davis, 1997). IRA는 평가자들이 동일하게 긍정적으로 평가한 항목의 수를 전체 항목 수로 나눈 값으로, IRA 값이 0이면 평가자들의 의견이 완전히 불일치하는 것이고, 1이면 평가자 간의 의견이 완전히 일치한다고 판단할 수 있다. 일반적으로 IRA값이 0.80 이상이면 평가의 일관성을 신뢰할 수 있다고 판단한다(Rubio et al, 2003). 또한, 각 평가 영역에 대한 개방형 서술 응답도 수집하였다.

제 4 장 연구 결과

제 1 절 초기 읽기 자료 개발 결과 및 전문가 검토 결과

초기 읽기 자료 결과물 중 최소대립쌍 결과물은 [표7]과 같으며, 총 120개의 결과물을 생성하였다. 그러나 이 결과물 내에는 최소대립쌍으로 성립하는 단어와, 최소대립쌍이 아닌 단어가 혼재되어 있다.

Difficulty	Minimal Pair 1	Minimal Pair 2
very_low_difficulty	do	go
very_low_difficulty	do	so
very_low_difficulty	go	so
very_low_difficulty	go	do
very_low_difficulty	so	go
very_low_difficulty	so	do
very_low_difficulty	bad	bag
very_low_difficulty	bad	dad
very_low_difficulty	bad	sad
very_low_difficulty	bad	bed
very_low_difficulty	bag	bad
very_low_difficulty	bag	big
very_low_difficulty	bed	bad
very_low_difficulty	bed	red
very_low_difficulty	big	bag

very_low_difficulty	big	pig
very_low_difficulty	bus	buy
very_low_difficulty	buy	bus
very_low_difficulty	cap	cup
very_low_difficulty	cap	map
very_low_difficulty	cap	cat
very_low_difficulty	cap	car
very_low_difficulty	cap	rap
very_low_difficulty	car	cap
very_low_difficulty	car	far
very_low_difficulty	car	cat
very_low_difficulty	cat	hat
very_low_difficulty	cat	cap
very_low_difficulty	cat	eat
very_low_difficulty	cat	car
very_low_difficulty	cup	cap
very_low_difficulty	dad	bad
very_low_difficulty	dad	sad
very_low_difficulty	dad	day
very_low_difficulty	day	say
very_low_difficulty	day	way
very_low_difficulty	day	dad
very_low_difficulty	eat	hat
very_low_difficulty	eat	cat
very_low_difficulty	fan	far
very_low_difficulty	fan	ran
very_low_difficulty	fan	fun
very_low_difficulty	fan	man
very_low_difficulty	far	fan
very_low_difficulty	far	car
very_low_difficulty	fun	sun
very_low_difficulty	fun	fan
very_low_difficulty	fun	run
very_low_difficulty	get	let
very_low_difficulty	get	got
very_low_difficulty	get	met

very_low_difficulty	got	get
very_low_difficulty	got	lot
very_low_difficulty	got	hot
very_low_difficulty	hat	eat
very_low_difficulty	hat	cat
very_low_difficulty	hat	hit
very_low_difficulty	hat	hot
very_low_difficulty	hit	hat
very_low_difficulty	hit	sit
very_low_difficulty	hit	hot
very_low_difficulty	hot	hat
very_low_difficulty	hot	how
very_low_difficulty	hot	lot
very_low_difficulty	hot	hit
very_low_difficulty	hot	got
very_low_difficulty	how	hot
very_low_difficulty	how	now
very_low_difficulty	job	joy
very_low_difficulty	joy	job
very_low_difficulty	let	get
very_low_difficulty	let	lot
very_low_difficulty	let	met
very_low_difficulty	lot	let
very_low_difficulty	lot	got
very_low_difficulty	lot	hot
very_low_difficulty	man	fan
very_low_difficulty	man	map
very_low_difficulty	man	ran
very_low_difficulty	man	men
very_low_difficulty	map	cap
very_low_difficulty	map	man
very_low_difficulty	map	rap
very_low_difficulty	men	met
very_low_difficulty	men	man
very_low_difficulty	met	let
very_low_difficulty	met	get

very_low_difficulty	met	men
very_low_difficulty	mom	tom
very_low_difficulty	new	now
very_low_difficulty	now	new
very_low_difficulty	now	how
very_low_difficulty	pig	big
very_low_difficulty	pop	top
very_low_difficulty	ran	fan
very_low_difficulty	ran	run
very_low_difficulty	ran	man
very_low_difficulty	ran	rap
low_difficulty	curry	hurry
low_difficulty	funny	sunny
low_difficulty	hurry	curry
low_difficulty	later	water
low_difficulty	sorry	worry
low_difficulty	sunny	funny
low_difficulty	water	later
low_difficulty	worry	sorry
low_difficulty	became	become
low_difficulty	become	became
low_difficulty	better	butter
low_difficulty	butter	better
low_difficulty	forget	forgot
low_difficulty	forgot	forget
low_difficulty	making	taking
low_difficulty	taking	making
low_difficulty	cooking	looking
low_difficulty	grandma	grandpa
low_difficulty	grandpa	grandma
low_difficulty	looking	cooking
low_difficulty	talking	walking
low_difficulty	walking	talking

[표 7] 초기 최소대립쌍 개발 결과

초기 읽기 자료 결과물 중 문장 생성 결과물은 [표8]과 같으며, 총 169

개의 짧은 문장을 생성하였다.

Sentence	TF-IDF Sum	Difficulty Score	Average Word Length
Go for it!	0.883449069	0.25	4
Take a break.	0.388348624	0.5	4
Here we go!	0.193000108	0.35	4
Clean the dishes.	0.111803399	0.375	4
Enjoy the moment.	0.111803399	0.475	4
Enjoy the sunshine.	0.111803399	0.475	4
Look up at the sky.	0.092474337	0.283333333	6
Go to the store.	0.091151001	0.3	5
Here I am.	0.041042593	0.475	4
I don't understand.	0.033451892	0.5	5
Nice to meet you.	0.002361821	0.32	5
I love to cook.	0.001708282	0.42	5
I love to laugh.	0.001708282	0.42	5
I like to dance.	0.001708282	0.5	5
I like to smile.	0.001708282	0.5	5
Go outside and play.	0.00120169	0.32	5
See you soon.	0	0.25	4
Eat healthy food.	0	0.275	4
Happy New Year!	0	0.3	4
Eat ice cream today.	0	0.34	5
The sun shines bright.	0	0.34	5
Take it easy.	0	0.35	4
Go away now.	0	0.375	4
Eat well today.	0	0.375	4
Go outside now.	0	0.375	4
Eat more fruits.	0	0.375	4
She smiled brightly.	0	0.375	4
He loves her.	0	0.375	4
I ate lunch.	0	0.4	4
Cats love milk.	0	0.4	4
Sunny days ahead.	0	0.4	4
Happy summer days!	0	0.4	4
Eat breakfast every day.	0	0.42	5
Let's eat pizza.	0	0.42	5
Let's go camping.	0	0.42	5
It's sunny today.	0	0.42	5

Let's go outside!	0	0.42	5
Let's go swimming.	0	0.42	5
Let's go shopping.	0	0.42	5
It's sunny outside.	0	0.42	5
I'm ready now.	0	0.44	5
I'm happy now.	0	0.44	5
Have a good day!	0	0.44	5
Have a great day.	0	0.44	5
I like ice cream.	0	0.44	5
I love ice cream.	0	0.44	5
I love sunny days.	0	0.44	5
I saw her smile.	0	0.44	5
It's a beautiful day.	0	0.45	6
I love you!	0	0.475	4
I like pie.	0	0.475	4
I love you.	0	0.475	4
I like you.	0	0.475	4
I love tea.	0	0.475	4
I love pie.	0	0.475	4
Blue sky today.	0	0.475	4
Make it happen.	0	0.475	4
I eat breakfast.	0	0.475	4
She smiled happily.	0	0.475	4
She laughed happily.	0	0.475	4
Eat healthily today.	0	0.475	4
Eat more vegetables.	0	0.475	4
He smiled happily.	0	0.475	4
He laughed happily.	0	0.475	4
I like cats.	0	0.5	4
I love cars.	0	0.5	4
Make a wish.	0	0.5	4
I like dogs.	0	0.5	4
I love cats.	0	0.5	4
I love dogs.	0	0.5	4
I love pets.	0	0.5	4
I like food.	0	0.5	4
I love food.	0	0.5	4
I love shoes.	0	0.5	4
I like books.	0	0.5	4

I love books.	0	0.5	4
I laugh often.	0	0.5	4
I love skiing.	0	0.5	4
I love spring.	0	0.5	4
Nice day today.	0	0.5	4
I will succeed.	0	0.5	4
Take action now.	0	0.5	4
Hope never dies.	0	0.5	4
I ate breakfast.	0	0.5	4
Make today count.	0	0.5	4
I laughed loudly.	0	0.5	4
Love never fails.	0	0.5	4
Cats love playing.	0	0.5	4
Play outside today.	0	0.5	4
Good morning everyone!	0	0.5	4
Good morning sunshine!	0	0.5	4
Happy weekend everyone!	0	0.5	4
Everything changes quickly.	0	0.5	4
I love it here.	0	0.5	5
It's raining outside.	0	0.5	5
It's beautiful outside.	0	0.5	5
I'm here now.	0	0.52	5
I love my dog.	0	0.52	5
I love summer days.	0	0.52	5
I love reading books.	0	0.52	5
I like reading books.	0	0.52	5
I enjoy reading books.	0	0.52	5
I'm happy today.	0	0.52	5
I love hot chocolate.	0	0.52	5
I love cake.	0	0.6	4
I love life.	0	0.6	4
I like cake.	0	0.6	4
I love music.	0	0.6	4
I love pizza.	0	0.6	4
I love Italy.	0	0.6	4
I love pasta.	0	0.6	4
I like pizza.	0	0.6	4
I love pizza!	0	0.6	4
I love tacos.	0	0.6	4

I smile often.	0	0.6	4
I love winter.	0	0.6	4
Take a chance.	0	0.6	4
I love cheese.	0	0.6	4
I like coffee.	0	0.6	4
I love apples.	0	0.6	4
I love summer.	0	0.6	4
I love coffee.	0	0.6	4
I love travel.	0	0.6	4
I love hiking.	0	0.6	4
I love nature.	0	0.6	4
I like apples.	0	0.6	4
I like cheese.	0	0.6	4
I like summer.	0	0.6	4
I like animals.	0	0.6	4
I love reading.	0	0.6	4
I like bananas.	0	0.6	4
I love England.	0	0.6	4
I love singing.	0	0.6	4
I love summers.	0	0.6	4
I love beaches.	0	0.6	4
I love sunsets.	0	0.6	4
I like turtles.	0	0.6	4
I like cookies.	0	0.6	4
I love bananas.	0	0.6	4
I like reading.	0	0.6	4
I love cookies.	0	0.6	4
I love turtles.	0	0.6	4
I love running.	0	0.6	4
I love kittens.	0	0.6	4
I like running.	0	0.6	4
I love animals.	0	0.6	4
I like singing.	0	0.6	4
I smile easily.	0	0.6	4
I like learning.	0	0.6	4
I smile happily.	0	0.6	4
I love swimming.	0	0.6	4
I love pancakes.	0	0.6	4
I love football.	0	0.6	4

I love sunshine.	0	0.6	4
I love learning.	0	0.6	4
Take care today.	0	0.6	4
I love laughter.	0	0.6	4
I smiled happily.	0	0.6	4
I love chocolate.	0	0.6	4
I love computers.	0	0.6	4
I love traveling.	0	0.6	4
I love basketball.	0	0.6	4
I love springtime.	0	0.6	4
I love summertime!	0	0.6	4
I love summertime.	0	0.6	4
Nice weather today.	0	0.6	4
I love my life.	0	0.6	5
I'm learning English.	0	0.6	5
I love peanut butter.	0	0.6	5

[표 8] 초기 문장 개발 결과

초기 개발 코드 및 텍스트에 대한 1차 타당화 결과는 [표 9]와 같다.

항목	전문가						평균	SD	CVI	IRA
	A	B	C	D	E	F				
타당성	3	3	4	2	3	3	3	0.63	0.83	
설명력	3	3	4	3	4	3	3.33	0.51	1	
유용성	4	4	4	3	4	4	3.83	0.40	1	0.8
보편성	4	4	4	4	4	3	3.83	0.40	1	
이해도	3	3	4	3	3	3	3.16	0.40	1	

[표 9] 1차 전문가 타당화 결과

전문가들은 초기 개발 결과물의 타당성, 설명력, 유용성, 보편성, 이해도의 각 항목에 대하여 모든 문항에서 내용 0.8이상의 내용 타당도 지수(CVI)를 보였다. 채점자 간 일치도 지수(IRA)는 0.8로 나타나서 해당 평가 결과의 신뢰도가 높다고 할 수 있다. 특히 유용성과 보편성 항목은

평균 3.83점으로 전문가들이 가장 긍정적인 반응을 보였다. 이는 본 연구 목표에 따른 결과물이 현장 교사들의 학습 부진 학생 지도에 유용할 것이며, 여러 교사들이 보편적으로 사용할 수 있는 읽기 자료라고 해석할 수 있다. 그러나 타당성 항목의 내용 타당도 지수(CVI)는 0.83으로 한 명의 전문가가 '타당하지 않음'(2점) 의견을 나타냈고, 구체적인 의견은 개방형 응답 결과에 자세히 기술하였다.

각 영역에 대한 전문가들의 개방형 응답 결과와 의견 반영 내용은 [표 10]과 같다.

영역	우수한 점	보완할 점	의견 반영 내용
타당성	• 텍스트 전처리 기법과 영어학적 지식을 혼합하여 텍스트 난이도를 설정함	• TF-IDF값과 난이도의 상관관계에 관한 연결성 부족, 각 수치를 별도로 산출하여 관계 비교 제안	• 난이도 재산출 및 기존 데이터와 비교
	• 단어 빈도 뿐 아니라 읽기 난이도에 대한 선행연구를 다양한 측면에서 분석함	• 연구에서 사용한 davinci-003 모델 외의 다른 모델 출력 결과 비교 제안	• gpt-3.5-turbo 모델 활용 텍스트 비교
		• 최소대립쌍에 해당하지 않는 단어 세트 삭제 필요	• 최소대립쌍 검토 후 부적절한 세트 삭제
	• 초등학교 및 중학교 교과서 전종을 활용하여 교육과정과의 연계성이 높음	• 개발 문장 구조를 다양하게 구성하는 것 제안	• 후속 연구 제안에 내용 포함
	• 현행 초, 중학교 교과서의 단어를 활용하였으므로 학습부진 학생을 위한 자료로서의 타당도가 높음	• 3~4 문장 정도의 짧은 단락 개발 후속 연구 제안	
		• 텍스트 구성 난이도 기준의 타당도 관련 근거 부족	• 선행연구 보완 및 텍스트 생성 기준 보완
설명력	• 연구의 배경, 목적, 설계 과정 설명의 명확성	• 데이터 프레임 생성시 단어의 난도 결정 요인 중 일부(TF-IDF값)는 선행연구를 통해 충분히 설명되지 않음	• 선행연구 설명 보완
		• 기존 학습 부진 학생 지도 자료의 한계 설명 보완	
		• 연구 절차에 활용된 각종	• 연구 절차 설

		용어 및 난이도 설정 기준에 대한 추가 설명 필요	명 보완
	• 어휘 읽기 지도 필요성의 설명의 논리성	• 읽기 자료를 활용한 수업 및 평가시 각 항목을 어떻게 활용할 수 있는지에 대한 안내 추가 필요	
		• 마지막 개발 단계에서 최소대립쌍 및 문장 수정 기준에 대한 설명 필요	
유용성	• 현장에서 학습부진 학생 지도에 유용하게 활용할 것으로 판단		
	• Reading Progress 프로그램에서 기존에 제공하던 텍스트가 한국 학습 부진 학생들에게 적절하지 않아서 발생하는 교사들의 자료 개발 부담을 크게 완화함		
	• 학생들의 음운 인식 및 발음 관련하여 구체적으로 어려운 점을 파악하고 지원할 수 있을 것으로 기대함		
보편성	• 한국 교육과정 내의 교과서에 포함된 고빈도 단어 대상이므로 많은 학교에서 학습부진 자료로서 활용도 높음	• 본 연구의 핵심은 텍스트 생성 기법(method)이며, 선행연구와 생성형 거대 언어 모델 기반으로 결과를 도출한 것이므로 소규모의 실증 추가 제안	• 후속연구 제안에 내용 포함
	• 현행 기초학력 진단도구와 함께 보조적 진단도구로 활용 또는 기초학력 지도 자료 등으로 광범위하게 사용될 수 있을 것으로 판단	• 후속 연구로 해당 자료 활용 수업 모형 제안	
이해도	• 개발된 읽기 자료를 AI기반 프로그램에서 활용할 수 있는 방법에 대한 추가 설명 필	• 영어과 최소한의 성취기준 관련 선행연구 보완	• 선행연구 보완
		• 문장의 난이도 위계 결정 기준에 대한 상세 설명 필요	• 연구 절차 설명 보완

	요	• ‘텍스트’라는 용어의 모호성	• ‘텍스트’ 용어를 ‘읽기 자료’로 변경

[표 10] 1차 전문가 검토 개방형 응답 의견 및 반영 사항

타당도와 관련된 개방형 응답에서 다수의 전문가들은 텍스트의 전처리 기법과 영어학적 지식의 혼합을 통한 난이도 설정에 대해 긍정적으로 평가하였다. 이를 통해 단어의 빈도 뿐 아니라 읽기 난이도에 대한 다양한 요소가 고려되었다는 점이 특히 주목받았다. 또한, 초등학교와 중학교 교과서를 전체적으로 활용해 교육과정과의 연계성을 높인 점, 현행 교과서의 단어를 이용한 점에서 학습 부진 학생들을 위한 자료로서의 타당도가 높게 판단되었다.

그러나, 일부 부분에서 보완이 필요하다는 의견이 제시되었다. 텍스트 구성 난이도 기준 선정에 대한 근거가 부족하다는 점과 TF-IDF 값과 난이도의 상관관계에 대한 연결성이 부족하다는 지적이 있었으며, 이에 대한 보다 명확한 근거 제시를 요구하였다. 추가로, davinci-003 모델 외의 다른 모델과의 출력 결과 비교를 제안하였다. 또한, 결과물 중 일부 단어 세트가 최소대립쌍에 해당하지 않음에도 불구하고 포함되어 있어 이 부분의 삭제가 필요하다는 지적이 있었다.

설명력에 관하여 전문가들은 연구의 배경, 목적, 설계 과정에 대한 설명이 명확하게 이루어졌다고 평가하였다. 그러나 데이터 프레임 생성시 단어의 난이도 결정 요인 중 일부인 TF-IDF 값에 대한 설명이 선행 연구를 통해 충분히 이루어지지 않았다는 지적이 있었다. 또한, 기존 학습 부진 학생 지도자료의 한계에 대한 설명이 보다 구체적으로 보완될 필요가 있음을 제언하였다.

연구 절차에 활용된 각종 용어 및 난이도 설정 기준에 대한 추가 설명이 필요하다는 의견도 있었다. 또한, 어휘 읽기 지도의 필요성에 대한 설

명의 논리성을 높이는 방향으로 보완할 필요가 있다고 제안하였다. 읽기 자료를 활용한 수업 및 평가 시 각 항목을 어떻게 활용할 수 있는지에 대한 안내를 추가적으로 제공하는 것이 필요하다는 의견도 있었다. 마지막으로, 마지막 개발 단계에서 최소대립쌍 및 문장 수정 기준에 대한 설명이 필요하다는 점을 지적하였다.

유용성과 관련하여서 전문가들은 AI를 활용한 읽기 텍스트 세트가 현장에서 학습부진 학생들의 지도에 유용하게 활용될 것으로 판단하였다. 특히 Reading Progress 프로그램에서 기존에 제공되던 텍스트가 한국 학습 부진 학생들에게 적절하지 않아서 발생하는 교사들의 자료 개발 부담을 크게 완화할 것으로 보인다는 점을 장점으로 판단하였다. 또한, 개발된 텍스트를 활용함으로써 학생들의 음운 인식 및 발음에 대한 구체적인 어려움을 파악하고, 이에 대한 지원이 가능할 것으로 기대한다는 의견이 제시되었다.

보편성 측면에서 전문가들은 교과서에 포함된 고빈도 단어를 대상으로 한 점이 많은 학교에서 학습 부진 학생 지도를 위한 자료로 활용성이 높을 것으로 평가하였다. 또한 전문가들은 본 연구의 핵심인 텍스트 생성 기법이 거대 언어 모델을 기반으로 한 점을 긍정적으로 평가하였다. 그러나, 전문가들은 이에 대한 신뢰성을 더 확보하기 위해 소규모의 실증 연구를 추가로 제안하였다. 개발된 텍스트를 현행 기초학력 진단 도구와 함께 보조적인 진단 도구로 활용하거나, 기초학력 지도를 위한 수업 자료 등으로 광범위하게 사용될 수 있을 것이라는 의견도 제시하였다.

이해도 영역에 대해서 전문가들은 개발된 읽기 자료를 AI 기반 프로그램에서 어떻게 활용할 수 있는지에 대한 추가적인 설명이 필요하다고 지적하였다. 또한, 영어과의 최소한의 성취기준과 관련된 선행연구가 보완되어야 한다는 의견을 제시하였다. 더불어, 문장의 난이도를 결정하는 기준에 대해 더 상세하게 설명하는 것이 필요하다고 지적하였다. 전문가

들은 또한 '텍스트'라는 용어에 대한 모호성을 지적하였으며, 이 부분에 대해 명확한 정의가 제시되어야 할 것으로 제안하였다. 마지막으로, '읽기 유창성'에 대한 조작적 정의를 추가하는 것이 필요하다고 전문가들은 평가하였다.

전문가 검토 의견에 따라 아래와 같이 연구를 수정하였다. 우선, 읽기 자료 개발 기준의 타당도를 높이기 이독성과 관련된 선행연구를 추가하고 텍스트 생성 기준을 더욱 구체화하였다. 또한, 읽기 자료 개발 단계에 활용되는 주요 개념에 대한 설명을 추가하였다. 단어의 난이도 산출 기준에는 단어의 TF-IDF값에 더하여 문서 내의 출현 빈도, 실생활 활용 빈도 데이터를 추가하였다. 단어의 실생활 활용 빈도 수치 비교를 위하여 Brown Corpus를 활용하여 각 단어의 'real_life_frequency' 값을 추가했다. Brown Corpus는 500개의 다양한 분야에서 사용되는 약 100만개의 단어로 구성되어 있으며, 미국 영어의 광범위한 사용 예시를 반영하고 있고, 각 텍스트는 품사 태깅(part-of-speech)가 적용되어있는 텍스트 데이터의 집합이다.

단어의 난이도 계산시에는 단어의 음절 수 기준에 더하여 단모음, 단자음, 이중모음, 이중자음의 결합 체계를 구분하여 단어 난이도 수치를 부여했다. 최소대립쌍 생성시에는 itertools라는 패키지를 활용하여 한 개의 음소가 다르게 발음되는 단어를 생성하였으며, 한 단어에 여러 개의 단어가 추천되도록 했다.

제 2 절 2차 개발 읽기 자료 및 전문가 검토 결과

2차 전문가 검토는 1차 검토와 동일한 전문가가 참여하였으며, 검토에 활용한 2차 개발 결과물은 [표11], [표12]와 같다. 생성된 문장에 대해 문장 구조, 단어의 개수, 단어의 길이, TF-IDF값, 단어의 실생활 빈도, 단어의 난이도, 과거 동사의 유무의 기준에 따라 난이도 점수를 수치화 하

여 부여했다. 실제 문서에 나타난 빈도를 추정하는 TF-IDF값과 함께 Brown corpus를 기준으로 산출한 real_life_frequency 수치를 반영할 때에는 실제 생활에서 빈번하게 사용하는 단어의 수치를 전체 난이도 수치에서 감하는 방식으로 문장 난이도를 산출하는 방식을 시도했다. 또한, spacy라는 NLP 라이브러리를 사용하여 문장 구조를 분석하였다. 동사의 경우 과거 동사의 유무를 파악하고, 현재형 동사를 활용하는 경우 과거형 동사에 비해 조금 더 쉬운 난이도 값이 부여되도록 하였다.

그 후 문장을 구성하는 단어들의 난이도 점수의 합과 문장 길이 수치를 5:5로 반영하고 정규화하여 새로운 문장 난이도 점수를 계산하고, 이를 기준으로 문장을 상, 중, 하 3가지 카테고리로 분류했다.

word	minimal_pairs	real_life_frequency	frequency	tf-idf	syllable
did	['dad']	994	0	0.0254989	1
way	['day', 'pay', 'say']	892	15	0.0254989	1
good	['food', 'wood']	767	25	0.0509978	1
men	['pen', 'man']	736	2	0.0254989	1
long	['song']	713	18	0.0509978	1
year	['hear', 'near', 'wear']	649	3	0.0509978	1
right	['light', 'night']	597	11	0.0509978	1
come	['home', 'came']	589	19	0.0509978	1
small	['smell']	518	15	0.0254989	1
went	['tent', 'want']	506	25	0.0254989	1
say	['day', 'pay', 'way']	495	3	0.0254989	1
got	['hot', 'lot', 'pot', 'god']	471	15	0.0254989	1
think	['thank', 'thing']	427	11	0.0509978	1
took	['book', 'cook', 'look']	425	9	0.0254989	1
hand	['band', 'land', 'sand']	418	6	0.0254989	1
told	['cold', 'hold']	411	6	0.0254989	1
far	['car']	409	3	0.0509978	1
set	['let', 'met', 'pet', 'wet']	408	2	0.0254989	1
house	['mouse']	400	18	0.0254989	1
night	['light', 'right']	398	10	0.0254989	1
best	['rest']	341	10	0.0509978	1

thing	['think']	331	2	0.0254989	1
power	['tower']	326	1	0.0254989	1
want	['went']	326	21	0.0509978	1
light	['night', 'right']	320	1	0.0254989	1
big	['pig', 'bag']	316	28	0.0254989	1
let	['met', 'pet', 'set', 'wet']	293	3	0.0509978	1
means	['beans', 'meals']	291	6	0.0254989	1
gave	['save', 'game']	285	9	0.0254989	1
line	['life', 'like', 'live']	282	4	0.0254989	1
car	['far']	270	5	0.0254989	1
tell	['fell', 'sell', 'tall']	245	5	0.0254989	1
wife	['life', 'wise']	227	7	0.0509978	1
love	['live']	214	33	0.0254989	1
run	['fun', 'sun', 'ran']	206	6	0.0254989	1
mean	['meal', 'meat']	199	2	0.0254989	1
land	['band', 'hand', 'sand']	199	6	0.0254989	1
play	['plan']	197	29	0.0254989	1
plan	['play']	196	4	0.0254989	1
cut	['cat', 'cup']	180	2	0.0254989	1
soon	['moon', 'noon']	176	3	0.0509978	1
book	['cook', 'look', 'took']	176	4	0.0254989	1
pay	['day', 'say', 'way']	169	2	0.0254989	1
read	['dead', 'head']	167	5	0.0254989	1
dead	['head']	166	2	0.0254989	1
class	['glass']	164	22	0.0509978	1
hold	['cold', 'told']	162	3	0.0509978	1
cold	['hold', 'told']	161	14	0.0254989	1
rest	['best']	160	4	0.0254989	1
hear	['near', 'year']	151	2	0.0254989	1
sat	['cat', 'hat', 'sit', 'sad']	149	3	0.0254989	1
bad	['dad', 'sad', 'bed', 'bag']	139	6	0.0254989	1
wall	['ball', 'fall', 'tall', 'walk']	139	7	0.0254989	1
fall	['ball', 'tall', 'wall', 'fell']	136	3	0.0254989	1
ran	['fan', 'man', 'run', 'rap']	132	8	0.0254989	1
met	['let', 'pet', 'set', 'wet', 'men']	129	3	0.0254989	1
bed	['fed', 'red']	126	4	0.0254989	1
lot	['got', 'hot', 'pot', 'let']	124	20	0.0254989	1

game	['came', 'gave']	122	19	0.0509978	1
hot	['got', 'lot', 'pot', 'hat', 'hit']	118	14	0.0254989	1
farm	['warm']	112	9	0.0254989	1
hit	['sit', 'hat', 'hot']	111	5	0.0254989	1
stop	['shop']	110	3	0.0254989	1
wish	['dish', 'fish', 'wash']	109	2	0.0254989	1
stay	['star']	107	4	0.0254989	1
sun	['fun', 'run']	102	7	0.0254989	1
race	['rice']	100	10	0.0254989	1
talking	['walking']	99	2	0.0254989	2
walk	['talk', 'wall']	96	8	0.0254989	1
project	['protect']	91	4	0.0254989	2
ship	['shop']	82	13	0.0254989	1
sea	['tea']	80	17	0.0254989	1
nice	['rice']	69	6	0.0509978	1
sit	['hit', 'sat', 'set']	65	10	0.0254989	1
warm	['farm']	65	8	0.0254989	1
grew	['drew']	64	2	0.0254989	1
nose	['note']	60	2	0.0254989	1
save	['gave', 'safe']	60	6	0.0509978	1
eat	['cat', 'hat', 'sat']	59	21	0.0254989	1
safe	['save']	58	2	0.0254989	1
song	['long', 'sang', 'sing']	56	3	0.0509978	1
shop	['stop', 'ship']	56	3	0.0254989	1
tree	['free', 'true']	56	6	0.0509978	1
hat	['cat', 'sat', 'hit', 'hot']	55	2	0.0254989	1
seat	['meat']	54	2	0.0254989	1
text	['tent']	54	2	0.0254989	1
draw	['drew']	53	3	0.0254989	1
tall	['ball', 'fall', 'wall', 'tell']	53	3	0.0254989	1
wet	['let', 'met', 'pet', 'set', 'web']	51	2	0.0254989	1
wood	['food']	51	2	0.0254989	1
moon	['noon', 'soon']	49	3	0.0254989	1
meat	['seat', 'meet', 'meal', 'mean']	45	4	0.0254989	1
ride	['rode', 'rice']	45	4	0.0254989	1
band	['hand', 'land', 'sand']	44	2	0.0254989	1

cup	['cut']	44	4	0.0254989	1
fun	['run', 'sun', 'fan']	43	11	0.0254989	1
bag	['big', 'bad']	42	2	0.0254989	1
fed	['bed', 'red']	40	2	0.0254989	1
seed	['need', 'sees']	40	3	0.0254989	1
sell	['fell', 'tell']	40	3	0.0254989	1
sad	['bad', 'dad', 'sam', 'sat']	34	2	0.0254989	1
wise	['wife']	33	3	0.0509978	1
wash	['wish']	33	5	0.0254989	1
pot	['got', 'hot', 'pet']	28	2	0.0254989	1
sand	['band', 'hand', 'land']	28	3	0.0254989	1
bath	['math']	25	2	0.0254989	1
king	['sing']	25	18	0.0509978	1
noon	['moon', 'soon']	24	2	0.0254989	1
rice	['nice', 'race', 'ride']	24	6	0.0254989	1
cook	['book', 'look', 'took']	24	8	0.0254989	1
tent	['went']	20	1	0.0254989	1
god	['got']	17	3	0.0254989	1
fan	['man', 'ran', 'fun']	17	3	0.0254989	1
dish	['fish', 'wish']	15	5	0.0254989	1
cake	['bake', 'lake', 'make', 'came']	13	6	0.0254989	1
map	['rap', 'man']	12	3	0.0254989	1
swim	['swam']	11	2	0.0254989	1
lake	['bake', 'cake', 'make', 'like', 'late']	11	5	0.0509978	1
fox	['box']	9	1	0.0254989	1
duck	['luck']	9	0	0.0254989	1
pet	['let', 'met', 'set', 'wet]	8	2	0.0254989	1
pig	['big']	8	5	0.0254989	1
web	['wet']	6	2	0.0254989	1
math	['bath']	3	7	0.0254989	1
dad	['bad', 'sad', 'did']	2	2	0.0509978	1
rap	['map', 'ran']	2	3	0.0254989	1
bake	['cake', 'lake', 'make']	2	3	0.0254989	1
grandma	['grandpa']	1	0	0.0254989	1

[표 11] 1차 전문가 검토 결과를 반영한 최소대립쌍 읽기 자료(2차)

sentence	difficulty	length	normalized_ score	category	contains_pre sent_verb
I win!	10	6	0.498228663	하	TRUE
I smile.	10	8	0.560728663	하	TRUE
I'm here.	10	9	0.591978663	하	TRUE
I'm late.	13	9	0.592461755	하	TRUE
I'm home.	13	9	0.592461755	하	TRUE
Good day!	7	9	0.591495572	하	FALSE
Have fun!	7	9	0.591495572	하	FALSE
Run fast!	7	9	0.591495572	하	FALSE
Let's go!	10	9	0.591978663	하	FALSE
I smiled.	12	9	0.592300725	하	FALSE
Thank you.	7	10	0.622745572	하	TRUE
Good luck!	7	10	0.622745572	하	FALSE
Happy day!	7	10	0.622745572	하	FALSE
Eat lunch.	7	10	0.622745572	하	FALSE
Sunny day.	7	10	0.622745572	하	FALSE
Let's eat.	10	10	0.623228663	하	FALSE
I like you.	-2442	11	0.259631643	하	TRUE
I love you.	-396	11	0.589100242	하	TRUE
I like snow.	-2442	12	0.290881643	하	TRUE
I like dogs.	-2442	12	0.290881643	하	TRUE
I like cake.	-2442	12	0.290881643	하	TRUE
I like cats.	-2442	12	0.290881643	하	TRUE
I like blue.	-2442	12	0.290881643	하	TRUE
I love dogs.	-396	12	0.620350242	하	TRUE
I love life.	-396	12	0.620350242	하	TRUE
I love cats.	-396	12	0.620350242	하	TRUE
I love food.	-396	12	0.620350242	하	TRUE
I like books.	-2442	13	0.322131643	하	TRUE
I like pizza.	-2442	13	0.322131643	하	TRUE
I like bikes.	-2442	13	0.322131643	하	TRUE
I like tacos.	-2442	13	0.322131643	하	TRUE
I like cheese.	-2442	14	0.353381643	하	TRUE
I like coffee.	-2442	14	0.353381643	하	TRUE
I like apples.	-2442	14	0.353381643	하	TRUE
I like cookies.	-2442	15	0.384631643	하	TRUE
I like turtles.	-2442	15	0.384631643	하	TRUE
I like reading.	-2442	15	0.384631643	하	TRUE

I like football.	-2442	16	0.415881643	하	TRUE
Have a good day!	-1506	16	0.56660628	하	FALSE
I like ice cream.	-2514	17	0.43553744	하	TRUE
We go now.	12	10	0.623550725	중	TRUE
I'm happy.	13	10	0.623711755	중	TRUE
I'm sorry.	13	10	0.623711755	중	TRUE
I'm tired.	13	10	0.623711755	중	TRUE
I'm ready.	13	10	0.623711755	중	TRUE
I'm happy!	13	10	0.623711755	중	TRUE
I'm Ready.	13	10	0.623711755	중	TRUE
He smiled.	12	10	0.623550725	중	FALSE
I'm hungry.	13	11	0.654961755	중	TRUE
It's sunny.	13	11	0.654961755	중	TRUE
Lovely day!	7	11	0.653995572	중	FALSE
Enjoy life.	7	11	0.653995572	중	FALSE
Eat dinner.	7	11	0.653995572	중	FALSE
Let's play.	10	11	0.654478663	중	FALSE
We laughed.	12	11	0.654800725	중	FALSE
I can help.	12	11	0.654800725	중	FALSE
I'm smiling.	10	12	0.685728663	중	TRUE
I'm excited.	13	12	0.686211755	중	TRUE
I see stars.	15	12	0.686533816	중	TRUE
I ate lunch.	-3	12	0.683635266	중	FALSE
Hello World!	7	12	0.685245572	중	FALSE
Eat healthy.	7	12	0.685245572	중	FALSE
Stay strong.	7	12	0.685245572	중	FALSE
Take a walk.	9	12	0.685567633	중	FALSE
I love books.	-396	13	0.651600242	중	TRUE
I love pizza.	-396	13	0.651600242	중	TRUE
Welcome home!	7	13	0.716495572	중	TRUE
I ate dinner.	-3	13	0.714885266	중	FALSE
Happy Friday!	7	13	0.716495572	중	FALSE
Good morning!	7	13	0.716495572	중	FALSE
Good evening.	7	13	0.716495572	중	FALSE
Take chances.	7	13	0.716495572	중	FALSE
Come quickly.	7	13	0.716495572	중	FALSE
Love forever.	7	13	0.716495572	중	FALSE
Happy Monday!	7	13	0.716495572	중	FALSE
Take a break.	9	13	0.716817633	중	FALSE

Take it easy.	9	13	0.716817633	중	FALSE
Take me home.	9	13	0.716817633	중	FALSE
See you soon.	9	13	0.716817633	중	FALSE
Walk the dog.	9	13	0.716817633	중	FALSE
I love summer.	-396	14	0.682850242	중	TRUE
I love coffee.	-396	14	0.682850242	중	TRUE
I love spring.	-396	14	0.682850242	중	TRUE
It's raining.	10	13	0.716978663	상	TRUE
I missed you.	17	13	0.718105878	상	FALSE
I smile often.	12	14	0.748550725	상	TRUE
The cat slept.	-12	14	0.74468599	상	FALSE
Eat breakfast.	7	14	0.747745572	상	FALSE
Eat healthily.	7	14	0.747745572	상	FALSE
Happy Tuesday!	7	14	0.747745572	상	FALSE
Stay positive.	7	14	0.747745572	상	FALSE
Take a chance.	9	14	0.748067633	상	FALSE
Open the door.	9	14	0.748067633	상	FALSE
The sun shines.	-174	15	0.749849034	상	TRUE
That's amazing!	13	15	0.779961755	상	TRUE
The cat jumped.	-12	15	0.77593599	상	FALSE
Happy birthday!	7	15	0.778995572	상	FALSE
Hello everyone!	7	15	0.778995572	상	FALSE
Happy holidays!	7	15	0.778995572	상	FALSE
Good afternoon.	7	15	0.778995572	상	FALSE
Eat more fruit.	9	15	0.779317633	상	FALSE
I love learning.	-396	16	0.745350242	상	TRUE
I love chocolate.	-396	17	0.776600242	상	TRUE
I smiled happily.	-104	17	0.823621176	상	FALSE
Love one another.	9	17	0.841817633	상	FALSE
Enjoy the moment.	9	17	0.841817633	상	FALSE
Sun shines bright.	15	18	0.874033816	상	TRUE
Carrots taste good.	-80	19	0.88998591	상	TRUE
I listen carefully.	-32	19	0.897715378	상	TRUE
Nice weather today!	-97	19	0.88724839	상	FALSE

[표 12] 1차 전문가 검토 결과를 반영한 문장 읽기 자료(2차)

2차 개발 코드 및 텍스트에 대한 1차 타당화 결과는 [표 13]와 같다. 전문가들의 2차 개발 코드 및 결과물에 대한 평가 결과 타당성, 설명력, 유용성, 보편성, 이해도의 모든 문항에서 내용 타당도(CVI) 지수는 1로

항목	전문가						평균	SD	CVI	IRA
	A	B	C	D	E	F				
타당성	4	3	4	3	4	4	3.66	0.51	1	
설명력	4	3	4	4	4	4	3.83	0.40	1	
유용성	4	4	4	3	4	4	3.83	0.40	1	1
보편성	4	4	4	3	4	4	3.83	0.40	1	
이해도	4	4	4	3	3	4	3.66	0.51	1	

[표 13] 2차 전문가 타당화 결과

산출되었다. 또한, 채점자 간 일치도 지수(IRA)는 1로 평가자간 의견이 완전히 일치한다. 전문가들은 설명력, 유용성, 보편성 항목에 대해 평균 3.83점으로 전문가들이 가장 긍정적인 반응을 보였다. 1차 전문가 검토때와 마찬가지로 유용성, 보편성에 긍정적인 의견을 보였으며, 2차 연구 결과물은 현장 교사들이 읽기 자료를 이해하는 측면에서 조금 더 개선되었음을 알 수 있다. 또한 타당성도 1차 전문가 검토 결과(CVI: 0.83)에 비하여 수치가 높아졌다. 구체적인 전문가 의견은 개방형 응답 결과인 [표 14]에 나타나 있다.

영역	우수한 점	보완할 점	의견 반영 내용
타당성	• 단어의 사용 빈도와 관련한 선행연구 분석을 추가함으로써 구인 타당도를 높임.	• 단어의 사용 빈도와 관련한 비교적 최근 연구 추가	• 난이도 산출 기준 관련 선행 연구 보완
	• 구체적인 이론적 근거와 설계 기준을 설명하여 타당도를 높임	• 문장 난이도 산출 과정에서 단어 난이도와 문장 길이 변수의 가중치 설정에 대한 근거 추가 필요	
		• 후속 연구로 짧은 스토리가 이어지는 단락 단위의 읽기 자료 개발 제안	• 후속 연구 제안
	• 2차 결과물에서 1차에 비해 다양한 어휘와 문장이 구성되어 학생들에게 다양한 발음의 조합을 읽어볼 수 있게 함	• 개발 결과물의 난이도 산출 기준에 대한 상세한 설명 보완	• 난이도 산출 기준 관련 선행 연구 보완

	• 1차 개발 자료에 비해 타당한 최소대립쌍이 다양하게 제시됨	• 상, 중, 하 단계로 문장을 구분하는 기준의 명확화	
설명력	• 학습 부진 학생에 대한 이해와 지도 방법에 대한 이론적 설명이 설득력이 있음.	• 데이터 테이블에 있는 각 칼럼 간의 관계의 모호성. 5:5 정규화에 대한 근거 추가 필요	• 정규화 기준 재검토 및 수정
		• 음수로 표현된 difficulty 수치에 대한 설명 또는 보정 필요	• 정규화 점수 보정
	• 음소에 초점을 둔 어휘나 문장 수준 발화에 대한 필요성을 효과적으로 설명	• 과거형 동사의 유무가 단어의 난이도와 어떠한 연관성이 있는지 설명 추가	• 선행 연구 추가
유용성	• 교과서를 코퍼스로 선정하여 교사들이 학습 부진 학생 지도에서 활용하기에 적절함.	• Reading Progress에 읽기 자료 투입하는 모형에 관한 선행연구 추가	• 선행 연구 추가
	• 다양한 AI 기반 프로그램에 활용할 수 있는 읽기 자료로서 유용함.		
	• AI를 영어교육에 적용하고자 하는 교사를 위한 읽기 자료 개발 가이드로서의 가치가 있음.		
보편성	• 학교현장에서 학습용 기기만 갖춰지면 자료를 활용하여 누구나 시도해 볼 수 있는 지도 방법임.	• 추후 연구 결과물을 교사들에게 어떤 형태로 공유할지에 대한 계획 추가	• 코드 공유 방법 추가
	• 연구는 중학교 학습부진 학생을 대상으로 하였으나, 읽기 자료 개발 방법은 다른 연령대 학습자를 위해서도 적용 가능함.	• 파일럿 테스트 후속 연구 필요	• 후속 연구 제안
이해도	• 인공지능의 원리를 잘 모르는 교사들도 이해하기 쉽게 결과물이 산출됨.	• 읽기 자료를 사용할 교사를 위한 읽기 자료에 대한 설명 추가	• 최종 읽기자료에 설명 추가

	• 자연어 처리 과정을 단계별로 나누어 설명하여 AI관련 전문 지식이 부족한 교사들도 이해하기 쉬움.		

[표 14] 2차 전문가 검토 개방형 응답 의견 및 반영 사항

타당성 측면에서 전문가들은 1차 개발 자료에 비하여 2차 개발 자료의 구인 타당도가 높아졌다고 평가했다. 비교적 최근의 연구를 추가하고, 이론적 근거와 설계 기준을 상세하게 설명함으로써 전체적인 타당성을 높였다고 평가했다. 또한 2차 결과물에서 1차에 비해 다양한 어휘와 문장을 추가하여 학생들이 다양한 발음 조합을 읽어볼 수 있도록 했다는 점을 높이 평가했다. 그러나 전문가들은 여전히 문장 난이도 산출 과정에서 단어 난이도와 문장 길이 변수의 가중치 설정에 대한 근거가 부족하다고 지적했다. 후속 연구로는 단락 단위의 읽기 자료 개발을 제안했다.

설명력 측면에서는 학습 부진 학생에 대한 이해와 지도 방법에 대한 이론적 설명과 음소에 초점을 둔 어휘나 문장 수준 발화에 대한 필요성을 효과적으로 설명하였다고 평가했다. 그러나 문장 난이도 산출시 단어의 난이도와 문장 길이를 5:5 비율로 정규화한 것에 대한 근거 추가와 음수로 표현된 난이도 수치에 대한 설명 또는 보정이 필요하다는 점을 지적했다.

유용성 측면에서는 전문가들은 교과서를 코퍼스로 선정함으로써 교사들이 학습 부진 학생 지도에서 활용하기에 적절하다고 평가했다. 연구의 결과물이 추후 다양한 AI 기반 프로그램에서 활용 가능하며, AI를 영어 교육에 적용하고자 하는 교사를 위한 읽기 자료 개발 가이드로서의 가치가 있음을 인정했다. 보완해야 할 점으로 Reading Progress 프로그램에 읽기 자료를 활용하는 과정에 관한 선행연구를 추가할 것을 제안했다.

보편성 측면에서 전문가들은 학교 현장에서 학습용 기기만 갖추면 누구나 해당 읽기 자료를 활용하여 학습부진 학생 지도를 시도해볼 수 있다고 평가했다. 그러나 추후 연구 결과물을 교사들에게 어떤 형태로 공

유할지에 대한 계획을 추가할 것이 필요하다고 지적했다. 이 외에도, 구체적으로 개발 코드를 공유할 방법을 추가할 것을 제안했다.

이해도 관점에서 전문가들은 인공지능의 원리를 잘 모르는 교사들도 이해하기 쉽게 결과물이 산출되었다고 평가했다. 그러나 읽기 자료를 사용할 교사를 위한 읽기 자료에 대한 설명을 추가하라는 제언을 했다.

위와 같은 전문가 의견을 반영하여 최종 읽기 자료 개발 단계에서는 읽기 자료 개발 시 정규화 기준을 재검토하고 수정하였으며, 난이도 산출 기준에 대한 상세한 설명을 보완하였다.

제 3 절 최종 개발 읽기 자료

1. 읽기 자료의 특징

최종 개발된 읽기 자료는 개발 과정에서 다음과 같은 요소를 고려하였다. 첫째, 단어의 빈도 기준을 교과서 텍스트를 기반으로 한 코퍼스에 나타난 빈도를 기준으로 빈도가 높은 단어를 1,500개 선정하여, 이를 기반으로 최소대립쌍과 문장을 생성하였다. 2차 전문가 검토시 활용하였던 Brown Corpus는 미국영어를 기준으로 한 일상 생활 기반 코퍼스여서 제2 언어 학습자 중 학습 부진을 겪는 학습자를 위한 자료 개발을 위한 어휘의 빈도 산출에는 적절하지 않다는 판단하에 최종 결과물 산출시에는 교과서 코퍼스를 기준으로 한 빈도값만 고려하였다.

둘째, 단어의 음절 수, 단어의 음절 구조의 결합 형태를 분석할 때 CMUDict의 음소 분석을 먼저 수행하도록 코드를 수정하였다. 이를 통해 실제 발음을 기반으로 단어의 음절 수를 산출하고, 음절의 결합 형태를 분석하여 각 단어의 읽기 난이도를 수치화하였다. 이후 음소를 기반으로 최소대립쌍을 생성하여 결과물의 타당도와 정확도를 높였다.

셋째, 학습 부진 대상 학생으로 한 학습자료 개발이 목적이었으므로 문장 생성의 단어 기준을 2~5단어로 제한하여 문장 구조에서 오는 난이도를 최소화하고자 하였다. 생성된 문장 세트는 문장 길이, 문장을 구성하는 단어의 음절수, 음절 구조 난이도를 통합하여 산출한 문장 읽기 난이도 값에 따라 하, 중, 상의 난이도를 구분하였다.

2. 최소대립쌍 읽기 자료

최종 개발된 최소대립쌍은 288개로 전체 텍스트에서 산출한 빈도가 높은 1,500개 단어 중 난이도 분류에서 아주 낮은 난이도 59개, 낮은 난이도 99개의 단어를 기준으로 생성되었다. 대상 단어 중 일부는 가능한 최소대립쌍이 생성되지 않았으며, 일부 단어는 한 개의 단어에 여러 개의 최소 대립쌍이 산출된 경우도 있다. 자음, 모음을 기준으로 생성된 최소대립쌍의 음소별 구분은 [표15]와 같다. 전체 최소대립쌍은 [부록1]과 같다.

Phoneme	Pairs	Count
g	(go, know), (go, show), (get, wet), (get, let), (get, met), (get, set), (got, lot), (got, hot), (got, pot), (bag, bad), (bag, bath), (bag, back), (grew, drew), (grew, true), (glass, class)	15
d	(day, way), (day, say), (day, pay), (bad, bath), (bad, back), (bad, bag), (god, got), (fed, fell), (mud, much), (sad, sat), (sad, sang), (dad, bad), (dad, sad), (dish, wish), (dish, fish), (drew, grew), (drew, true), (spend, spent)	18
w	(one, sun), (one, run), (one, fun), (way, say), (way, pay), (way, day), (wet, let), (wet, get), (wet, met), (went, tent), (well, sell), (well, tell), (walk, talk), (wall, fall), (wall, ball), (wall, call), (wish, fish), (wish, dish)	18
b	(big, pig), (buy, high), (bad, dad), (bed, red), (bed, fed), (web, wet), (web, well), (ball, wall), (ball, fall), (ball, call), (ball, tall), (blue, flew), (best, rest), (band, land), (band, sand), (band, hand), (bath, math)	17
n	(new, two), (man, map), (man, math), (ran, rap), (win, wish), (ten, tell), (men, met), (pen, pet), (next, text), (neck, check), (know, go), (know, show)	12
l	(lot, got), (lot, hot), (lot, pot), (let, wet), (let, get), (let, met), (let, set), (long, wrong), (long, song), (last, fast), (well, wet), (well, web), (wall, walk), (land, band), (land, sand), (land, hand), (tell, ten), (fell, fed), (flea, free), (tall, talk), (sell, says), (sell, set), (hill, hit), (glass, grass)	24
t	(lot, lock), (got, god), (sit, sick), (sit, sing), (ate, age), (top, shop), (hit, hill), (met, men), (sat, sang), (ten, men), (ten, pen), (pet, pen), (wet, well), (wet, web), (set, says), (cut, cup), (two, new), (talk, walk), (start, stars), (start, smart), (told, hold), (told, cold), (tree, free), (true, drew), (true, grew), (tell, well), (tell, sell), (tell, fell), (tall, wall), (tall, fall), (tall, ball), (tall, call), (part,	37

	park), (tips, chips), (text, next), (spent, spend), (tent, went)	
h	(hot, got), (hot, lot), (hot, pot), (hit, sit), (hat, sat), (hat, cat), (hat, chat), (hand, band), (hand, land), (hand, sand), (high, buy), (hold, cold), (hold, told)	13
f	(fun, sun), (fun, one), (fun, run), (fan, ran), (fan, man), (far, car), (fed, red), (fed, bed), (fish, wish), (fish, dish), (find, kind), (fast, last), (free, tree), (fell, well), (fell, sell), (fall, wall), (fall, ball), (fall, call), (fall, tall), (flew, blue)	20
ɹ	(red, bed), (red, fed), (ran, man), (ran, fan), (run, sun), (run, one), (run, fun), (rap, map), (free, flea), (wrong, song), (wrong, long), (rest, best), (grass, glass)	13
s	(sit, hit), (sun, one), (sun, run), (sun, fun), (sat, cat), (sat, hat), (sat, chat), (say, way), (say, pay), (say, day), (sad, bad), (sad, dad), (set, wet), (set, let), (set, get), (sing, thing), (sing, king), (sell, well), (sell, tell), (sell, fell), (song, wrong), (song, long), (sand, band), (sand, land), (sand, hand)	25
k	(cat, sat), (cat, hat), (cat, chat), (car, far), (class, glass), (king, sing), (king, thing), (talk, tall), (cold, hold), (cold, told), (back, bad), (back, bath), (back, bag), (park, part), (walk, wall), (sick, sing), (sick, sit), (kind, find), (lock, lot), (call, wall), (call, fall), (call, ball), (call, tall)	23
m	(man, ran), (man, fan), (met, wet), (met, let), (met, get), (met, set), (map, rap), (men, pen), (men, ten), (math, bath), (smart, start)	11
p	(pig, big), (cup, cut), (map, man), (map, math), (rap, ran), (pet, wet), (pet, let), (pet, get), (pot, got), (pot, lot), (pot, hot), (pay, way), (pay, say), (pay, day), (pen, men), (pen, ten), (pink, think)	17
dʒ	(age, ate)	1
z	(says, sell), (says, set), (stars, start)	3
tʃ	(much, mud), (check, neck), (chips, tips), (chat, sat), (chat, cat), (chat, hat)	6
θ	(think, pink), (math, map), (math, man), (bath, bad), (bath, back), (bath, bag), (thing, sing), (thing, king)	8
ʃ	(show, know), (show, go), (shop, top), (wish, win)	4
ŋ	(sang, sat), (sang, sad), (sing, sick)	3
합계		288

[표 15] 최종 최소대립쌍 음소별 분류

3. 문장 읽기 자료

최종 개발된 문장은 중복을 제외하고 총 112문장이며, 하 수준 44개, 중 수준 58개, 상 수준 10개로 이루어져 있다. 난이도별 문장은 [표16]과 같다. 구체적인 수치와 함께 제시된 전체 읽기 자료는 [부록2]와 같다.

읽기 난이도	문장	
하	• Go now. • Go out. • Go away. • Go home. • I agree. • I smile. • Look up. • Eat cake. • Eat well. • Good day! • Have fun! • I'm fine. • I'm here. • I'm late. • I'm lost. • I'm okay. • Let's go! • Love you. • Nice day! • Pick one. • Eat lunch. • Good luck! • I'm happy.	• I'm ready. • I'm sorry. • I'm tired. • Live life. • Love wins. • Play ball. • Rainy day. • Ready, go. • Stay home. • Stay safe. • Take care! • Thank you. • Turn left. • Walk away. • I'm hungry. • It's rainy. • It's sunny. • Take risks. • Eat healthy. • Yes, please. • Take chances.
중	• Eat pizza. • I love it. • I love you • Eat dinner. • Enjoy life. • Go outside. • Happy life. • I like you. • I miss you. • Hello world! • Best wishes. • Eat cookies. • Go outdoors. • Go shopping. • Hello again. • I ate lunch. • I like cats. • I like dogs. • I live here. • I love cats. • I love food. • Stay strong. • Take action. • Take breaks. • Good morning! • Happy Monday! • Happy summer! • I like books. • I like music. • I like pizza.	• I understand. • Keep smiling. • Life goes on. • Love forever. • Nice weather. • Play outside. • Take it easy. • Welcome home. • Beautiful day. • Coffee please. • Eat breakfast. • Happy Tuesday! • Happy weekend! • I believe you. • I like coffee. • Meet me there. • Never give up. • Stay positive. • Take a chance. • Dinner's ready. • Good afternoon. • Happy Birthday! • Happy holidays. • I love cookies. • Nice day today! • Sleeping early. • You guessed it. • Happy Wednesday!

상	• Take a break. • I love summer. • Open the door. • Eat more fruit. • Happy New Year!	• I smile happily. • Have a nice day! • Take action now. • Hope you're well. • Take a deep breath.

[표 16] 최종 문장 읽기 난이도별 분류

제 5 장 논의 및 결론

제 1 절 결론 및 제언

본 연구는 학습 부진 학생들의 영어 읽기 능력 향상을 목표로 자연어 처리 기술을 기반으로 교사들이 학습 부진 학생의 읽기 유창성 향상 지도를 AI기반 프로그램을 활용하여 진행할 때 사용할 수 있는 텍스트 제작에 초점을 맞추었다. 연구를 통해 개발된 단계별 읽기 자료는 AI기술 기반 읽기 학습 프로그램인 Reading Progress를 모델로 설정하여 개발하였으며, 해당 프로그램과 유사한 기능을 하는 다양한 학습 프로그램에 활용할 수 있을 것으로 기대한다.

본 연구에서 제작한 텍스트는 현행 교과서 읽기 지문에 나타난 단어들의 출현 빈도, 단어의 음절 수, 음절 결합 형태, 음소 구성에 근거하여 개발하였다는 점에서 의의를 가진다.

개발된 결과물은 현장에서 학습 부진 학생들을 지도하는 교사들이 겪는 어려움을 완화하는 데에 도움이 될 것으로 기대한다. 특히 AI기술 기반 학습 프로그램을 활용하여 학생들의 음소 인식을 지도하거나, 문장 단위 읽기 유창성 향상을 위한 반복 학습을 진행할 때 즉시 활용할 수 있는 읽기 자료로서 기능할 것이다.

본 연구에서 개발하여 활용한 코드를 활용한다면, 교사들이 학교급이나 수업 목적에 맞춰서 다른 영어 코퍼스를 활용하여 빈도나 음절 기반의 데이터를 추출할 수 있을 것으로 예상한다.

그러나 본 연구의 방법론에는 여러 한계점이 있다. 첫째, 문장 난이도 산출을 위한 읽기 난이도 산출에서 단어의 음절 수와 음절 결합 구조만

을 고려하였다. 읽기 자료 개발 시에는 대상 단어의 발음을 분석하기 위해 CUMDct를 활용하였는데, 이는 단어가 문장에서 활용될 때 나타나는 연음, 억양 등에 대한 난이도는 고려하지 못한다는 한계가 있다. 이러한 한계는 추후 기술의 발전을 통해 문장 단위 음절 분석이 가능해진다면 조금 더 정확한 난이도 산출이 가능해질 것으로 기대한다.

둘째, 생성형 AI 모델을 활용하여 문장을 생성하는 데 있어 기술적 한계로 인해 문장의 다양성에 제한이 있었다. 교과서 텍스트 코퍼스에서 선정한 1,500개의 단어를 활용하여 문장 생성 과정에서 단어 리스트에 존재하는 단어의 비율을 0.8, 기능어의 비율을 0.2로 임의로 설정하였다. 이러한 접근방식은 현재 활용 가능한 생성형 엔진의 기술적 한계로 인한 것으로, 단어 리스트의 비율을 최대한 유지하면서도 문법적인 문장을 만들어지는 비율을 경험적으로 찾아낸 결과이다. 추후 API 엔진 성능이 향상된다면 주어진 단어의 비율을 최대한으로 유지하면서도 다양한 구조의 긴 문장 및 문단 단위의 글 생성까지 가능해질 것으로 예상한다.

위와 같은 한계점에도 불구하고, 본 연구는 자연어 처리 기법과 생성형 AI를 활용하여 주어진 코퍼스를 활용하여 빈도와 음절 난이도 기반으로 단어를 추출하여 텍스트를 생성하고자 하는 시도를 하였다. 이는 AI를 활용하여 교육적 목적에 맞는 텍스트를 개발하고자 하는 교사들에게 자연어 처리 기술을 활용한 처리 방법을 참고할 수 있는 사례로 기능할 것으로 기대한다.

후속 연구로는 연구 결과물의 효과성을 확인하는 현장 연구, 길이가 긴 문장 또는 문단 길이의 텍스트 생성, 해당 읽기 자료를 활용한 AI 도구 활용 수업모형 개발 등을 제안한다. 이를 통해 학습 과정에서 어려움을 겪는 학습 부진 학생 지도에 도움이 되는 다양한 자료가 개발될 수 있을 것으로 기대한다.

참 고 문 헌

교육부. "제1차 기초학력 보장 종합계획(2023~2027)". (2022)

교육부(2022). 보도자료. (2022.10.11.) “국가-교육청-학교가 함께 모든 학생의 기초학력을 보장합니다.”

김성혜, 조윤동, 김명화, 박태준, 조호제. 초·중학교 수학, 영어 학습 부진 학생을 위한 교육과정 개선 방안 탐색 / 김성혜 연구책임 ; 조윤동 [외]공동연구 (2019). Web.

김은정, 이병민. "영어 읽기 능력에 관한 초등학교 6학년 학생들의 잠재 프로파일 분석." 초등영어교육 27.2 (2021): 75-93. Web.

서울특별시교육청 기초학력 진단-보정 시스템. (https://s-basic.sen.go.kr/)

김태은. "4차 산업혁명시대에 요구되는 광의의 기초학력 개념 정립 연구 (CRI 2019-7).“, 2019.

김태은. 느린 학습자 선별을 위한 체크리스트 개발. 충청북도: 한국교육과정평가원, 2020.

김태은. 초,중학교 학습 부진학생의 성장 과정에 대한 연구(Ⅳ) (RRI 2020-5). 한국교육과정평가원, 2020.

서울특별시교육청. (2022). 2022 서울교육 주요업무. 서울특별시교육청 보도자료(2022.01.05.), https://enews.sen.go.kr/news/view.do?bbsSn=174751

신종호. "저성취 학습 부진아동을 위한 학습전략 프로그램에 대한 분석적 고찰." 아시아교육연구 3.1 (2002): 63-88. Web.

양원나, 이동주. "중등학교 영어 교과서 읽기 지문의 코퍼스 기반 어휘 뭉치 분석." 교원교육 36.4 (2020): 283-305. Web.

엄월영, 김혜련. "학습전략을 활용한 초등영어 학습 부진아의 수준별 읽기 지도 효과." Foreign languages education 21.4 (2014): 173-200.

이윤. "초등영어 교과서 읽기 텍스트의 어휘 난이도." Foreign languages education 23.1 (2016): 221-242.

이혜원, 임희정. "읽기 학습 전략 훈련을 통한 초등학교 영어 학습 부진아의 초기 읽기 능력 향상 연구." 영어어문교육 18.2 (2012):

171-87. Web.
임영진, 조윤경, 정영경. "초등학교 6학년과 중학교 1학년 영어 교과서 읽기 지문 연계성 분석." Foreign Languages Education 22.1 (2015): 125-48. Web.
정지현, 김혜련. "다중지능이론을 활용한 초등영어 학습 부진아의 초기 읽기 지도 효과 연구." 초등영어교육 23.2 (2017): 69-90. Web.
조형숙. "영어학습 부진아 지도를 위한 중등 예비교사 양성 사례연구: 유네스코의 '모두를 위한 교육(Education for All)'의 관점에서." 교사교육연구 57.2 (2018): 155-68. Web.
차경환, 손석완, 김동희, 전종호, 유경아. "중학교 영어 교과서 Readability 측정." 한국교육문제연구 -.11 (1996): 149-168.
최수진, 이윤. "반복 읽기에서 음소 지도가 초등학생의 영어 읽기 유창성에 미치는 효과." 영어교과교육 14.1 (2015): 135-54. Web.
홍유정. "심층인공 신경망 모형을 활용한 기초학력 미달 학생 예측연구." 국내 박사 학위 논문 서울대학교 대학원, 2022. 서울
Ayres, K. M. (1998). Effects of a phonological awareness program on English
Bygate, M. (1996). Effects of task repetition: appraising the developing language of learners. In: Willis, J., Willis, D. (Eds.), Challenge and Change in Language Teaching. Heinemann, Oxford, pp. 136 - 146.
Bygate, M. (2001). Effects of task repetition on the structure and control of language. In: Bygate, M., Skehan, P., Swain, M. (Eds.), Task-based Learning: Language Teaching, Learning, and Assessment. Longman, London, pp. 23 - 48.
Chall, Jeanne Sternlicht. Stages of reading development. New York(State): McGraw-Hill, 1983.
Dale, E., & Chall, J. (1949). The concept of readability. Elementary English, 26(1), 19-26

DeKeyser, R. (1998). Beyond focus on form: cognitive perspectives on learning and practicing second language grammar. In: Doughty, C., Williams, J. (Eds.), Focus on Form in Classroom Second Language Acquisition. Cambridge University Press, Cambridge, pp. 42 - 63.

Ellis, R. (2003). Task-based Language Learning and Teaching. Oxford University Press, Oxford.

Graves, M. F., & Graves, B. B. (2003). Scaffolding reading experiences: Designs for student success (2nd ed.). Norwood, MA: Christopher-Gordon.

Hawkes, M. (2012). Using task repetition to direct learner attention and focus on form. ELT J. 66, 327 - 336.

Introducing Reading Coach for personalized practice, and other major updates to Reading Progress (https://techcommunity.microsoft.com/t5/education-blog/introducing-reading-coach-for-personalized-practice-and-other/ba-p/3223533)

Klare, G. R. and B. Buck, 1954. Know your reader, the scientific approach to readability. New York: Hermitage House.

Klare, G. R. 1968. "The role of word frequency in readability." Elementary English, 45:12-22.

Levelt, W. (1989). Speaking: from Intention to Articulation. The MIT Press, MA.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), Word recognition in beginning literacy (pp. 89 - 120). Lawrence Erlbaum Associates.

Patanasorn, C. (2010). Effects of Procedural Content and Task Repetition on Accuracy and Fluency in an EFL Contexts. PhD dissertation. Northern Arizona University, Flagstaff.

Reading Progress in Microsoft Teams to improve student reading fluency. (https://techcommunity.microsoft.com/t5/education-blog/reading-progress-in-microsoft-teams-to-improve-student-reading/ba-p/2675732)

Spear-Swerling, Louise, and Robert J. Sternberg. "The Road Not Taken." Journal of Learning Disabilities 27.2 (1994): 91-103. Web.

Takimoto, M. (2012). Assessing the effects of identical task repetition and task-type repetition on learners' recognition and production of second language request downgraders. J. Pragmatics 9, 71 - 96.

Abstract

Development of AI-Based Reading Materials for Enhancing Reading Fluency in Underachieving Middle School English Learners through Reading-Aloud Activities

Eun Ju, Lee
Department of AI-Integrated Education
The Graduate School
Seoul National University

The aim of this study is to use AI technology to develop reading materials designed to assist struggling students in enhancing their English reading proficiency. To achieve this, a corpus was developed from English reading texts of textbooks for sixth-grade elementary school and first-grade middle school students, based on the 2015 revised curriculum of Korea. Utilizing natural language processing techniques and a generative AI engine, minimal pairs and sentence

reading materials were created for use in struggling students' reading aloud activities. The developed reading materials were formatted for immediate application in AI-based reading learning programs like Reading Progress, easily accessible in school settings.

The specific stages of reading material development include: First, the entire text under investigation was amalgamated into a single document to create a corpus. Natural language processing was then used to extract the top 1,500 words based on frequency from this unified text. These words were subsequently quantified according to real-life frequency, frequency within the curriculum, syllable count, and reading difficulty, resulting in a data frame. Second, using the quantified difficulty data, the generative AI created minimal pairs of the key words. Third, short sentences using the 1,500 key words were generated, duplicate sentences were eliminated, and the text was then classified by reading difficulty.

To ensure the validity of the developed text, the process, criteria, and results of reading material development underwent a review by six expert teachers specializing in English education, AI-integrated education, and natural language processing. The final text was developed by integrating expert feedback from two rounds of reviews.

The final minimal pairs were developed using 59 words classified as 'very low difficulty' and 99 words classified as 'low difficulty' from the 1,500 high-frequency words identified in the entire text.. A total of 288 minimal pairs were created for 20 phonemes, and after eliminating duplicates, 112 sentences were finalized, comprising 44 at a low level, 58 at an intermediate level, and 10 at a high level.

The significance of this study is rooted in the development of reading materials for struggling students, considering the frequency of word appearances in current textbook readings, the syllable count of

words, syllable combinations, and phoneme compositions. Furthermore, the produced material can be effectively utilized for guiding underperforming middle school students using AI-based learning programs. The developed code can generate similar results using other corpora, thus providing practical assistance.

However, there are limitations. The phoneme-based analysis tool used for evaluating sentence difficulty fails to consider sentence-level phenomena such as liaison or intonation. Technical constraints of the generative AI engine also posed limitations in creating a wide variety of sentences.

Despite these limitations, this study's attempts to develop learning materials using natural language processing techniques and a generative AI can serve as a useful reference for educators in creating AI-enhanced instructional materials.

Future research is suggested to conduct field studies for confirming the effectiveness of these findings, generate longer sentences or paragraph-length texts, and develop an AI-assisted teaching model using the produced reading materials. It is anticipated that such endeavors will lead to the development of diverse materials that can assist in guiding students with various difficulties using AI.

keywords : Learning difficulties, Basic Skills Instruction, Reading Fluency, AI-assisted Education, Natural Language Processing

Student Number **: 2021-22351**

[부록1] 최종 개발 읽기 자료(최소대립쌍)

Word	Minimal Pairs	Frequency	Syllable Count	Pattern	Difficulty	Difficulty_Group
go	['know', 'show']	34	1	CV	0.1	아주 낮은 난이도
us	['ice']	20	1	VC	0.1	아주 낮은 난이도
day	['way', 'say', 'pay']	42	1	CVC	0.2	아주 낮은 난이도
one	['win', 'sun', 'run', 'fun']	34	1	VCV	0.2	아주 낮은 난이도
big	['pig', 'bag']	28	1	CVC	0.2	아주 낮은 난이도
new	['know', 'two']	23	1	CVC	0.2	아주 낮은 난이도
get	['wet', 'got', 'let', 'met', 'set']	20	1	CVC	0.2	아주 낮은 난이도
lot	['got', 'let', 'lock', 'hot', 'pot']	20	1	CVC	0.2	아주 낮은 난이도
saw	['say']	15	1	CVC	0.2	아주 낮은 난이도
way	['say', 'pay', 'day']	15	1	CVC	0.2	아주 낮은 난이도
got	['get', 'lot', 'hot', 'pot', 'god']	15	1	CVC	0.2	아주 낮은 난이도
hot	['got', 'hat', 'lot', 'hit', 'pot']	14	1	CVC	0.2	아주 낮은 난이도
put	['pet', 'pot']	14	1	CVC	0.2	아주 낮은 난이도
fun	['sun', 'one', 'run', 'fan']	11	1	CVC	0.2	아주 낮은 난이도
red	['bed', 'fed']	11	1	CVC	0.2	아주 낮은 난이도
sit	['sat', 'sick', 'sing', 'set', 'hit']	10	1	CVC	0.2	아주 낮은 난이도
ate	['age']	9	1	VCV	0.2	아주 낮은 난이도
cat	['sat', 'hat', 'cut', 'chat']	9	1	CVC	0.2	아주 낮은 난이도
man	['men', 'ran', 'map', 'fan', 'math']	9	1	CVC	0.2	아주 낮은 난이도
boy	['buy']	9	1	CVC	0.2	아주 낮은 난이도
ice	['us']	8	1	VCV	0.2	아주 낮은 난이도
buy	['boy', 'high']	8	1	CVC	0.2	아주 낮은 난이도
ran	['rap', 'run', 'man', 'fan']	8	1	CVC	0.2	아주 낮은 난이도
sun	['sign', 'one', 'run', 'fun']	7	1	CVC	0.2	아주 낮은 난이도
bad	['bath', 'bed', 'back', 'bag', 'dad']	6	1	CVC	0.2	아주 낮은 난이도
run	['sun', 'ran', 'one', 'fun']	6	1	CVC	0.2	아주 낮은 난이도
car	['far']	5	1	CVC	0.2	아주 낮은 난이도
top	['shop']	5	1	CVC	0.2	아주 낮은 난이도
pig	['big']	5	1	CVC	0.2	아주 낮은 난이도
hit	['hat', 'hill', 'hot', 'sit']	5	1	CVC	0.2	아주 낮은 난이도
bed	['bad', 'red', 'fed']	4	1	CVC	0.2	아주 낮은 난이도
cup	['cut']	4	1	CVC	0.2	아주 낮은 난이도
met	['men', 'wet', 'let', 'get', 'set']	3	1	CVC	0.2	아주 낮은 난이도
let	['wet', 'get', 'met', 'lot', 'set']	3	1	CVC	0.2	아주 낮은 난이도
map	['rap', 'man', 'math']	3	1	CVC	0.2	아주 낮은 난이도
sat	['cat', 'hat', 'chat', 'set', 'sang']	3	1	CVC	0.2	아주 낮은 난이도
fan	['ran', 'man', 'fun']	3	1	CVC	0.2	아주 낮은 난이도
god	['got']	3	1	CVC	0.2	아주 낮은 난이도
far	['car']	3	1	CVC	0.2	아주 낮은 난이도
win	['wish', 'one']	3	1	CVC	0.2	아주 낮은 난이도
rap	['ran', 'map']	3	1	CVC	0.2	아주 낮은 난이도
say	['way', 'pay', 'saw', 'day']	3	1	CVC	0.2	아주 낮은 난이도
bag	['bad', 'bath', 'big', 'back']	2	1	CVC	0.2	아주 낮은 난이도
ten	['town', 'men', 'turn', 'pen',	2	1	CVC	0.2	아주 낮은 난이도

	'tell']					
pet	['wet', 'put', 'let', 'get', 'pen']	2	1	CVC	0.2	아주 낮은 난이도
wet	['well', 'let', 'get', 'web', 'met']	2	1	CVC	0.2	아주 낮은 난이도
fed	['red', 'bed', 'fell']	2	1	CVC	0.2	아주 낮은 난이도
mud	['much']	2	1	CVC	0.2	아주 낮은 난이도
age	['ate']	2	1	VCV	0.2	아주 낮은 난이도
men	['pen', 'ten', 'met', 'man']	2	1	CVC	0.2	아주 낮은 난이도
sad	['bad', 'sat', 'sang', 'dad']	2	1	CVC	0.2	아주 낮은 난이도
pot	['got', 'put', 'lot', 'pet', 'hot']	2	1	CVC	0.2	아주 낮은 난이도
hat	['sat', 'cat', 'chat', 'hot', 'hit']	2	1	CVC	0.2	아주 낮은 난이도
pay	['way', 'say', 'day']	2	1	CVC	0.2	아주 낮은 난이도
pen	['men', 'ten', 'pet']	2	1	CVC	0.2	아주 낮은 난이도
web	['wet', 'well']	2	1	CVC	0.2	아주 낮은 난이도
set	['says', 'wet', 'sat', 'let', 'get']	2	1	CVC	0.2	아주 낮은 난이도
cut	['cat', 'cup']	2	1	CVC	0.2	아주 낮은 난이도
dad	['bad', 'sad']	2	1	CVC	0.2	아주 낮은 난이도
two	['new']	21	1	CCV	0.3	낮은 난이도
went	['tent', 'want']	25	1	CVCC	0.4	낮은 난이도
class	['glass']	22	1	CCVCC	0.4	낮은 난이도
want	['went']	21	1	CVCC	0.4	낮은 난이도
next	['text']	19	1	CVCC	0.4	낮은 난이도
long	['wrong', 'song']	18	1	CVCC	0.4	낮은 난이도
king	['sing', 'thing']	18	1	CVCC	0.4	낮은 난이도
ball	['wall', 'fall', 'bowl', 'call', 'tall']	17	1	CVCC	0.4	낮은 난이도
says	['sell', 'set']	17	1	CVCC	0.4	낮은 난이도
blue	['flew']	16	1	CCVV	0.4	낮은 난이도
small	['smell']	15	1	CCVCC	0.4	낮은 난이도
much	['mud']	15	1	CVCC	0.4	낮은 난이도
talk	['walk', 'tall']	14	1	CVCC	0.4	낮은 난이도
cold	['hold', 'told']	14	1	CVCC	0.4	낮은 난이도
ship	['shop']	13	1	CCVC	0.4	낮은 난이도
work	['walk']	12	1	CVCC	0.4	낮은 난이도
back	['bad', 'bath', 'bag']	11	1	CVCC	0.4	낮은 난이도
think	['thank', 'pink']	11	1	CCVCC	0.4	낮은 난이도
show	['know', 'go']	11	1	CCVC	0.4	낮은 난이도
park	['part']	11	1	CVCC	0.4	낮은 난이도
fish	['wish', 'dish']	11	1	CVCC	0.4	낮은 난이도
neck	['check']	11	1	CVCC	0.4	낮은 난이도
find	['kind']	10	1	CVCC	0.4	낮은 난이도
know	['new', 'go', 'show']	10	1	CCVC	0.4	낮은 난이도
best	['rest']	10	1	CVCC	0.4	낮은 난이도
last	['lost', 'fast']	9	1	CVCC	0.4	낮은 난이도
start	['stars', 'smart']	9	1	CCVCC	0.4	낮은 난이도
well	['wet', 'wall', 'web', 'sell', 'tell']	9	1	CVCC	0.4	낮은 난이도
walk	['talk', 'work', 'wall']	8	1	CVCC	0.4	낮은 난이도
turn	['town', 'ten']	8	1	CVCC	0.4	낮은 난이도
fast	['last']	8	1	CVCC	0.4	낮은 난이도
free	['tree', 'flea']	7	1	CCVV	0.4	낮은 난이도
math	['bath', 'map', 'man']	7	1	CVCC	0.4	낮은 난이도

wall	['well', 'fall', 'walk', 'ball', 'call']	7	1	CVCC	0.4	낮은 난이도
told	['hold', 'cold']	6	1	CVCC	0.4	낮은 난이도
tree	['free', 'true']	6	1	CCVV	0.4	낮은 난이도
true	['drew', 'tree', 'grew']	6	1	CCVV	0.4	낮은 난이도
land	['band', 'sand', 'hand']	6	1	CVCC	0.4	낮은 난이도
hand	['band', 'land', 'sand']	6	1	CVCC	0.4	낮은 난이도
town	['turn', 'ten']	5	1	CVCC	0.4	낮은 난이도
tell	['well', 'ten', 'tall', 'sell', 'fell']	5	1	CVCC	0.4	낮은 난이도
city	['sits']	5	1	CVCC	0.4	낮은 난이도
wash	['wish']	5	1	CVCC	0.4	낮은 난이도
smart	['start']	5	1	CCVCC	0.4	낮은 난이도
high	['buy']	5	1	CVCC	0.4	낮은 난이도
dish	['wish', 'fish']	5	1	CVCC	0.4	낮은 난이도
sang	['song', 'sat', 'sing', 'sad']	5	1	CVCC	0.4	낮은 난이도
fell	['well', 'fall', 'full', 'sell', 'fed']	5	1	CVCC	0.4	낮은 난이도
grow	['grew']	5	1	CCVC	0.4	낮은 난이도
thank	['think']	5	1	CCVCC	0.4	낮은 난이도
stars	['start']	4	1	CCVCC	0.4	낮은 난이도
sick	['sing', 'sit']	4	1	CVCC	0.4	낮은 난이도
wrong	['song', 'long']	4	1	CCVCC	0.4	낮은 난이도
drew	['draw', 'grew', 'true']	4	1	CCVC	0.4	낮은 난이도
sing	['song', 'sick', 'thing', 'sang', 'king']	4	1	CVCC	0.4	낮은 난이도
flea	['free', 'flew']	4	1	CCVV	0.4	낮은 난이도
rest	['best']	4	1	CVCC	0.4	낮은 난이도
full	['fall', 'fell']	4	1	CVCC	0.4	낮은 난이도
spend	['spent']	4	1	CCVCC	0.4	낮은 난이도
kind	['find']	4	1	CVCC	0.4	낮은 난이도
tall	['talk', 'wall', 'fall', 'ball', 'call']	3	1	CVCC	0.4	낮은 난이도
fall	['wall', 'ball', 'full', 'call', 'tall']	3	1	CVCC	0.4	낮은 난이도
swam	['swim']	3	1	CCVC	0.4	낮은 난이도
sell	['says', 'well', 'tell', 'set', 'fell']	3	1	CVCC	0.4	낮은 난이도
draw	['drew']	3	1	CCVC	0.4	낮은 난이도
song	['wrong', 'long', 'sing', 'sang']	3	1	CVCC	0.4	낮은 난이도
luck	['lock']	3	1	CVCC	0.4	낮은 난이도
smell	['small']	3	1	CCVCC	0.4	낮은 난이도
sand	['band', 'land', 'hand']	3	1	CVCC	0.4	낮은 난이도
part	['park']	3	1	CVCC	0.4	낮은 난이도
tips	['chips']	3	1	CVCC	0.4	낮은 난이도
hold	['cold', 'told']	3	1	CVCC	0.4	낮은 난이도
sits	['city']	3	1	CVCC	0.4	낮은 난이도
shop	['top', 'ship']	3	1	CCVC	0.4	낮은 난이도
sign	['sun']	3	1	CVCC	0.4	낮은 난이도
grew	['drew', 'grow', 'true']	2	1	CCVC	0.4	낮은 난이도
flew	['blue', 'flea']	2	1	CCVC	0.4	낮은 난이도
band	['land', 'sand', 'hand']	2	1	CVCC	0.4	낮은 난이도
check	['neck']	2	1	CCVCC	0.4	낮은 난이도

lost	['last']	2	1	CVCC	0.4	낮은 난이도
bowl	['ball']	2	1	CVCC	0.4	낮은 난이도
grass	['glass']	2	1	CCVCC	0.4	낮은 난이도
bath	['bad', 'back', 'bag', 'math']	2	1	CVCC	0.4	낮은 난이도
hill	['hit']	2	1	CVCC	0.4	낮은 난이도
chips	['tips']	2	1	CCVCC	0.4	낮은 난이도
text	['next']	2	1	CVCC	0.4	낮은 난이도
thing	['sing', 'king']	2	1	CCVCC	0.4	낮은 난이도
glass	['grass', 'class']	2	1	CCVCC	0.4	낮은 난이도
spent	['spend']	2	1	CCVCC	0.4	낮은 난이도
lock	['luck', 'lot']	2	1	CVCC	0.4	낮은 난이도
swim	['swam']	2	1	CCVC	0.4	낮은 난이도
call	['wall', 'fall', 'ball', 'tall']	2	1	CVCC	0.4	낮은 난이도
wish	['win', 'wash', 'fish', 'dish']	2	1	CVCC	0.4	낮은 난이도
pink	['think']	1	1	CVCC	0.4	낮은 난이도
chat	['sat', 'cat', 'hat']	1	1	CCVC	0.4	낮은 난이도
tent	['went']	1	1	CVCC	0.4	낮은 난이도

[부록2] 최종 개발 읽기 자료(문장)

Sentence	Length	Syllables	Word Difficulty	Sentence Difficulty	Difficulty Category
Go now.	2	2	2.3	1.4	하
Go out.	2	2	1.7	1	하
Go away.	2	2	1.4	1.6	하
Go home.	2	2	2.1	0.9	하
I agree.	2	2	3	1.7	하
I smile.	2	2	2.4	1.5	하
Look up.	2	2	3	0.8	하
Eat cake.	2	2	3.1	1.2	하
Eat well.	2	2	4.3	1.7	하
Good day!	2	2	4.5	1.5	하
Have fun!	2	2	2.5	1.9	하
I'm fine.	2	2	2.4	0.8	하
I'm here.	2	2	2.4	0.8	하
I'm late.	2	2	2.4	0.8	하
I'm lost.	2	2	4.3	1.3	하
I'm okay.	2	2	3.1	1.5	하
Let's go!	2	2	2.9	1.7	하
Love you.	2	2	2.4	1	하
Nice day!	2	2	2.5	1.9	하
Pick one.	2	2	3.1	1.9	하
Eat lunch.	2	2	4.3	1.7	하
Good luck!	2	2	5	1.5	하
I'm happy.	2	3	4.3	1.3	하
I'm ready.	2	2	5.2	1.4	하
I'm sorry.	2	3	4.3	1.3	하
I'm tired.	2	2	3.1	1.5	하
Live life.	2	2	2.1	1.4	하
Love wins.	2	2	2.5	1.9	하
Play ball.	2	2	4.3	2.4	하
Rainy day.	2	2	4.5	2.5	하
Ready, go.	2	2	3.2	1.8	하
Stay home.	2	2	3.5	1.9	하
Stay safe.	2	2	3.5	1.9	하
Take care!	2	2	2.1	1.4	하
Thank you.	2	2	4.5	2.6	하
Turn left.	2	2	4.3	2.4	하
Walk away.	2	2	3.1	2.6	하
I'm hungry.	2	3	4.3	1.3	하
I'm hungry.	2	3	4.3	1.3	하
It's rainy.	2	2	5.2	1.4	하
It's sunny.	2	3	4.3	1.3	하
Take risks.	2	2	2.5	1.9	하
Eat healthy.	2	2	5.8	1.8	하
Yes, please.	2	2	3.9	2.9	하
Take chances.	2	2	4	3	하
Eat pizza.	2	3	3.8	2.4	중
I love it.	3	3	3.1	1.3	중

I love you	3	3	3.4	1	중
Eat dinner.	2	3	3.8	2.8	중
Enjoy life.	2	3	4.5	2.9	중
Go outside.	2	3	2.4	2.2	중
Happy life.	2	3	3.1	1.9	중
I like you.	3	3	3.8	1.1	중
I miss you.	3	3	6	1.6	중
Hello world!	2	3	2.3	3.5	중
Best wishes.	2	3	3.8	3.5	중
Eat cookies.	2	3	4.6	2	중
Go outdoors.	2	3	3.1	2.9	중
Go shopping.	2	3	3.8	2.5	중
Hello again.	2	3	2.8	3	중
I ate lunch.	3	3	2.5	1.8	중
I like cats.	3	3	3.5	2	중
I like dogs.	3	3	3.5	2	중
I live here.	3	3	3.1	1.5	중
I love cats.	3	3	3.5	2	중
I love food.	3	3	4.9	2.1	중
Stay strong.	2	2	4.3	5.5	중
Take action.	2	3	3.5	3.4	중
Take breaks.	2	2	6	3.4	중
Good morning!	2	3	4.5	2.6	중
Happy Monday!	2	4	3.8	3.5	중
Happy summer!	2	4	3.8	3.5	중
I like books.	3	3	4.9	2.1	중
I like music.	3	4	3.1	2.2	중
I like pizza.	3	4	3.5	2.7	중
I understand.	2	4	4.3	4.4	중
Keep smiling.	2	3	3.8	2.4	중
Life goes on.	3	3	4.3	1.5	중
Love forever.	2	4	2.1	2.1	중
Nice weather.	2	3	4.3	3.7	중
Play outside.	2	3	4.3	3.2	중
Play outside.	2	3	4.3	3.2	중
Take it easy.	3	3	3.5	1.7	중
Welcome home.	2	3	2.1	2.8	중
Beautiful day.	2	4	4	1.9	중
Coffee please.	2	3	5	3.5	중
Eat breakfast.	2	3	6.1	4.6	중
Happy Tuesday!	2	4	6.1	4.2	중
Happy weekend!	2	4	3.9	2.6	중
I believe you.	3	4	3.8	1.1	중
I like coffee.	3	4	4.2	3	중
Meet me there.	3	3	5.2	1.9	중
Never give up.	3	4	3.1	1.9	중
Stay positive.	2	4	3.5	1.9	중
Take a chance.	3	3	4	3	중
Dinner's ready.	2	3	6.6	3.6	중
Good afternoon.	2	4	4.1	4.4	중

Happy Birthday!	2	4	4.3	5	중
Happy birthday.	2	4	4.3	5	중
Happy holidays.	2	5	3.1	2.6	중
I love cookies.	3	4	4.5	2.3	중
Nice day today!	3	4	3.5	2.6	중
Sleeping early.	2	4	6	3.2	중
You guessed it.	3	3	4.1	2.9	중
Happy Wednesday!	2	4	3.8	3.5	중
Sunshine smiles.	2	3	3.1	5.2	중
Walking outside.	2	4	6.3	4.3	중
Take a break.	3	3	5.5	3.4	상
I love summer.	3	4	3.5	3.1	상
Open the door.	3	3	5.6	3	상
Eat more fruit.	3	3	6.4	3.9	상
Happy New Year!	3	4	7.6	3	상
Have a nice day!	4	4	2.5	2.6	상
I smile happily.	3	5	3.5	3.8	상
Take action now.	3	4	5.8	3.5	상
Hope you're well.	3	3	4.3	3.5	상
Take a deep breath.	4	4	6	3.7	상

[부록3] 1차 전문가 검토지

중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 텍스트 개발에 대한 전문가 질문지

□ 안내문

본 질문지는 중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 텍스트 개발 과정 및 결과물에 대한 타당도를 검증하기 위해 제작되었습니다.

본 질문지는 연구의 개요, 텍스트 도출과정, 텍스트, 활용 방법에 대한 의견을 묻는 질문지로 구성되어 있습니다. 질문지를 읽어보시고 타당도에 따라 해당하는 점수를 표시하여 주시기 바랍니다.

본 질문지를 작성하는데 약 20분 정도 소요될 것으로 예상되며, 응답 시 궁금한 사항은 연구자에게 질문하여 주시기 바랍니다.

본 조사에서 작성하신 기본 정보 및 응답 내용은 연구 목적 이외에는 절대 사용되지 않으며, 비밀을 보장합니다. 논문에서는 전문가임을 증명하기 위해 전공분야, 최종학력, 소속, 경력 부분만 언급됨을 사전에 알려드립니다.

바쁘신 와중에 본 연구에 협조하여 주셔서 진심으로 감사합니다.

2023년 6월

서울대학교 사범대학 AI융합교육학과
지도교수: 김선희 교수
연구자: 석사과정 이은주
이메일: lej3687@snu.ac.k

□ 전문가 프로필

성함	직업	경력(년)	최종 학력	전공 분야
(기재 예시) *홍길동*	*중학교 교사*	*15*	*석사*	*교육 공학/국어 교육/AI 융합 교육*

연구개요

1. 연구의 배경

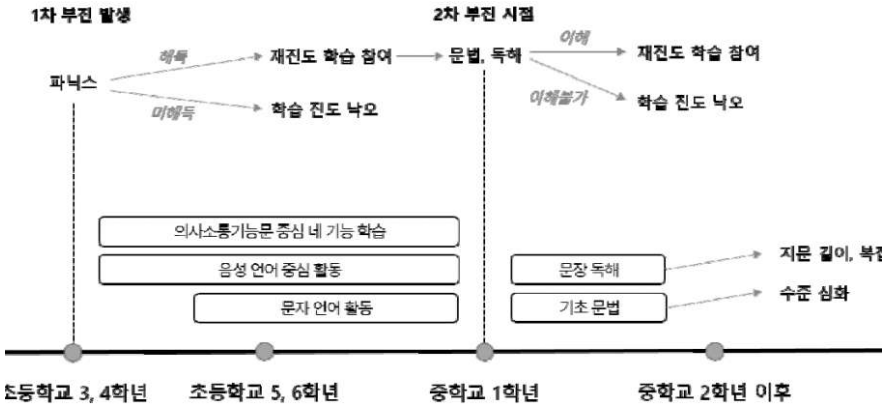
해가 거듭될수록 학생 간 학습격차는 심화되어 기초학력이 부족한 학생들을 위한 정책의 필요성이 더욱 대두되고 있다. 교육부는 제1차 기초학력 보장 종합계획(2023~2027)을 발표하여 코로나로 인한 비대면 수업기간 동안 심화된 학생들의 학습 결손과 정서 안정을 지원하는 국가 단위의 기초학력 보장 체계를 마련하고자 했다. [그림1]은 2017년과 2021년도의 국가수준 학업성취도 평가에서 기초학력 미달 학생들의 비율을 나타내는데, 코로나 이전에 비하여 기초학력 학생의 비율이 모든 과목에서 급격히 증가했음을 알 수 있다.

[그림 6] 2017년 · 2021년 국가수준 학업성취도 평가 1수준(기초학력 미달) 비율 추이(교육부, 2022)

교육부는 이번 종합계획에서 인공지능(AI)기반 진단체계 구축을 강화하여 에듀테크 기반 학생 개별화 학습 지원을 강조했다. 이를 위해 국가기초학력지원센터를 운영하며, 2022 개정 교육과정에 따른'최소한의 성취기준'을 개발하고, 국가수준의 진단 도구 및 교수·학습 자료를 보급하고자 한다(교육부, 2022).

2. 연구의 필요성

현재 서울특별시교육청은 기초학력 진단-보정 시스템(s-basic)홈페이지를 통해 학년 초 학생들의 교과학습능력(국어, 영어, 수학 중심) 진단검사를 진행한다. 영어 진단평가의 경우 4기능(듣기, 말하기, 읽기, 쓰기)를 모두 포함하고 있으나, 말하기 영역은 학생의 발화를 직접 녹음하여 분석하는 형태가 아닌, 맥락에 맞는 적절한 응답 고르기의 형태로 제공한다.



[그림 7] 영어 학습 부진 심화 패턴(김태은, 2020)

김태은(2020)은 기초학력 진단-보정 시스템이 유일한 기초 영어 능력 검사도구라는 점을 강조하며, [그림2]에서와 같이 초등학교에서 발생하는 1차 부진(파닉스) 영역에 대한 진단 도구의 부족을 지적했다. 파닉스 능력 진단과 선제적 지원을 제공 한다면 학습 부진이 발생하는 초기에 교사가 개입하여 학습 부진을 예방할 수 있다는 것이다.

중학교 영어 학습 부진 학생들을 지도하기 위해 서울특별시교육청이 교사들에게 제공하는 자료는 각 학년별 의사소통영역 중심으로 이루어진 영어 '늘품이'교재와 음소별로 학습이 가능한 '15분 파닉스'교재이다. '15분 파닉스'교재는 온라인 단어 학습 사이트인 클래스카드에 단어 세트를 업로드하여 교사들이 사용할 수 있도록 하고 있다. 하지만 중학교 기초학력 미달 학생들의 음성 발화를 분석하고 즉시 피드백을 줄 수 있는 AI기반 프로그램과 연계하여 활용할 텍스트는 별도로 개발되어 있지 않다.

3. 연구의 목적

본 연구는 중학교 영어 학습 부진 학생들을 대상으로 한 읽기 유창성 향상을 위한 소리내어 읽기(reading-aloud)활동에서 활용할 텍스트 세트를 제작하는 것을 목적으로 한다. 2015 개정 교육과정의 초등학교 6학년교과서 5종, 중학교 1학년 교과서 9종의 교과서 읽기 본문 텍스트를 통합하고, 자연어 처리 기술을 활용하여 통합된 전체 텍스트의 주요 어휘 1000개를 빈도 기반으로 추출히여 주요 어휘의 음절 수와 음절 난이도 값을 수치화하고, 생성형AI를 활용하여 해당 단어를 활용한 읽기 자료를 개발하는 것을 목적으로 한다. 가장 낮은 단계

의 읽기 자료로는 가장 많은 빈도수를 가진 단어들의 최소 대립쌍(minimal pair)를, 중간 단계로는 빈도가 높고 음절 난이도가 낮은 단어들로 이루어진 5단어 내의 문장을, 가장 높은 난이도로는 음절 난이도가 높은 단어들로 이루어진 10단어 내의 문장을 생성한다. 개발된 텍스트는 1차 전문가 타당화를 거친 뒤 수정하고, 2차 전문가 타당화를 거쳐 영어 학습 부진 학생들의 읽기 유창성 개발을 위한 AI기반 프로그램 활용 소리내어 읽기(reading-aloud) 수업 과정에서의 활용 타당도를 알아보고자 한다.

4. 이론적 배경 (요약)

1. 영어 학습 부진 학생

1.1 학습 부진 학생의 정의

기초학력 부진 학생들은 학습 속도가 느리고 어휘력, 언어이해, 언어 표현이 부족하여 낮은 자존감과 낮은 학습 동기, 불안, 위축 등의 특성을 보인다(김태은 외, 2020). 본 연구에서는 학습 부진 학생을 기초학력미달 또는 기초학습 부진학생과 구분하여 영어 교과 교육과정의 성취기준에 도달하지 못한 학습 부진 학생을 대상으로 한다. 김성혜(2019)의 연구에 따르면 중학교 영어 성취수준이 '하'수준인 학생의 경우 20~50% 미만의 학생들이 수업 내용을 잘 이해하지 못하며, 교육과정 성취기준을 미충족한다는 것을 보여준다.

1.2 영어 학습 부진 학생의 읽기 지도 필요성

초등학교의 영어 수업 시간은 읽기 학습에 충분하지 않아 일부 학생들은 단어 해독 능력이 있음에도 문장이나 문단 이해에 어려움을 겪는다. 따라서, 읽기 발달 초기 단계에 있는 학생들의 읽기 유창성을 향상시키기 위한 노력이 필요하다(최수진, 이윤, 2015). 중학생 중에서 영어 학습에 어려움을 겪는 학생들의 영어 읽기 능력 부족은 기본적인 음철법의 적용에서부터 동료들의 영어 독해 이해도 및 독해 전략의 활용에 이르기까지 그 범위가 광범위하며, 이는 초등학생들에 비해 더 큰 편차를 보인다. 이러한 중학생들의 영어 읽기 능력 부족 문제는 반드시 해결해야 하는 중요한 과제이다. 영어 학습의 어려움은 읽기 능력과 깊은 연관이 있으므로, 이러한 학습 어려움을 해결하기 위한 교육과정 개선은 문자의 노출 양과 시기, 방법뿐만 아니라 문자언어 관련 성취기준의 연계성을 강화하는 것이 필요하다.(한국교육과정평가원, 2019)

1.3. 영어 학습 부진 학생 지도의 어려움

중학교 1학년부터 영어 교과는 점차 음성언어에서 문자 언어 중심으로 전환되며, 학습 어휘의 양이 급증하고 언어 형식이 강화되는 시기이며, 영어 교사들은 학습 부진 학생들이 '어휘'와 '읽기' 영역을 우선적으로 학습해야 한다는 입장을 보였다.(김성혜, 2019) 해당 연구에 참여한 교사들은 문자 언어와 관련된 교육과정의 학습 수준이나 학습량 적정화보다는, 기존의 어휘 및 언어형식을 제시하는 방법의 위계성이 부족하다는 점과 주제별로 구성된 교과서에 나타난 어휘 체계의 설계 기준이 모호함을 지적하며, 교사들의 기본학력 학생 지도의 어려움을 호소했다. 김성혜(2019)는 2015 개정 교육과정 영어과의 성취기준, 어휘, 언어형식이 학교급별로 구분되어 있어 학습 부진 학생 지도에 활용도가 떨어짐을 지적하며 일상생활에서 사용 빈도가 높은 낱말과 의사소통 기능 표현을 순차적으로 익힐 수 있도록 학습 자료가 세분화되어야 한다는 것을 강조했다. 또한, 필수 성취기준과 필수 어휘, 의사소통기능, 언어

형식을 기초군으로 분류하여 교육과정에 제시해야 함을 강조했고, 초기 읽기 능력 습득이 학습 부진의 경계가 되는 지점이므로 고빈도어를 토대로 한 기초 어휘목록이 교육과정에서 제시되어야 한다고 주장했다.

1.4 3. 음운론적 인식 및 읽기 유창성의 학습 필요성

엄월영, 김혜련(2014)은 영어 학습 부진아들이 반복 학습 및 효과적인 학습 방법 등을 통해 보다 높은 학습 수준에 도달할 수 있다고 강조하였다. 김은정과 이병민(2021)은 초등학교 6학년 학생들의 영어 읽기 능력을 세분화하여 분석하였는데, 영어는 문자와 파닉스가 일정하지 않음에도 음운법 규칙을 통해 어휘의 84%를 규칙성을 적용해 읽을 수 있고, 알지 못하는 어려운 단어를 유추하는 데에 유용하기 때문에 학생들에게 맞춤형 읽기 보충학습은 필수적이라고 주장했다. 이를 위해서는 단어와 문장 유창성과 관련된 다양한 영어 독서 자료와 함께 해독 능력을 향상하기 위한 교육이 제공되어야 하며, 학생들 간의 개인차는 정규 영어 수업 또는 방과후 교육 등을 통해 공교육 내에서 처치가 이루어져야 한다고 강조했다.

2. 이독성과 텍스트 난이도

1.1 이독성의 정의

이독성(Readability)는 주어진 읽기 자료 내의 모든 요소들에 대해 글을 읽는 사람의 이해도, 최적의 속도로 읽는 능력, 그리고 그 읽기 자료에 대한 흥미 정도에 대한 종합적인 개념이다(Dale & Chall, 1949). Graves와 Graves(2003)는 이독성에 영향을 미치는 두 가지 요인으로 '읽기자료에 내재된 요인들(factors largely inherent in the text itself)'과 '독자 및 읽기자료와 관계있는 요인들(factors involving the reader and the text)을 제시했다. 내재된 요인으로는 어휘, 문장 구조, 문장의 길이, 풍부한 설명(elaboration), 일관성(coherence), 텍스트 구조가 있으며, 외재적 요인으로는 내용 친숙도, 배경지식, 독자의 인지적 수준, 글의 특징, 독자의 흥미가 있다. 최근의 연구들이 제시하는 이독성 공식들은 단어의 음절이 길수록, 단어의 철자 수가 많을수록, 저빈도(low frequency)의 단어가 많을수록, 한 문장 내의 단어가 많을수록 그 읽기 자료는 이해하는 것이 어렵다고 가정한다(임영진, 조윤경, and 정영경, 2015). 이처럼 이독성에는 다양한 원인이 영향을 미치지만 본 연구에서는 초기 읽기 학습자를 위한 텍스트 개발이라는 목적과 관련하여 이윤(2016)이 제시한 요소를 기준으로 삼았다. 이윤(2016)은 읽기 교재가 어떤 어휘로 구성되는가에 따라 텍스트의 난이도가 달라지므로 학생들의 읽기 능력에 적합한 텍스트를 제공하는 것이 읽기 학습에서 중요한 역할을 한다고 했다. 그러나 음운 규칙 기반 어휘로만 텍스트를 구성하는 것은 어려운 일이며, 영어의 특성상 초기에 습득해야 하는 어휘들 중 다수가 음운 규칙이 적용되지 않는 의미 중심 어휘이므로, 읽기 학습 초기에 이 특성을 고려하여 난이도를 조정하는 것이 필요하다고 강조했다.

1.2 영어 교과서 이독성 수준

영어 교과서의 이독성 수준이 학습자의 이해 능력을 초과하는 경우 지식 전달 과정에서 공백이 생기게 되어 학습 의욕 저하를 유발하므로 학습자의 이해력에 맞는 교재를 활용해야 한다(차경환 and 손석완, 2006). 특히 중학교로 올라가면서 학습 부진의 편차가 커지는 것은 읽기 영역의 학습 부진에서 기인한 것으로 문자와 관련된 영역의 교과 내용의 적정성을 제고할 필요가 있다. 우리나라 2015 개정 교육과정의 초등학교 교과서 읽기 지문의 가독성 지

수를 분석한 결과, 5학년과 6학년 사이의 가독성은 미국 학제 기준으로 한 학기 정도의 차이가 났지만(황은경, 박광현, 2019), 초등학교 6학년과 중학교 1학년 읽기 지문 사이에는 미국 학제 기준으로 한 학년 이상의 가독성 수치 차이가 나타난다(배지영, 2019). 임영진, 조윤경, and 정영경(2015)은 초등학교 6학년과 중학교 영어 교과서의 이독성과 어휘 난이도를 분석였는데, 이독성 지수는 학교급간에 1점 이상 증가하였으며, 중학교 교과서의 어휘 수는 초등학교 교과서 대비 세 배 이상 증가하였다. 무작위로 증감하는 이독성 수치는 학습자에게 어려움을 줄 수 있으므로, 학습 자료는 저난이도에서 고난이도, 적은 학습량에서 많은 학습량으로 조직화되어야 한다는 결론을 도출하였다. 이윤(2016)은 초등영어 4종의 교과서를 분석해 본 결과, 공통적으로 어휘 반복율, 어휘 난이도 지수 측면에서 모두 읽기 발달 초기 학습자의 특성을 고려하지 못하는 것을 발견했다. 따라서, 중학교 영어 학습 부진 학생은 교과서를 활용한 읽기 활동에 어려움을 겪을 수 있으며, 초등학교 교과서 텍스트를 활용한 읽기 학습을 하더라도, 여전히 학습부진 학생을 읽기 능력 발달을 위한 학습에 적절한 텍스트라고 보기 어렵다.

3. ESL학습에서의 음절 연계 단어 학습 중요성

여러 연구들에서 음절 연계 단어 학습의 효과성을 연구하였다. Metsala와 Walley(1998)은 "lexical restructuring hypothesis"를 제시하여 어휘 성장과 음운인식, 독서 발달 간의 밀접한 관계를 주장하였다. 이 연구에서는 어린이들의 어휘가 확장되고 음운적으로 유사해짐에 따라 음절적 표현이 발달되며, 이러한 음절적 표현이 음운인식의 발달에 기여하고 독서와 쓰기 학습에 영향을 미친다고 언급했다. 이와 함께, Ayres(1998)는 음운인식 프로그램이 영어를 학습하는 어린이들에게 어휘 학습을 촉진시킨다는 것을 밝혔다. 이 연구에서는 음운인식 프로그램과 어휘 학습 활동을 통합하여 구성된 교육 프로그램이 어린이들의 단어 분절(segmentation) 능력에 긍정적인 영향을 미친다고 결과를 제시했다. 이러한 선행 연구들은 ESL 학생의 음절 연계 단어 학습의 중요성을 강조하며, 구두 언어 능력과 어휘 수준을 고려하여 교육 프로그램을 설계해야 한다는 것을 나타낸다.

4. 영어 학습자의 반복 읽기 효과성

여러 학자들은 연구를 통해 언어학습자들이 언어학습을 반복(task-repetition)할 때 언어 사용의 복잡성(complexity), 정확성(accuracy), 그리고 유창성(fluency) 등이 제2언어(L2) 학습 및 발달에 어떻게 영향을 미치는지를 탐구해왔다(Hawkes, 2012; Takimoto, 2012). Bygate (1996, 2001)는 Levelt의(1989) 연설 생산 모델을 이용해 작업 반복의 잠재적인 심리언어학적 이점을 설명하였다. 이 모델에 따르면 언어 사용은 개념화(conceptualization), 구현(formulation), 및 발성(articulation)의 세 단계를 거친다. 작업을 반복함으로써 학습자들은 이전의 처리 단계(즉, 개념화, 구현 및 발성)를 기억 저장소에 보관하고 다음 번에 재사용할 수 있다. 최수진, 이윤(2015)는 초등학생들의 영어 읽기 유창성 향상을 위해 반복적인 읽기 학습에 음소 지도를 도입하여 음소 인식을 포함한 반복 읽기 방법이 단어 인식의 정확성, 운율, 이해력 증진 등에서 더 큰 향상을 보였음을 확인하였고, 반복적인 읽기가 읽기 유창성과 독해력 향상에 효과적임을 발견했다.

5. 맞춤형 수업을 위한 AI프로그램 활용

5.1 학습부진 학생 지도 과정에서 AI프로그램 활용의 효과

교과 학습 부진 학생을 지도하기 위해서는 각 학생에 맞는 개별 보충 지도를 제공해야 하므로 다인수 학급에서는 적절한 수업이 이루어지기 어렵다. 대상 학생 간에도 학습 부진의 정도에 있어 차이를 보이기 때문에 학생들의 영어 수준에 따라 다른 종류의 보충학습이 제공되어야 한다(엄월영, 김혜련, 2014). 김태은(2020)의 연구에 따르면 학습 부진 학생들의 성장을 위해서는 개별 학생의 학습 과정과 어려움의 원인을 '들여다보기', 학습 단계와 자료를 세밀하게 '세분화하기', 반복적이고 꾸준하게 반복하는 '기초 근육 만들어주기', 꾸준히 성장 과정을 지켜보며 감정적인 지원을 제공하는 '견뎌주기' 과정이 필요하다. 김태은(2020)은 효과적인 학습 부진 지도를 위해서는 교사의 지도계획과 실행이 현실적으로 적정화 되어야 하며, 그 과정에서 AI기반 프로그램을 활용하는 것이 필요하다고 강조했다. 이는 개별지도 효과를 높이면서도 교사의 부담을 줄이고, 실수를 두려워하는 학생들이 정의적 여과(affective filter)가 낮은 상태에서 학습에 임할 수 있도록 하기 때문이다.

5.2 텍스트 활용을 위한 AI기반 도구

본 연구에서는 Microsoft의 교육 플랫폼인 Teams의 과제 기능인 Reading Progress 프로그램을 활용하여 개발된 텍스트를 과제로 부여하는 형태로 단계별 소리 내어 읽기(reading-aloud) 활동을 진행한다. Reading Progress 프로그램은 교사가 업로드한 텍스트를 학생이 소리 내어 읽으면, 학생의 수행 결과를 자동으로 분석한 결과를 [그림3]과 같이 보여준다.

[그림 8] Reading Coach 기능

분석 결과에는 텍스트를 읽는 학생의 영상이 좌측에 나타나고, 학생의 발화가 녹음되어 저장된다. 중앙에 있는 결과 리포트 상단에는 1분당 읽은 올바른 단어 수(correct word per minute)와 정확도 비율(accuracy rate)이 숫자로 표기되어 나타난다. 학생의 발화 오류는 자동으로 다섯 가지 영역(mispronunciation, omission, self-correction, insertion, repetition)으로

분석되어 텍스트에 각기 다른 색깔로 표기된다.

[그림 9] Insights기능을 활용한 학생 수행 결과 자동 분석(Microsoft, 2022)

교사는 [그림4]와 같이 Insights기능을 활용하여 학생의 과제별 수행 결과 그래프를 볼 수 있다. 그래프 하단에는 학생이 부정확하게 읽은 단어가 자동으로 분석되어 워드 클라우드의 형태로 제공된다. 기간을 설정하여 그룹 평균 수행 대비 해당 학생의 수행 결과를 비교하여 볼 수도 있다.

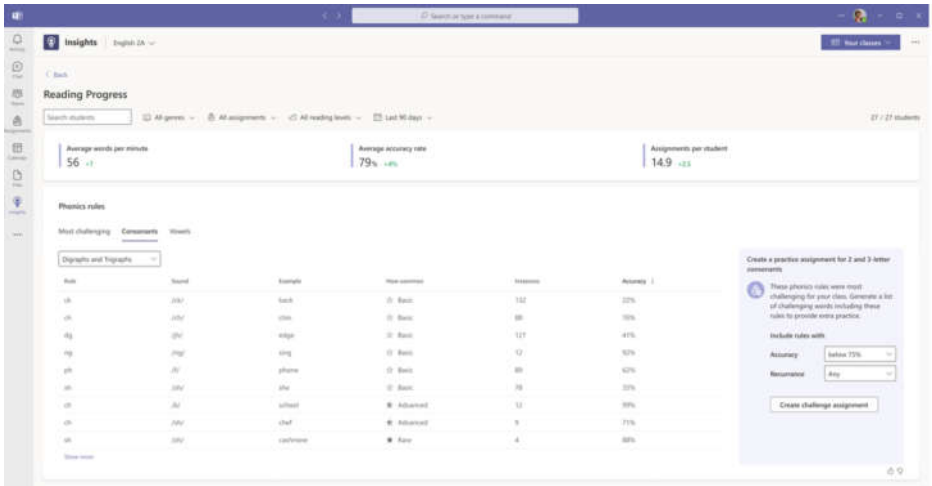
과제 설정 시 교사는 읽기 후에 리딩 코치(reading coach) 기능을 추가할 수 있다. 리딩 코치 기능은 [그림5]와 같이 학생이 글을 읽을 때 부정확하게 발음한 단어를 자동으로 분석하여 제시하며, 올바른 발음을 음절별로 다시 들을 수 있도록 음원을 제공한다. 학생은 리딩 코치가 제공하는 발음을 듣고 다시 단어를 올바르게 읽도록 연습하며, 일정 수준 이상의 정확도에 도달하면 다음 단어로 넘어갈 수 있다.

모든 과제가 끝나면 학생이 정확히 읽지 못한 모든 단어가 자동으로 분석되며, 교사는 개별 학생의 이름을 클릭하여 해당 학생이 어떤 음소를 부정확하게 발음했는지를 [그림6]의 화면을 통해 파악할 수 있다. 오류 비율이 높은 음소를 교사가 선택하여 추가 도전 과제 만들기 버튼을 클릭하면 해당 학생에게만 개별화된 단어 읽기 과제가 부여된다.

5.3. 읽기 텍스트 구성 기준

이윤(2016)은 교과서 텍스트는 영어 학습자의 학습 내용이자 도구이므로, 초기 읽기 단계에서 학생들이 알파벳 원리를 익혀서 단어 인지 과정을 학습하도록 교재를 체계적으로 구성해야 함을 강조했다. 현재 교과서 텍스트의 경우 내용만 고려하여 어휘 선정을 하는 경우, 등장 인물이나 장소의 이름이 발음하기 어렵거나 길이가 긴데, 이것은 읽기 난이도를 높이는 주 원인임을 지적했다. 따라서 낮은 난이도의 텍스트에서는 소리와 철자의 대응이 명확한 어휘와 고빈도어 중심으로 텍스트를 구성하거나 특정 모음 패턴을 반복적으로 제시하는 등의

방법을 통해 읽기 텍스트가 학생들의 읽기 학습을 돕도록 특별하게 체계화되어야 한다고 강



[그림 10] 학생의 오류 음소 패턴 자동 분석 및 개별화된 추가 과제 자동 생성

조했다.

6. 연구 대상

텍스트 개발을 위한 연구 대상은 2015 개정 교육과정 초등 6학년 5종, 중학교 1학년 9종 교과서의 읽기(reading) 본문을 대상으로 한다.

학교급	출판사	저자
초등	동아	박기화 외
	대교	이재근 외
	천재	함순애 외
	YBM(김)	김혜리 외
	YBM((최)	최희경 외
중등	능률(김)	김성곤 외
	동아(윤)	윤정미 외
	동아(이)	이병민 외
	미래엔(최)	최연희 외
	비상(김)	김진완 외
	지학사	민찬규 외
	천재(이)	이재영 외
	천재(정)	정사열 외
	YBM(송)	송미정 외

7. 연구 절차

1) 코퍼스 수집

먼저, 텍스트 개발 절차에서 주요한 첫 단계는 코퍼스를 수집하는 것이다. 본 연구에서는 2015 개정 교육과정의 초등학교 6학년 교과서 5종과 중학교1학년 교과서 9종의 읽기 지문 텍스트를 추출하여 하나의 통합된 파일로 작성했다. 이렇게 수집된 코퍼스는 본 연구의 주요 데이터 소스이며, 이는 읽기 지문 텍스트를 분석하고 토큰화하는 데 사용되었다.

2) 텍스트 전처리 및 고빈도 단어 추출

다음 단계는 텍스트 전처리와 고빈도 단어 추출이다. 이 단계에서는 Google Colab을 활용하여 통합된 텍스트에 나타난 단어를 토큰화하고, 문장부호, 숫자 등의 불용어를 처리했다. 토큰화란 텍스트를 개별 단어나 표현으로 분리하는 과정을 말하며, 불용어는 분석에서 제외하는 덜 중요한 단어들을 말한다. 불용어 처리 후, 빈도수가 가장 높은 단어를 추출하기 위해 역문서빈도(TF-IDF)값이 높은 상위 단어 1000개를 추출했다. TF-IDF는 단어의 중요성을 판단하는 수치로, 해당 단어가 문서 내에서 얼마나 자주 등장하는지와 동시에 전체 문서에서 얼마나 희귀하게 등장하는지를 모두 고려한 값이다.

3) 단어의 읽기 난이도 분석

단어의 읽기 난이도 매핑 과정에서는 음운론적 복잡성(phonological complexity)을 중심으로 하여 낮은 난이도부터 어려운 난이도를 구분하였다. 이러한 난이도 기준을 추가한 이유는 문장 생성 과정에서 어휘의 빈도와 함께 및 음운론적 요소를 고려하기 위함이다.

우선 앞서 분석한 텍스트에서 가장 높은 빈도 순으로 단어 1000개를 추출한 후, 이 단어들의 품사를 태깅하여 각 단어의 품사 정보를 단어와 짝짓는 과정을 거쳤다. 품사 태깅은 단어가 문장에서 어떤 역할을 하는지 식별하는 것이다.

다음으로, 음절을 분리하는 cmudict 패키지를 설치하여 추출된 단어의 음절을 분석했다. CMUDict는 Carnegie Mellon University Pronouncing Dictionar의 약어로 134,000개 이상의 단어와 그 발음을 포함하는 북미 영어용 오픈 소스 기계 판독식 발음 사전으로 총 39개의 음소를 기준으로 주어진 단어의 음절을 구분할 수 있다. 음절은 1음절인 경우 '매우 낮은 난이도'(very-low-difficulty), 2음절일 경우 '낮은 난이도'(low-difficulty), 3음절일 경우 '중간 난이도'(mid-difficulty), 4음절일 경우 '높은 난이도'(high-difficulty), 5음절 이상일 경우 '매우 높은 난이도'(very-high difficulty)로 설정하였다.

음절 수 분석에 더하여 난이도 패턴을 매핑하여 추가적으로 난이도 지수를 세분화 하였다. 단어의 자음과 모음의 조합 구조 역시 텍스트의 난이도를 좌우하는 주 원인이므로 교과서의 단어들이 어떤 음운 구조인지를 분석하여 난이도를 설정할 필요가 있다(이윤, 2016). 영어 자음과 모음 결합을 기분으로 'CV': 0.1, 'VC': 0.1, 'CVC': 0.2, 'VCV': 0.2, 'CCV': 0.3, 'VCC': 0.3, 'CCVC': 0.4, 'CCVCC': 0.4, 'CCCVCC': 0.5, 'CCCVCCCC': 0.5의 난이도 수치를 추가하여 난이도 위계를 설정하였다.

이후 각 난이도별로 단어들의 음절수와 음운론적 특성을 분석하여 낮은 난이도에서 어려운 난이도로 주어진 단어들을 재배열하였으며, tf-idf값에 기반한 빈도값을 동시에 고려하여 데이터 프레임을 생성하였다.

4) 최소대립쌍 및 문장 생성

최소대립쌍(minimal pair) 생성은 난이도 분석을 통해 분류된 단어를 기반으로 아주 낮은 난이도 그룹과 낮은 난이도 그룹으로 구분된 단어들만을 활용하여 최소대립쌍 세트를 제작했다. 이 때 생성하는 최소대립쌍은 기준 단어에서 하나의 음절만 다른 단어를 찾아서 세트로 반환하도록 했다.

문장 생성을 위해서는 Open AI의 API를 연결하여 GPT-3와 같은 생성형 AI를 기반으로 기존에 추출한 1000개 단어의 빈도와 음절 난이도를 고려하여 3 단어에서 10단어 사이의 길이의 짧은 문장을 생성하도록 코드를 작성했다. Text-davinci-003엔진을 활용하였고, 문장 생성에 사용되는 기존에 추출한 1000개 단어의 활용 비율을 최대 0.8로, 그 외 be동사, do동사, 전치사 등의 간단한 추가 단어 리스트의 비율을 0.2로 지정하여 자연스러운 문장을 생성하도록 했다. 최대 1000번의 문장 생성을 시도하도록 하였고, 비문법적인 문장과 중복 문장은 제외하여 총 300개의 문장을 생성하도록 코드를 작성하였다.

문장이 생성된 뒤에는 각 문장을 구성하는 단어에 매핑된 빈도 수치(tf-idf값)와 음절 구조의 난이도의 위계를 바탕으로 각 문장의 난이도 수치를 계산하여 문장을 쉬운 순부터 어려운 순으로 재배열하였다. 낮은 난이도에서는 간단한 음절 구조를 사용한 짧은 문장을, 어려운 난이도로 갈수록 더 복잡한 음절 구조와 긴 길이의 문장을 배치하였다. 이렇게 함으로써, 문장이 발음하기 쉬운 구조에서 점점 어려운 구조로 진행되도록 하였다.

마지막으로, 생성된 최소대립쌍과 문장을 평가하여 각 난이도에 적합한지 확인하였다. 필요한 경우, 문장을 수정하거나 삭제하여 텍스트 자료를 개발했다.

최종 개발 텍스트

1. 최소대립쌍

Difficulty	Minimal Pair 1	Minimal Pair 2
very_low_difficulty	do	go
very_low_difficulty	do	so
very_low_difficulty	go	so
very_low_difficulty	go	do
very_low_difficulty	so	go
very_low_difficulty	so	do
very_low_difficulty	bad	bag
very_low_difficulty	bad	dad
very_low_difficulty	bad	sad
very_low_difficulty	bad	bed
very_low_difficulty	bag	bad
very_low_difficulty	bag	big

very_low_difficulty	bed	bad
very_low_difficulty	bed	red
very_low_difficulty	big	bag
very_low_difficulty	big	pig
very_low_difficulty	bus	buy
very_low_difficulty	buy	bus
very_low_difficulty	cap	cup
very_low_difficulty	cap	map
very_low_difficulty	cap	cat
very_low_difficulty	cap	car
very_low_difficulty	cap	rap
very_low_difficulty	car	cap
very_low_difficulty	car	far
very_low_difficulty	car	cat
very_low_difficulty	cat	hat
very_low_difficulty	cat	cap
very_low_difficulty	cat	eat
very_low_difficulty	cat	car
very_low_difficulty	cup	cap
very_low_difficulty	dad	bad
very_low_difficulty	dad	sad
very_low_difficulty	dad	day
very_low_difficulty	day	say
very_low_difficulty	day	way
very_low_difficulty	day	dad
very_low_difficulty	eat	hat
very_low_difficulty	eat	cat
very_low_difficulty	fan	far
very_low_difficulty	fan	ran
very_low_difficulty	fan	fun
very_low_difficulty	fan	man
very_low_difficulty	far	fan
very_low_difficulty	far	car
very_low_difficulty	fun	sun
very_low_difficulty	fun	fan
very_low_difficulty	fun	run
very_low_difficulty	get	let
very_low_difficulty	get	got
very_low_difficulty	get	met
very_low_difficulty	got	get
very_low_difficulty	got	lot
very_low_difficulty	got	hot
very_low_difficulty	hat	eat
very_low_difficulty	hat	cat
very_low_difficulty	hat	hit

very_low_difficulty	hat	hot
very_low_difficulty	hit	hat
very_low_difficulty	hit	sit
very_low_difficulty	hit	hot
very_low_difficulty	hot	hat
very_low_difficulty	hot	how
very_low_difficulty	hot	lot
very_low_difficulty	hot	hit
very_low_difficulty	hot	got
very_low_difficulty	how	hot
very_low_difficulty	how	now
very_low_difficulty	job	joy
very_low_difficulty	joy	job
very_low_difficulty	let	get
very_low_difficulty	let	lot
very_low_difficulty	let	met
very_low_difficulty	lot	let
very_low_difficulty	lot	got
very_low_difficulty	lot	hot
very_low_difficulty	man	fan
very_low_difficulty	man	map
very_low_difficulty	man	ran
very_low_difficulty	man	men
very_low_difficulty	map	cap
very_low_difficulty	map	man
very_low_difficulty	map	rap
very_low_difficulty	men	met
very_low_difficulty	men	man
very_low_difficulty	met	let
very_low_difficulty	met	get
very_low_difficulty	met	men
very_low_difficulty	mom	tom
very_low_difficulty	new	now
very_low_difficulty	now	new
very_low_difficulty	now	how
very_low_difficulty	pig	big
very_low_difficulty	pop	top
very_low_difficulty	ran	fan
very_low_difficulty	ran	run
very_low_difficulty	ran	man
very_low_difficulty	ran	rap
low_difficulty	curry	hurry
low_difficulty	funny	sunny
low_difficulty	hurry	curry
low_difficulty	later	water

low_difficulty	sorry	worry
low_difficulty	sunny	funny
low_difficulty	water	later
low_difficulty	worry	sorry
low_difficulty	became	become
low_difficulty	become	became
low_difficulty	better	butter
low_difficulty	butter	better
low_difficulty	forget	forgot
low_difficulty	forgot	forget
low_difficulty	making	taking
low_difficulty	taking	making
low_difficulty	cooking	looking
low_difficulty	grandma	grandpa
low_difficulty	grandpa	grandma
low_difficulty	looking	cooking
low_difficulty	talking	walking
low_difficulty	walking	talking

2. 문장

Sentence	TF-IDF Sum	Difficulty Score	Average Word Length
Go for it!	0.883449069	0.25	4
Take a break.	0.388348624	0.5	4
Here we go!	0.193000108	0.35	4
Clean the dishes.	0.111803399	0.375	4
Enjoy the moment.	0.111803399	0.475	4
Enjoy the sunshine.	0.111803399	0.475	4
Look up at the sky.	0.092474337	0.283333333	6
Go to the store.	0.091151001	0.3	5
Here I am.	0.041042593	0.475	4
I don't understand.	0.033451892	0.5	5
Nice to meet you.	0.002361821	0.32	5

I love to cook.	0.001708282	0.42	5
I love to laugh.	0.001708282	0.42	5
I like to dance.	0.001708282	0.5	5
I like to smile.	0.001708282	0.5	5
Go outside and play.	0.00120169	0.32	5
See you soon.	0	0.25	4
Eat healthy food.	0	0.275	4
Happy New Year!	0	0.3	4
Eat ice cream today.	0	0.34	5
The sun shines bright.	0	0.34	5
Take it easy.	0	0.35	4
Go away now.	0	0.375	4
Eat well today.	0	0.375	4
Go outside now.	0	0.375	4
Eat more fruits.	0	0.375	4
She smiled brightly.	0	0.375	4
He loves her.	0	0.375	4
I ate lunch.	0	0.4	4
Cats love milk.	0	0.4	4
Sunny days ahead.	0	0.4	4
Happy summer days!	0	0.4	4
Eat breakfast every day.	0	0.42	5
Let's eat pizza.	0	0.42	5
Let's go camping.	0	0.42	5
It's sunny today.	0	0.42	5
Let's go outside!	0	0.42	5
Let's go swimming.	0	0.42	5
Let's go shopping.	0	0.42	5
It's sunny outside.	0	0.42	5
I'm ready now.	0	0.44	5
I'm happy now.	0	0.44	5
Have a good day!	0	0.44	5
Have a great day.	0	0.44	5
I like ice cream.	0	0.44	5
I love ice cream.	0	0.44	5
I love sunny days.	0	0.44	5
I saw her smile.	0	0.44	5
It's a beautiful day.	0	0.45	6
I love you!	0	0.475	4
I like pie.	0	0.475	4
I love you.	0	0.475	4
I like you.	0	0.475	4

I love tea.	0	0.475	4
I love pie.	0	0.475	4
Blue sky today.	0	0.475	4
Make it happen.	0	0.475	4
I eat breakfast.	0	0.475	4
She smiled happily.	0	0.475	4
She laughed happily.	0	0.475	4
Eat healthily today.	0	0.475	4
Eat more vegetables.	0	0.475	4
He smiled happily.	0	0.475	4
He laughed happily.	0	0.475	4
I like cats.	0	0.5	4
I love cars.	0	0.5	4
Make a wish.	0	0.5	4
I like dogs.	0	0.5	4
I love cats.	0	0.5	4
I love dogs.	0	0.5	4
I love pets.	0	0.5	4
I like food.	0	0.5	4
I love food.	0	0.5	4
I love shoes.	0	0.5	4
I like books.	0	0.5	4
I love books.	0	0.5	4
I laugh often.	0	0.5	4
I love skiing.	0	0.5	4
I love spring.	0	0.5	4
Nice day today.	0	0.5	4
I will succeed.	0	0.5	4
Take action now.	0	0.5	4
Hope never dies.	0	0.5	4
I ate breakfast.	0	0.5	4
Make today count.	0	0.5	4
I laughed loudly.	0	0.5	4
Love never fails.	0	0.5	4
Cats love playing.	0	0.5	4
Play outside today.	0	0.5	4
Good morning everyone!	0	0.5	4
Good morning sunshine!	0	0.5	4
Happy weekend everyone!	0	0.5	4
Everything changes quickly.	0	0.5	4
I love it here.	0	0.5	5
It's raining outside.	0	0.5	5
It's beautiful outside.	0	0.5	5
I'm here now.	0	0.52	5
I love my dog.	0	0.52	5

I love summer days.	0	0.52	5
I love reading books.	0	0.52	5
I like reading books.	0	0.52	5
I enjoy reading books.	0	0.52	5
I'm happy today.	0	0.52	5
I love hot chocolate.	0	0.52	5
I love cake.	0	0.6	4
I love life.	0	0.6	4
I like cake.	0	0.6	4
I love music.	0	0.6	4
I love pizza.	0	0.6	4
I love Italy.	0	0.6	4
I love pasta.	0	0.6	4
I like pizza.	0	0.6	4
I love pizza!	0	0.6	4
I love tacos.	0	0.6	4
I smile often.	0	0.6	4
I love winter.	0	0.6	4
Take a chance.	0	0.6	4
I love cheese.	0	0.6	4
I like coffee.	0	0.6	4
I love apples.	0	0.6	4
I love summer.	0	0.6	4
I love coffee.	0	0.6	4
I love travel.	0	0.6	4
I love hiking.	0	0.6	4
I love nature.	0	0.6	4
I like apples.	0	0.6	4
I like cheese.	0	0.6	4
I like summer.	0	0.6	4
I like animals.	0	0.6	4
I love reading.	0	0.6	4
I like bananas.	0	0.6	4
I love England.	0	0.6	4
I love singing.	0	0.6	4
I love summers.	0	0.6	4
I love beaches.	0	0.6	4
I love sunsets.	0	0.6	4
I like turtles.	0	0.6	4
I like cookies.	0	0.6	4
I love bananas.	0	0.6	4
I like reading.	0	0.6	4
I love cookies.	0	0.6	4
I love turtles.	0	0.6	4
I love running.	0	0.6	4

I love kittens.	0	0.6	4
I like running.	0	0.6	4
I love animals.	0	0.6	4
I like singing.	0	0.6	4
I smile easily.	0	0.6	4
I like learning.	0	0.6	4
I smile happily.	0	0.6	4
I love swimming.	0	0.6	4
I love pancakes.	0	0.6	4
I love football.	0	0.6	4
I love sunshine.	0	0.6	4
I love learning.	0	0.6	4
Take care today.	0	0.6	4
I love laughter.	0	0.6	4
I smiled happily.	0	0.6	4
I love chocolate.	0	0.6	4
I love computers.	0	0.6	4
I love traveling.	0	0.6	4
I love basketball.	0	0.6	4
I love springtime.	0	0.6	4
I love summertime!	0	0.6	4
I love summertime.	0	0.6	4
Nice weather today.	0	0.6	4
I love my life.	0	0.6	5
I'm learning English.	0	0.6	5
I love peanut butter.	0	0.6	5

■ 작성 코드

https://colab.research.google.com/drive/1Imlud9qLULPYp4Q705W0mdKXBEN1fnkA?usp=sharing

[부록4] 2차 전문가 검토지

중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 텍스트 개발에 대한 전문가 질문지(2차)

□ **안내문**

본 질문지는 중학교 영어 학습 부진 학생의 읽기 유창성 향상을 위한 AI기반 영어 소리 내어 읽기(reading-aloud) 텍스트 개발 과정 및 결과물에 대한 타당도를 검증하기 위해 제작되었습니다.

본 질문지는 연구의 개요, 텍스트 도출과정, 텍스트, 활용 방법에 대한 의견을 묻는 질문지로 구성되어 있습니다. 질문지를 읽어보시고 타당도에 따라 해당하는 점수를 표시하여 주시기 바랍니다.

본 질문지를 작성하는데 약 20분 정도 소요될 것으로 예상되며, 응답 시 궁금한 사항은 연구자에게 질문하여 주시기 바랍니다.

본 조사에서 작성하신 기본 정보 및 응답 내용은 연구 목적 이외에는 절대 사용되지 않으며, 비밀을 보장합니다. 논문에서는 전문가임을 증명하기 위해 전공분야, 최종학력, 소속, 경력 부분만 언급됨을 사전에 알려드립니다.

바쁘신 와중에 본 연구에 협조하여 주셔서 진심으로 감사합니다.

2023년 6월

서울대학교 사범대학 AI융합교육학과
지도교수: 김선희 교수
연구자: 석사과정 이은주
이메일: lej3687@snu.ac.k

□ **전문가 프로필**

성함	직업	경력(년)	최종 학력	전공 분야
(기재 예시) *홍길동*	*중학교 교사*	*15*	*석사*	*교육 공학/국어 교육/AI 융합 교육*

연구개요

1. 연구의 배경

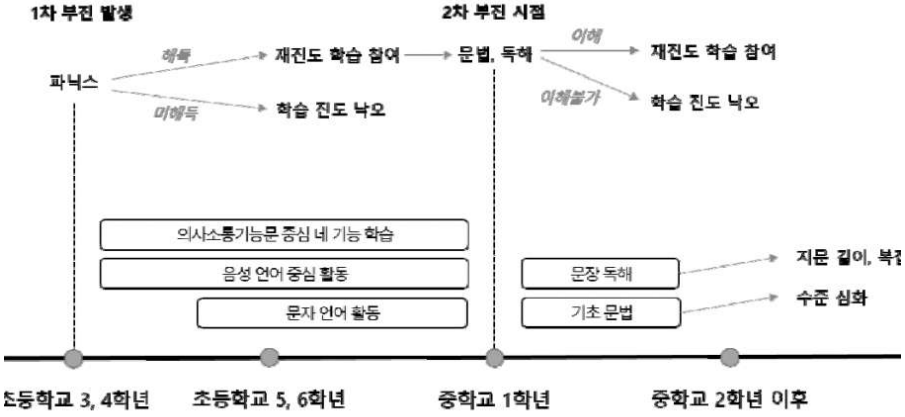
해가 거듭될수록 학생 간 학습격차는 심화되어 기초학력이 부족한 학생들을 위한 정책의 필요성이 더욱 대두되고 있다. 교육부는 제1차 기초학력 보장 종합계획(2023~2027)을 발표하여 코로나로 인한 비대면 수업기간 동안 심화된 학생들의 학습 결손과 정서 안정을 지원하는 국가 단위의 기초학력 보장 체계를 마련하고자 했다. [그림1]은 2017년과 2021년도의 국가수준 학업성취도 평가에서 기초학력 미달 학생들의 비율을 나타내는데, 코로나 이전에 비하여 기초학력 학생의 비율이 모든 과목에서 급격히 증가했음을 알 수 있다.

[그림 11] 2017년 · 2021년 국가수준 학업성취도 평가 1수준(기초학력 미달) 비율 추이(교육부, 2022)

교육부는 이번 종합계획에서 인공지능(AI)기반 진단체계 구축을 강화하여 에듀테크 기반 학생 개별화 학습 지원을 강조했다. 이를 위해 국가기초학력지원센터를 운영하며, 2022 개정 교육과정에 따른'최소한의 성취기준'을 개발하고, 국가수준의 진단도구 및 교수·학습 자료를 보급하고자 한다(교육부, 2022).

2. 연구의 필요성

현재 서울특별시교육청은 기초학력 진단-보정 시스템(s-basic)홈페이지를 통해 학년 초 학생들의 교과학습능력(국어, 영어, 수학 중심) 진단검사를 진행한다. 영어 진단평가의 경우 4기능(듣기, 말하기, 읽기, 쓰기)를 모두 포함하고 있으나, 말하기 영역은 학생의 발화를 직접 녹음하여 분석하는 형태가 아닌, 맥락에 맞는 적절한 응답 고르기의 형태로 제공한다.



[그림 12] 영어 학습 부진 심화 패턴(김태은, 2020)

김태은(2020)은 기초학력 진단-보정 시스템이 유일한 기초 영어 능력 검사도구라는 점을 강조하며, [그림2]에서와 같이 초등학교에서 발생하는 1차 부진(파닉스) 영역에 대한 진단 도구의 부족을 지적했다. 파닉스 능력 진단과 선제적 지원을 제공한다면 학습 부진이 발생하는 초기에 교사가 개입하여 학습 부진을 예방할 수 있다는 것이다.

중학교 영어 학습 부진 학생들을 지도하기 위해 서울특별시교육청이 교사들에게 제공하는 자료는 각 학년별 의사소통영역 중심으로 이루어진 영어 '늘품이'교재와 음소별로 학습이 가능한 '15분 파닉스'교재이다. '15분 파닉스'교재는 온라인 단어 학습 사이트인 클래스카드에 단어 세트를 업로드하여 교사들이 사용할 수 있도록 하고 있다. 하지만 중학교 기초학력 미달 학생들의 음성 발화를 분석하고 즉시 피드백을 줄 수 있는 AI기반 프로그램과 연계하여 활용할 텍스트는 별도로 개발되어 있지 않다.

3. 연구의 목적

본 연구는 중학교 영어 학습 부진 학생들을 대상으로 한 읽기 유창성 향상을 위한 소리내어 읽기(reading-aloud)활동에서 활용할 텍스트 세트를 제작하는 것을 목적으로 한다. 2015 개정 교육과정의 초등학교 6학년교과서 5종, 중학교 1학년 교과서 9종의 교과서 읽기 본문 텍스트를 통합하고, 자연어 처리 기술을 활용하여 통합된 전체 텍스트의 주요 어휘 1000개를 빈도 기반으로 추출히여 주요 어휘의 음절수와 음절 난이도 값을 수치화하고, 생성형AI를 활용하여 해당 단어를 활용한 읽기 자료를 개발하는 것을 목적으로 한다. 가장 낮은 단계의 읽기 자료로는 가장 많은 빈도수를 가진 단어들의 최소 대립쌍(minimal pair)를, 중간 단계로는 빈도가 높고 음절 난이도가 낮은 단어들로 이루어진 5단어 내의 문장을, 가장 높은 난이도로는 음절 난이도가 높은 단어들로 이루어진 10단어 내의 문장을 생성한다. 개발된 텍스트는 1차 전문가 타당화를 거친 뒤 수정하고, 2차 전문가 타당화를 거쳐 영어 학습 부진 학생들의 읽기 유창성 개발을 위한 AI기반 프로그램 활용 소리내어 읽기(reading-aloud) 수업 과정에서의 활용 타당도를 알아보고자 한다.

4. 이론적 배경 (요약)

1. 영어 학습 부진 학생

1.1 학습 부진 학생의 정의

기초학력 부진 학생들은 학습 속도가 느리고 어휘력, 언어이해, 언어 표현이 부족하여 낮은 자존감과 낮은 학습 동기, 불안, 위축 등의 특성을 보인다(김태은 외, 2020). 본 연구에서는 학습 부진 학생을 기초학력미달 또는 기초학습 부진학생과 구분하여 영어 교과 교육과정의 성취기준에 도달하지 못한 학습 부진 학생을 대상으로 한다. 김성혜(2019)의 연구에 따르면 중학교 영어 성취수준이 '하'수준인 학생의 경우 20~50% 미만의 학생들이 수업 내용을 잘 이해하지 못하며, 교육과정 성취기준을 미충족한다는 것을 보여준다.

1.2 영어 학습 부진 학생의 읽기 지도 필요성

초등학교의 영어 수업 시간은 읽기 학습에 충분하지 않아 일부 학생들은 단어 해독 능력이 있음에도 문장이나 문단 이해에 어려움을 겪는다. 따라서, 읽기 발달 초기 단계에 있는 학생들의 읽기 유창성을 향상시키기 위한 노력이 필요하다(최수진, 이윤, 2015). 중학생 중에서 영어 학습에 어려움을 겪는 학생들의 영어 읽기 능력 부족은 기본적인 음철법의 적용에서부터 동료들의 영어 독해 이해도 및 독해 전략의 활용에 이르기까지 그 범위가 광범위하며, 이는 초등학생들에 비해 더 큰 편차를 보인다. 이러한 중학생들의 영어 읽기 능력 부족 문제는 반드시 해결해야 하는 중요한 과제이다. 영어 학습의 어려움은 읽기 능력과 깊은 연관이 있으므로, 이러한

학습 어려움을 해결하기 위한 교육과정 개선은 문자의 노출 양과 시기, 방법뿐만 아니라 문자언어 관련 성취기준의 연계성을 강화하는 것이 필요하다.(한국교육과정평가원, 2019)

1.3. 영어 학습 부진 학생 지도의 어려움

중학교 1학년부터 영어 교과는 점차 음성언어에서 문자 언어 중심으로 전환되며, 학습 어휘의 양이 급증하고 언어 형식이 강화되는 시기이며, 영어 교사들은 학습 부진 학생들이 '어휘'와 '읽기' 영역을 우선적으로 학습해야 한다는 입장을 보였다.(김성혜, 2019) 해당 연구에 참여한 교사들은 문자 언어와 관련된 교육과정의 학습 수준이나 학습량 적정화보다는, 기존의 어휘 및 언어형식을 제시하는 방법의 위계성이 부족하다는 점과 주제별로 구성된 교과서에 나타난 어휘 체계의 설계 기준이 모호함을 지적하며, 교사들의 기본학력 학생 지도의 어려움을 호소했다. 김성혜(2019)는 2015 개정 교육과정 영어과의 성취기준, 어휘, 언어형식이 학교급별로 구분되어 있어 학습 부진 학생 지도에 활용도가 떨어짐을 지적하며 일상생활에서 사용 빈도가 높은 낱말과 의사소통 기능 표현을 순차적으로 익힐 수 있도록 학습 자료가 세분화되어야 한다는 것을 강조했다. 또한, 필수 성취기준과 필수 어휘, 의사소통기능, 언어형식을 기초군으로 분류하여 교육과정에 제시해야 함을 강조했고, 초기 읽기 능력 습득이 학습 부진의 경계가 되는 지점이므로 고빈도어를 토대로 한 기초 어휘목록이 교육과정에서 제시되어야 한다고 주장했다.

1.4 3. 음운론적 인식 및 읽기 유창성의 학습 필요성

엄월영, 김혜련(2014)은 영어 학습 부진아들이 반복 학습 및 효과적인 학습 방법 등을 통해 보다 높은 학습 수준에 도달할 수 있다고 강조하였다. 김은정과 이병민(2021)은 초등학교 6학년 학생들의 영어 읽기 능력을 세분화하여 분석하였는데, 영어는 문자와 파닉스가 일정하지 않음에도 음운법 규칙을 통해 어휘의 84%를 규칙성을 적용해 읽을 수 있고, 알지 못하는 어려운 단어를 유추하는 데에 유용하기 때문에 학생들에게 맞춤형 읽기 보충학습은 필수적이라고 주장했다. 이를 위해서는 단어와 문장 유창성과 관련된 다양한 영어 독서 자료와 함께 해독 능력을 향상하기 위한 교육이 제공되어야 하며, 학생들 간의 개인차는 정규 영어 수업 또는 방과후 교육 등을 통해 공교육 내에서 처치가 이루어져야 한다고 강조했다.

2. 이독성과 텍스트 난이도

1.1 이독성의 정의

이독성(Readability)는 주어진 읽기 자료 내의 모든 요소들에 대해 글을 읽는 사람의 이해도, 최적의 속도로 읽는 능력, 그리고 그 읽기 자료에 대한 흥미 정도에 대

한 종합적인 개념이다(Dale & Chall, 1949). Graves와 Graves(2003)는 이독성에 영향을 미치는 두 가지 요인으로 '읽기자료에 내재된 요인들(factors largely inherent in the text itself)'과 '독자 및 읽기자료와 관계있는 요인들(factors involving the reader and the text)을 제시했다. 내재된 요인으로는 어휘, 문장 구조, 문장의 길이, 풍부한 설명(elaboration), 일관성(coherence), 텍스트 구조가 있으며, 외재적 요인으로는 내용 친숙도, 배경지식, 독자의 인지적 수준, 글의 특징, 독자의 흥미가 있다. 최근의 연구들이 제시하는 이독성 공식들은 단어의 음절이 길수록, 단어의 철자 수가 많을수록, 저빈도(low frequency)의 단어가 많을수록, 한 문장 내의 단어가 많을수록 그 읽기 자료는 이해하는 것이 어렵다고 가정한다(임영진, 조윤경, and 정영경, 2015). 이처럼 이독성에는 다양한 원인이 영향을 미치지만 본 연구에서는 초기 읽기 학습자를 위한 텍스트 개발이라는 목적과 관련하여 이윤(2016)이 제시한 요소를 기준으로 삼았다. 이윤(2016)은 읽기 교재가 어떤 어휘로 구성되는가에 따라 텍스트의 난이도가 달라지므로 학생들의 읽기 능력에 적합한 텍스트를 제공하는 것이 읽기 학습에서 중요한 역할을 한다고 했다. 그러나 음운 규칙 기반 어휘로만 텍스트를 구성하는 것은 어려운 일이며, 영어의 특성상 초기에 습득해야 하는 어휘들 중 다수가 음운 규칙이 적용되지 않는 의미 중심 어휘이므로, 읽기 학습 초기에 이 특성을 고려하여 난이도를 조정하는 것이 필요하다고 강조했다.

1.2 영어 교과서 이독성 수준

영어 교과서의 이독성 수준이 학습자의 이해 능력을 초과하는 경우 지식 전달 과정에서 공백이 생기게 되어 학습 의욕 저하를 유발하므로 학습자의 이해력에 맞는 교재를 활용해야 한다(차경환 and 손석완, 2006). 특히 중학교로 올라가면서 학습 부진의 편차가 커지는 것은 읽기 영역의 학습 부진에서 기인한 것으로 문자와 관련된 영역의 교과 내용의 적정성을 제고할 필요가 있다. 우리나라 2015 개정 교육과정의 초등학교 교과서 읽기 지문의 가독성 지수를 분석한 결과, 5학년과 6학년 사이의 가독성은 미국 학제 기준으로 한 학기 정도의 차이가 났지만(황은경, 박광현, 2019), 초등학교 6학년과 중학교 1학년 읽기 지문 사이에는 미국 학제 기준으로 한 학년 이상의 가독성 수치 차이가 나타난다(배지영, 2019). 임영진, 조윤경, and 정영경(2015)은 초등학교 6학년과 중학교 영어 교과서의 이독성과 어휘 난이도를 분석였는데, 이독성 지수는 학교급간에 1점 이상 증가하였으며, 중학교 교과서의 어휘 수는 초등학교 교과서 대비 세 배 이상 증가하였다. 무작위로 증감하는 이독성 수치는 학습자에게 어려움을 줄 수 있으므로, 학습 자료는 저난이도에서 고난이도, 적은 학습량에서 많은 학습량으로 조직화되어야 한다는 결론을 도출하였다. 이윤(2016)은 초등영어 4종의 교과서를 분석해 본 결과, 공통적으로 어휘 반복율, 어휘 난이도 지수 측면에서 모두 읽기 발달 초기 학습자의 특성을 고려하지 못하는 것을

발견했다. 따라서, 중학교 영어 학습 부진 학생은 교과서를 활용한 읽기 활동에 어려움을 겪을 수 있으며, 초등학교 교과서 텍스트를 활용한 읽기 학습을 하더라도, 여전히 학습부진 학생을 읽기 능력 발달을 위한 학습에 적절한 텍스트라고 보기 어렵다.

3. 영어 학습에서의 음절 연계 단어 학습 중요성

여러 연구들에서 음절 연계 단어 학습의 효과성을 연구하였다. Metsala와 Walley(1998)은 "lexical restructuring hypothesis"를 제시하여 어휘 성장과 음운인식, 독서 발달 간의 밀접한 관계를 주장하였다. 이 연구에서는 어린이들의 어휘가 확장되고 음운적으로 유사해짐에 따라 음절적 표현이 발달되며, 이러한 음절적 표현이 음운인식의 발달에 기여하고 독서와 쓰기 학습에 영향을 미친다고 언급했다. 이와 함께, Ayres(1998)는 음운인식 프로그램이 영어를 학습하는 어린이들에게 어휘 학습을 촉진시킨다는 것을 밝혔다. 이 연구에서는 음운인식 프로그램과 어휘 학습 활동을 통합하여 구성된 교육 프로그램이 어린이들의 단어 분절(segmentation) 능력에 긍정적인 영향을 미친다고 결과를 제시했다. 이러한 선행 연구들은 ESL 학생의 음절 연계 단어 학습의 중요성을 강조하며, 구두 언어 능력과 어휘 수준을 고려하여 교육 프로그램을 설계해야 한다는 것을 나타낸다.

4. 영어 학습자의 반복 읽기 효과성

여러 학자들은 연구를 통해 언어학습자들이 언어학습을 반복(task-repetition)할 때 언어 사용의 복잡성(complexity), 정확성(accuracy), 그리고 유창성(fluency) 등이 제2언어(L2) 학습 및 발달에 어떻게 영향을 미치는지를 탐구해왔다(Hawkes, 2012; Takimoto, 2012). Bygate (1996, 2001)는 Levelt의(1989) 연설 생산 모델을 이용해 작업 반복의 잠재적인 심리언어학적 이점을 설명하였다. 이 모델에 따르면 언어 사용은 개념화(conceptualization), 구현(formulation), 및 발성(articulation)의 세 단계를 거친다. 작업을 반복함으로써 학습자들은 이전의 처리 단계(즉, 개념화, 구현 및 발성)를 기억 저장소에 보관하고 다음 번에 재사용할 수 있다. 최수진, 이윤(2015)는 초등학생들의 영어 읽기 유창성 향상을 위해 반복적인 읽기 학습에 음소 지도를 도입하여 음소 인식을 포함한 반복 읽기 방법이 단어 인식의 정확성, 운율, 이해력 증진 등에서 더 큰 향상을 보였음을 확인하였고, 반복적인 읽기가 읽기 유창성과 독해력 향상에 효과적임을 발견했다.

5. 맞춤형 수업을 위한 AI프로그램 활용

5.1 학습부진 학생 지도 과정에서 AI프로그램 활용의 효과

교과 학습 부진 학생을 지도하기 위해서는 각 학생에 맞는 개별 보충 지도를 제공

해야 하므로 다인수 학급에서는 적절한 수업이 이루어지기 어렵다. 대상 학생 간에도 학습 부진의 정도에 있어 차이를 보이기 때문에 학생들의 영어 수준에 따라 다른 종류의 보충학습이 제공되어야 한다(엄월영, 김혜련, 2014). 김태은(2020)의 연구에 따르면 학습 부진 학생들의 성장을 위해서는 개별 학생의 학습 과정과 어려움의 원인을 '들여다보기', 학습 단계와 자료를 세밀하게 '세분화하기', 반복적이고 꾸준하게 반복하는 '기초 근육 만들어주기', 꾸준히 성장 과정을 지켜보며 감정적인 지원을 제공하는 '견뎌주기' 과정이 필요하다. 김태은(2020)은 효과적인 학습 부진 지도를 위해서는 교사의 지도계획과 실행이 현실적으로 적정화 되어야 하며, 그 과정에서 AI기반 프로그램을 활용하는 것이 필요하다고 강조했다. 이는 개별지도 효과를 높이면서도 교사의 부담을 줄이고, 실수를 두려워하는 학생들이 정의적 여과(affective filter)가 낮은 상태에서 학습에 임할 수 있도록 하기 때문이다.

5.2 텍스트 활용을 위한 AI기반 도구

본 연구에서는 Microsoft의 교육 플랫폼인 Teams의 과제 기능인 Reading Progress 프로그램을 활용하여 개발된 텍스트를 과제로 부여하는 형태로 단계별 소리 내어 읽기(reading-aloud) 활동을 진행한다. Reading Progress 프로그램은 교사가 업로드한 텍스트를 학생이 소리 내어 읽으면, 학생의 수행 결과를 자동으로 분석한 결과를 [그림3]과 같이 보여준다.

[그림 13] Insights기능을 활용한 학생 수행 결과 자동 분석(Microsoft, 2022)

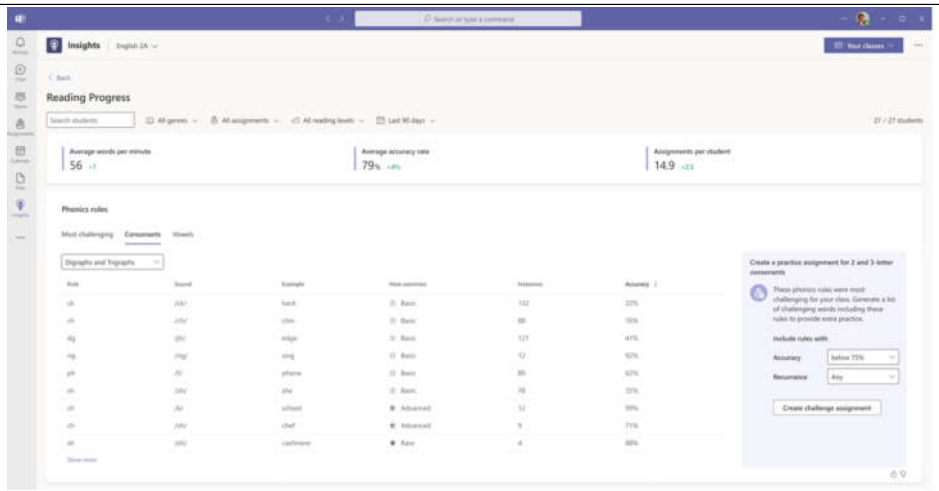
분석 결과에는 텍스트를 읽는 학생의 영상이 좌측에 나타나고, 학생의 발화가 녹음되어 저장된다. 중앙에 있는 결과 리포트 상단에는 1분당 읽은 올바른 단어 수(correct word per minute)와 정확도 비율(accuracy rate)이 숫자로 표기되어 나타난다. 학생의 발화 오류는 자동으로 다섯 가지 영역(mispronunciation, omission, self-correction, insertion, repetition)으로 분석되어 텍스트에 각기 다른 색깔로 표기된다.

교사는 [그림4]와 같이 Insights기능을 활용하여 학생의 과제별 수행 결과 그래프를 볼 수 있다. 그래프 하단에는 학생이 부정확하게 읽은 단어가 자동으로 분석되어 워드 클라우드의 형태로 제공된다. 기간을 설정하여 그룹 평균 수행 대비 해당 학생의 수행 결과를 비교하여 볼 수도 있다.

과제 설정 시 교사는 읽기 후에 리딩 코치(reading coach) 기능을 추가할 수 있다. 리딩 코치 기능은 [그림5]와 같이 학생이 글을 읽을 때 부정확하게 발음한 단어를 자동으로 분석하여 제시하며, 올바른 발음을 음절별로 다시 들을 수 있도록 음원을 제공한다. 학생은 리딩 코치가 제공하는 발음을 듣고 다시 단어를 올바르게 읽도록 연습하며, 일정 수준 이상의 정확도에 도달하면 다음 단어로 넘어갈 수 있다.

[그림 14] Reading Coach 기능

모든 과제가 끝나면 학생이 정확히 읽지 못한 모든 단어가 자동으로 분석되며, 교사는 개별 학생의 이름을 클릭하여 해당 학생이 어떤 음소를 부정확하게 발음했는지를 [그림6]의 화면을 통해 파악할 수 있다. 오류 비율이 높은 음소를 교사가 선택하여 추가 도전 과제 만들기 버튼을 클릭하면 해당 학생에게만 개별화된 단어 읽기 과제가 부여된다.



[그림 15] 학생의 오류 음소 패턴 자동 분석 및 개별화된 추가 과제 자동 생성

5.3. 읽기 텍스트 구성 기준

이윤(2016)은 교과서 텍스트는 영어 학습자의 학습 내용이자 도구이므로, 초기 읽기 단계에서 학생들이 알파벳 원리를 익혀서 단어 인지 과정을 학습하도록 교재를 체계적으로 구성해야 함을 강조했다. 현재 교과서 텍스트의 경우 내용만 고려하여 어휘 선정을 하는 경우, 등장 인물이나 장소의 이름이 발음하기 어렵거나 길이가 긴데, 이것은 읽기 난이도를 높이는 주 원인임을 지적했다. 따라서 낮은 난이도의 텍스트에서는 소리와 철자의 대응이 명확한 어휘와 고빈도어 중심으로 텍스트를 구성하거나 특정 모음 패턴을 반복적으로 제시하는 등의 방법을 통해 읽기 텍스트가 학생들의 읽기 학습을 돕도록 특별하게 체계화되어야 한다고 강조했다. Klare(1954)은 이독성과 관련하여 자주 사용되는 두 변수인 '의미 측정'(어휘 난이도 등)과 '구문 측정'(평균 문장 길이 등)이 텍스트 난이도의 가장 효과적인 예측 요인임을 연구를 통해 입증했고, 이 두 변수를 통해 본문의 읽기 난이도를 추정하였다. 또한, Klare(1968)의 연구에 따르면, 사람들은 일부 단어를 다른 단어보다 훨씬 더 자주 사용하며, 이런 빈번하게 사용하는 단어들을 더 빨리 인식하고 선호하며, 더욱 쉽게 이해하고 배운다. 그래서 이 '단어의 사용 빈도'라는 변수가 텍스트의 읽기 쉬운 정도를 결정하는데에 중요한 역할을 한다고 주장했다. 따라서 본 연구는 연구의 대상인 교과서 텍스트를 분석할 때 문서에 나타난 단어의 빈도, 일상 생활에서의 빈도, 음절의 길이, 음절의 구성(단모음, 단자음, 이중모음, 이중자음)에 대한 난이도 수치를 부여하고, 합산된 난이도를 기준으로 최소대립쌍과 짧은 문장을 생성하고자 한다.

6. 연구 대상

텍스트 개발을 위한 연구 대상은 2015 개정 교육과정 초등 6학년 5종, 중학교 1학년 9종 교과서의 읽기(reading) 본문을 대상으로 한다.

학교급	출판사	저자
초등	동아	박기화 외
	대교	이재근 외
	천재	함순애 외
	YBM(김)	김혜리 외
	YBM((최)	최희경 외
중등	능률(김)	김성곤 외
	동아(윤)	윤정미 외
	동아(이)	이병민 외
	미래엔(최)	최연희 외
	비상(김)	김진완 외
	지학사	민찬규 외
	천재(이)	이재영 외
	천재(정)	정사열 외
	YBM(송)	송미정 외

7. 연구 절차

1) 코퍼스 수집

먼저, 텍스트 개발 절차에서 주요한 첫 단계는 코퍼스를 수집하는 것이다. 본 연구에서는 2015 개정 교육과정의 초등학교 6학년 교과서 5종과 중학교1학년 교과서 9종의 읽기 지문 텍스트를 추출하여 하나의 통합된 파일로 작성했다. 이렇게 수집된 코퍼스는 본 연구의 주요 데이터 소스이며, 이는 읽기 지문 텍스트를 분석하고 토큰화하는 데 사용되었다.

2) 텍스트 전처리 및 고빈도 단어 추출

다음 단계는 텍스트 전처리와 고빈도 단어 추출이다. 이 단계에서는 Google Colab을 활용하여 통합된 텍스트에 나타난 단어를 토큰화하고, 문장부호, 숫자 등의 불용어를 처리했다. 토큰화란 텍스트를 개별 단어나 표현으로 분리하는 과정을 말

하며, 불용어는 분석에서 제외하는 덜 중요한 단어들을 말한다. 불용어 처리 후, 빈도수가 가장 높은 단어를 추출하기 위해 역문서빈도(TF-IDF)값이 높은 상위 단어 1000개를 추출했다. TF-IDF는 단어의 중요성을 판단하는 수치로, 해당 단어가 문서 내에서 얼마나 자주 등장하는지와 동시에 전체 문서에서 얼마나 희귀하게 등장하는지를 모두 고려한 값이다.

3) 단어의 읽기 난이도 분석

단어의 읽기 난이도 매핑 과정에서는 음운론적 복잡성(phonological complexity)을 중심으로 하여 낮은 난이도부터 어려운 난이도를 구분하였다. 이러한 난이도 기준을 추가한 이유는 문장 생성 과정에서 어휘의 빈도와 함께 및 음운론적 요소를 고려하기 위함이다.

우선 앞서 분석한 텍스트에서 가장 높은 빈도 순으로 단어 1000개를 추출한 후, 이 단어들의 품사를 태깅하여 각 단어의 품사 정보를 단어와 짝짓는 과정을 거쳤다. 품사 태깅은 단어가 문장에서 어떤 역할을 하는지 식별하는 것이다.

다음으로, 음절을 분리하는 cmudict 패키지를 설치하여 추출된 단어의 음절을 분석했다. CMUDict는 Carnegie Mellon University Pronouncing Dictionar의 약어로 134,000개 이상의 단어와 그 발음을 포함하는 북미 영어용 오픈 소스 기계 판독식 발음 사전으로 총 39개의 음소를 기준으로 주어진 단어의 음절을 구분할 수 있다. 음절은 1음절인 경우 '매우 낮은 난이도'(very-low-difficulty), 2음절일 경우 '낮은 난이도'(low-difficulty), 3음절일 경우 '중간 난이도'(mid-difficulty), 4음절일 경우 '높은 난이도'(high-difficulty), 5음절 이상일 경우 '매우 높은 난이도'(very-high difficulty)로 설정하였다.

음절 수 분석에 더하여 난이도 패턴을 매핑하여 추가적으로 난이도 지수를 세분화 하였다. 단어의 자음과 모음의 조합 구조 역시 텍스트의 난이도를 좌우하는 주원인이므로 교과서의 단어들이 어떤 음운 구조인지를 분석하여 난이도를 설정할 필요가 있다(이윤, 2016). 영어 자음과 모음 결합을 기분으로 'CV': 0.1, 'VC': 0.1, 'CVC': 0.2, 'VCV': 0.2, 'CCV': 0.3, 'VCC': 0.3, 'CCVC': 0.4, 'CCVCC': 0.4, 'CCCVCC': 0.5, 'CCCVCCCC': 0.5의 난이도 수치를 추가하여 난이도 위계를 설정하였다.

이후 각 난이도별로 단어들의 음절수와 음운론적 특성을 분석하여 낮은 난이도에서 어려운 난이도로 주어진 단어들을 재배열하였으며, tf-idf값에 기반한 빈도값을 동시에 고려하여 데이터 프레임을 생성하였다.

4) 최소대립쌍 및 문장 생성

최소대립쌍(minimal pair) 생성은 난이도 분석을 통해 분류된 단어를 기반으로 아주 낮은 난이도 그룹과 낮은 난이도 그룹으로 구분된 단어들만을 활용하여 최소대

립쌍 세트를 제작했다. 이 때 생성하는 최소대립쌍은 기준 단어에서 하나의 음절만 다른 단어를 찾아서 세트로 반환하도록 했다.

문장 생성을 위해서는 Open AI의 API를 연결하여 GPT-3와 같은 생성형 AI를 기반으로 기존에 추출한 1000개 단어의 빈도와 음절 난이도를 고려하여 3 단어에서 10단어 사이의 길이의 짧은 문장을 생성하도록 코드를 작성했다. Text-davinci-003 엔진을 활용하였고, 문장 생성에 사용되는 기존에 추출한 1000개 단어의 활용 비율을 최대 0.8로, 그 외 be동사, do동사, 전치사 등의 간단한 추가 단어 리스트의 비율을 0.2로 지정하여 자연스러운 문장을 생성하도록 했다. 최대 1000번의 문장 생성을 시도하도록 하였고, 비문법적인 문장과 중복 문장은 제외하여 총 300개의 문장을 생성하도록 코드를 작성하였다.

문장이 생성된 뒤에는 각 문장을 구성하는 단어에 매핑된 빈도 수치(tf-idf값)와 음절 구조의 난이도의 위계를 바탕으로 각 문장의 난이도 수치를 계산하여 문장을 쉬운 순부터 어려운 순으로 재배열하였다. 낮은 난이도에서는 간단한 음절 구조를 사용한 짧은 문장을, 어려운 난이도로 갈수록 더 복잡한 음절 구조와 긴 길이의 문장을 배치하였다. 이렇게 함으로써, 문장이 발음하기 쉬운 구조에서 점점 어려운 구조로 진행되도록 하였다.

마지막으로, 생성된 최소대립쌍과 문장을 평가하여 각 난이도에 적합한지 확인하였다. 필요한 경우, 문장을 수정하거나 삭제하여 텍스트 자료를 개발했다.

8. 1차 전문가 검토 의견 반영 수정사항

전문가 검토 의견에 따라 아래와 같이 연구를 수정하였다. 우선, 읽기 자료 개발 기준의 타당도를 높이기 텍스트 난이도와 관련된 선행연구를 보완하였다. 이를 통해 텍스트 생성 기준을 더욱 구체화하고, 읽기 자료 개발 단계에 활용되는 주요 개념에 대한 설명을 추가하였다. 단어의 난이도 산출 기준에는 단어의 TF-IDF값에 더하여 문서 내의 출현 빈도, 실생활 활용 빈도 데이터를 추가하였다. 단어의 실생활 활용 빈도 수치 비교를 위하여 Brwon Corpus를 다운받아서 각 단어의 'real_life_frequency'값을 추가했다. Brown Corpus는 500개의 다양한 분야에서 사용되는 약 100만개의 단어로 구성되어 있으며, 미국 영어의 광범위한 사용 예시를 반영하고 있고, 각 텍스트는 품사 태깅(part-of-speech)가 적용되어있는 텍스트 데이터의 집합이다. 단어의 난이도 계산시에는 단어의 음절 수 기준에 더하여 단모음, 단자음, 이중모음, 이중자음의 결합 체계를 구분하여 단어간 상대적인 난이도 수치를 임의로 부여했다. 최소대립쌍 생성시에는 itertools라는 패키지를 활용하여 한 개의 음소가 다르게 발음되는 단어를 생성하였으며, 한 단어에 여러개의 단어가 추천되도록 했다. 문장 생성시에는 전문가 제안 의견에 따라 최신 엔진인 gpt-3.5-turbo를 활용하여 생성된 텍스트를 비교 분석하고자 하였으나, gpt-3.5-turbo활용시 토큰 수

제약으로 문장 생성이 불가능하거나, 일부 생성된 문장 역시 비문법적인 문장이 산출되는 결과가 반복되어 기존의 text-davinci-003엔진을 사용하는 방법을 유지하였다. 결과물로 산출된 빈도가 높은 단어들의 최소대립쌍을 검토하여 중복되거나 최소대립쌍으로 부적절하다고 판단된 단어는 삭제하였다. 생성된 문장은 문장의 평균 길이, 평균 단어 난이도를 합산한 수치에 따라 A, B, C그룹으로 분류하여 각 그룹 내의 문장을 길이 순으로 배열하여 산출하였다. 문장 역시 중복된 문장은 삭제하였다.

읽기 자료(2차)

1. 최소대립쌍

word	minimal_pairs	real_life_frequency	frequency	tf-idf	syllable
did	['dad']	994	0	0.0254989	1
way	['day', 'pay', 'say']	892	15	0.0254989	1
good	['food', 'wood']	767	25	0.0509978	1
men	['pen', 'man']	736	2	0.0254989	1
long	['song']	713	18	0.0509978	1
year	['hear', 'near', 'wear']	649	3	0.0509978	1
right	['light', 'night']	597	11	0.0509978	1
come	['home', 'came']	589	19	0.0509978	1
small	['smell']	518	15	0.0254989	1
went	['tent', 'want']	506	25	0.0254989	1
say	['day', 'pay', 'way']	495	3	0.0254989	1
got	['hot', 'lot', 'pot', 'god']	471	15	0.0254989	1
think	['thank', 'thing']	427	11	0.0509978	1
took	['book', 'cook', 'look']	425	9	0.0254989	1
hand	['band', 'land', 'sand']	418	6	0.0254989	1
told	['cold', 'hold']	411	6	0.0254989	1
far	['car']	409	3	0.0509978	1
set	['let', 'met', 'pet', 'wet']	408	2	0.0254989	1
house	['mouse']	400	18	0.0254989	1
night	['light', 'right']	398	10	0.0254989	1
best	['rest']	341	10	0.0509978	1
thing	['think']	331	2	0.0254989	1
power	['tower']	326	1	0.0254989	1

want	['went']	326	21	0.0509978	1
light	['night', 'right']	320	1	0.0254989	1
big	['pig', 'bag']	316	28	0.0254989	1
let	['met', 'pet', 'set', 'wet']	293	3	0.0509978	1
means	['beans', 'meals']	291	6	0.0254989	1
gave	['save', 'game']	285	9	0.0254989	1
line	['life', 'like', 'live']	282	4	0.0254989	1
car	['far']	270	5	0.0254989	1
tell	['fell', 'sell', 'tall']	245	5	0.0254989	1
wife	['life', 'wise']	227	7	0.0509978	1
love	['live']	214	33	0.0254989	1
run	['fun', 'sun', 'ran']	206	6	0.0254989	1
mean	['meal', 'meat']	199	2	0.0254989	1
land	['band', 'hand', 'sand']	199	6	0.0254989	1
play	['plan']	197	29	0.0254989	1
plan	['play']	196	4	0.0254989	1
cut	['cat', 'cup']	180	2	0.0254989	1
soon	['moon', 'noon']	176	3	0.0509978	1
book	['cook', 'look', 'took']	176	4	0.0254989	1
pay	['day', 'say', 'way']	169	2	0.0254989	1
read	['dead', 'head']	167	5	0.0254989	1
dead	['head']	166	2	0.0254989	1
class	['glass']	164	22	0.0509978	1
hold	['cold', 'told']	162	3	0.0509978	1
cold	['hold', 'told']	161	14	0.0254989	1
rest	['best']	160	4	0.0254989	1
hear	['near', 'year']	151	2	0.0254989	1
sat	['cat', 'hat', 'sit', 'sad']	149	3	0.0254989	1
bad	['dad', 'sad', 'bed', 'bag']	139	6	0.0254989	1
wall	['ball', 'fall', 'tall', 'walk']	139	7	0.0254989	1
fall	['ball', 'tall', 'wall', 'fell']	136	3	0.0254989	1
ran	['fan', 'man', 'run', 'rap']	132	8	0.0254989	1
met	['let', 'pet', 'set', 'wet', 'men']	129	3	0.0254989	1
bed	['fed', 'red']	126	4	0.0254989	1
lot	['got', 'hot', 'pot', 'let']	124	20	0.0254989	1
game	['came', 'gave']	122	19	0.0509978	1
hot	['got', 'lot', 'pot', 'hat', 'hit']	118	14	0.0254989	1
farm	['warm']	112	9	0.0254989	1
hit	['sit', 'hat', 'hot']	111	5	0.0254989	1

stop	['shop']	110	3	0.0254989	1
wish	['dish', 'fish', 'wash']	109	2	0.0254989	1
stay	['star']	107	4	0.0254989	1
sun	['fun', 'run']	102	7	0.0254989	1
race	['rice']	100	10	0.0254989	1
talking	['walking']	99	2	0.0254989	2
walk	['talk', 'wall']	96	8	0.0254989	1
project	['protect']	91	4	0.0254989	2
ship	['shop']	82	13	0.0254989	1
sea	['tea']	80	17	0.0254989	1
nice	['rice']	69	6	0.0509978	1
sit	['hit', 'sat', 'set']	65	10	0.0254989	1
warm	['farm']	65	8	0.0254989	1
grew	['drew']	64	2	0.0254989	1
nose	['note']	60	2	0.0254989	1
save	['gave', 'safe']	60	6	0.0509978	1
eat	['cat', 'hat', 'sat']	59	21	0.0254989	1
safe	['save']	58	2	0.0254989	1
song	['long', 'sang', 'sing']	56	3	0.0509978	1
shop	['stop', 'ship']	56	3	0.0254989	1
tree	['free', 'true']	56	6	0.0509978	1
hat	['cat', 'sat', 'hit', 'hot']	55	2	0.0254989	1
seat	['meat']	54	2	0.0254989	1
text	['tent']	54	2	0.0254989	1
draw	['drew']	53	3	0.0254989	1
tall	['ball', 'fall', 'wall', 'tell']	53	3	0.0254989	1
wet	['let', 'met', 'pet', 'set', 'web']	51	2	0.0254989	1
wood	['food']	51	2	0.0254989	1
moon	['noon', 'soon']	49	3	0.0254989	1
meat	['seat', 'meet', 'meal', 'mean']	45	4	0.0254989	1
ride	['rode', 'rice']	45	4	0.0254989	1
band	['hand', 'land', 'sand']	44	2	0.0254989	1
cup	['cut']	44	4	0.0254989	1
fun	['run', 'sun', 'fan']	43	11	0.0254989	1
bag	['big', 'bad']	42	2	0.0254989	1
fed	['bed', 'red']	40	2	0.0254989	1
seed	['need', 'sees']	40	3	0.0254989	1
sell	['fell', 'tell']	40	3	0.0254989	1
sad	['bad', 'dad', 'sam', 'sat']	34	2	0.0254989	1

wise	['wife']	33	3	0.0509978	1
wash	['wish']	33	5	0.0254989	1
pot	['got', 'hot', 'pet']	28	2	0.0254989	1
sand	['band', 'hand', 'land']	28	3	0.0254989	1
bath	['math']	25	2	0.0254989	1
king	['sing']	25	18	0.0509978	1
noon	['moon', 'soon']	24	2	0.0254989	1
rice	['nice', 'race', 'ride']	24	6	0.0254989	1
cook	['book', 'look', 'took']	24	8	0.0254989	1
tent	['went']	20	1	0.0254989	1
god	['got']	17	3	0.0254989	1
fan	['man', 'ran', 'fun']	17	3	0.0254989	1
dish	['fish', 'wish']	15	5	0.0254989	1
cake	['bake', 'lake', 'make', 'came']	13	6	0.0254989	1
map	['rap', 'man']	12	3	0.0254989	1
swim	['swam']	11	2	0.0254989	1
lake	['bake', 'cake', 'make', 'like', 'late']	11	5	0.0509978	1
fox	['box']	9	1	0.0254989	1
duck	['luck']	9	0	0.0254989	1
pet	['let', 'met', 'set', 'wet]	8	2	0.0254989	1
pig	['big']	8	5	0.0254989	1
web	['wet']	6	2	0.0254989	1
math	['bath']	3	7	0.0254989	1
dad	['bad', 'sad', 'did']	2	2	0.0509978	1
rap	['map', 'ran']	2	3	0.0254989	1
bake	['cake', 'lake', 'make']	2	3	0.0254989	1
grandma	['grandpa']	1	0	0.0254989	1

2. 문장

<구성 단계>

1) 문장 생성: OpenAI의 GPT-3를 활용하여 "2에서 10단어 이내의 짧은 영어 문장을 생성하라"는 프롬프트로 총 300개의 문법적인 문장이 생성될 때까지 또는 최대 1,000번 시도할 때까지 반복하였다.

2) 난이도 점수 부여: 생성된 각 문장에 대해, 문장 구조, 단어의 개수, 단어의 길이, 단어의 실생활 빈도, 단어의 난이도, 과거 동사의 유무의 기준에 따라 난이도 점수를 수치화 했다. 또한, spacy라는 NLP 라이브러리를 사용하여 문장 구조를 분석하고, 과거 동사의 유무를 파악하였다. Brown corpus를 기준으로 산출한

real_life_frequency 수치를 반영할 때에는 실제 생활에서 빈번하게 사용하는 단어는 보다 쉬운 단어로 간주하기 위하여 난이도 산출시 해당 수치를 감하도록 조정하여 일부 문장의 difficulty가 음수로 산출되었다.

3) 난이도에 따른 정렬: 계산된 난이도 점수와 문장의 길이를 기준으로 문장을 정렬했다.

4) 정규화 및 카테고리 분류: 문장을 구성하는 단어들의 난이도 점수의 합과 문장 길이 수치를 5:5로 반영하고 정규화하여 새로운 문장 난이도 점수를 계산하고, 이를 기준으로 문장을 상, 중, 하 3가지 카테고리로 분류했다.

sentence	difficulty	length	normalized_ score	category	contains _present _verb
I win!	10	6	0.498228663	하	TRUE
I smile.	10	8	0.560728663	하	TRUE
I'm here.	10	9	0.591978663	하	TRUE
I'm late.	13	9	0.592461755	하	TRUE
I'm home.	13	9	0.592461755	하	TRUE
Good day!	7	9	0.591495572	하	FALSE
Have fun!	7	9	0.591495572	하	FALSE
Run fast!	7	9	0.591495572	하	FALSE
Let's go!	10	9	0.591978663	하	FALSE
I smiled.	12	9	0.592300725	하	FALSE
Thank you.	7	10	0.622745572	하	TRUE
Good luck!	7	10	0.622745572	하	FALSE
Happy day!	7	10	0.622745572	하	FALSE
Eat lunch.	7	10	0.622745572	하	FALSE
Sunny day.	7	10	0.622745572	하	FALSE
Let's eat.	10	10	0.623228663	하	FALSE
I like you.	-2442	11	0.259631643	하	TRUE
I love you.	-396	11	0.589100242	하	TRUE
I like snow.	-2442	12	0.290881643	하	TRUE
I like dogs.	-2442	12	0.290881643	하	TRUE
I like cake.	-2442	12	0.290881643	하	TRUE
I like cats.	-2442	12	0.290881643	하	TRUE
I like blue.	-2442	12	0.290881643	하	TRUE
I love dogs.	-396	12	0.620350242	하	TRUE
I love life.	-396	12	0.620350242	하	TRUE

I love cats.	-396	12	0.620350242	하	TRUE
I love food.	-396	12	0.620350242	하	TRUE
I like books.	-2442	13	0.322131643	하	TRUE
I like pizza.	-2442	13	0.322131643	하	TRUE
I like bikes.	-2442	13	0.322131643	하	TRUE
I like tacos.	-2442	13	0.322131643	하	TRUE
I like cheese.	-2442	14	0.353381643	하	TRUE
I like coffee.	-2442	14	0.353381643	하	TRUE
I like apples.	-2442	14	0.353381643	하	TRUE
I like cookies.	-2442	15	0.384631643	하	TRUE
I like turtles.	-2442	15	0.384631643	하	TRUE
I like reading.	-2442	15	0.384631643	하	TRUE
I like football.	-2442	16	0.415881643	하	TRUE
Have a good day!	-1506	16	0.56660628	하	FALSE
I like ice cream.	-2514	17	0.43553744	하	TRUE
We go now.	12	10	0.623550725	중	TRUE
I'm happy.	13	10	0.623711755	중	TRUE
I'm sorry.	13	10	0.623711755	중	TRUE
I'm tired.	13	10	0.623711755	중	TRUE
I'm ready.	13	10	0.623711755	중	TRUE
I'm happy!	13	10	0.623711755	중	TRUE
I'm Ready.	13	10	0.623711755	중	TRUE
He smiled.	12	10	0.623550725	중	FALSE
I'm hungry.	13	11	0.654961755	중	TRUE
It's sunny.	13	11	0.654961755	중	TRUE
Lovely day!	7	11	0.653995572	중	FALSE
Enjoy life.	7	11	0.653995572	중	FALSE
Eat dinner.	7	11	0.653995572	중	FALSE
Let's play.	10	11	0.654478663	중	FALSE
We laughed.	12	11	0.654800725	중	FALSE
I can help.	12	11	0.654800725	중	FALSE
I'm smiling.	10	12	0.685728663	중	TRUE
I'm excited.	13	12	0.686211755	중	TRUE
I see stars.	15	12	0.686533816	중	TRUE
I ate lunch.	-3	12	0.683635266	중	FALSE
Hello World!	7	12	0.685245572	중	FALSE
Eat healthy.	7	12	0.685245572	중	FALSE

Stay strong.	7	12	0.685245572	중	FALSE
Take a walk.	9	12	0.685567633	중	FALSE
I love books.	-396	13	0.651600242	중	TRUE
I love pizza.	-396	13	0.651600242	중	TRUE
Welcome home!	7	13	0.716495572	중	TRUE
I ate dinner.	-3	13	0.714885266	중	FALSE
Happy Friday!	7	13	0.716495572	중	FALSE
Good morning!	7	13	0.716495572	중	FALSE
Good evening.	7	13	0.716495572	중	FALSE
Take chances.	7	13	0.716495572	중	FALSE
Come quickly.	7	13	0.716495572	중	FALSE
Love forever.	7	13	0.716495572	중	FALSE
Happy Monday!	7	13	0.716495572	중	FALSE
Take a break.	9	13	0.716817633	중	FALSE
Take it easy.	9	13	0.716817633	중	FALSE
Take me home.	9	13	0.716817633	중	FALSE
See you soon.	9	13	0.716817633	중	FALSE
Walk the dog.	9	13	0.716817633	중	FALSE
I love summer.	-396	14	0.682850242	중	TRUE
I love coffee.	-396	14	0.682850242	중	TRUE
I love spring.	-396	14	0.682850242	중	TRUE
It's raining.	10	13	0.716978663	상	TRUE
I missed you.	17	13	0.718105878	상	FALSE
I smile often.	12	14	0.748550725	상	TRUE
The cat slept.	-12	14	0.74468599	상	FALSE
Eat breakfast.	7	14	0.747745572	상	FALSE
Eat healthily.	7	14	0.747745572	상	FALSE
Happy Tuesday!	7	14	0.747745572	상	FALSE
Stay positive.	7	14	0.747745572	상	FALSE
Take a chance.	9	14	0.748067633	상	FALSE
Open the door.	9	14	0.748067633	상	FALSE
The sun shines.	-174	15	0.749849034	상	TRUE
That's amazing!	13	15	0.779961755	상	TRUE
The cat jumped.	-12	15	0.77593599	상	FALSE
Happy birthday!	7	15	0.778995572	상	FALSE
Hello everyone!	7	15	0.778995572	상	FALSE
Happy holidays!	7	15	0.778995572	상	FALSE

Good afternoon.	7	15	0.778995572	상	FALSE
Eat more fruit.	9	15	0.779317633	상	FALSE
I love learning.	-396	16	0.745350242	상	TRUE
I love chocolate.	-396	17	0.776600242	상	TRUE
I smiled happily.	-104	17	0.823621176	상	FALSE
Love one another.	9	17	0.841817633	상	FALSE
Enjoy the moment.	9	17	0.841817633	상	FALSE
Sun shines bright.	15	18	0.874033816	상	TRUE
Carrots taste good.	-80	19	0.88998591	상	TRUE
I listen carefully.	-32	19	0.897715378	상	TRUE
Nice weather today!	-97	19	0.88724839	상	FALSE

▪ 작성 코드

https://colab.research.google.com/drive/1levUkEKr-GYyaH7epdyC_eQgO9YROyDJ?usp=sharing

[부록5] 최종 개발 코드

https://github.com/Eungzz/public.git