

AI 도구를 활용한 중등교사 영어 출제 연수 사례 연구: ChatGPT를 중심으로

신동광[†]
광주교육대학교

A Case Study on English Test Item Development Training for Secondary School Teachers Using AI Tools: Focusing on ChatGPT

Dongkwang Shin [†]
Gwangju National University of Education

ABSTRACT

This study analyzed the use of various AI tools (e.g., ChatGPT) by ten teachers who participated in a secondary teacher training program hosted by University A in January 2023. The training program was conducted over three days, and after learning how to use AI tools, the teachers were asked to apply the AI tools in the actual practice of test item development, particularly for reading test items in the College Scholastic Ability Test (CSAT). The study divided the test item development process into five stages: 1) passage creation; 2) test item creation; 3) test item selection; 4) vocabulary level control; and 5) refining details such as styling, choice length, and so on. At each stage, the cases showed that AI tools could significantly contribute to improving the efficiency of developing test items. The teachers also responded very positively to the training program, and the conclusion section discussed future challenges and further research.

Keywords: generative AI, test item development, ChatGPT, ATG, AIG, teacher training

1. 서 론

최근 OpenAI에서 개발하여 공개한 챗봇형 인공지능(AI)인 ChatGPT(Chat Generative Pretrained Transformer)에 관한 관심이 폭발적이다. 2022년 11월 30일에 첫 출시된 후 5일 만에 100만 명, 그리고 2주일 만에 200만 명의 사용자를 기록하였고

[†] Corresponding author: sdhera@gmail.com



지금까지 100명의 사용자를 최단기간 달성한 기록은 페이스북(현 메타)으로 10개월이 소요되었고 이 기록과 비교하면 ChatGPT가 국내외에 광풍을 일으키고 있다고 해도 과언이 아니다(Kim, T., 2023).

사실 ChatGPT는 검색엔진과 유사하다. 하지만 생성형 AI인 ChatGPT는 기존의 검색방식과 차별화된 서비스를 제공한다. 먼저 검색엔진은 키워드를 기반으로 정보를 검색하며 입력한 키워드가 포함된 많은 정보를 찾아 제시한다. 반면 ChatGPT는 질문의 맥락을 이해하여 마치 요청한 정보를 가지고 있는 인간이 그에 대해 설명하듯 사용자가 원하는 정보만을 찾아 제시한다는 점에서 추가적인 정보검색을 요구하지 않는다는 점에서 편의성이 탁월하다. 또한, 정보가 정확하지 않을 때는 역시 인간과 상호작용하듯 조건을 추가하여 재요구할 수 있다는 점은 상호작용을 통한 친근감과 정보습득의 효율성에 모두 기여할 수 있다(Lee & Shin, 2020). 무엇보다도 기존 검색엔진과 비교하여 ChatGPT가 가진 가장 큰 차별성은 콘텐츠 생성에 있다. 초거대언어모델(Large Language Model, LLM)인 GPT-3.5에 기반한 ChatGPT는 생성할 콘텐츠의 조건을 입력하면 소설, 기사, 대화문, 에세이, 블로그, 계약서 등 다양한 장르와 유형의 글을 작성할 수 있다. 뿐만 아니라 파이썬(Python)과 같은 프로그래밍 코드의 생성 및 검토와 상담, 여행 계획 수립, 평가 문항생성까지 지원할 수 있다. 이는 현재 AI 기술이 인간의 단순한 업무를 대체하는 수준이라고 믿던 통념을 깨고 고도의 전문성을 요하는 수준까지 지원 또는 대체할 수 있음을 의미하기 때문에 ChatGPT에 대한 관심이 그만큼 뜨거울 수밖에 없다. 결국 이와 같은 AI 기술의 진화는 교육을 포함하여 사회 전반에 큰 변화를 가져올 수밖에 없을 것이다. 즉 정보를 가지는 것이 경쟁력이었던 지식기반 사회에서는 가장 중요한 교육의 목표 또한 지식암기에 있었지만 이제는 더 이상 전화번호를 암기할 필요가 없듯 지식에 대한 관점도 암기(기억)가 아닌 응용과 재창조의 관점으로 변화하고 있는 것이다.

교육 분야에 한정하여 본다면 교사는 AI 도구를 활용하여 교육의 준비, 수행, 평가, 관리라는 업무를 보다 효율적으로 수행할 수 있고 이러한 교육의 효율성 제고에 힘입어 학생과의 상호작용을 통한 심층적인 교육활동에 보다 집중할 수 있을 것이다. 물론 최근 학교의 과제 제출 시 ChatGPT를 활용한 부정행위가 사회적 이슈로 대두되는 등 부작용도 보고되고 있지만 AI 기술의 일상화 또는 대중화(Artificial General Intelligence, AGI)는 이제 거스를 수 없는 대세가 되리라는 전망이 우세하다(Simon, 2022). 본 연구는 기본적으로 AI 기술을 긍정적인 측면에서 교육에 접목하고자 하는 시도이며 이에 중등 영어 교사들을 대상으로 시행된 출제 연수 프로그램을 중심으로 AI 기술이 실제 영어 문항 출제에 어떻게 활용될 수 있고 어떠한 장점이 있는지 사례 중심으로 살펴보고자 한다.

2. 이론적 배경 및 선행연구

Sireci와 Zenisky(2006)에 따르면 인터넷 기반 컴퓨터 평가는 기존의 지필 평가에 비

해 많은 장점이 있다. 이러한 장점에는 디지털 방식의 혁신적인 문항유형의 적용, 보다 복잡한 지식과 그와 관련된 역량 측정(Bartram & Hambleton, 2005), 대규모 자동채점 (Shermis & Burstein, 2013), 학생에 대한 즉각적인 피드백 제공(Attali & Powers, 2010), 그리고 적응형/맞춤형(adaptive) 평가(van der Linden & Glas, 2010) 등의 기술을 접목할 수 있다는 점이 포함된다. 하지만 이러한 인터넷 기반 컴퓨터 평가가 실제 시행되기 위해서는 대규모 문제은행 구축이라는 전제가 따르고 이에 따라 더 많은 양의 문항을 개발해야 하는 부담이 발생하게 되었다(Downing & Haladyna, 2006). 이에 연구자들은 자연스레 자동문항생성(Automated Item Generation, AIG) 기술로 눈을 돌리게 되었다. 기존의 자동문항생성 기술은 다양한 콘텐츠 영역과 형식의 아이템을 생성하는 데 사용되기는 하였지만 비교적 모델링하기 쉬운 콘텐츠 영역에 국한되었다(Gierl & Haladyna, 2013).

Kurdi et al.(2020)에 따르면 자동문항생성 방식은 크게 템플릿 기반 방식, 규칙 기반 방식, 그리고 통계 기반 방식이 있다. 먼저, 템플릿 기반 방식에서 ‘템플릿’은 ‘모델’과 같은 의미로 가장 일반적으로 사용된다. 이 방식은 전형적인 문항의 포맷에 따라 문항을 생성하는 방식이다. 즉, 문항에 필요한 정보는 템플릿에 미리 정해져 있으며, 그 정보를 채워넣는 것으로 문항을 생성한다. 그러나 이 방식은 문항유형이 매우 한정적이며, 새로운 유형을 추가하거나 문항을 수정할 때마다 템플릿을 다시 수정해야 하는 번거로움이 있다. 규칙 기반 방식은 구문 또는 의미론적 언어자질을 활용하여 문장에 주석(annotation)을 단 다음, 문항유형별 적절한 규칙을 사용하여 문항을 생성하는 방식이다. 이 방식도 템플릿 기반 방식과 마찬가지로 문항유형이 매우 한정적이며, 새로운 유형을 추가하거나 문항을 수정할 때마다 규칙을 다시 설계해야 하는 한계가 있다. 반면에, 통계 기반 방식은 기존의 두 가지 방식과 차별화된다. 이 방식에서는 문항생성을 위한 규칙을 정교화된 언어모델을 통해 학습시킨다는 점이 특징이다. 딥러닝과 같은 기계학습 기술을 이용해 많은 양의 문항 데이터를 학습하면서, 문항생성을 위한 패턴과 규칙을 찾아내어 자동으로 문항을 생성할 수 있다. 이 방식은 문항유형이나 포맷에 대한 제약이 적으며, 새로운 유형의 문항을 생성하는 것도 쉽다. 또한, 통계 기반 방식은 템플릿 기반 방식과 규칙 기반 방식보다 자연스러운 문항을 생성할 수 있다는 장점이 있다. 하지만, 통계 기반 방식도 문항 데이터의 양이 충분하지 않으면 언어모델이 제대로 학습되지 않을 수 있고, 데이터에 있는 패턴만을 따라 문항을 생성하게 되므로, 문항의 다양성이나 새로운 아이디어의 도출이 어려울 수 있다. 또한, 언어모델이 학습한 데이터에 편향성이 존재할 수 있어서, 다양한 문화나 배경의 학생들에게는 적절하지 않은 문항이 생성될 수도 있다는 제한점이 있다.

그럼에도 언어모델링은 Ney et al.(1994)의 초기 연구를 시작으로 수년 동안 자연어 처리 연구의 주요 초점이 되어 왔다. 그리고 Radford et al.(2019)의 연구와 같이 트랜스포머(Transformer) 네트워크 방식은 최근의 혁신적인 문항생성 기술 발전을 잘 보여준다. 트랜스포머 방식은 이전의 신경망(Neural Network) 방식과 비교하여 문장 구성

시 많은 맥락적 정보를 활용하여 특정 단어 다음에 올 단어를 보다 정확하게 예측하기 때문에 더 긴 맥락에서 자연스러운 텍스트 산출을 기대할 수 있다. OpenAI의 GPT-3 모델 또한 트랜스포머 방식을 따르고 있고 GPT-3은 주석 태깅, 모델 개발, 학습 과정과 같은 명시적인 모델 업데이트 없이 몇 가지 샘플 모델만으로도 새롭고 일관된 콘텐츠를 생성할 수 있다(Brown et al., 2020). GPT-3는 열거형 목록, 입력된 텍스트의 포맷과 스타일, 이해도 평가 문항, 대칭 구조의 문단 심지어 구조화된 HTML 코드까지 모방하여 유사한 결과물을 산출할 수 있다.

자동문항생성의 활용 사례를 살펴보면, von Davier(2018)는 장단기기억(LSTM) 네트워크라고 하는 일종의 순환신경망(RNN) 방식 딥러닝 기법을 활용하여 감정, 태도, 행동 및 성격 특성과 같은 비인지적인 측면을 측정하는 문항생성에 성공하였고, Hommel et al.(2021)은 트랜스포머 기반 모델을 사용하여 유사한 문항을 생성하였다. Kumar et al.(2018)은 LSTM 모델을 기반으로 입력한 문장을 분석하여 이에 관한 질문-답변 쌍을 생성하기도 하였다. Khan et al.(2020)이 개발한 스피크스(Sphinx)라는 프로그램은 읽기 텍스트를 자동으로 생성하는 하이브리드 시스템으로서 간단한 문법 문항은 자동으로 생성하고 독해력 측정과 같이 다소 복잡한 문항은 인간 전문가가 담당하는 방식을 채택하기도 하였다. 국내에서는 Shin(2022)의 연구에서 CopyAI라는 자동지문 생성기와 ATM이라는 자동문항생성기를 활용하여 문항을 생성해 봄으로써 영어 평가에서 이 두 프로그램의 활용가능성을 탐색하였다. 하지만 자동문항생성이라는 분야는 아직까지 관련 상용 프로그램의 개발과 활용에 관한 많은 연구가 필요한 미개척 분야라고 할 수 있다.

3. 연구방법

3.1. 연구대상

본 연구는 A 대학의 주관으로 2023년 1월 중에 실시한 중등(중학교, 고등학교)교사 출제 연수 프로그램에 참가한 10명의 연수 참가 교사들을 대상으로 진행되었다. 교사들의 교육경력은 대부분은 5년 이내의 교사들이었고 교육경력이 15-20년인 교사도 2명 포함되어 있었다. 연수는 3일간 진행되었고 1일차의 경우 출제 및 평가 이론 강의, 2일과 3일차는 실제 출제 실습으로 구성되었고 본 연구는 2, 3일차의 출제 실습 과정 중 수능 유형의 읽기 문항 출제를 위해 연수 참가 교사들이 활용한 다양한 AI 도구들의 적용 사례 분석에 초점을 두었다.

3.2. 문항 개발용 AI 도구

본 연구는 앞서 언급한 바와 같이 연수 참가 교사 개인별로 수능 읽기 문항 중 하나를

문항유형을 선택하여 한 개 문항을 출제하고 수정하여 완성하는 과정을 분석하였다. 또한, 본 연구에서는 기존의 출제 관련 연구자료와 연구물(예, Gyeongnam Office of Education, 2015; National Board of Educational Evaluation, 1992; Ulsan Metropolitan Office of Education, 2014)을 바탕으로 다음의 표 1과 같이 5개의 문항 개발 단계를 나누어 연수 참가 교사들에게 AI 도구를 소개하고 활용하도록 하였다.

표 1. 문항 개발 단계별 활용가능한 AI 도구

문항 개발 단계	AI 도구	특징
1. 지문생성	ChatGPT	https://chat.openai.com/ 웹버전, 무료, Plus 버전은 유료
	CopyAI	https://www.copy.ai/ 웹버전, 무료+유료
	YouWrite	https://you.com/search?q=how+to+write+well&&tbm=youwrite&cf=write 웹버전, 무료
2. 문항생성	ATM	https://atm.lxper.ai/ 웹버전, 무료+유료
	ChatGPT	https://chat.openai.com/ 웹버전, 무료, Plus 버전은 유료
3. 문항선제	Google Translate	https://translate.google.com/ 웹버전+앱버전 무료
4. 어휘통제	AntWordProfiler	https://www.laurenceanthony.net/software/antwordprofiler/releases/AntWordProfiler140/AntWordProfiler.exe exe 실행 프로그램
	ChatGPT	https://chat.openai.com/ 웹버전, 무료, Plus 버전은 유료
	CopyAI	https://www.copy.ai/ 웹버전, 무료+유료
5. 문체, 선택지 길이 등 세부 수정	Text Helper	https://www.er-central.com/text-helper/ 웹버전, 무료
	ChatGPT	https://chat.openai.com/ 웹버전
	CopyAI	https://www.copy.ai/ 웹버전, 무료+유료

ChatGPT는 2022년 11월에 OpenAI가 공개한 대화형 인공지능 챗봇으로, GPT-3 (Generative Pretrained Transformer-3)의 개선판인 GPT-3.5를 기반으로 개발되었다. ChatGPT는 먼저 AI 어시스턴트 역할을 한 데모 데이터로 학습하고, 그 다음 해당 데이터에 대한 모델의 산출물에 대한 만족도를 인간이 순위를 매겨 다시 학습하였으며, 강화 학습이라는 최적화 과정을 통해 최대한 인간과 가까운 답변을 산출하도록 개발되었다. 이러한 사전 학습과정을 거쳐 ChatGPT는 다양한 지식 분야에서 상세한 응답과 정교한

답변이 가능하게 되었다(Shin, Jung, & Lee, 2023). ChatGPT는 다양한 생성 기능을 가지고 있지만 특히 지문 생성의 조건을 입력하면 여러 장르의 글을 생성할 수 있다는 것이 장점이다. 이러한 이유로 ChatGPT, 그리고 GPT-3에 기반하여 개발된 CopyAI (Yacoubian, 2020)와 YouWrite(McCann & Socher, 2022)를 1단계의 지문생성에 활용하도록 소개하였다. CopyAI와 YouWrite는 텍스트 생성에 최적화된 모델(AI Content Generator)로 활용이 쉽다. 하지만 CopyAI의 경우는 유료이고 YouWrite 또한 텍스트 생성의 포맷이 정해져 있는 만큼 사용자에게 편리성을 제공하지만 동시에 지시문 (Prompt) 입력의 유연성에는 제약이 따른다.

2단계에서는 AIG(Automated Item Generation) 기능을 포함한 AI 도구들로 구성하였다. 먼저 ATM(AI Test Maker)은 LXPOR Inc.(Lee, 2018)가 서비스하고 있는 인공지능 기반 영어 학습 플랫폼의 기능 중 하나이며 ATM은 자연어 처리를 기반으로 하는 자동문항생성을 지원한다. 특히 수능 영어 읽기 영역에서 출제되는 9개 유형의 문항(예, 빈칸 추론, 주제 추론, 문장 배열 등)에 기반한 문항생성을 지원한다. 또한, ChatGPT는 지문생성후 선택지(multiple choice)의 개수 및 문항유형을 기술하여 문항생성을 요구하면 문항을 즉시 생성해 준다. 이때 중요한 점은 지시문의 구체성과 명료성이다. 대화형 챗봇인 ChatGPT는 지시문의 형식과 질이 산출물의 질에 큰 영향을 미친다.

3단계에서는 출제문항 초안을 가지고 수정단계로 넘어갈 것인지 아니면 탈락시켜 재출제할 것인지를 결정하는 선제과정을 거쳤다. 선제과정에서 사용한 AI 도구는 번역기였다. 번역기의 2개 창 중 왼쪽 입력창에는 영문 문항을, 오른쪽은 출력창에는 한글번역본을 동시에 제시하고 전체 연수 참가자들이 참여하여 문항선제를 위한 토론을 진행하였다. 번역기를 활용한 이유는 내용 파악의 시간을 줄이고 전문용어 등 번역기의 번역이 어색하다면 실제 비영어권인 수험자들도 동일한 오역으로 인해 어려움을 겪을 수도 있기 때문에 번역이 어색한 부분을 중점적으로 검토하였다. 이 단계에서 사용한 번역기는 2016년 신경망 번역기로 업그레이드된 Google Translate(Wikipedia_a, n.d.)인데 Google Translate보다 더 정확한 DeepL 번역기¹⁾가 있음에도 불구하고 Google Translate를 사용한 이유는 너무 자연스러운 번역물을 산출하면 앞서 언급한 수험자의 수준에서 지문의 특성을 파악하기 어렵다는 점을 고려하여 적절한 수준의 번역 오류를 포함한 Google Translate를 활용하였다.

4단계에서는 시험에서 지정한 특정 어휘수준으로 어휘를 통제하는 과정으로 과거 수능 영어 문항의 출제에서는 수능 어휘목록으로 지문의 어휘수준을 통제하였다. 하지만 현재는 원문의 취지를 최대한 유지하기 위해 수능 어휘수준을 벗어난 단어는 주석으로 처리하고 있다. 본 연구의 출제 연수에서는 모의 수능 문항 출제를 목표로 하였지만

1) DeepL 번역기는 2017년 DeepL SE가 출시한 신경망 기계 번역 서비스로 2023년 한국어 버전을 추가하여 현재 31개 언어에 대한 서비스를 제공하고 있다. 웹버전은 물론 앱버전도 제공되고 있으며 번역 후 사용자가 직접 단어/구/문장을 쉽게 수정 또는 교체할 수 있도록 다양한 대안 표현을 제공하며 주석을 추가할 수도 있다(Wikipedia_b, n.d.)는 점에서 교수학습의 측면에서는 더 유용할 것으로 판단된다.

현재 수능을 제외한 대부분의 국가단위 영어 시험에서는 어휘수준을 통제하고 있다는 만큼, 이번 출제 연수에서는 Nation(2016)이 개발한 BNC/COCA 25000에서 상위 4,000개의 어휘족(word family)을 어휘 상한선으로 하여 어휘를 통제하고자 하였다. 어휘통제에는 어휘목록을 탑재하여 지문에서 어휘목록을 벗어난 항목만 하이라이트해주는 AntWordProfiler(Anthony, 2013)와 특정 어휘수준을 벗어난 항목에 대하여 자동으로 영문 주석을 제공해 주는 Text Helper(Extensive Reading Central, n.d.)를 소개하였다. ChatGPT에서도 ‘easy/advanced/sophisticated words’란 용어를 사용하여 ‘paraphrase’를 요구하거나 CopyAI의 ‘Sentence Rewrite’ 기능을 활용해도 정확한 어휘수준은 아니더라도 다른 대체 표현을 추천받을 수 있기 때문에 이 두 프로그램 또한 소개하였다.

끝으로 5단계에서는 문체, 선택지 길이 등과 같은 세부수정에 활용할 수 있는 AI 도구로 CopyAI와 ChatGPT를 소개하였다. CopyAI의 ‘Tone Changer’ 기능을 활용하면 지문 전체의 문체를 friendly/luxury/witty/professional 등으로 일괄 교체할 수 있으며 CopyAI의 작업창에 지문을 붙여넣고 특정 문단에서 ‘Tab’ 버튼을 누르면 글을 확장해서 작성할 수도 있다.

4. 연구결과

연구의 목적에서도 밝힌 바와 본 연구에서는 앞서 정리한 5단계의 문항 개발 절차에 따라 AI 도구의 활용 사례를 분석하였다. 연수 참가 교사들은 ChatGPT 지시문 작성 요령 등 AI 도구의 기본적인 사용법에 대한 교육을 받고 다양한 수능 영어 읽기 문항생성을 시도하였다.

4.1. 지문생성

ChatGPT의 경우 한글을 인식할 수 있지만 한글의 경우 데이터 처리 속도가 늦고 아직은 데이터가 부족하여 정확하지 않은 표현들이 나타날 수 있어 출제 연수에서는 영문 지시문(Prompt)의 사용 방법만 소개하였다. ChatGPT는 선다형 문항(Multiple Choice Question)을 MCQ로 인식하는 것을 고려하여 교사들은 ‘climate change(기후 변화)’를 주제로 하여 ‘글의 목적 파악’ 유형의 문항 개발을 한다고 가정할 때 다음과 같은 지시문을 ChatGPT의 대화창에 입력하였다.

“Make a new passage about recent ‘climate change’ and make a MCQ (including 5 choices) to test the understanding of the purpose of the passage”

 Climate change has become one of the most significant global issues of our time, with its impacts being felt in different parts of the world. Recent studies have shown that the earth's average temperature has risen by about 1.1 degrees Celsius since the pre-industrial era. This warming trend is attributed to the increasing concentrations of greenhouse gases in the atmosphere, primarily carbon dioxide, emitted by human activities such as burning fossil fuels, deforestation, and industrial processes.

The effects of climate change are being experienced in different ways, including extreme weather events, rising sea levels, ocean acidification, and melting glaciers. These changes have led to environmental, social, and economic consequences, such as food and water insecurity, displacement of communities, and increased economic costs.

Efforts are being made to address climate change, including mitigation measures such as reducing greenhouse gas emissions and adaptation measures such as developing resilient infrastructure and technologies. The United Nations Framework Convention on Climate Change (UNFCCC) and the Paris Agreement are two international treaties aimed at addressing the challenges of climate change.

MCQ: What is the purpose of this passage?
 A. To describe the effects of climate change
 B. To highlight the causes of climate change
 C. To discuss efforts being made to address climate change
 D. To criticize the UNFCCC and Paris Agreement
 E. To provide a historical perspective on climate change

그림 1. ChatGPT를 활용한 지문 및 문항생성 사례(교사4)

ChatGPT는 문항생성 시 정답을 주기도 하고 주지 않는 때도 있는데 정답에 대한 질문을 하면 정답의 근거로 다음과 같이 제시하였다.

“The answer is C. To discuss efforts being made to address climate change. The passage discusses recent climate change and its impacts, but the main focus is on the measures being taken to mitigate and adapt to it. The passage mentions the United Nations Framework Convention on Climate Change (UNFCCC) and the Paris Agreement as international treaties aimed at addressing the challenges of climate change.”

여러 테스트를 확인한 결과, ChatGPT는 특정 어휘수준으로 지문을 통제할 수는 없었고 지문의 길이를 단어 수로 제한하고자 할 때도 정확한 통제는 되지 않았다. 한 가지 중요한 발견 및 시사점 중 하나는 지시문에서 지문과 문항을 동시에 생성해 달라고 지시하기보다는 지문을 먼저 생성하고 이후 문항 개발을 지시하는 할 때 ChatGPT의 산출물은 보다 사용자의 의도에 가깝게 나타났다. 또한, 지문의 주제를 제시하고 지문을 생성하는 대신 수능 지문과 같은 예시 지문을 함께 입력하면 기존 문항의 지문 포맷에 맞게 지문을 생성할 수 있었다.

<p>수능 문항 모델 지문</p>	<p>27. Green Tea Packaging Design Competition에 관한 다음 안내문의 내용과 일치하지 않는 것은?</p> <div style="border: 1px solid black; padding: 5px; margin: 10px auto; width: 80%;"> <p style="text-align: center;">Green Tea Packaging Design Competition</p> <p>Take the opportunity to design the packaging box for brand-new green tea products of TIIS Tea in the competition!</p> <p>Deadline: December 2, 2019, 6:00 p.m. Participants: Lokota County residents only</p> <p>Details</p> <ul style="list-style-type: none"> • Our company name "TIIS Tea" should appear on the design. • The competition theme is "Go Green with Green Tea." • Entries (JPG format only) should be submitted by email to designmanager@tiisteacom. <p>Evaluation Criteria</p> <ul style="list-style-type: none"> • Functionality • Creativity • Eco-friendliness <p>Awards</p> <ul style="list-style-type: none"> • 1st place: \$1,000 • 2nd place: \$500 • 3rd place: \$250 (The first-place winner's signature will be printed on the packaging box.) <p>Please visit www.tiisteacom to learn more about the competition.</p> </div> <ol style="list-style-type: none"> ① 신제품 녹차를 위한 포장 상자 디자인 대회이다. ② Lokota County 주민들만 참가할 수 있다. ③ 출품작은 직접 방문하여 제출해야 한다. ④ 평가 기준에 창의성이 포함된다. ⑤ 1등 수상자의 서명이 포장 상자에 인쇄될 것이다. 	
<p>지시문</p>	<p>A. Make a similar passage with a similar topic as follows: Green Tea Packaging Design Competition~(위 수능문항의 지문 삽입)</p>	<p>B. Make a new passage with a different topic as follows: Green Tea Packaging Design Competition~(위 수능문항의 지문 삽입)</p>
<p>ChatGPT 생성 지문</p>	<p>Organic Coffee Packaging Design Contest Design the packaging for our new line of organic coffee products and showcase your creativity and skills! Deadline: August 31, 2023, 11:59 p.m. Participants: Open to all residents of the United States</p> <p>Details</p> <ul style="list-style-type: none"> - Our company name "Organic Java Co." should be included in the design - The competition theme is "Sustainability and Natural Goodness." - Entries (PNG or PDF format) must be submitted via email to designcontest@organicjava.com. <p>Evaluation Criteria</p> <ul style="list-style-type: none"> - Overall appeal and originality of the design - Creativity and innovation - Consistency with the theme <p>Awards</p> <ul style="list-style-type: none"> - First place: \$1,500 and the winning design will be printed on our new line of organic coffee products. - Second place: \$750 - Third place: \$500 <p>Visit www.organicjava.com for more details about the competition and our products.</p>	<p>Healthy Recipe Contest Share your favorite healthy recipe and enter the contest to win exciting prizes and recognition! Deadline: September 30, 2023, 11:59 p.m. Participants: Open to all residents of the United States</p> <p>Details</p> <ul style="list-style-type: none"> - Your recipe should be original, healthy, and easily replicable - Entries should be emailed to recipe@healthfoodcontest.com in a Word document - Please include your name, address, and contact information in the email <p>Evaluation Criteria</p> <ul style="list-style-type: none"> - Nutritional value of the recipe - Creativity and originality - Clarity of instructions <p>Awards</p> <ul style="list-style-type: none"> - First place: \$1,000 and a feature of the winning recipe on our website and social media channels. - Second place: \$500 - Third place: \$250 <p>Visit www.healthfoodcontest.com to learn more about the competition and for some inspiration for your healthy recipe.</p>

그림 2. ChatGPT를 활용한 수능 영어지문 생성 사례(교사2)

위의 그림 2와 같이 두 가지 방식의 지시문을 선택적으로 사용하면 보다 효과적으로 목적에 부합한 수능 영어 지문을 생성할 수 있다. 먼저 수능 기출 문항의 지문과 “a similar topic”으로 생성해 달라고 요청하면 동일한 주제에 대하여 기존 지문을 마치 ‘paraphrase’하듯 생성할 수 있어 주제만 연계하는 EBS 연계 문항을 개발할 때 유용하게 활용될 수 있을 것이라 판단된다. 반면 “a new passage with a different topic”이란 문구를 활용하면 다른 주제에 대하여 수능 문항에 사용된 동일한 포맷의 지문을 생성할

수 있다. ChatGPT에게 최근의 분야별 키워드를 요청하고 이 주제를 활용하여 지문을 생성하면 최근의 사회적 경향성을 반영한 지문생성도 가능하다.

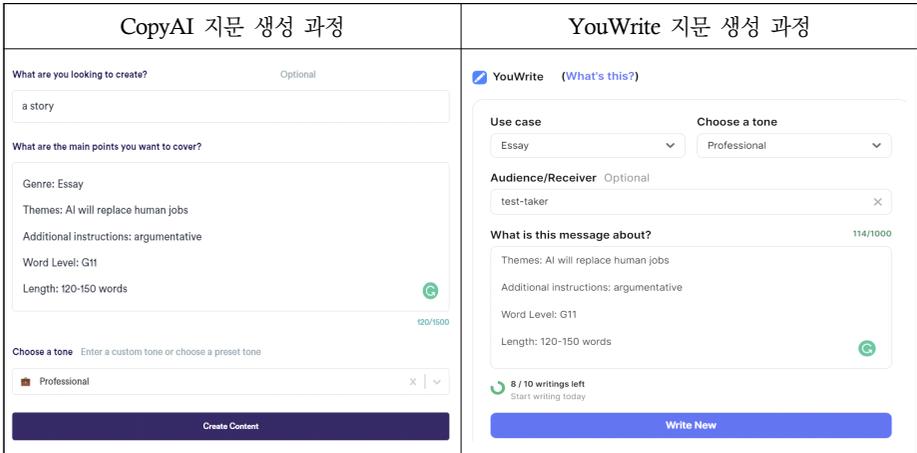


그림 3. CopyAI와 YouWrite를 활용한 지문생성 과정 비교(교사1)

연수 과정 중 ChatGPT의 사용자 증가로 접속이 원활하지 않은 경우²⁾는 CopyAI와 YouWrite를 사용하기도 하였다. 두 프로그램은 사용 방법이 유사하며 지문생성 전용 도구이다 보니 ChatGPT를 활용 때처럼 지시문 작성 시 문장 단위의 입력이 필요하지 않고 단순하게 지문과 관련한 핵심 정보(단어, 구, 문장 모두 가능)만 입력하면 된다는 장점이 있다. 입력 정보에 대한 포맷의 예시는 제공되지만 자유로운 형식의 입력도 가능하다. 이 두 프로그램은 각각 장단점이 있지만 CopyAI가 생성한 지문의 길이에 대부분은 부합했지만 YouWrite는 요청 길이의 두 배 정도 길게 작성하는 경향을 보였다. 어휘 난이도 통제에의 경우는 두 프로그램 모두 정확하지 않았고, YouWrite가 상대적으로 좀 더 어렵게 작성하는 경향을 보였다(예, 위 조건(G11). 입력 시 이독지수인 Flesch-Kincaid Grade Level 기준으로 CopyAI는 G12 Level, YouWrite는 College Level).

4.2. 문항생성

연수에 참가한 교사들은 각각 1개의 수능 영어 읽기 문항에 대하여 지문을 생성한 후 문항생성을 시도하였고 대체로 선택지의 형식이 문항유형에 큰 영향을 미치는 않는

2) 인터넷 브라우저 중 Chrome을 통한 ChatGPT 접속이 많기 때문에 Firefox나 NAVER Whale과 같은 브라우저를 설치하여 접속하면 상대적으로 접속이 원활하고 유료 버전인 ChatGPT Plus인 경우는 접속이 안정적이다.

‘글의 목적/요지/의도/주제/제목 찾기’나 ‘세부내용 일치/불일치 찾기’는 다음의 표 2와 같은 지시문 형식으로 쉽게 문항생성이 가능하였다.

표 2. ChatGPT 활용 수능 영어 읽기 문항생성 일반 형식

공통) Make a new passage with a different topic and make a MCQ (including 5 choices)
 1) to test the understanding of the purpose/gist/main idea/best title of the passage
 2) to confirm the agreement with the details of the passage
 as follows: 수능 지문 및 문항 삽입

※ ‘세부 내용일치/불일치 찾기’의 경우 “Only one answer choice is correct/incorrect.”와 같은 추가 조건을 선택적으로 입력

‘세부 내용일치/불일치 찾기’ 유형의 경우에는 추가적으로 “Only one answer choice is correct/incorrect.”와 같은 조건을 입력해야 원하는 형식의 선택지 구성이 가능하였다. 하지만 문항유형, 특히 선택지의 구성이 지문의 특성과 밀접하게 연계되어 있는 경우에는 문항별로 차별화된 지시문이 필요하였다. 다음의 표 3은 연수 참가 교사들의 다양한 문항생성 시도를 바탕으로 수능 영어 읽기 문항유형에 따른 성공적인 지시문만을 발췌하여 정리한 것이다.

표 3. 수능 영어 읽기 문항유형별 ChatGPT 활용 문항생성 절차

수능 문항유형	지시문(Prompt)
1. 심경·분위기 파악하기	Make a new passage with a different topic with a similar MCQ as follows: Which of the following best describes Jonas's change in emotion? (발문 영문 번역 후 입력) 수능 기출 문항 지문 및 선택지 삽입
2. 밑줄 친 표현의 맥락적 의미 파악하기	1) Make a new passage with a different topic as follows: 수능 기출 문항 지문 삽입 2) Underline one phrase and create five choices for what that phrase implies in the following passage: 생성 지문 삽입
3. 빈칸 추론(구문)하기	Make a new passage with a different topic as follows and make a < > by deleting one phrase, and make 5 choices for < >: 수능 기출 문항 지문 삽입 (빈칸 밑줄 대신 < > 또는 ()으로 교체)
4. 글 흐름에 부적절한 문장 찾기	Make a new passage with a different topic, and keeping the first two sentences of the passage, then assign 5 numbers before some sentences that make up the rest part of the passage, additionally only one of the 5 sentences should be unsuitable for the flow of the text as follows: 수능 기출 문항 지문 삽입

표 3. 계속

수능 문항유형	지시문(Prompt)
5. 글 순서 맞추기	1) Make a new passage with a different topic as follows: 수능 기출 문항 지문 삽입 2) Keep the first sentence in the following passage and shuffle paragraphs (A), (B), (C): 생성 지문 삽입 3) Revise the above passage by adjusting the amounts in paragraphs (A), (B), and (C) to be equal. (필요 시)
6. 글 흐름에 맞게 문장 끼워넣기	Paragraph (A) is taken from one of (1), (2), (3), (4), (5) in the following passage, and make a new passage with a different topic as the following passage and extract one paragraph like (A) at the top of the story and make 5 ()s as options: 수능 기출 문항 지문 삽입 (원문자를 () 형식으로 교체)
7. 어법 상 틀린 표현 찾기	1) Make a new passage with a new topic as the following passage and make five < >s containing a word: 수능 기출 문항 지문 삽입 2) Change one of the words in the five < > into an unsuitable or ungrammatical word out of context of the following passage: 생성 지문 삽입
8. 지문 요약에서의 빈칸 추론(단어)하기	1) Make a new passage with a different topic as follows: 수능 기출 문항 지문 삽입 2) Summarize the following passage in one sentence or so: 생성 지문 삽입 3) Make it shorter. (필요 시 입력) 4) Make Blanks (A) and (B) for two key words in the summary. 5) Make 5 word pairs as options for (A) and (B)

예를 들어, 등장인물의 심경 또는 분위기를 파악하는 문항은 2017학년도 수능부터 하나의 감정이나 분위기를 묻는 문항에서 심경의 변화를 묻는 문항으로 고정적으로 출제되고 있다. 이 문항의 경우 지문생성 단계에서부터 두 가지 심경이 제시되도록 조건화되고 선택지의 구성 또한 기출 문항의 형식을 따라야 하기 때문에 표 2의 첫 번째 지시문과 같이 수능 기출 문항의 발문을 영문으로 번역한 후 예시 문항에서 이를 대체하여 다음의 그림 4와 같이 문항생성을 시도하였다.

그림 4와 같이 ChatGPT가 생성한 ‘심경·분위기 파악하기’ 문항은 기출 문항과 유사한 형식으로 출제된 것을 확인할 수 있다. 물론 앞서 언급한 바와 같이 특정 어휘목록을 기준으로 통제하는 것은 아직 불가능하여 ‘muttered’나 ‘devastated’란 어휘는 교체하거나 주석으로 처리가 필요해 보인다. 특히 정답인 “③ elated → devastated”의 경우 두 단어 모두 수능 영어의 수준을 넘어서는 단어이며 지문에 사용된 어휘를 그대로 차용했다는 점에서 expectant/hopeful → frustrated/disappointed 정도로 수정이 필요해 보인다.

지 시 문	<p>Make a new passage with a different topic with a similar MCQ as follows: Which of the following best describes Jonas's change in emotion?</p> <p>Looking out the bus window, Jonas could not stay calm. He had been looking forward to this field trip. It was the first field trip for his history course. His history professor had recommended it to the class, and Jonas had signed up enthusiastically. He was the first to board the bus in the morning. The landscape looked fascinating as the bus headed to Alsace. Finally arriving in Alsace after three hours on the road, however, Jonas saw nothing but endless agricultural fields. The fields were vast, but hardly appealed to him. He had expected to see some old castles and historical monuments, but now he saw nothing like that awaiting him. "What can I learn from these boring fields?" Jonas said to himself with a sigh.</p> <p>① excited → disappointed ② indifferent → thrilled ③ amazed → horrified ④ surprised → relieved ⑤ worried → confident</p>
생 성 된 문 항	<p>Which of the following best describes Sarah's reaction to the news?</p> <p>Sarah had been waiting anxiously for her exam results to be released. She had studied tirelessly for months and had even enrolled in an extra tutoring course. When the results were finally released, Sarah eagerly logged into her student portal. However, as she read through the results, her face fell. She had not done as well as she had hoped. Her grades were average, and she had barely passed some of her subjects. Sarah was devastated. She had expected to do better and had even envisioned receiving high marks. "What am I going to do now?" Sarah muttered to herself, feeling disheartened.</p> <p>① thrilled → disappointed ② indifferent → excited ③ elated → devastated ④ surprised → content ⑤ anxious → confident</p>

그림 4. '심경·분위기 파악하기' 유형에 대한 문항생성 사례(교사2)

물론 만족스런 지문과 문항이 생성될 때까지 재생성하는 것도 하나의 대안이 될 수 있다.

그 밖의 다른 수능 문항유형 또한 표 2에 제시된 지시문을 활용하여 문항생성이 가능하였고 현재의 ChatGPT가 시각자료는 인식하지 못하는 관계로 '시각자료(예, 그래프) 기술하기'에 관련한 문항만 생성할 수 없었다. 다만 그래프의 수치를 글로 풀어 설명하면 이 문항유형 또한 간접적인 생성은 가능하다. 단 문항유형 6과 7처럼 '글의 흐름에 맞지 않는' 또는 '어법에 맞지 않는'이라는 지시를 하기 위해 'inappropriate' 또는 'illogical'이라는 단어를 사용할 경우 AI 윤리규정과 관련하여 허위사실 등의 제작을 예방하기 위해 잘못된 부분의 생성을 무시하고 올바른 문장으로만 지문을 구성할 수도 있다. 이것이 표 3의 지시문에 'inappropriate' 또는 'illogical' 대신 'unsuitable'이라는 단어를 사용한 이유이다.

4.3. 문항선제

문항선제 과정에는 그림 5와 같이 가독성을 높이기 위해 Google Translate를 사용하여 한글 번역을 참고하여 검토하였다.

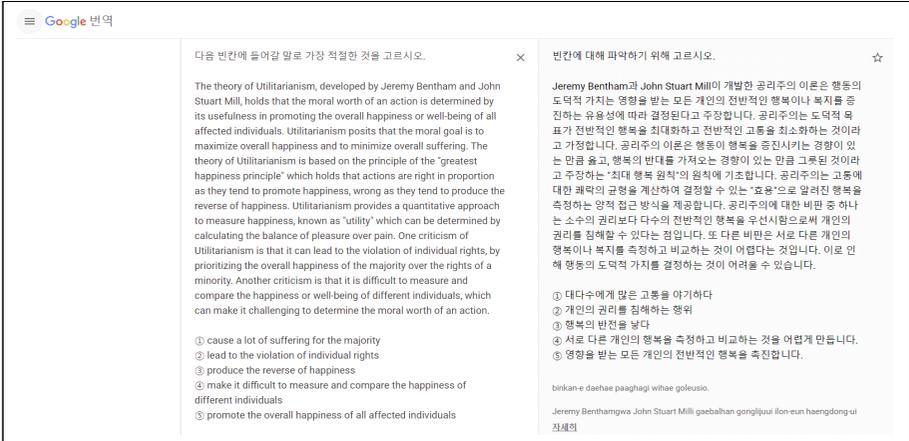


그림 5. 문항 선제를 위한 Google Translate 활용 사례(교사7)

다음의 표 4는 그림 5에서 살펴본 문항의 지문 일부를 발췌하여 Google Translate와 DeepL 번역기로 비교·분석한 결과이다.

표 4. Google Translate와 DeepL 번역기의 번역 비교

원문	The theory of Utilitarianism is based on the principle of the "greatest happiness principle" which holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness.
Google Translate 번역	공리주의 이론은 행동이 행복을 증진시키는 경향이 있는 만큼 옳고, 행복의 반대를 가져오는 경향이 있는 만큼 그릇된 것이라고 주장하는 "최대 행복 원칙"의 원칙에 기초합니다.
DeepL 번역	공리주의 이론은 행복을 증진하는 경향이 있는 행동은 그에 비례하여 옳고, 행복의 반대를 초래하는 행동은 그에 비례하여 그르다는 '최대 행복 원칙'에 기초하고 있습니다.

앞서 연구방법에서도 언급한 바와 같이 번역기의 결과물만을 비교해 보면 DeepL 번역기가 Google Translate의 번역과 비교하여 더 자연스럽다는 것을 확인할 수 있다. 하지만 영어 능숙도면에서 미숙할 수 있는 수험자들의 입장에서는 위의 원문 표현을 Google Translate처럼 번역할 가능성이 있다는 점을 고려하였다. “in proportion as”라

는 표현을 Google Translate는 “~만큼”으로 번역하고 DeepL 번역기는 “그에 비례하여”라고 번역하고 있는데 “~만큼”의 경우는 마치 ‘조건’, 즉 ‘~때문에’란 의미를 내포하고 있는 것처럼 해석될 수도 있어서 이러한 표현들이 실제 수험자들의 입장에서 어떻게 해석되고 문항의 난이도에 어떤 영향을 미치는지 문항선제 시 검토의 대상이 될 수 있다고 판단하였다.

4.4. 어휘통제

본 연구에서 분석한 출제 연수는 수능 영어 읽기 문항을 기준으로 다양한 활동이 진행되었지만 출제 전반에 관한 교육인 만큼 학교급별 수준에 따라 지문의 어휘를 통제하면서 지문의 표현을 수정하는 교육을 함께 진행하였다. 수능의 영어 어휘목록이 공개되지 않는 만큼 그와 유사한 수준인 Nation(2016)의 대표적인 어휘목록인 BNC/COCA25000 가운데 상위빈도 4,000개 어휘족을 어휘통제에 활용하였다. 물론 ChatGPT나 CopyAI 활용 시에도 학습자의 연령 또는 학년 정보, 그 밖에도 [easy/difficult/advanced/sophisticated] [words/vocabulary]란 표현을 사용하면 대략적인 어휘통제는 가능하지만 이는 특정 어휘목록으로 지문의 어휘를 통제하는 국가시험의 문항 개발에는 여전히 한계가 있다.³⁾ 이에 출제 연수에서는 Anthony(2013)의 AntWordProfiler에 BNC/COCA25000의 상위 4,000개 어휘족을 탑재하여 어휘를 통제하였다.

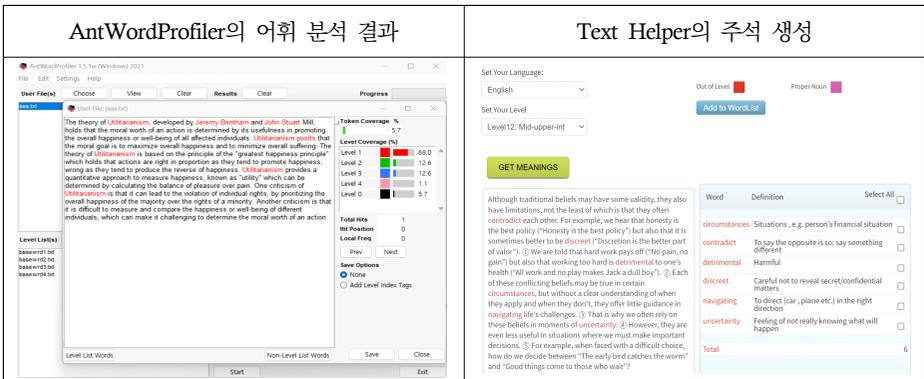


그림 6. 지문의 어휘통제 사례(교사 7)

3) GPT-3.5 모델에 기반한 ChatGPT는 한 번에 입력할 수 있는 데이터의 양이 3,000개 정도로 제한되어 수능 어휘목록(word family) 등을 모두 입력하여 어휘를 통제하는 것이 불가능했지만 최신의 GPT-4 모델에 기반한 ChatGPT PLUS는 입력 가능한 데이터의 양이 25,000개 단어로 확장되었다. 하지만 여전히 ChatGPT PLUS의 경우도 500개 정도의 어휘목록을 입력하여 어휘통제가 가능한 수준이다. 하지만 앞으로 개선될 버전에서는 입력 데이터의 규모가 확장되며 자연스럽게 해결될 문제로 판단된다.

AntWordProfiler를 활용한 어휘통제 절차는 Shin(2020)가 제시한 방법을 따랐으며 어휘 수정 및 교체에는 Extensive Reading Central(n.d.)에서 서비스하는 Text Helper 라는 주석 생성기를 활용하였다. Text Helper는 지문을 입력하고 특정 어휘수준을 설정 하면 설정한 수준을 벗어난 단어에 대한 주석을 그림 6과 같이 생성해 준다.

4.5. 문체, 선택지 길이 등 세부 수정

출제 연수의 절차에는 연수 참가 교사들간의 문항 교차검토 시간이 포함되어 있었다. 교사8이 받은 피드백 중 하나는 지문의 문체가 너무 비격식체(informal)라는 의견이었다. 일반적으로 이러한 피드백을 수용하고자 한다면 기존 지문을 교체하여 새로 출제할 수밖에 없다. 하지만 교사8은 CopyAI에 지문을 입력하고 ‘Tone Changer’ 기능을 이용하여 ‘tone’을 ‘professional’로 교체하여 간단히 문체를 수정하였다. 다음의 표 5는 CopyAI의 Tone Changer’ 기능을 활용하여 수정한 사례이다.

표 5. CopyAI의 ‘tone’ 변경 기능을 활용한 문체 수정 사례(교사8)

원본	Heat is energy and when you add energy to any system changes occur. Because all systems in the global climate system are connected, adding heat energy causes the global climate as a whole to change.
수정본	Heat is a sort of energy and change occurs when energy is added to any system. A rise of temperature in one part of the globe will affect other parts since all climate systems are interconnected.

원본의 “all ~ connected”는 “interconnected”로, “adding”은 수동태인 “is added to”로 수정되었고 “causes the global climate as a whole to change”로 “affect other parts”로 수정되어 보다 격식있는(formal) 문체로 수정된 것을 확인할 수 있다.

ChatGPT를 활용하면 보다 다양한 형식의 문체로 수정이 가능하다. 다음의 표 6은 표 5의 원본을 미국의 Obama 대통령과 Trump 대통령의 문체로 수정한 사례이다.

유명인의 글이나 연설 등의 데이터는 이미 ChatGPT의 개발에 사용된 언어모델인 GPT-3.5에 포함되어 있기 때문에 이러한 특정인의 문체로도 변경도 가능하다. Obama 대통령의 경우 격식있고 고급스런 단어를 구사하는 반면 Trump 대통령의 문체는 비격식의 구어체, 보다 직설적이라는 것을 알 수 있다.

표 6. ChatGPT를 활용한 문체 변경 사례

<p>Obama 대통령 문체로 수정</p>	<p>It is a fundamental truth that heat, at its essence, is energy — a potent force that possesses the ability to initiate changes within any given system. It is important to note that in our world, every system within the global climate operates as an interconnected and interdependent entity. Thus, when we introduce additional heat energy into this delicate web of systems, it inevitably sets off a chain reaction, causing the global climate, as a whole, to undergo a transformation.</p>
<p>Trump 대통령 문체로 수정</p>	<p>Look, folks, let me tell you something — heat, it's energy, okay? And when you add that energy to any system, changes are bound to happen. That's just science, it's a fact. Now, here's the thing — every single system in the global climate is connected, they're all working together, for better or for worse. So when you add heat energy into that mix, it's gonna cause the whole global climate to change. And let me tell you, nobody knows more about change than me, believe me.</p>

끝으로 ‘글의 순서 맞추기’ 유형의 경우 도입부에 연결되는 (A), (B), (C) 문단의 분량이 유사해야 하는데 ChatGPT를 활용하여 지문을 생성했을 때 (A), (B), (C)의 분량이 달라 문제가 될 수도 있다. 다음의 그림 7은 CopyAI를 문단의 확장에 적용한 사례이다.

<p>원문 CopyAI 작업창에 입력하기</p>	
<p>확장하려는 문장의 끝에서 Tab 버튼 클릭하여 확장 버전 선택하기</p>	

그림 7. 문단 확장 사례(교사10)

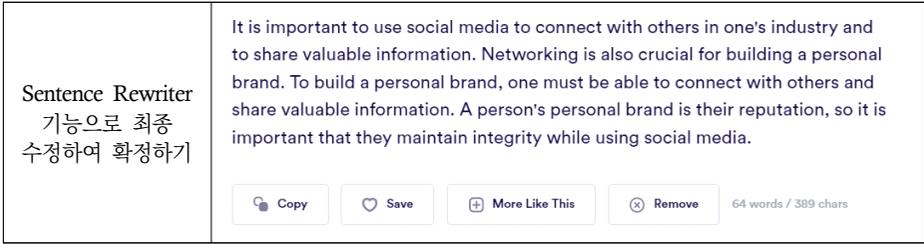


그림 7. 계속

교사10의 경우 문단 (A)가 문단 (B)와 (C)에 비해 상대적으로 짧아 CopyAI의 문단 확장 기능을 사용하여 지문을 확장하고 이를 다시 ‘Sentence Rewriter’ 기능으로 다듬어서 자연스럽게 좀 더 긴 문단으로 수정하였다.

5. 결론 및 시사점

본 연구에서는 10명의 중등 영어 교사를 대상으로 수능 영어 읽기 문항의 출제 연수 과정 중에 학습한 다양한 AI 도구의 활용 사례를 분석하였다. 사례분석은 문항의 5개 출제 단계, 즉 1) 지문생성, 2) 문항생성, 3) 문항선제, 4) 어휘통제, 5) 문체, 선택지 길이 등 세부 수정으로 나누어 살펴보았다.

물론 AI 도구를 활용하여 개발한 문항에 대한 질적 평가는 실험 연구를 통해 기존의 방식대로 출제된 문항과 비교·분석해야 정확히 판단할 수 있겠지만 적어도 문항 출제의 효율성이나 만족도 면에서는 출제 연수에 참가한 교사들 모두 기대 이상이었다는 반응을 보였다. “요즘 언론에서 AI 시대가 도래했다는 말만 들었지 정말 이 정도인지는 몰랐어요. 이제 교사들은 뭘해야 할까요?”(교사2), “어휘통제나 문항의 난이도 조정은 아직 좀 필요하지만 이제 굳이 수능 출제는 장기 합숙으로 할 필요는 없을 것 같아요.”(교사 9), “무엇보다도 ChatGPT는 정말 놀랍습니다. 이제 교육현장에 정말 큰 변화가 올 것 같아요.”(교사 3) 등의 인터뷰 반응만으로도 ‘놀람’, ‘혼란’, ‘걱정’, ‘기대’ 등이 뒤섞인 복합적인 감정을 확인할 수 있었다. ChatGPT와 같은 AI 도구의 등장은 새로운 교육 기회 창출과 함께 교사들의 입장에서는 충분히 위협적인 대상으로 받아들여질 수도 있을 것이다. 전 세계적으로 ChatGPT의 교육적 활용에 대해 상반된 시각을 갖는 이유도 이와 맥락을 같이 한다(Kim, D., 2023). 하지만 본 연구의 사례를 바탕으로 비추어 문항 출제에 AI 도구를 본격적으로 도입할 경우 다음과 같은 효과가 기대된다.

첫째, AI 도구를 활용한 문항 출제는 동일한 지문 및 문항생성 알고리즘을 적용함으로써 평가의 타당성과 신뢰성 면에서 일정 수준을 유지하는데 큰 기여를 할 수 있을 것이다.

둘째, 지금까지 국가단위의 시험 문항에는 막대한 시간과 비용이 소요되었음에도 불구하고 그러한 출제의 노하우는 매우 제한적인 전문가들만이 소유할 수 있었다. 하지만 AI 도구의 활용은 평가 전문성의 대중화는 물론 출제 시간과 비용의 효율성 면에 획기적인 개선을 가져올 것이라 기대된다.

셋째, AI 도구의 활용은 학습자 개별 맞춤형 평가는 물론 즉각적인 피드백 제공까지 가능하게 하여 단순히 평가에 그치는 수준이 아닌 교수학습에 선순환되는 효과를 기대할 수 있다.

그럼에도 불구하고 평가에 있어 AI 도구의 활용 효과를 극대화하기 위해서는 몇 가지 해결해야 할 과제가 있다. 본 연구에서 소개한 AI 도구는 문항 출제 전용으로 개발된 프로그램은 아니다. 그러한 이유로 문항별 잘 고안된 지시문이 필요했고 어휘수준이나 문항 난이도의 정교한 통제는 쉽지 않았다. 따라서 ChatGPT API 등을 활용한 문항 출제에 최적화된 전용 프로그램을 개발하여 교육자는 물론 학습자들이 사용하기에도 쉬운 프로그램들이 개발·보급될 필요가 있다. 물론 최근 ChatGPT를 기반으로 하여 개발된 응용 플랫폼인 GENQUE(Kim, K., 2023)는 지문생성은 물론 문항생성의 편의성을 반영하였지만 어휘수준의 통제나 문항 난이도 조절 면에서는 여전히 정교성이 떨어지며 문항의 완성도 또한 검증되지 못했다. 따라서 후속 연구에서는 ChatGPT나 GENQUE와 같은 AI 도구를 활용해 출제한 문항의 질적 완성도를 분석하여 추후 개선을 위한 시사점을 도출할 필요가 있다.

References

- Amodei, D. (2020). Language models are few-shot learners. *arXiv* (preprint). Retrieved from <https://arxiv.org/abs/2005.14165>
- Anthony, L. (2013). *AntWordProfiler Version 1.4.0w* [Computer Software]. Waseda University. Retrieved from <https://www.laurenceanthony.net/software/antwordprofiler/releases/AntWordProfiler140/AntWordProfiler.exe>
- Attali, Y., & Powers, D. (2010). Immediate feedback and opportunity to revise answers to open-ended questions. *Educational and Psychological Measurement, 70*, 22-35.
- Bartram, D., & Hambleton, R. (2005). *Computer-based testing and the Internet: Issues and advances*. Hoboken, NJ: JohnWiley & Sons.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv* (preprint). Retrieved from <https://arxiv.org/abs/2005.14165>
- DeepL. (n.d.). *The world's most accurate and nuanced machine translation*. Retrieved from <https://www.deepl.com/en/whydeepl>
- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development*. Mahwah, NJ:

- Lawrence Erlbaum Associates Publishers.
- Extensive Reading Central. (n.d.) *Text helper* [Computer Software]. Retrieved from <https://www.er-central.com/text-helper/>
- Gierl, M. J., & Haladyna, T. M. (2013). *Automatic item generation: Theory and practice*. New York, NY: Routledge.
- Gyeongnam Office of Education. (2015). *2015 competency-building training for process-oriented assessment (2015-183)*. Changwon: Elementary Education Division, Gyeongnam Office of Education.
- Google. (n.d.). *Google Translate* [Computer Software]. Retrieved from <https://translate.google.com/>
- Hommel, B. E., Wollang, F.-J. M., Kotova, V., Zacher, H., & Schmukle, S. C. (2021). Transformer-based deep neural language modeling for construct-specific automatic item generation. *Psychometrika*, 87, 1-24.
- Khan, S., Huang, Y., Pu, S., Tarasov, V., Andrade, A., Meisner, R., ... & von Davier, A. A. (2020). *Sphinx: An automated generation system for English reading comprehension assessment*. Paper presented at the International Conference on Learning Analytics and Knowledge (Online).
- Kim, D. (2023, February). 'AI textbooks' in which the level of learning content goes up as the learner studies well are coming soon. *THE AI*. Retrieved from https://newsthai.com/site/data/html_dir/2023/02/24/2023022480178.html
- Kim, K. (2023). *GENQUE* [Computer Software]. Artificial Society. Retrieved from <https://www.genque.ai/>
- Kim, T. (2023). *Can ChatGPT be an innovative tool?: The use cases and prospects of ChatGPT (NIA_The AI Report 2023-1)*. NIA AI Future Strategy Center.
- Kumar, V., Boorla, K., Meena, Y., Ramakrishnan, G., & Li, Y.-F. (2018). *Automating reading comprehension by generating question and answer pairs*. Paper presented at the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Melbourne, Australia.
- Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 30, 121-204.
- Lee, H. (2018). *AI Test Maker (ATM)* [Computer Software]. LXPEN Inc. Retrieved from <https://atm.lxper.ai/>
- Lee, Y., & Shin, D. (2020). A study on the online assessment using artificial intelligence for distance education. *Journal of Learner-Centered Curriculum and Instruction*, 20(14), 389-407.
- McCann, B., & Socher, R. (2022). *YouWrite* [Computer Software]. You.com. Retrieved from <https://you.com/search?q=how+to+write+well&&tbm=youwrite&cfr=write>
- Nation, I. S. P. (2016) *Making and using word lists for language learning and teaching*. Amsterdam, Netherland: John Benjamins Publishing Company.
- National Board of Educational Evaluation. (1992). *Methods and procedures for developing test items to assess academic achievement*. Seoul: Far Eastern Cultural History.

- Ney, H., Essen, U., & Kneser, R. (1994). On structuring probabilistic dependences in stochastic language modelling. *Computer Speech and Language*, 8, 1-38.
- OpenAI. (2022). *ChatGPT: Optimizing language models for dialogue* [Computer Software]. Retrieved from <https://chat.openai.com/>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1, 9.
- Shermis, M. D., & Burstein, J. (2013). *Handbook of automated essay evaluation: Current applications and new directions*. New York, NY: Routledge.
- Shin, D. (2020). An empirical study on paraphrasing patterns of Korean pre-service teachers in the process of controlling words that make up the passage for test items in English reading assessment. *Multimedia-Assisted Language Learning*, 23(3), 157-179.
- Shin, D. (2022). What is the limit for using AI in English education?: Feasibility of using automated text generation and automated item generation. *Journal of the Korea English Education Society*, 21(3), 147-163.
- Simon, C. (2022, May). It's time to "push the envelope" on AI and prepare for the AGI era. *InfoWorld*. Retrieved from <https://www.itworld.co.kr/opinion/235494#csidx7b60556267fb24a9b47cec45b44c91c>
- Sireci, S. G., & Zenisky, A. L. (2006). Innovative item formats in computerbased testing: In pursuit of improved construct representation. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 329-347). Mahwah, NJ: Erlbaum.
- Ulsan Metropolitan Office of Education. (2014). *Student assessment support materials for education that nurtures dreams and aspirations* (2014-62). Ulsan: Ulsan Metropolitan Office of Education, Curriculum Operation Division.
- van der Linden, W. J., & Glas, C. A. (2010). *Elements of adaptive testing*. New York, NY: Springer.
- von Davier, M. (2018). Automated item generation with recurrent neural networks. *Psychometrika*, 83, 847-857.
- Wikipedia_a (n.d.). *Google Translate: Neural machine translation*. Retrieved from https://en.wikipedia.org/wiki/Google_Translate
- Wikipedia_b (n.d.). *DeepL Translator*. Retrieved from https://en.wikipedia.org/wiki/DeepL_Translator
- Yacobian, P. (2020). *CopyAI* [Computer Software]. CopyAI Inc. Retrieved from <https://www.copy.ai/>

신동광
부교수
영어교육과
광주교육대학교
61204 광주광역시 북구 필문대로 55
E-mail: sdhera@gmail.com

접수일자 : 2023. 2. 28
수정본 접수 : 2023. 4. 6
게재결정 : 2023. 4. 20