

ChatGPT 4.0의 중등교원 임용시험 전공영어  
기출문제 풀이수행능력 탐색 연구\*

Exploring ChatGPT 4.0' Ability to Solve Secondary School English Teacher  
Certification Exam Questions

김경민<sup>1</sup>, 이장호<sup>2</sup>  
Kyungmin Kim, Jang Ho Lee

Kim, Kyungmin & Lee, Jang Ho. (2024). Exploring ChatGPT 4.0' ability to solve secondary school English teacher certification exam questions. *Multimedia-Assisted Language Learning*, 27(2), 1-22.

In recent years, ChatGPT has integrated itself into applied linguistics and language education research, with studies proposing different ways to use this new technology in second language teaching. As part of these efforts, the present study aimed to investigate ChatGPT's ability to solve the domain-specific questions of high-stakes exams. To this end, the dataset for this study consisted of multiple-choice questions taken from the teacher certification exams for prospective English teachers, developed by the Korea Institute for Curriculum and Evaluation between 2009 and 2013. The authors then asked ChatGPT to solve each question by entering them as prompts. The results showed that ChatGPT's average percentage of correct answers was 66%, with the percentage slightly lower for the English Literature and Linguistics subsections than for the English Education Theories and General English subsections. The qualitative analysis also revealed four different patterns in the errors made by ChatGPT when answering the target questions, which we illustrate with specific examples. Based on these findings, we have provided implications for pedagogy and future research.

**Key words:** ChatGPT 4.0, domain-specific knowledge, teacher certification exam, English education  
**Applicable level:** tertiary education

\* This research was supported by the Chung-Ang University Research Scholarship Grants in 2023.

<sup>1</sup> Kyungmin Kim (Graduate Student, 1<sup>st</sup> author, Department of English Education, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea; E-mail: teacherkm2016@naver.com)

<sup>2</sup> Jang Ho Lee (Professor, Corresponding author, Department of English Education, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea; E-mail: jangholee@cau.ac.kr)

Received: April 30, 2024; Reviewed: May 31, 2024; Accepted: June 15, 2024

## I. 서론

2022년 11월 OpenAI사에서 처음 출시되었던 ChatGPT는 현 시점까지 다양하고 많은 분야에서 활용되고 있다. 출시 두 달 만에 1억 명이 넘는 사용자가 생겼을 만큼 ChatGPT가 미쳤던 과급력은 대단하다고 볼 수 있다(Jang & So, 2023). 최근 출시된 ChatGPT 4.0의 경우에는 이전 버전보다 데이터 베이스가 광범위해졌으며 응집력 있고 맥락적으로 관련성이 있는 텍스트를 생성할 수 있게 되었다. 뿐만 아니라 이미지까지 분석이 가능한 시스템을 탑재하고 있어 기존의 텍스트를 직접 입력해야 상호작용이 가능하였던 ChatGPT 3.5와는 차별화된 편리성을 지니고 있다. ChatGPT 4.0은 이전 버전에 비해 추상적이고 깊이 있는 맥락의 이해를 요구하는 과업 수행 능력이 향상되었으며, 이는 ChatGPT 4.0의 데이터베이스 내에 매개변수가 더 늘어남에 따라서 언어의 복잡한 패턴을 학습할 수 있음에 기인한다(Caramancion, 2023).

최근 ChatGPT의 교육적 활용에 대하여 학술적인 관심 및 논의가 점차 증가하고 있다. Jang과 So(2023)의 연구에 의하면 현재까지 진행되어온 ChatGPT 관련 연구는 ChatGPT의 개념에 관한 연구가 가장 많았으며 ChatGPT의 성능에 대한 경험적인 연구 역시 활발하게 진행되고 있는 것을 확인할 수 있다. ChatGPT가 거대언어모델(Large Language Model, 이하 LLM)이라는 점에서 주로 언어 영역에 관련된 연구가 주로 수행되고 있으며 그 중 국내 연구 중에서는 영어 교과 관련 연구가 2위로 높은 순위에 위치하고 있음을 확인할 수 있다(Lee & Song, 2023).

영어교육 분야의 경우 다양한 주제의 ChatGPT 관련 연구가 제시되고 있는 추세인데, 가령 ChatGPT를 교실 현장에서 활용하였을 때 학습자들의 인식을 살펴보거나(Lee et al., 2023), 학생들을 대상으로 한 평가 도구로서의 잠재적 가치를 살펴본 연구(Jeon & Lee, 2023), 그리고 문항 생성 기능 탐색(Shin & Lee, 2023) 등 전반적으로 해당 연구들은 영어교육에 있어 ChatGPT의 활용 가능성에 대해 긍정적으로 평가하고 있다. 이 중, ChatGPT의 영어 문제풀이 수행능력에 대한 연구 또한 주목을 받고 있다. Kwon과 Lee(2023)는 ChatGPT의 문제풀이 수행능력을 활용하여 학습자들이 자기주도적 학습을 실시할 수 있을 것으로 전망하였고, 이러한 목적으로 ChatGPT를 활용하기 위해서는 ChatGPT의 문제풀이 수행능력을 검증할 필요성이 있음을 밝혔다. 그럼에도 불구하고 ChatGPT의 문제 풀이 수행 능력을 확인하고자 하였던 국내 영어교육 분야 연구는 많지 않은 편이며, 기존 연구들(Ahn, 2023; Kwon & Lee, 2023)은 모두 수능시험의 영어 읽기 영역과 같은 일반영어에 대한 ChatGPT의 수행 능력을 살펴보았다는 점을 알 수 있다. 즉, 임용시험 등과 같이 영역-특수적(domain-specific) 지식을 요구하는 문항들에 대한 ChatGPT의 문제풀이 수행능력을 살펴본 연구는 그 수가 현저하게 적음을 알 수 있다.

따라서 본 연구는 고부담 시험 중 하나인 중등교원 임용시험 전공 영어 문항에 대한 ChatGPT 4.0의 문항 풀이 수행 능력을 확인하고자 하였으며, 정답이 제시된 2009년부터 2013년까지 총 5개년의 시험을 대상으로 진행하였다. 이러한 본 연구의 결과는 특수목적 영어 분야에서의 ChatGPT의 활용과 관련하여 다양한 함의를 제시할 수 있을 것으로 생각된다.

## II. 이론적 배경 및 선행 연구

### 1. ChatGPT의 교육적 활용

현 시점까지 ChatGPT의 성능 탐색 및 특정 분야에서의 수행 능력 평가에 관하여 진행된 연구는 다양하다. Kohnke 외(2023)은 ChatGPT가 다양한 방식으로 언어 교수 및 학습에 있어 도움을 줄 수 있음을 시사한 바 있다. 예를 들어 ChatGPT는 발화 상황에서 해당 단어의 쓰임이 어떠한지 알려주며 예시 문장 및 다양한 의미를 제시함으로써 어휘학습 측면에서 학습자들이 ChatGPT를 활용할 수 있다고 하였다. 교수자의 입장에서는 교수 목적으로 활용할 수 있는 지문 생성과 더불어 관련 문항들을 만들 수 있는 성능을 가지고 있음을 밝혔다. Li 외(2023)은 학습자의 영어 작문에 피드백을 제공하는 도구로서 ChatGPT의 잠재성을 탐색하였다. 해당 연구에서는 수준이 서로 다른 두 학급을 대상으로 ChatGPT를 영어 작문 보조도구로서 활용한 결과, ChatGPT가 수준이 상대적으로 낮은 학급에서는 대화 상대로서의 역할을 수행하여 학생들이 대화형 의사소통을 연습할 수 있게 도와주었다. 수준이 상대적으로 높은 학급에서는 적절한 피드백을 제공하였으며 학습자들의 문법 학습 및 작문 능력을 향상시키는데 긍정적인 영향을 미칠 수 있다는 점을 밝혔다.

앞서 언급하였던 ChatGPT의 지문 생성 능력은 외국어 학습 자료 준비로 인하여 교사들이 느낄 수 있는 부담감을 경감시킬 수 있다. 이와 관련하여 Lee 외(2023)은 초등 영어 학습자들을 대상으로 특정 키워드를 입력하면 자동으로 다양한 장르의 지문을 형성해주는 ‘CopyAI’라는 프로그램을 활용하여 읽기 수업을 진행한 후, 학생들의 인식변화를 확인하였다. 연구 결과 AI 기술을 활용하여 스스로 직접 생각한 키워드들로 지문을 생성하는 학습 활동은 기존의 영어 읽기 수업 방식과 비교하였을 때 초등 학습자들의 영어 학습 동기에 더 긍정적인 영향을 미치는 것으로 확인되었다. 더불어 AI를 통하여 학생들의 영어 읽기 수업에 대한 흥미 수준이 증진되었음을 확인할 수 있었다.

ChatGPT의 성능 중 하나인 문항 생성 능력을 조사한 연구들도 진행된 바 있다. Shin(2023)과 Shin과 Lee(2023)의 연구에서는 여러 영역들 중 영어 읽기 시험 문항 제작에 대하여 살펴보았다. Shin(2023)은 영어 교사들을 대상으로 ChatGPT를 활용하여 제작한 문항에 대한 인식 조사를 하였고, 대부분의 교사들은 ChatGPT의 문항 제작 능력에 대하여 긍정적인 평가를 한 것으로 확인되었다. Shin(2023)은 ChatGPT를 활용하여 문항을 제작할 경우 문항의 유형에 맞도록 특정 프롬프트를 입력해야 함을 강조하였으며 해당 프롬프트들의 예시를 제공하였다. Shin과 Lee(2023)는 현직 교사들과 예비 영어 교사들을 대상으로 5개 유형의 수능시험 영어 독해 문항과 ChatGPT를 활용하여 제작한 동일한 유형의 문항들을 대상으로 블라인드 테스트를 진행하였다. 그 결과 문항 내 글 흐름의 자연스러움 그리고 표현의 자연스러움에 관련하여서는 수능 문항과

ChatGPT 생성 문항 간에 통계적으로 유의미한 차이가 없는 것으로 확인되었다. 그러나 보기의 매력도와 문항의 완성도 측면에서는 수능문항이 ChatGPT가 제작한 문항에 비해 교사들에게 더 긍정적으로 인식되었음을 확인할 수 있었다. 이러한 다양한 기능들을 언급하며 Jeon과 Lee(2023)는 ChatGPT가 교실 내에서 담화자, 콘텐츠 제공자, 평가자, 교사보조 등 다양한 역할을 수행할 수 있다는 점을 주장하였고 향후 효과적인 교수-학습을 위해서는 ChatGPT와 인간의 협업에 관련한 연구 수행이 필요함을 강조하였다.

최근에는 ChatGPT의 언어교육 평가도구로서의 잠재성을 조사한 연구도 진행되었다. Yi 외(2023)은 ChatGPT 3.5와 4.0을 대상으로 회귀 분석을 실시하여 문항 난이도 예측 도구로서의 활용성을 탐색하고자 하였다. 연구 결과 ChatGPT 4.0은 특히 0.5-0.7 범위에서 양의 상관관계를 보인 바 있다고 밝혔으며 이를 통하여 문항 난이도를 예측할 수 있을 뿐 아니라 정답률 또한 예상이 가능하다고 전망하였다.

위의 관련 선행연구 고찰을 통하여 언어 교육 분야에서 ChatGPT가 다양한 목적(학생들의 영어 작문에 대한 피드백 제공, 읽기 학습을 위한 지문 생성, 읽기 시험 문항 제작 등)으로 활용될 수 있다는 점을 살펴볼 수 있었다. 그러나 대부분의 연구는 일반영어학습의 내용을 다루고 있어 특수목적 영어 교수 및 학습을 위한 ChatGPT의 활용 연구는 그 수가 매우 부족하다는 점을 알 수 있다. 이에 본 연구는 중등교원 임용시험이라는 영역-특수적 지식을 필요로 하는 시험에 대한 ChatGPT의 문항 풀이 수행 능력에 대해 살펴보고자 한다.

## 2. ChatGPT의 문항 풀이 수행 능력

현재 ChatGPT가 교육 현장에서 다양한 역할을 수행할 수 있는 잠재적 가치를 가지고 있음에는 선행연구들을 통하여 확인이 가능하였다. 특히 평가 도구로서 활용될 수 있는 측면은 향후 교육현장에서 교수자들에게 큰 도움이 될 수 있을 것으로 예상된다. 그러나 전문적인 기능을 수행하기 위해서는 ChatGPT가 높은 정확도를 가지고 문항들을 풀이할 수 있는지 확인할 필요성이 제기된다.

이와 관련하여 Ahn(2023)은 고부담 시험들 중 하나인 대학수학능력시험 영어 읽기 영역 문항들에 대한 ChatGPT 3.5의 문제풀이 수행능력이 어떠한지 연구하였다. 연구 결과 ChatGPT 3.5는 2023학년도 대학수학능력시험 중 영어 과목의 읽기 영역에서 총 74%의 정답률을 보였으며 다양한 문항 유형들 중 특히 세부내용 파악과 주제 찾기 유형에 강점을 보이는 것으로 확인되었다. 또한 시험에 출제되는 다양한 질문들을 대부분 이해할 수 있는 것으로 확인되었고 선택한 답안에 대하여 충분한 근거를 제공하였다. 그러나 순서 배열 및 맥락 파악이 필요한 문항들에 대한 정확도는 상대적으로 낮은 편이었다. 이를 보완하기 위해서 Ahn(2023)은 ChatGPT 활용 전 사전 훈련(pre-training)이 필수적이라는 점을 제시하였다. 더불어 고부담 시험에 대하여 ChatGPT가 제공하는 해설의 정확도와 신뢰도를 향상시키기 위해서는 인간의 개입이 필요함을 강조하였다. 같은 맥락으로 Kwon과 Lee(2023)는 ChatGPT 3.5와 4.0의 대학수학능력시험 영어 과목

읽기 영역 문항에 대한 문제풀이 수행능력을 탐색하였다. 연구 대상은 2021학년도부터 2023학년도까지 총 3개년의 대학수학능력시험이었으며 추가적으로 TOEFL 시험도 분석 대상에 포함시켰다. 연구 결과 74%의 정답률을 나타내었던 ChatGPT 3.5와는 다르게, ChatGPT 4.0은 93%의 평균 정답률을 보였다. 뿐만 아니라 ChatGPT 3.5가 오류를 범하였던 문항들에 대해서도 ChatGPT 4.0은 대부분 정답을 제공하였으며, 특히 사실적 이해와 빈칸추론 및 요약하기 문항에서 높은 정답률을 보였다. 즉, 추론적 이해 영역에 약점을 보였던 ChatGPT 3.5와 비교되는 성능의 발전을 확인할 수 있었다. 더불어 TOEFL 독해 영역에서도 대학수학능력시험과 비슷한 양상을 보였다. 연구 결과를 통하여 Kwon과 Lee(2023)는 ChatGPT가 학습자의 자기주도적 학습에 긍정적인 도움을 줄 수 있을 것으로 전망하였으며, 특히 ChatGPT의 상호작용능력을 통해 학습자가 단순히 정답과 오답을 확인하는 것이 아닌 구체적인 설명도 제공받을 수 있다는 점에서 ChatGPT의 교육적 가치가 높다고 평가하였다.

비록 그 수가 많지 않지만, 영역-특수적 지식 기반 시험에서의 ChatGPT의 문제풀이 수행능력을 살펴본 연구 또한 진행된 바 있다. Gilson 외(2022)의 연구는 미국 의사 자격증 시험에서의 ChatGPT 3.5의 문제풀이 수행능력을 살펴보고, 연구 결과 ChatGPT는 3학년 의과 대학생 수준의 의학 지식을 가지고 있는 것으로 밝혀졌다. 그러나 영역-특수적 지식이 필요한 시험에서의 ChatGPT 수행 능력을 시험한 연구들은 대부분 ChatGPT 3.5를 활용하였으며 보다 고도화된 버전인 ChatGPT 4.0의 수행 능력을 탐색한 연구들은 많지 않다. 따라서 본 연구는 영역-특수적 지식을 요구하는 고부담 시험 중 하나인 중등교원 임용시험 전공영어 분야에서 ChatGPT 4.0의 문항풀이 수행능력이 어떠한지, 그리고 이를 통하여 ChatGPT 4.0이 수험자들의 시험 준비에 잠재적인 활용가치가 있는지 확인하고자 한다. 본 연구에서 제시하는 연구 질문은 다음과 같다.

첫째, 중등교원 임용시험 전공영어 영역에서의 ChatGPT 4.0의 과업 수행능력은 얼마나 우수한가?

둘째, ChatGPT 4.0이 문항 풀이 과정에서 오류를 범하는 문항의 유형은 무엇이며, 그 이유는 무엇인가?

### III. 연구 방법

#### 1. 연구 도구

##### 1) OpenAI ChatGPT 4.0

본 연구는 실험 도구로서 OpenAI사의 ChatGPT 4.0을 사용하였다. ChatGPT 4.0은 기존에 널리 알려진 ChatGPT 3.5의 고도화된 버전으로서 이전 모델과는 다르게 텍스트뿐 아니라 이미지를 첨부하여 분석을 가능케 하는 기능이 추가되었다. 본 연구에서는 ChatGPT 4.0의 해당 기능을 활용하여 PDF 파일 형태로 저장된 임용시험 시험 문제 양식을 이미지로 캡처한 후 ChatGPT 4.0에 업로드한 후 분석하게 하였다.

## 2) 중등교사임용시험 전공영어 영역 5개년 기출문제

본 연구에서 ChatGPT 4.0의 성능 분석 대상으로 사용된 데이터는 2009년부터 2013까지 총 5개 년의 중등교사임용시험 전공영어 기출문제 중 듣기 문항(1번-8번)을 제외한 문항들로 구성하였다. 문항들은 크게 4개의 영역(일반영어, 영어교육론, 영어학, 영문학)으로 구분되어 있었으며 각 영역의 문항은 일반영어 6문항, 영어교육론 14문항, 영어학 8문항, 그리고 영미문학 4문항으로 연도별로 전반적으로 동일하였다(단, 2009학년도 임용시험은 일반영어가 5문항, 영어교육론이 15문항으로 구성).

## 2. 자료 수집 방법

진술한 바와 같이 본 연구는 자료를 수집하기 위하여 한국교육과정평가원(<https://www.kice.re.kr/>)에 PDF형태로 등재되어 있는 중등교사임용시험 5개년 기출문제들을 ChatGPT 4.0에 이미지 파일로 입력하여 분석하도록 하였다. 제공된 이미지 파일을 ChatGPT가 분석하지 못할 경우를 방지하기 위하여 연구자들은 추가적인 지시문으로 이미지에 수록되어 있는 발문들을 텍스트로 한 번 더 언급하였다.

자료 수집 도중 일부 문항 내 발문에서의 요구 조건 중 여러 개의 진술들을 나열한 후 옳은 보기를 골라야 하는 문항들이 있었는데 ChatGPT가 간혹 선다형에 없는 보기들의 조합을 정답으로써 도출하는 경우가 있었다. 따라서 연구자들은 이를 방지하기 위하여 추가적인 지시문으로 “You must only choose an answer based on 5 multiple choices in the picture. Read multiple choices really carefully and never make mistakes. Also do not make your own combination. Make a combination based on the multiple choices and follow the order in the picture that I provided.” 를 추가하였다. ChatGPT가 문항에 대하여 응답을 하지 못하는 경우에는 프롬프트를 부분적으로 수정 및 추가하여(예: “You are making same mistakes. Read the question again.”) 최대 3번의 기회를 제공하여 답을 도출하게 유도하였다. ChatGPT가 각 문항을 분석한 후 연구자들은 한국교육과정평가원에서 제시한 정답과 비교하여 해당 결과를 기록하였으며 ChatGPT가 도출한 답안과 해설을 Microsoft Excel을 활용하여 정리하였다. 정리된 파일에 근거하여 연구자들은 정성적 분석과 정량적 분석을 실시하였다.

### 3. 자료 분석 방법

연구자들은 자료분석을 위해 우선 ChatGPT 4.0의 임용시험 영어 문항 분석 결과를 연도별, 그리고 영역별(일반영어, 영어교육론, 영어학, 영문학)로 구분하여 정리하였다. 이후 ChatGPT가 제공한 답변에 대하여 정량적인 분석을 위해 정답률을 계산하였다. 그 후 ChatGPT가 오답을 도출한 문항들에 한하여 ChatGPT의 해설을 면밀히 검토하여 ChatGPT가 올바른 정답을 도출하지 못하였는지에 대한 원인에 대해 정성적으로 분석하였다.

## IV. 연구 결과 및 논의

### 1. ChatGPT의 전공영어 문항풀이에 대한 정답률

2009년부터 2013년도까지 5개년도 중등교사임용시험 전공영어 영역에 대한 ChatGPT의 문항풀이에 대한 결과를 4개 영역별로 구분하여 정답률을 정리한 결과는 다음 <Table 1>과 같다.

**Table 1**

The Answer Rate of ChatGPT's Performance on the Five-Year English Teacher Certification Exam

Year	General English	English education theories	English linguistics	English literature	Overall
2009	3/5 (60%)	12/15 (80%)	5/8 (62.5%)	1/4 (25%)	21/32 (66%)
2010	4/6 (67%)	12/14 (86%)	3/8 (38%)	2/4 (50%)	21/32 (66%)
2011	6/6 (100%)	10/14 (71%)	3/8 (38%)	3/4 (75%)	22/32 (69%)
2012	3/6 (50%)	10/14 (71%)	3/8 (38%)	1/4 (25%)	17/32 (53%)
2013	5/6 (83%)	9/14 (64%)	6/8 (75%)	4/4 (100%)	24/32 (75%)
Average answer rate	21/29 (72%)	53/71 (75%)	20/40 (50%)	11/20 (55%)	105/160 (66%)

<Table 1> 에서 볼 수 있듯이 ChatGPT의 중등교사임용시험 전공영어 영역에서의 문항풀이 수행능력은 문항 전체의 약 66%의 정답률(160문항 중 105문항)을 나타내고 있음을 확인할 수 있었다. 영역별로 살펴보자면 영어교육론이 가장 높은 정답률(75%)을 보였으며 그 후 일반영어(72%), 영미 문학(55%), 영어학(50%) 순서대로 높았다. 전공 영역 중 영어교육론의 경우 비교적 높은 정답률(75%)을 보였으나 영어학(50%)과 영미 문학(55%)의 정답률 등을 전체적으로 고려할 때 Ahn(2023)의 연구결과와 마찬가지로 특정 영역에서의 ChatGPT의 수행능력이 다소

떨어진다는 점을 확인할 수 있었다. 일반영어의 정답률도 72퍼센트의 비교적 낮은 정답률을 보여주었는데, 해당 영역의 문항의 성격이 대학수학능력 시험의 읽기 영역과 유사하다는 점을 고려할 때, 이는 ChatGPT 4.0이 대학수학능력 영어시험의 읽기 영역에서 90퍼센트 이상의 높은 정답률을 보여준 Kwon과 Lee(2023)의 연구 결과와 대비되는 것으로 볼 수 있다.

## 2. ChatGPT의 오답 패턴 분석

ChatGPT 4.0의 영역별 정답률을 확인한 후, 오답을 제시한 문항들에 대해 ChatGPT가 제시한 설명을 분석하여 총 4개의 범주로 구분하였으며 각 범주에 대한 결과와 예시는 다음과 같다.

### 1) 문항의 발문을 부분적으로 읽은 경우

연구자들은 ChatGPT 4.0이 문제 풀이 과정 속에서 제공되어 있는 문항의 발문 또는 조건을 모두 활용하지 않는 경우를 확인할 수 있었다.

[A]	
<p>The two sentences below stand in a semantic relation.          (1) Mary sings arias beautifully. (2) Mary sings arias.          Whenever sentence (1) is true, sentence (2) is also true; whenever sentence (2) is false, (1) is also false. This relation is called entailment: sentence (1) entails sentence (2).</p>	
[B]	
<p>(a) X: John met with some of the delegates. Y: John did not meet with <u>all of</u> the delegates.</p> <p>(b) X: John does not eat vegetables. Y: John does not eat broccoli.</p> <p>(c) X: John regretted telling a lie. Y: John told a lie.</p>	<p>(d) X: John was not present at the reception. Y: John was absent from the reception.</p> <p>(e) X: John bought a fake leather jacket. Y: John bought a leather jacket.</p>
<p>Based on &lt;A&gt;, choose all the pairs in &lt;B&gt; in which X entails Y.</p> <p>① a, b ② a, b, d ③ b, c, d ④ b, d ⑤ c, d, e (정답: ④, ChatGPT가 제시한 답: ③)</p>	

Figure 1

Question #35 on the 2010 Teacher Certification Exam: English Major (Korea Institute for Curriculum and Evaluation, 2009)

<Figure 1>에 제시된 2010학년도 35번 문항은 영어학의 화용론 중 함의(entailment)에 대한 내용으로 <A>에 제시되어 있는 조건들을 모두 활용하여 <B>에 있는 문장들이 <A>의 조건들에 부합하는지 확인하여 고르는 것이었다. <A>에는 2개의 문장을 예시로 제공한 후 해당 문장들 사이의 sentence entailment라는 개념을 설명하고 있다. 다음은 ChatGPT가 문항을 풀이한 후 제시한



해설의 내용이다.

Entailment in semantics is a relationship between two sentences where the truth of the first (X) necessitates the truth of the second (Y), but not necessarily the other way around. Let's determine which pairs in box <B> exhibit entailment. Looking at the pairs:

b. If X is true ("John does not eat vegetables"), it must mean Y is true ("John does not eat broccoli"), because broccoli is a vegetable. So, X entails Y.

c. If X is true ("John regretted telling a lie"), it must mean Y is true ("John told a lie"), because regretting telling a lie implies that a lie was told. So, X entails Y.

d. If X is true ("John was not present at the reception"), it must mean Y is true ("John was absent from the reception"), because not being present is the same as being absent. So, X entails Y.

Based on the entailment relationship described the pairs where X entails Y are b, c, and d. From the given multiple-choice options, the one that includes b, c, and d is: b, c, d

ChatGPT의 해설을 분석해본 결과 ChatGPT는 <A>에서 부여한 조건 2개 중 첫 번째 조건 (Whenever sentence (1) is true, sentence (2) is also true.) 만 활용하여 문항을 풀이한 것을 확인할 수 있었다. 위 문항의 경우 (c)의 두 문장의 경우 함의가 아닌 전제(presupposition)관계에 있는데, 함의와 마찬가지로 첫 번째 문장이 참일 경우 두 번째 문장이 참이 성립하지만 함의와 다르게 두 번째 문장이 거짓일 경우 첫 번째 문장의 참, 거짓 여부가 불확실해지는 것을 알 수 있다(Saeed, 2009). 따라서 ChatGPT는 이러한 두 번째 조건을 활용하지 않아 전제관계인 (c)번 문장을 함의로 판단한 것으로 보인다. 그러나 비록 2개의 조건 중 하나만 사용하여 문항을 풀이하였음에도 불구하고 첫 번째 조건을 올바르게 이해하여 다른 예시 문장들에 적용하여 올바른 해설을 제공하였다는 것은 복잡한 과업을 요구하는 문항(조건 제시 후 문제 적용)들을 풀이하는데 어려움을 느꼈던 이전 ChatGPT 3.5와 달리(Ahn, 2023) 전적인 오류가 아닌 부분적인 오류만을 발생시킨다는 점에서 ChatGPT 4.0의 기능을 재조명할 수 있었다.

[A]

Consider the sentences in (1).  
 (1) a. John is sure that he will succeed.  
       b. John is sure to succeed.  
       c. \*John is sure his success.

The example (1c) is ungrammatical, though its underlined complement is legitimate. The ungrammaticality of (1c) is attributable to the following constraint on NPs: NPs in sentences should have a case either inherently or through being assigned by finite inflection, a transitive verb, or a preposition. Sentences with a caseless NP are ruled out. In (1c) the underlined complement lacks its case assigner.

[B]

a. *The child is difficult to travel alone.	d. *Not surprisingly did they miss the train.
b. *The city's recent dumping the garbage made people upset.	e. *It turned out the suspect to be innocent.
c. *I was pleased my sister to be pregnant.	(*' indicates that the sentence is ungrammatical.)

Choose all the sentences in <B> which have a caseless overt NP.

① a, c, d   ② a, c, e   ③ b, c, e   ④ b, e   ⑤ d, e (정답: ③, ChatGPT: ②)

Figure 2

Question #31 on the 2010 Teacher Certification Exam: English Major (Korea Institute for Curriculum and Evaluation, 2009)

Based on the explanation in box <A>, a caseless overt NP is a noun phrase that doesn't have a case assigned to it, which is necessary in English for the NP to be grammatically correct. Now, examining the sentences in box <B> for caseless overt NPs:

- a. "The child is difficult to travel alone." - The sentence is ungrammatical because "the child" does not receive a case from the infinitive "to travel."
- b. "The city's recent dumping the garbage made people upset." - This sentence appears to be grammatically correct as "the city's recent dumping the garbage" functions as the subject with an implied "of" after "dumping."
- c. "I was pleased my sister to be pregnant." - The sentence is ungrammatical; "my sister" lacks a case assigner for the infinitive "to be."
- e. "It turned out the suspect to be innocent." - The sentence is ungrammatical because "the suspect" does not have a case marker for the infinitive "to be."

The sentences that have a caseless (overt NP, based on the criteria given, would be a, c, and e. From the provided multiple-choice options, the one that includes a, c, and e is: a, c, e

<Figure 2>에 제시된 2010학년도 31번은 영어학 영역 중 통사론에 관련된 내용의 문항으로, 명사구(noun phrase)의 문법성을 설명하는 격이론(case theory) (Chomsky, 1986)에 대한 지식을 묻고 있다. 해당 문항은 격을 부여 받지 못한 명사구(caseless noun phrase)가 포함이 된 비문을 고르는 문항인데, 격이론과 관련된 배경지식으로 <A>에서는 정형굴절요소, 타동사, 전치사 등의 요소가 명사구에 격을 제공할 수 있는 것으로 제시하고 있다. ChatGPT는 <B>에서 제시된 보기 중 a. "The child is difficult to travel alone."의 경우 비문인 이유를 명사구인 the child가 비정형동사인 to

travel로부터 격을 제공받지 못해서라고 제시하고 있는데, 해당 문장에서 the child는 be동사와 결합된 정형굴절요소로부터 격을 부여 받을 수 있는 것을 알 수 있다. 즉, 해당 문항 풀이의 오류는 ChatGPT가 문항의 발문을 부분적으로 반영하여 일어난 것이라고 볼 수 있다. 해당 문항의 비문성은 격보다는 “It is difficult to travel with the child alone.” 등의 통사적 형태로 제시함으로써 해결될 수 있을 것이다.

## 2) 문항에서 요구하는 영역-특수적 지식의 결여

기존 ChatGPT 3.5와 마찬가지로 ChatGPT 4.0 또한 영역-특수적 지식이 부족하여 문제 풀이에 난항을 겪는 경우가 있었음을 확인할 수 있었다.

[A]

(In a class, a teacher and students are talking about what makes them unhappy.)

S1: When it rain I am unhappy.

T: Why does rain make you unhappy?

S1: When I walk to school, I wet.

T: Oh, I see. Do you live close to school?

S1: Yeah. 10 minute.

(S2 raises a hand.)

S2: When mosquitoes ... eat me.

T: Oh, mosquitoes bite me. Bite me!

S2: Oh, OK. When mosquitoes bite me. They bite me everyday's evening.

(S3 interrupts the conversation.)

S3: Yeah, me, too. Everyday's evening, mosquitoes bite me, too.

T: That's terrible. I think mosquitoes make many people unhappy. T: teacher, S: student

[B]

a. Display questions are used by the teacher.

b. An uptake occurs after the teacher's corrective feedback.

c. Students are exposed to the interlanguage of other students.

d. A clarification request is used by the teacher for meaning negotiation.

Choose all and only the correct statements about the classroom interaction in <A> from the list in <B>.

① a, b ② a, c ③ b, c ④ b, d ⑤ c, d (정답: ③, ChatGPT: ②)

Figure 3

Question #28 on the 2011 Teacher Certification Exam: English Major (Korea Institute for Curriculum and Evaluation, 2010)

<Figure 3>에 제시된 2011학년도 28번 문항은 <A>에 제시되어 있는 교실 상황에서의 교사와 학생 사이의 대화를 분석하여 <B>의 각 선택지에서 설명하고 있는 영어교육론의 특정 개념들(예: display question, uptake, corrective feedback 등)을 이해하고 있는지에 대해 묻는 문항이다. 아래는 ChatGPT의 해설이다:

From the classroom interaction described in box <A>, let's assess the statements in box <B> to identify the correct ones:

a. Display questions are used by the teacher.

This is correct. The teacher asks questions such as “Why does rain make you unhappy?” which are display questions meant to elicit specific language practice or check comprehension. (이하 생략)

ChatGPT의 해설을 분석해본 결과 ChatGPT는 <B>에 제시되어 있는 영어교육론 개념들 중 하나인 전시형 질문(display question)에 대하여 정확하게 알고 있지 못하는 것을 확인할 수 있었다. 영어교육론적인 측면에서 전시형 질문이란, 교수자는 미리 답을 알고 있으나 학습자가 정답에 대하여 인지하고 있는지 또는 이를 대답할 수 있는지 확인하고자 하는 질문을 일컫는다(Chaudron, 1988). ChatGPT는 <A>의 대화에 교사가 학생에게 특정 표현을 연습시키고 학습자가 이를 이해하고 있는지 확인하기 위하여 전시형 질문을 사용하였다고 진술하였다(The teacher asks questions such as “Why does rain make you unhappy?” which are display questions meant to elicit specific language practice or check comprehension). 그러나 <A>에서 교사의 질문은 학생의 지식을 측정하기보다는 학생의 생각 등을 묻기 위한 참조형 질문(referential question)으로 볼 수 있다. 따라서 본 문항은 ChatGPT가 특정 영역에서의 지식이 부재하여 발생한 오답 관련 예시라고 볼 수 있다.

[A]	
When neighboring sounds mutually affect each other to merge into a third sound, the process is called coalescent assimilation, which can be found in palatalization in English. This coalescent assimilation or palatalization occurs when ‘a morpheme-final or word-final alveolar obstruent’ is followed by ‘the palatal glide’ in English, merging the two sounds into ‘a palatalized fricative or affricate’. For example, within a word as in <i>architecture</i> , the underlying /t/ in the morpheme final position of <i>architect-</i> and the initial /j/ in the suffix <i>-ure</i> affect each other to merge into the palatalized affricate [tʃ]. This process can also be found across words as in phrase <i>kiss you</i> in fast, casual speech of North American English.	
[B]	
a. You don't accept your failure easily, do you?	d. We should cut down on our spending next year.
b. You seem to be under the delusion that he follows you.	e. He checked his yacht before his departure for Australia.
c. The old class divisions had begun to melt down.	f. After six years, her gracious demeanor became known to everybody.
Which of the following lists all and only sentences that contain the expressions which can undergo the coalescent assimilation described in <A> both within a word and across words?	
① a, c, d ② a, d, e ③ b, c, f ④ b, e, f ⑤ c, d, e (정답: ④, ChatGPT: ②)	

Figure 4

Question #33 on the 2012 Teacher Certification Exam: English Major (Korea Institute for Curriculum and Evaluation, 2011)

<Figure 4>에 제시된 2012학년도 33번 문항은 <A>에서 설명하는 상호동화(coalescent assimilation) 또는 구개음화 (palatalization)의 조건을 읽고 <B>에 제시되어 있는 문장들 중에서 전술한 음운변동이 단어 안(within a word) 그리고 단어와 단어 사이(across words)에 둘 다 일어나는 문장만을 고르는 것이었다. ChatGPT의 해설은 다음과 같았다:

Coalescent assimilation, as described in box <A>, is a phonological process where neighboring sounds affect each other and merge into a third sound. This is common in palatalization in English when a morpheme-final or word-final alveolar obstruent (like /t/ or /d/) is followed by a palatal glide (/j/, the sound of “y” in “yes”), resulting in a palatalized fricative or affricate (like [tʃ], the sound of “ch” in “church”).

ChatGPT의 해설을 분석해본 결과 ChatGPT는 <A>의 설명에 기반하여 <B>의 문장들에 잘 적용시키는 것처럼 보였다. 또한 음절의 마지막 또는 단어의 마지막 치경음 중 폐쇄음(alveolar obstruent)이 구개음 중 활강음(palatal glide)을 만났을 때 구개음화 현상이 일어나 [j], [tʃ] 와 같은 마찰음(fricative) 또는 파찰음(affricate)으로 발음된다는 원리를 파악하고 있었다. 치경음 중 폐쇄음에는 /t/와 /d/와 같은 파열음(stop sound)뿐만 아니라 /s/와 같은 마찰음도 속한다. 그러나 ChatGPT의 해설에 의하면 ChatGPT는 치경음 중 폐쇄음에 대한 개념 중 일부(/t/, /d/)만 적용하여 오답을 제시한 것으로 판단해볼 수 있다.

### 3) 풀이한 해설과 다른 답을 고른 경우

ChatGPT 4.0가 정답을 도출할 때 문항에 대하여 자체적으로 제공한 해설과는 다른 선지를 답으로 고르는 경우를 발견하였다.

The motor road hastily joins the railroad and runs beside it for a quarter of a mile, so as to shrink away from a certain desolate area of land. This is a valley of ashes — a fantastic farm where ashes grow like wheat into ridges and hills and grotesque gardens. Occasionally a line of grey cars crawls along an invisible track, gives out a ghastly creak, and comes to rest, and immediately the ash-grey men swarm up with leaden spades and stir up an impenetrable cloud, which screens their obscure operations from your sight. But above the grey land and the spasms of the bleak dust which drift endlessly over it, you perceive, after a moment, the eyes of Doctor Emerson. They are blue and gigantic — their retinas are one yard high. They look out of no face, but, instead, from a pair of enormous yellow spectacles which pass over a non-existent nose. Evidently some wild wag of an oculist set them there to fatten his practice, and then sank down himself into eternal blindness, or forgot them and moved away. But his eyes, dimmed a little by many painless days, under sun and rain, brood on over the solemn dumping ground. The valley of ashes is bounded on one side by a small foul river, and, when the drawbridge is up to let barges through, the passengers on waiting trains can stare at the dismal scene for as long as half an hour. There is always a halt there of at least a minute, and it was because of this that I first met Tom's girlfriend. "We're getting off," he insisted. "I want you to meet my girl." I followed him over a low-whitewashed railroad fence, and we walked back a hundred yards along the road under Doctor Emerson's persistent stare. The only building in sight was a small block of yellow brick sitting on the edge of the wasteland, a sort of compact Main Street ministering to it, and contiguous to absolutely nothing.

Which of the following is NOT a proper understanding of the passage?

① The first paragraph describes a dumping ground.  
 ② 'A fantastic farm' is an example of verbal irony.  
 ③ 'The eyes of Doctor Emerson' refers to an advertisement.  
 ④ 'The only building' is connected with the image of ashes.  
 ⑤ The passage above shows a Gothic atmosphere. (정답: ⑤, ChatGPT 답: ③)

Figure 5

Question #39 on the 2010 Teacher Certification Exam: English Major (Korea Institute for Curriculum and Evaluation, 2009)

<Figure 5>에 제시된 2010학년도 39번 문항은 소설 '위대한 개츠비' 중 2번째 장인 'The valley of ashes'의 내용 중 일부를 발췌한 영미문학 문제이다. 해당 문항은 수험생들이 문학 작품을 읽고 작품 속 등장하는 다양한 키워드들의 함축적인 의미를 문맥을 통하여 올바르게 해석할 수 있는 능력을 가지고 있는지 묻는 내용이다. 해당 문항은 5개의 보기 중 지문에 대해 적절하지 않은 해석을 나타낸 보기를 고르는 문항인데, ChatGPT는 정답인 (5)번 "The passage above shows a Gothic atmosphere"가 아닌 'The eyes of Doctor Emerson' refers to an advertisement.' 를 정답으로 제시하였는데 관련 설명은 다음과 같다.

This is because the text describes "the eyes of Doctor Emerson" as "blue and gigantic" and suggests that they are not attached to any face but instead, they look out of no face but, instead, from a pair of enormously yellow spectacles, indicating they are likely part of a billboard or similar advertisement rather than a literal description of a person's eyes. Additionally, it mentions that an oculist set them there, which further implies that it is an advertisement for an eye doctor's practice.

‘The eyes of Doctor Emerson’은 소설 속 내용에서 광고 표지판을 뜻한다. ChatGPT가 제공한 해설에 의하면 ‘The eyes of Doctor Emerson’ 구절을 말 그대로의 Emerson 박사의 눈이 아닌 이를 과랴고 거대하게 (blue and gigantic) 광고판에서 묘사한 광고 표지판을 표현한 것이라고 하였다. 즉, ChatGPT는 정확하게 작품을 해석한 것으로 추측된다. 하지만 ChatGPT는 올바른 해석을 하였음에도 불구하고 선지 3번을 옳지 않은 문항으로 선택하였다. 이는 ChatGPT가 답안을 제시하는 과정에서 착오가 발생하여 생긴 오류로 추측된다.

[A]

The adverbials briefly and wisely in sentences (1a) and (1b) are not integrated into the sentences in which they occur and they are understood as being superordinate to the sentences. Thus, sentences (1a) and (1b) can be paraphrased as (2a) and (2b), respectively.

(1) a. Briefly, there is nothing I can do to help.  
       b. Wisely, Mrs. Jensen consulted her lawyer.

(2) a. If I put it briefly, there is nothing I can do to help.  
       b. It was wise for Mrs. Jensen to consult her lawyer.

Adverbials of this kind are called DISJUNCTS. Notice, however, that the way the sentences are paraphrased is not the same. As we can see in (2), the adverbial briefly conveys the speaker’s comment on the style and form of what is being said, while the adverbial wisely conveys the speaker’s comment on the actual content of the utterance and on its truth condition. Adverbials of the former type are called STYLE disjuncts and adverbials of the latter are called CONTENT disjuncts.

[B]

a. With respect, he doesn’t deserve the award.	d. Fortunately, his application has arrived in time.
b. To my regret, he wasn’t hesitant to help her.	e. Stupidly, Bill dropped out of college.
c. Honestly, the movie was a little boring.	

Which of the following lists all and only sentences with a style disjunct adverbial in <B>?

① a, c ② a, d ③ b, c, e ④ b, d, e ⑤ d, e (정답: ①, ChatGPT 답: ④)

Figure 6

Question #30 on the 2012 Teacher Certification Exam: English Major (Korea Institute for Curriculum and Evaluation, 2011)

<Figure 6>에 제시된 2012학년도 30번 문항은 위 문항은 <A>에서 설명하는 이접어(disjunct)가 무엇인지, 이접어의 두 가지 종류인 문체적 이접어(style disjunct)와 내용적 이접어(content disjunct)에 대한 개념을 올바르게 이해하여야 풀 수 있는 문항이다. 이해를 돕기 위하여 <A>에서는 문체적 이접어의 예시인 briefly를 제시하였고 내용적 이접어의 예시로는 wisely를 제시하였다. 또한 <A>에서 예문을 들며 문체적 이접어인 briefly같은 경우 화자의 말투가 ‘간결하다는 것’을 전달하고 있지만 내용적 이접어인 wisely와 같은 경우 문장의 내용이 ‘현명하다는 것’을 전달하고

있으므로 이 둘의 차이점을 파악해야 풀이할 수 있는 문항이었다. 해당 문항과 관련된 ChatGPT의 해설은 다음과 같았다:

Based on the explanation in box <A>, style disjuncts are adverbials that convey the speaker's comment on the style and form of what is being said. Content disjuncts, on the other hand, comment on the actual content of the utterance and its truth condition.

Now, looking at the sentences in box <B>, we need to identify which ones contain a style disjunct:

- a. "With respect, he doesn't deserve the award." - This could be seen as commenting on the manner in which the speaker is addressing the issue, thus could be a style disjunct.
- b. "To my regret, he was hesitant to help her." - This is more about the speaker's feelings about the content of what is being said, so it is likely a content disjunct.
- c. "Honestly, the movie was a little boring." - This is commenting on the speaker's sincerity about the content, making it a content disjunct.
- d. "Fortunately, his application has arrived in time." - This is expressing the speaker's attitude towards the event, which could be considered content-related, but it does not comment on the style or manner of expression itself.
- e. "Stupidly, Bill dropped out of college." - This seems to be commenting on the manner of the action taken by Bill, thus could be a style disjunct.

Based on this analysis, the sentences in box <B> that contain a style disjunct are likely a and e.

Therefore, the correct choice would be: 4) b, d, e

ChatGPT의 해설을 살펴보면, ChatGPT는 두 가지 종류의 이접어들(문체적 이접어, 내용적 이접어)의 개념을 구분하여 이해하고 있는 것을 확인할 수 있다. 또한 <B>의 a문장에서의 with respect가 문체적 이접어인 것 또한 정확하게 분석하였다. 그럼에도 불구하고 답에는 a를 포함하지 않았으며 오히려 내용적 이접어로 직접 확인한 b문장과 d문장을 답으로 도출하며 오류를 범하였다. 물론 a에서 e까지의 문장들 속에 제시된 이접어들의 종류를 모두 정확하게 파악한 것은 아니지만(예: e문장의 stupidly는 문체적 이접어가 아닌 내용적 이접어임), 풀이 과정에서 문체적 이접어로 판단한 문장들(예: a, e 문장)을 답으로 설정하였어야 한다. 위 문항의 성격이 선다형 문항임을 고려한다면 ChatGPT가 직접 풀이하여 추려낸 답안들이 선다형에 포함되어 있지 않아 답안 도출 과정에서 생긴 오류로 보인다.

#### 4) 비유적 표현의 함축적인 의미를 파악하지 못한 경우

문학적 해석을 요구하는 '시' 또는 '소설'과 같은 영미문학 문항을 풀이하는 과정에서 연구자들은 ChatGPT 4.0이 난항을 겪는 경우를 발견할 수 있었다.



Up to the age of five I enjoyed the privileges of myopia, seeing the world in a glorious haze like an impressionist painting. My contemporaries appeared obscure and attractive, grown-ups loomed in vague magnificence. I went daily to school and kept my eyes politely on the blackboard where I could see only chalky confusion. After a year of this my mother noticed that my education was at a standstill and sent me to the oculist: The world sprang at me in hideous reality, full of people with open pores, blackheads and impetigo. A deep-focus moustache appeared on an art mistress whom I had considered beautiful. Flinching from this unusual clarity I went to school and sat in my usual place at the morning assembly, unrecognizable in a nose-pinching pair of wire-framed specs. The headmaster, whose awareness of his pupils was always somewhat vague, thought that this bespectacled intruder was a new boy. As I was too shy to disillusion him, I was put back in the bottom class to restart my unpromising academic career. I suppose I had become a new person, one who looked on life and actually saw it; but when faced with anything I am really reluctant to see, like an animal killed and plastered across the road, I still have the defence of taking off my glasses and returning the world to the safe blur of childhood.

Which of the following images would most likely be related to the underlined part?

- ① open pores ② wire-framed specs ③ a deep-focus moustache ④ an impressionist painting ⑤ an animal killed (정답: ④, ChatGPT 답: ②)

Figure 7

Question #39 on the 2009 Teacher Certification Exam: English Major (Korea Institute for Curriculum and Evaluation, 2008)

<Figure 7>에 제시된 2009학년도 39번 문항은 John Mortimer가 저술한 ‘Clinging to the wreckage’의 내용 중 일부를 발췌한 영미문학 문제이다. 이는 화자가 안경을 쓰기 전, 근시를 가진 눈으로 바라보았던 세상의 심상과 안경을 쓴 후 변화된 시야와 심상으로 세상을 바라볼 때 나타난 화자의 심경을 묘사하고 있다. 내용 속 화자는 안경을 쓰기 전 본인의 어렴풋한 시각으로 바라본 세상을 긍정적으로 생각하고 있으며 그 시점을 그리워하고 있다. 위 문제는 전술한 내용에 기반하여 밑줄 친 부분 ‘the safe blur of childhood’의 의미에 대한 해석한 후 함축적으로 같은 심상을 가지고 있는 단어를 올바르게 연관 지을 수 있는 능력을 가지고 있는지 묻고 있다. ChatGPT는 이에 대해 다음과 같은 해설을 제시하였다:

In the story provided, the underlined part refers to the moment the narrator got glasses and saw the world in “hideous reality.” The images that would be most related to this part of the text are those that depict things one could see clearly and in detail with glasses, as opposed to the blurry vision one would have without them. Therefore, the image most likely related to the underlined part would be wire-framed specs because it is the glasses that corrected the narrator’s vision, allowing them to see the world with this new, sharp focus that they found so jarring.

ChatGPT는 밑줄 친 부분을 내용 속 화자가 안경을 쓴 후 세상을 바라보는 시점 이후라고 해석하였다. 그러나 바로 앞 문장(I still have the defence of taking off my glasses and returning the world to)을 살펴보면 안경을 쓰기 전 시점인 어린 유년기(밑줄 친 부분) 시점을 그리워하는 모습을 확인할 수 있다. 따라서 정답은 지문의 첫 문장에 포함되어 있으며 화자가 안경을 쓰기 전

바라본 세상을 묘사한 표현인 ④번 선지(an impressionist painting)이다. 비록 ChatGPT는 밑줄 친 부분의 해석을 글의 내용과는 정반대로 해석하였지만, 주목할만한 점은 밑줄 친 부분을 부정적인 심상으로 해석한 후 내용 속 관련된 또 다른 부정적인 심상을 나타내는 표현과 연관시킬 수 있는 능력을 가지고 있다는 것이었다.

[A]

Every morning, I lay on the floor in the front parlor watching her door. The blind was pulled down within an inch of the sash so that I could not be seen. In the classroom her image came between me and the page I strove to read, and yet her name was like a summons to all my foolish blood. On Saturday evenings when my aunt went marketing I had to go to carry some of the parcels. We walked through the flaring streets, jostled by drunken men and bargaining women, the shrill litanies of shopboys, and the nasal chanting of street singers. These noises converged in a single sensation of life for me: I imagined that I bore my chalice safely through the throng of foes. Her name sprang to my lips at moments in strange prayers and praises which I myself did not understand. My eyes were often full of tears and at times a floor from my heart seemed to pour itself out into my bosom. I did not know whether I would ever speak to her or not or, if I spoke to her, how I could tell her of my confused adoration. But my body was like a harp and her words were like fingers running upon the wires. One evening I went into the back-drawing room. It was a dark rainy evening and there was no sound in the house. Through one of the broken panes I heard the rain impinge upon the earth, the fine incessant needles of water playing in the sodden beds. Some distant lighted window gleamed below me. I was thankful that I could see so little. All my senses seemed to desire to veil themselves and, feeling that I was about to slip from them, I pressed the palms of my hands together until they trembled, murmuring: *O love! O love!* many times.

[B]

narration	interior monologue	a soliloquy or a confession of "I"'s love for "her"
protagonist	the narrator	a shy schoolboy
figurative expression	(1)	"I"'s love for "her"

Which of the following is the most appropriate for (1) in <B>?

① the front parlor ② the flaring streets ③ the shrill litanies of shopboys ④ my chalice ⑤ the broken panes (정답: ④, ChatGPT 답: ③)

**Figure 8**

Question #37 on the 2012 Teacher Certification Exam: English Major (Korea Institute for Curriculum and Evaluation, 2011)

<Figure 8>에 제시된 2012 학년도 37 번 문항은 James Joyce 의 소설 ‘Dubliners(더블린 사람들)’ 속 ‘Araby(애러비)’의 내용 중 일부를 발췌한 문항이다. 소설 속 주인공인 소년의 Mangan 누나에 대한 사랑을 상징적으로 나타낸 표현을 <A> 부분에서 찾아낼 수 있는지 묻는 문항이었다. <A>에서 소년은 소녀에 대한 은밀한 사랑을 키워 가는 과정에서 소년이 자신을 낭만적인 이야기의 주인공으로 상상하고 있다. 내용에 따르면 소년이 아주머니를 따라 시장에 갈 때면, 술주정꾼들(drunken men)과 상인들(bargaining women), 지저분하고 소란한 시장 거리(the flaring

street)를 걸으며 소년은 자신이 성배(chalice)를 운반하고 있다고 생각한다. 소년은 시장에 등장하는 다양한 사람들을 적으로 여기며 이들로부터 성배를 지켜야 한다고 생각한다. 여기서 성배를 조심스럽게 다루는 소년의 모습을 통하여 이는 소년의 소녀에 대한 사랑을 함축적으로 나타내고 있음을 유추해야 풀 수 있는 문항이었다. 이에 대해 ChatGPT 는 다음과 같은 해설을 제시하였다:

After carefully reading the excerpt in box <A> and the information in box <B>, it is clear that the most appropriate option for (1) in box <B> is: the shrill litanies of shopboys

This phrase from box <A> is a figurative expression that reflects the protagonist's heightened emotional state and fits the context of an interior monologue filled with figurative language about the narrator's feelings for "her." It suggests a background cacophony that contrasts with his focused adoration.

ChatGPT는 ③번 선지(the shrill litanies of shopboys)가 소년이 소녀를 통해 감정이 고조되었음을 나타내고 있다고 설명하고 있다. 그러나 전술한 바와 같이 ③번 선지는 소년이 품고 있는 성배에 위협이 되는 존재이다. 흥미로운 사실은 ChatGPT의 해설 중 ③번 선지가 소년의 소녀에 대한 집중된 사랑과는 대조되는 불협화음의 배경을 나타내고 있다(It suggests a background cacophony that contrasts with his focused adoration)는 부분에서 문학적으로 대조되는 것을 제시하여 강조하고자 하는 대상을 부각시키는 반어법에 대한 설명과 유사하다. 물론 ChatGPT의 해석은 정확하지 않았지만 문학에서 자주 등장하는 표현법인 반어에 대한 개념을 부분적으로 알고 있는 것으로 추측되었다. 사전 훈련을 통하여 문학적인 예문이 충분히 숙지가 된다면 정확도가 높아질 것으로 예상된다.

## V. 결론 및 제언

본 연구는 새롭게 개발된 ChatGPT 4.0의 성능을 파악하고자 하였다. 특히 ChatGPT 4.0이 가지고 있는 여러가지 특성들 중 영역-특수적 지식을 요구하는 영역 관련 공신력 있는 시험의 문제 풀이 능력을 측정하고자 하였다. 따라서 본 연구에서는 고부담 시험 중 하나인 중등교원 임용시험 중 전공영어 영역의 문항들로 데이터셋을 구성하였다.

연구 결과를 살펴보면 ChatGPT 4.0은 여전히 영역-특수적 지식에 기반을 둔 문항을 풀이하는데 비교적 제한적인 능력을 가지고 있는 것으로 확인할 수 있었다. ChatGPT 4.0이 발생시키는 오류들은 크게 4가지로 범주화 할 수 있었는데 각 범주들은 향후 ChatGPT 4.0의 능력을 향상시키는데 있어 충분히 고려되어야 할 부분임을 시사한다. 그럼에도 불구하고 ChatGPT 4.0은 일부 문항에서는 상당히 높은 추론 능력을 발휘하는 경우도 확인할 수 있었다. 뿐만 아니라, 본 연구의 연구자들은 ChatGPT가 인문학적인 소양을 필요로 하는 영미 문학 풀이 수행능력은 현저하게 떨어질 것으로 전망하였다. 비록 ChatGPT가 명확한 정답을 제시하지 못하였음에도

불구하고 설명문 이외의 문학 문제 풀이 능력도 일부 갖추고 있음을 확인할 수 있었다. 심지어 일부 문항은 일부 발췌문만 주어졌음에도 불구하고 작품이 무엇인지 명확하게 파악하고 정확한 해설을 제공하며 풀어내기도 하였다. 따라서 연구자들의 예상과는 다르게 ChatGPT가 설명문에 기반을 둔 독해 문항뿐만 아니라 비교적 인문학적인 해석이 필요한 영미 문학 풀이 수행능력을 부분적으로 갖추고 있음을 알 수 있었다.

본 연구의 결과로 미루어 보아 ChatGPT 4.0은 향후 영역-특수적 지식을 필요로 하는 시험의 타당도를 검증할 수 있는 도구로 활용될 수 있는 잠재성을 지니고 있다는 것을 알 수 있었다. ChatGPT 4.0이 오류를 범하는 문항들의 경우 대부분 높은 수준의 추론 능력이 필요한 문항들이었으며 이는 수험자들이 느끼기에도 높은 난이도일 가능성이 농후하다. 더불어 명확한 해설을 제공하지 않는 임용 시험의 특성에 착안하여 ChatGPT 4.0의 특정 문항들에 대한 설명은 시험의 해설 초안으로서 활용될 수 있음을 시사한다. 그러나 문항 풀이 과정에서 범하는 오류들이 아직 빈번하며 정답률 또한 높은 편에 속하지 않기 때문에 ChatGPT의 데이터베이스를 구성하는 모든 절차에는 인간의 개입 및 검토가 여전히 필요함을 알 수 있었다(Ahn, 2023; Shin, 2023). 이러한 본 연구의 결과에 기반하여, 다른 특수목적 영어 분야(항공 영어, 군사 영어)의 경우에도 영역-특수적 지식을 필요로 하는 시험의 개발과 검증에 ChatGPT를 활용할 수 있을 것으로 기대된다.

본 연구는 중등교원 임용 시험의 다양한 영역들 중 구체적인 한 가지 영역(전공영어)만 대상으로 연구를 진행하였다는 점에서 종합적인 결론을 도출하기에는 다소 한계가 있음을 밝힌다. 더불어 분석에 사용된 시험들의 문항이 전부 객관식으로 구성되어 있었기 때문에 다른 형태의 문항 유형(예: 서답형, 주관식)에서의 정답률을 종합적으로 확인할 수 없었다. 따라서 보다 일반적인 결과를 도출하기 위하여 비교적 넓은 유형의 정답률을 파악할 수 있는 후속 연구가 진행되기를 기대한다.

전술하였던 한계점에도 불구하고 본 연구는 영역-특수적 지식이 필요한 영역에서 ChatGPT 4.0의 수행능력을 조사한 최초의 연구들 중 하나라는 의의가 있으며 본 연구의 결과에 기반을 둔 ChatGPT 4.0의 문항 분석 및 풀이 능력을 살펴보는 후속 연구들이 많이 등장하기를 기대한다.

## REFERENCES

- Ahn, Y. Y. (2023). Performance of ChatGPT 3.5 on CSAT: Its potential as a language learning and assessment tool. *Journal of the Korea English Education Society*, 22(2), 119-145.
- Caramancion, K. M. (2023). News verifiers showdown: A comparative performance evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and bard in news fact-checking. *arXiv preprint arXiv:2306.17176*. <https://doi.org/10.48550/arXiv.2306.17176>

- Chaudron, C. (1988). *Second language classroom: Research on teaching and learning*. Cambridge University Press.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use*. Praeger.
- Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2022). How does ChatGPT perform on the medical licensing exams? The implications of large language models for medical education and knowledge assessment. *MedRxiv: The Preprint Server for Health Sciences*, 1-9. <https://doi.org/10.1101/2022.12.23.22283901>
- Jang, H., & So, H-J. (2023). The analysis of research trends and topics about the educational use of ChatGPT. *Journal of Research in Curriculum & Instruction*, 27(4), 387-401. <https://doi.org/10.24231/rici.2023.27.4.387>
- Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28, 15873-15892. <https://doi.org/10.1007/s10639-023-11834-1>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>
- Korea Institute for Curriculum and Evaluation. (2008). *2009 Teacher certification exams: English*. Korea Institute for Curriculum and Evaluation.
- Korea Institute for Curriculum and Evaluation. (2009). *2010 Teacher certification exams: English*. Korea Institute for Curriculum and Evaluation.
- Korea Institute for Curriculum and Evaluation. (2010). *2011 Teacher certification exams: English*. Korea Institute for Curriculum and Evaluation.
- Korea Institute for Curriculum and Evaluation. (2011). *2012 Teacher certification exams: English*. Korea Institute for Curriculum and Evaluation.
- Kwon, S-K., & Lee, Y. T. (2023). Investigating the performance of generative AI ChatGPT's reading comprehension ability. *Journal of the Korea English Education Society*, 22(2), 147-172.
- Lee, J. H., Shin, D., & Noh, W. (2023). Artificial intelligence-based content generator technology for young English-as-a-Foreign-Language learners' reading enjoyment. *RELC Journal*, 54(2), 508-516. <https://doi.org/10.1177/00336882231165060>
- Lee, S., & Song, K. (2023). Exploration of domestic research trends on educational utilization of generative artificial intelligence. *The Journal of Korean Association of Computer Education*, 26(6), 15-27. <https://doi.org/10.32431/kace.2023.26.6.002>
- Li, J., Ren, X., Jiang, X., & Chen, C. H. (2023). Exploring the Use of ChatGPT in Chinese Language Classrooms. *International Journal of Chinese Language Teaching*, 4(3), 36-55. <https://doi.org/10.46451/ijclt.20230303>
- Saeed, J. I. (2009). *Semantics* (3rd ed.). Wiley-Blackwell.
- Shin, D. (2023). A case study on English test item development training for secondary school teachers using AI tools: Focusing on ChatGPT. *Language Research*, 59(1), 21-42. <https://doi.org/10.30961/lr.2023.59.1.21>
- Shin, D., & Lee, J. H. (2023). Can ChatGPT make reading comprehension testing items on par with human experts?

*Language Learning & Technology*, 27(3), 27-40. <https://hdl.handle.net/10125/73530>

Yi, M., Lee, Y., Choi, H., & Choi, I-C. (2023). Predicting the difficulty of college scholastic ability test based on ChatGPT and corpus linguistic features. *Multimedia-Assisted Language Learning*, 26(4), 29-50. <http://doi.org/10.15702/mall.2023.26.4.29>