

수학 문장제 해결과 관련한 ChatGPT의 교수학적 활용 방안 모색

강 윤 지 (서울홍연초등학교, 교사)

최근 인공지능 언어 모델의 다양한 활용에 대한 관심이 높아지면서 수학교육에서의 교수학적 활용 방안 모색에 대한 필요성이 강조되고 있다. 인공지능 언어 모델은 자연어 처리가 가능하다는 특징으로 인하여 수학 문장제 해결과 관련된 활용이 기대된다. 인공지능 언어 모델 중 하나인 ChatGPT의 성능을 확인하기 위하여 초등학교 교과서에 제시된 문장제를 해결하도록 지시하였으며 풀이 과정 및 오류를 분석하였다. 분석 결과, 인공지능 언어 모델은 81.08%의 정답률을 나타내었으며 문제 이해 오류, 식 수립 오류, 계산 오류 등이 발생하였다. 이러한 문장제 해결 과정 및 오류 유형의 분석을 바탕으로 인공지능 언어 모델의 교수학적 활용 방안과 관련된 시사점을 제안하였다.

I. 서론

생성형 인공지능 모델(generative AI model)은 입력된 데이터를 학습하여 새로운 데이터를 생성하는 기술로 다양한 분야에서 활용되고 있다. 그중 ChatGPT는 OpenAI에서 개발한 인공지능 챗봇으로 자연어 처리가 가능하며 여러 인공지능 언어 모델 중 사회, 경제, 의료, 마케팅 등 다양한 분야에서 주목받고 있다. ChatGPT의 성능은 간단한 질의응답에서 상세한 연구에 이르기까지 다양한 측면에서 분석되고 있으며 이를 바탕으로 여러 활용 방향이 기대되고 모색되고 있다. 아울러 인공지능 언어 모델이 다양한 영역을 보조할 수 있다는 가능성에 따라 교육계에서도 ChatGPT 활용 방안에 대한 요구가 높아지고 있다.

최근 몇 년 동안 인공지능은 특정 종류의 언어 작업에서 놀라운 성공을 거두었고, 수학과 관련된 작업에서 주목할 만한 성공을 거두었다(Davis, 2023). 수학 교과는 학생들이 수학적 개념과 방법을 배우는 과정에서 언어, 논리, 관계, 수치적 연결의 활용을 요구하며, 인공지능은 관계와 논리를 감지하는 특징 때문에 수학교육과 학습을 위한 보완 도구로 다양한 가능성이 주목받고 있다. 예를 들어, 일대일 멘토링은 학생 학습에 긍정적인 영향을 미칠 수 있다고 알려져 있으나, 공립학교에서는 학생-교사 비율이 높아 시행이 어렵다(Zong & Krishnamachari, 2023). 인공지능 언어 모델은 개별 기기를 이용한 맞춤형 학습 지원이 가능하기에 이러한 현실적 한계를 보완할 수 있는 대안이 될 수 있다.

ChatGPT의 핵심 기술은 자연어 처리와 생성이며, 대부분 딥러닝 기반의 모델로 구현되며 데이터를 바탕으로 훈련되어 자연어 처리와 생성의 성능을 갖추고 있다. 다만 이러한 능력에도 불구하고 수학교육과 관련하여 몇 가지 한계가 존재한다. 인간과 인공지능 언어 모델이 수학을 주제로 소통하는 상황이라면 단순히 수학을 주제로 대화하는 것에서 더 나아가 수학적 논리와 절차의 적절성 및 정확성까지 점검해야 하기 때문이다. 다시 말하면 수학교육과 관련하여 단순히 대화의 자연스러움이나 유창성만을 고려하는 것이 아니라 그러한 대화의 내용에서 생성되는 자연어가 논리적으로 정확하고 일관성이 있는지 검증해야 한다는 것이다. 예를 들어, ChatGPT가 수학 문제를 풀거나 설명하는 경우 답이 올바른지, 과정이 타당한지, 표기법이 표준에 부합하는지 등을 확인할 필요가 있다. 대표적인 예 중 하나는 수학 표현식의 값을 계산하는 능력이다. ChatGPT는 수학 문제 풀이를 제시할 수

* 접수일(2024년 2월 5일), 심사(수정)일(2024년 3월 2일), 게재확정일(2024년 3월 11일)

* MSC2000분류 : 97U70

* 주제어 : ChatGPT, 문장제, 수학교육, 인공지능

있으며 수학 기호와 공식을 표현할 수 있다(권오남 외, 2023). 그러나 ChatGPT는 큰 수를 곱하거나, 값을 계산하는 등 간단한 연산이 가능하지만, 무리수에 대한 덧셈셈에 어려움이 있으며 복잡한 수학 표현식의 결과로 단지 숫자의 표본을 제시하는 것에 그칠 수도 있다는 한계가 있다(Azaria, 2022). 따라서, ChatGPT가 수학적 개념이나 용어를 사용하는 경우, 정의가 명확한지, 논리가 타당한지, 모순이 없는지 등을 검증하여야 한다(권오남 외, 2023).

더욱이 교육 분야에서 ChatGPT 등 인공지능 언어 모델의 활용 가능성이 기대되는 만큼 인공지능 언어 모델의 교수학적 활용 방안 모색에 대하여 깊은 논의가 필요하다. ChatGPT의 수학 문제해결 기능에 대한 보완이 이루어지고 있고(OpenAI, 2023) 여러 선행연구에서 ChatGPT를 수학교육에 활용하고자 다양한 활용 방안을 모색하고 있으나(손태권, 2023; 강윤지, 2023; 권오남 외, 2023) 수학 교수학습과 ChatGPT의 활용에 대한 기존 선행연구가 부족할 뿐 아니라 도입에 앞서 충분한 신뢰도와 타당도가 확보되지 않았다는 우려가 존재한다(강윤지, 2023; 권오남 외, 2023). 수학교육 현장에서의 적절한 활용을 위하여 ChatGPT의 수학적 성능과 교수학적 활용방안에 대한 점검과 연구가 충분히 뒷받침되어야 한다.

ChatGPT와 같은 인공지능 언어 모델이 텍스트로 접근한다는 특징을 고려하였을 때 수학교육에서 문장제 해결에 대한 활용을 기대할 수 있다. 수학 문장제는 수식이나 기호로만 표현한 유형의 문제가 아니라 문제의 상황을 문장으로 표현한 문제를 의미한다(박장희 외, 2012). 문장제는 문장 독해 과정과 연산 과정을 순차적으로 밟아서 해결해야 하며, 초등학생들의 경우 형식적인 계산을 요구하는 사칙 연산의 오류보다 문장제의 해결 과정에서 상대적으로 많은 오류가 발생한다(이병욱, 안병근, 2008). 이때 문장제 해결의 어려움은 언어적 표현의 이해에 국한되는 것은 아니며 수학적 요인 및 문장제의 맥락 등 문장제를 구성하는 여러 요소가 복합적인 영향을 미친다. 문장제의 수학적 요인은 문장제가 요구하는 수학적 내용, 원리 등을 의미하며 맥락은 문장제의 소재 및 내용과 연결된다. 따라서 이러한 특징을 고려한다면 자연어 처리가 가능한 인공지능 언어 모델이 도움을 줄 수 있을 것이라는 예상이 가능하다. 예를 들어, 문장제의 문장을 학생이 그대로 읽었을 때 이해가 어려운 부분이 있을 경우 해당 문장제의 어떠한 요소가 학생에게 걸림돌이 되는지 파악할 수 있다. 더 나아가 인공지능 언어 모델이 문장제의 언어적 표현을 개별 학생의 학습 성취에 알맞게 변환하여 줄 수 있다.

본 연구는 ChatGPT는 자연어 처리를 수행하는 인공지능 언어 모델이라는 측면에서 접근하여 수학 문장제 해결에 대한 성능을 분석하고 교수학적 활용 방안을 모색하고자 하였다. 기술의 발전으로 인하여 인공지능 언어 모델이 문장제를 해결할 수 있으며 그러한 문제해결 과정에서 나타나는 오류 유형을 분류하고 분석하는 과정을 바탕으로 문장제 해결 학습과 관련하여 교수학적 시사점을 모색할 수 있기 때문이다. 이를 위하여 ChatGPT의 정답률, 정확도, 오류 유형을 확인하고, 이를 수학교육에서 활용하기 위한 교수학적 방안을 모색하고자 하였다.

II. 연구의 배경

1. 이론적 배경

가. 수학교육에서의 인공지능 언어 모델의 활용 가능성 탐색

ChatGPT와 같은 인공지능 언어 모델의 성능을 평가한 선행연구들은 대부분 정확도를 성능 평가의 지표로 활용해 왔다. 이러한 연구는 인공지능 언어 모델의 성능을 확인하기 위하여 고정된 답이 있는 문제를 질문하고 그 답변의 정확도와 풀이 과정을 확인하는 형태로 진행되었다. 수학 문제해결과 관련하여 ChatGPT의 성능을 평가한 국외 연구는 다음과 같다.

Azaria(2022)의 연구는 대형 언어 모델인 ChatGPT가 긴 수학식을 계산하기 어렵다는 것을 보여준다. 하지만

동시에 답변을 제공하기 위해 추가 정보를 요청할 수 있고 응답에 대한 설명을 요청할 수 있다는 가능성을 제시하였으며 질문에 대한 사소한 변경으로 인해 모순되는 응답이 발생할 수 있음을 보여준다. Frieder 외 연구진(2023)의 연구는 수학자의 일상적인 활동에서 발생하는 사용 사례를 시뮬레이션하는 데이터셋을 활용하여 ChatGPT의 수학적 성능을 검증하였다. 그 결과, ChatGPT를 학부 수준의 수학을 수행하는 과정에 추가적인 사용이 가능할 것이라고 예상하였으나 대학원 수준의 난이도에는 어려움이 있다고 하였다. 이를 바탕으로 ChatGPT가 수학 보조 도구로 활용이 가능하지만 전반적인 수학적 성능은 대학원생 수준보다 부족하다는 한계를 확인하였다.

Remoto(2023)는 학생들의 학습과 관련하여 인공지능 모델의 도움과 한계를 분석하였다. 연구 결과, ChatGPT와 같은 인공지능 언어 모델은 문제 해결, 증명, 설명에 관련된 채팅 프롬프트에 응답하는 것 등의 활동을 수행할 수 있었으며 수학적 언어와 적분, 도함수, 극한, 분수, 지수, 구간에 대한 기호의 사용 등을 이해할 수 있었다. 다만, 부정확한 결과를 제공하거나 해결을 위한 잘못된 방법을 제공하기도 하였다. 이러한 결과를 바탕으로 인공지능 모델이 학생들의 학습을 돕고 그들의 학습 진행을 모니터링할 수 있을 것이라고 예상하였다.

국내에서 진행된 연구 중 손태권(2023)의 연구는 분수 문제에 대한 학생의 산출물과 예비 교사와 학생과의 담화 자료를 사례로 선정하여 ChatGPT를 통해 분석하고 수학교육전문가의 분석 결과와 비교하였다. 분석 결과, 학생 자료에 대한 분석은 수학교육전문가의 분석 결과와 유사하게 나타났으며 예비 교사의 자료와 비교하였을 때는 유사점과 차이점이 함께 나타났다. 이러한 결과를 바탕으로 ChatGPT가 수학교육에 제공할 수 있는 잠재력과 제한점을 제시하였다.

강윤지(2023)의 연구는 수학교육의 교수학적 실천과 연구 측면에서 AI 언어 모델 활용 방안을 탐색하고자 국어사전과 AI 언어 모델이 제시하는 수학 용어의 의미 간 일치율을 ChatGPT를 이용하여 비교 분석하였다. 분석 결과를 바탕으로 교수학습 설계 및 평가, 연구에 대한 AI 언어 모델의 다양한 활용 가능성을 인정하였으나 AI 언어 모델의 변동성을 고려하여 연구의 타당도와 신뢰도 확보를 위한 방안을 모색하여야 한다고 하였다.

권오남 외(2023)은 국가수준 학업성취도 평가 및 대학수학능력시험 문제에 대한 ChatGPT의 응답을 정답률, 풀이 과정의 정확도, 오류 유형으로 나누어 분석하였다. ChatGPT의 정답률은 국가수준 학업성취도 평가 37.1%, 대학수학능력시험은 15.97%로 나타났고, 풀이 과정의 정확도는 각각 5점 만점 중 3.44점, 2.49점으로 산출되었으며 절차적 오류와 기능적 오류가 나타났다. 분석 결과를 바탕으로 ChatGPT의 수학적 성능을 판단하는 기준으로 정답률뿐 아니라 풀이 과정의 정확도나 오류 유형까지 복합적으로 고려해야 함을 주장하였다.

이러한 선행연구는 수학교육에서의 ChatGPT의 적용 가능성을 다양한 측면에서 모색하고 있으며 여러 가지 한계가 있음에도 불구하고 수학교육에서의 다양한 활용 가능성이 있음을 인정하고 교수학적 활용 방안을 탐색하고 있다는 공통점이 있다.

나. 수학 문장제 학습과 오류 유형

전통적으로 문장제는 기존에 학교에서 배운 수학적 개념, 규칙 또는 기법을 적용하여 해결할 수 있는 텍스트 문제로 간주된다(Verschafel et al., 2000). 수학에서의 문장제는 언어 기능과 산술 기능이 통합되어 수행되도록 구성된 과제이다(박장희 외, 2012). 문장제는 학교 수학에서 항상 중요한 부분을 차지해 왔으며 수학 학습자들이 직면하는 가장 어려운 유형의 문제 중 하나이다(Verschafel et al., 2020).

수학 문장제는 문제 상황 속에서 하나 이상의 미지의 양에 대한 질문을 제기하는 문장의 형태를 취하며, 일반적인 일상생활 상황을 반영하는 맥락에서 주어진 정보를 사용하여 식 또는 답을 도출할 것을 요구하는 형태로 제시된다. 이러한 문제를 해결하려면 세 종류의 지식이 필요하다(Davis, 2023). 첫째, 기초 수학이다. 둘째, 언어이다. 만약 문장제가 다른 언어를 사용하는 사람에게 제시되면 아무리 수학을 잘 안다고 해도 문제를 해결할 수 없다. 셋째, 세계에 대한 기초 지식과 세계가 수학적으로 어떻게 특징지어지는지에 대한 이해가 필요하다. 이

는 문장제의 맥락 이해에 대한 중요성을 보여준다. 이러한 관점에서 접근하면 인공지능 언어 모델은 문장제를 해결하기 위하여 언어적 표현 속에서 수학적 정보와 개념, 맥락을 추출하여 문제에서 요구하는 유효한 수학 표현으로 번역하는 과정을 거쳐야 한다는 결론이 가능하다.

문장제를 해결하기 위하여 언어적으로 문제의 구조를 인지한 다음 적절한 알고리즘을 선택하여 문제해결 과정에 적용하여 빠르게 계산하기까지의 과정이 요구된다. 즉, 문장제를 해결하기 위하여 문제의 내용을 분석하고, 그 상황을 수식으로 표현하고 연산을 수행하는 능력이 필요하다. 이를 해결하기 위해서는 수학적 개념 숙지, 연산 능력, 수학적 사고력 등 종합적인 문제해결력이 요구된다. 그렇기에 이러한 문장제는 일반적인 수식이나 기호만으로 구성된 문제에 비하여 해결 과정 중 오류가 발생할 가능성이 높다. 문장제의 해결 과정 중 발생하는 오류는 한 번 발생하면 후속 학습으로 이어지기가 어렵고, 학생들에게 수학에 대한 자신감과 학습 의욕을 상실시킬 수 있다. 따라서 교사가 자주 발생하는 오류의 유형을 미리 파악하여 학생들에게 적절한 교수-학습 방법을 계획하여 지도한다면 오류를 예방할 수 있고, 학생들에게 효과적인 학습이 이루어질 것이다(박장희 외(2012)). 문장제의 해결 과정 중 나타나는 오류와 관련된 선행연구는 다음과 같다.

석경희, 백석운(2004)의 연구는 학생들의 문장제 해결 과정에서 어떠한 오류가 나타나는지를 알아보기 위해 초등학교 6학년 학생의 문제 풀이 과정을 분석하였다. 오류의 유형은 문제의 문장에 대한 오해, 전략의 성급한 선택, 잘못 이해된 수학 내용의 사용, 계산 과정의 오류, 풀이를 시도하지 않음, 검토되지 않은 해답 등의 6가지로 나타났다. 이를 바탕으로 효과적인 문장제 해결 학습을 위하여 해를 구하는 것뿐만 아니라 문제를 구조화하고 바람직한 전략을 생각하여 실행하도록 이끌어야 한다고 하였다.

이병옥, 안병곤(2008)의 연구는 초등학교 2학년 수학 교과서에 수록된 문장제의 문장 구조와 2학년 학생의 오류 경향을 분석하였다. 분석 결과, 문장제의 문장 구조는 주로 간단하고 반복적인 어휘를 사용하며, 실생활 소재가 반복적으로 등장하였다. 문장제의 오류 중 문장 해석상의 오류가 계산상의 오류보다 많이 나타났으며 이는 역연산 유형, 문장의 길이, 사용된 숫자의 양 등이 해석상의 오류에 영향을 미친 것으로 분석되었다. 이러한 결과를 바탕으로 학생들의 시각에서 문장제 구성을 고려할 필요가 있으며 기초적인 문장 해석 지도가 필요함을 확인하였다.

박장희 외(2012) 연구진은 중학교 3학년의 문장제 해결 과정에서 발생하는 오류 유형 및 발생 원인을 분석하였다. 연구 결과, 문항 이해의 부족, 풀이 과정의 오류, 정리나 정의에 대한 왜곡된 이해, 이기과정의 오류, 기술적 오류, 풀이 과정 생략 등의 오류가 나타났으며 일부 학생들은 문장제 해결에 대한 부담감으로 인해 문제 해결을 포기하는 경향을 보였다. 이를 바탕으로 문장제 해결을 위하여 문제에 대한 충분한 이해가 필요하다고 주장하였다.

이러한 선행연구는 학생들의 문장제 해결 과정 중 나타나는 오류 유형을 살펴봄으로써 효과적인 문장제 학습을 이끌고자 하였다. 선행연구를 통하여 교사가 학생들이 문장제 문제 해결 과정에서 발생하는 오류의 원인과 유형을 파악하여 이를 보완할 수 있는 방안을 고안한다면 오류의 발생을 줄이는 방향으로 학습을 지도할 수 있으리라 예상할 수 있다.

다. 수학 문장제 학습과 인공지능 언어 모델의 활용

문장제는 일반적으로 학문적 환경에서 제시되는 문제 상황에 대한 언어적 설명으로 정의되며, 문제 진술에서 사용할 수 있는 수치 데이터 또는 이로부터 파생된 수치 데이터에 대한 수학적 연산을 적용하여 답을 얻을 수 있는 하나 이상의 질문을 제기한다(Verschafel et al., 2000). 학생들은 일반적으로 자연어 표현을 하나 이상의 대수 방정식으로 변환하고, 언급된 질문에 대한 답을 찾아야 한다(Zong & Krishnamachari, 2023).

인공지능 언어 모델은 웹에서 다운로드한 방대한 텍스트 데이터를 활용한다. 이 데이터에 대한 자세한 설명은 발표되지 않았지만, 대부분이 다양한 종류의 영어 문서로 구성되어 있으며 다른 언어, 대중적인 컴퓨터 언어

로 작성된 소프트웨어, 수학 표기법의 수학, 이미지 및 기타 종류의 데이터로 된 상당한 양의 텍스트도 포함되어 있다(Davis, 2023). 그 결과, Chat GPT와 같은 대형 언어 모델은 문장제를 구성하는 세 가지 요소인 수학, 언어, 맥락에 어느 정도의 지식을 가지고 있다. 그러나 이러한 요소를 종합하여 문제를 해결하는 것은 여전히 도전적인 과제이다.

Davis(2023)의 연구는 수학 문장제를 해결하기 위해 인공 지능을 사용하는 것에 대하여 다루면서, AI 기술을 수학 문장제에 적용하는 최근 시스템에 초점을 맞추고 있다. AI에게 수학 문장제의 풀이를 요구하면 간단한 수학에서는 잘 작동할 수 있지만, 좀 더 복잡한 수학이 필요할 때는 적절한 답변을 구하기 어려울 수 있다고 언급하였다. 이를 바탕으로 현재의 AI 기술로는 기초 지식과 상식적 추론을 결합한 단어 문제를 안정적으로 해결할 수 없다고 하였다.

Zong & Krishnamachari(2023)의 연구는 ChatGPT의 활용을 통해 선형방정식과 관련된 도전 과제를 평가하였다. 분석 결과, ChatGPT는 문장제를 분류하는 데 높은 정확도를 보이며, 제공된 예제 수에 따라 문장제를 방정식 체계로 변환할 수 있었고 주어진 예시를 사용하여 유사한 문제를 생성할 수 있었다. 이를 바탕으로 하여 문장제와 관련된 인공지능 언어 모델의 활용 가능성에 대하여 긍정적으로 평가하였다.

문장제와 관련된 이러한 선행연구는 문장제 학습과 관련된 인공지능 언어 모델의 가능성과 함께 문장제와 인공지능 언어 모델을 관련짓는 연구의 필요성에 대하여 주장하고 있다. 문장제는 언어적 기술에 근거하여 필요한 연산의 표상을 구성해서 문제해결을 하도록 지시하며 이때 문제 해결은 하나의 과정 또는 문제를 해결하는 데 사용되는 일련의 행동을 가리킨다(박장희 외, 2012). 인공지능 언어 모델은 실시간으로 발전하고 있으며 변화하고 있다. 본 연구는 이러한 측면에서 대표적인 인공지능 언어 모델 중 하나인 ChatGPT를 활용하여 문장제의 해결 과정 및 결과를 탐색하고 교수학적 활용 방안을 모색하고자 한다.

2. 연구방법 및 절차

가. 연구 대상

2015 개정 교육과정에서 반영된 국정 교과서 1종, 검정 교과서 10종에서 제시하는 비와 비율, 비례식과 비례배분 단원에 해당하는 단원평가 문장제를 대상으로 선정하였다. 여러 가지 영역 중 비와 관련된 단원을 선정한 것은 해당 단원 고유의 특성을 고려하였기 때문이다. 수와 연산 단원의 경우 인공지능 언어 모델이 문장제에 포함된 수학적 정보만을 추출하여 기계적으로 계산할 우려가 있다. 도형 및 측정, 자료와 가능성 단원의 경우 시각적 표현을 동반하는 경우가 많으며 삽화, 표 등의 시각적 표현이 문제해결을 위한 결정적 정보를 담고 있는 등 텍스트 기반의 인공지능 언어 모델에게 문제 풀이를 지시하기 적절하지 않다. 반면, 비와 비율과 관련된 문제는 오롯이 문장만으로 문제의 내용을 서술할 수 있으며 해당 문장제에 대한 충분한 이해 없이 문제를 풀기 어렵다는 특징이 있다.

해당 단원의 평가 문제 중 언어로 서술되어 있으며 문제에서 각각의 맥락을 포함하는 것을 대상으로 선별하였다. 이때, 텍스트를 대상으로 하는 인공지능 언어 모델의 특성을 고려하여 삽화나 표를 포함하고 있으며 해당 문장제를 해결하기 위하여 삽화나 표가 중요한 수학적 정보를 내포하고 있는 경우는 대상에서 제외하였다. 이는 [그림 II-1]과 같이 삽화에 대한 파악 없이 문제의 해결이 불가능한 상황을 의미한다. 반면 [그림 II-2]와 같이 문제에서 삽화를 제시하더라도 문제의 해결이 가능한 경우는 대상에 포함하였다.



[그림 II-1] 국정 6학년 2학기 수학 교과서 4.비례식과 비례배분 5번(교육부, 2020b, p.87)

[그림 II-2] G출판사 6학년 2학기 수학 교과서 4.비례식과 비례배분 6번(장혜원 외, 2023, p.99)

연구 대상으로 선정한 비와 비율 및 비례식과 비례배분 단원의 문장제는 <표 II-1>과 같다. A 출판사의 교과서와 C 출판사의 교과서에서는 비와 비율 단원에서 연구 대상에 포함되는 문제가 제시되지 않았다. 이는 해당 출판사의 교과서에서 비와 비율 및 비례식과 비례배분 단원에서 문장제를 제시하지 않은 것이 아니라 해당 교과서에 제시된 문장제가 삽화 및 표에 결정적 정보를 제시하고 있어 연구 대상의 조건에 부합하지 않았기 때문이다. 공정성을 확보하기 위하여 출판사명은 익명으로 표기하였다.

<표 II-1> 비와 비율 및 비례식과 비례배분 단원의 분석 문항 수

	국정	A	B	C	D	E	F	G	H	I	J	계
비와비율	3	0	2	0	1	1	1	2	1	2	3	16
비례식과 비례배분	1	3	2	2	2	2	2	2	2	1	2	21
	4	3	4	2	3	3	3	4	3	3	5	37

나. 연구 방법

인공지능 언어 모델을 이용하여 문장제를 해결하는 방법은 세 가지가 있다(Davis, 2023). 첫 번째는 가장 간단한 방법으로 단순히 인공지능 언어 모델에게 문제에 대한 답을 제공하도록 요청하는 것이다. 두 번째는 인공지능 언어 모델이 컴퓨터 코드를 생성하도록 하는 것이며, 이를 실행하여 답을 계산할 수 있다. 세 번째는 자연어로 주어진 수학적 사양을 공식 사양으로 자동 변환하여 공식 검증 시스템에 제공하는 과정인 자동 공식화를 중심으로 구축하는 것이다. 본 연구에서는 더욱 대중적인 접근과 활용을 기대하기에 첫 번째 방법을 선택하였다.

ChatGPT 3.5에게 각 문항에 대한 답변을 5회 요구하였다. 이는 ChatGPT를 비롯한 대형 언어 모델의 답변이 실시간으로 구성되는 특성으로 인하여 필연적으로 변동성을 내포하고 있음을 고려한 것이다. 연구의 신뢰도를 높이기 위하여 각각의 문제는 5회 질문되었고 각각의 답변을 확인하였으며 각각의 풀이에 대한 정답률, 풀이 과정의 정확도 및 오류 유형을 분석하였다. 이는 정답률만이 ChatGPT의 수학적 성능을 판단하는 기준이 되어서는 안 되며, 풀이 과정의 정확도나 오류 유형까지도 복합적으로 고려해야 한다는 선행연구(권오남 외, 2023)의 관점을 고려한 것이다. 현재 ChatGPT 4가 개발되어 사용이 가능하지만 해당 모델의 경우 추가 비용의 지출이 필요하다. 따라서 더 많은 사람이 추가 비용 없이 사용할 수 있는 대중성을 고려하여 ChatGPT 3.5를 연구 도구로 활용하였다.

비와 비율, 비례식과 비례배분의 두 단원으로 나누어 각각의 정답률을 확인하였으며 답이 옳은 경우는 풀이가 옳은지, 풀이가 옳지만 6학년 교육과정을 벗어나지 않았는지, 풀이가 옳지 않은지 등을 고려하였다. 답이 옳지 않은 경우도 동일하게 세 가지 경우를 고려하였다. 다만, 답이 옳지 않으며 풀이가 옳지 않은 경우는 선행연구(석경희, 백석운, 2004; 박장희 외, 2012)를 참고하여 오류 유형을 범주화하였다. 문제 이해 오류는 문제의 정보나 문제에서 요구하는 바를 바르게 파악하지 못하여 발생하는 오류이다. 바르지 못한 전략을 선택하는 경우도 문제 이해 오류에 포함하였다. 식 세우기 오류는 문제를 이해하였더라도 문제에 포함된 수학적 정보나 개념, 지식을 바르게 활용하지 못하여 식을 바르게 수립하지 못하여 발생하는 오류이다. 계산 오류는 식의 해결 과정에서 숫자를 바꾸어 쓰거나 사칙연산의 결과가 그릇되게 나타나는 등 계산 과정에서 발생하는 오류이다.

5회 조사 중 답이 옳지 않은 경우의 풀이에서 한 가지 이상의 오류가 발생하는 경우가 나타났다. 예를 들어, 문제를 이해하지 못하여 식을 잘못 세우거나, 식을 잘못 세우고 식의 계산까지 오류가 발생하는 경우가 나타났다. 이처럼 한 가지 이상의 오류가 나타나는 경우 이를 복수로 인정하여 각각의 개수를 모두 포함하였다. 분석 문항의 평가 기준을 정리하면 <표 II-2>와 같다.

<표 II-2> 분석 문항의 평가 기준

답이 옳은 경우(T)	풀이가 옳은 경우	T1	
	풀이가 옳으나 초등학교 6학년 교육과정을 벗어나는 경우	T2	
	풀이가 옳지 않은 경우	T3	
답이 옳지 않은 경우(F)	풀이가 옳은 경우	F1	
	풀이가 옳지만 초등학교 6학년 교육과정을 벗어나는 경우	F2	
	풀이가 옳지 않은 경우	문제 이해 오류	F31
		식 세우기 오류	F32
계산 오류		F33	

III. 연구 결과 및 논의

1. ChatGPT의 정답률 분석

연구 대상으로 선정된 비와 비율 및 비례식과 비례배분 단원 내 37개의 문장제를 ChatGPT를 활용하여 5회 반복하여 해결하였으며 총 횟수는 185회로 나타났다. 단원별로 ChatGPT의 정답률을 정리하여 나타낸 결과는 <표 III-1>과 같다.

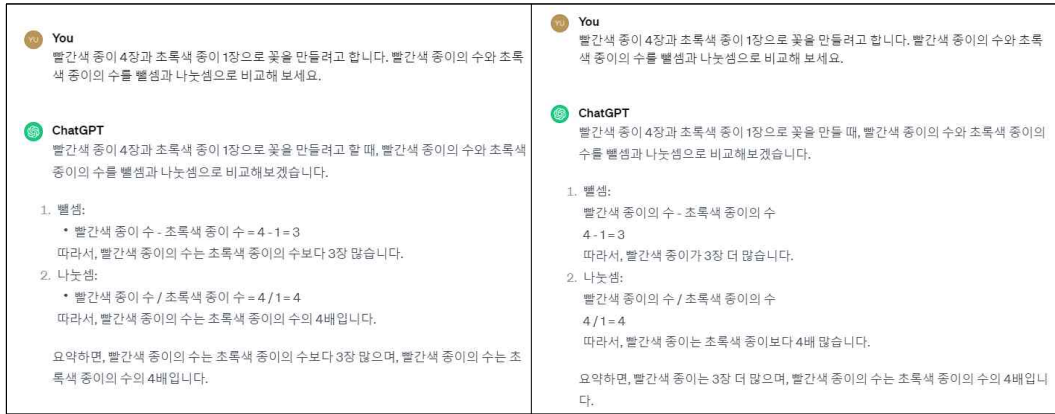
<표 III-1> 비와 비율 및 비례식과 비례배분 단원 문항에 대한 ChatGPT의 정답률

	답이 옳은 경우(T)	답이 옳지 않은 경우(F)	계
비와비율	72(90%)	8(10%)	80(100%)
비례식과 비례배분	78(74.29%)	27(25.71%)	105(100%)
	150(81.08%)	35(18.92%)	185(100%)

단원별로 살펴보았을 때 비와 비율 단원에서 나타난 정답률(90%)보다 비례식과 비례배분 단원의 정답률(74.29%)이 상대적으로 낮게 나타났다. 비와 비율 단원에서는 비와 비율과 관련된 개념이 처음으로 도입되며 단원 평가의 문장제 또한 개념을 확인할 수 있도록 구성되어 있다. 반면, 비례식과 비례배분 단원에서는 비와 비율 단원에서 학습한 내용을 바탕으로 내용이 전개되고 있으며 복잡한 맥락의 문제가 등장하였던 것이 영향을 미친

것으로 보인다.

ChatGPT는 [그림 III-1]에서 교과서의 문장제가 요구하는 대로 빨간색 종이의 수와 초록색 종이의 수를 뺄셈과 나눗셈으로 각각 해결하였다. 뺄셈으로 비교한 경우 빨간색 종이의 수는 초록색 종이의 수보다 3장 많았으며 나눗셈으로 비교한 경우 빨간색 종이의 수는 초록색 종이의 수의 4배로 나타났다. 동일한 문제를 해결한 [그림 III-2]에서도 문장제를 뺄셈과 나눗셈으로 해결하는 흐름이나 풀이의 내용이 유사함을 알 수 있다. 언어적 표현의 미세한 차이는 존재하지만 문제를 해결하기 위하여 필요한 요소를 파악하고 해결하는 과정이 비슷하다.



[그림 III-1] 국정 6학년 1학기 수학 교과서 비와 비율 1번(교육부, 2020a, p.88) 옳게 해결한 경우(1)

[그림 III-2] 국정 6학년 1학기 수학 교과서 비와 비율 1번(교육부, 2020a, p.88) 옳게 해결한 경우(2)

이외에도 문장제를 바르게 해결하였지만 초등학교 6학년 교육과정의 범위를 벗어나는 방법으로 해결하는 경우가 나타났다. [그림 III-3]은 비례식의 성질을 이용하여 3000원으로 사과를 2개 살 수 있다면 7500원으로는 사과를 몇 개 살 수 있는지 구해 보라고 지시하고 있다. 이 때 인공지능 언어 모델은 초등학교 6학년의 범위를 벗어나는 수학 기호와 방법을 사용하여 문제를 해결하고 있다. 학생들은 비와 비율, 비례식과 비례배분에 대하여 학습하였더라도 비례 기호 ‘ \propto ’를 활용하여 문제를 해결하는 방법을 학습하지 않는다. 학생들은 ‘가격= $k \times$ 수량’의 식을 활용하여 문제를 해결하는 것이 아니라 관계를 이용하여 문제를 해결할 수 있도록 학습한다. [그림 III-4]에서도 7500원을 주고 사과를 다섯 개 샀을 때, 3000원으로 살 수 있는 사과의 수를 구하는 과정에서 ‘ \approx ’라는 기호를 사용하여 나타내고 있다. 이 또한 초등 과정에서 사용하지 않는 수학 기호를 활용하여 문제를 해결한 방법이다.

문제를 해결하는 과정이 옳거나 옳지 않은 경우는 초등 과정에서 사용하지 않는 개념이나 기호, 방법 등을 활용하는 풀이가 나타났다. 이처럼 ChatGPT의 문제해결 과정에 초등 교육과정의 범위를 벗어난 방법이 등장하는 것은 인공지능 언어 모델의 수준이 초등학생보다 높으며 그로 인하여 더 다양하거나 수준 높은 방법을 활용하는 것이 가능하다는 것을 보여준다. 이러한 경우 인공지능 언어 모델에게 명령 프롬프트를 활용하여 초등학교 6학년 학생의 역할을 부여하거나 초등 수학교육과정 범위 내에서 제시된 문제를 해결하라는 등의 추가 지시를 통하여 문제해결 방법의 방향과 종류를 제한할 수 있다.

<p>You 3000원으로 사과를 2개 살 수 있습니다. 7500원으로는 사과를 몇 개 살 수 있는지 비례식의 성질을 이용하여 구해 보세요.</p> <p>ChatGPT 비례식은 두 양이 비례할 때 다음과 같이 나타낼 수 있습니다: 가격 \propto 수량 이를 수식으로 표현하면: 가격 = $k \times$ 수량 여기서 k는 비례 상수입니다. 주어진 정보에 따라 첫 번째 상황에서는 3000원으로 2개의 사과를 살 수 있습니다. 따라서: $3000 = k \times 2$ 이를 풀면 $k = 1500$이 됩니다. 이제 두 번째 상황에서의 수량을 찾아보겠습니다. 두 번째 상황에서의 가격은 7500원이므로: $7500 = 1500 \times$ 수량 이를 풀면: 수량 = $\frac{7500}{1500} = 5$ 따라서, 7500원으로는 사과를 5개 살 수 있습니다.</p>	<p>You 7500원을 주고 사과를 5개 샀습니다. 2000원으로는 사과를 몇 개 살 수 있는지 비례식을 세워 구해 보세요.</p> <p>ChatGPT 비례식을 세우기 위해서는 주어진 정보를 이용하여 비례식의 상수 비율을 찾아야 합니다. 주어진 정보에 따르면 7500원으로 5개의 사과를 샀습니다. 이를 이용하여 비례식을 세울 수 있습니다. 사과의 가격을 P라고 하면, $\frac{7500}{P} = \frac{5}{1}$ 이 비례식을 풀면, $P = \frac{7500 \times 1}{5} = 1500$ 따라서, 사과 한 개의 가격은 1500원입니다. 이제 주어진 2000원으로 몇 개의 사과를 살 수 있는지 구해보겠습니다. $\frac{2000}{P} = \frac{X}{1}$ 이 비례식을 풀면, $X = \frac{2000 \times 1}{1500}$ $X \approx 1.33$ 따라서, 2000원으로는 약 1.33개의 사과를 살 수 있습니다. 일반적으로 사과의 개수는 정수이므로, 가장 근사한 값으로 반올림하여 1개의 사과를 살 수 있습니다.</p>
--	---

[그림 III-3] A출판사 6학년 2학기 수학 교과서 비례식과 비례배분 5번(박만구 외, 2023, p.89) 바르게 풀었으나 교육과정 범위를 벗어난 경우

[그림 III-4] G출판사 6학년 2학기 수학 교과서 비례식과 비례배분 6번(장혜원 외, 2023, p.99) 바르게 풀었으나 교육과정 범위를 벗어난 경우

[그림 III-5]는 문제를 바르게 이해하였으나 전체 학생 수를 구하는 식 세우기 과정에서 오류가 나타난 것이다. 식을 바르게 세우지 못하였기 때문에 계산이 바르게 진행되었어도 틀린 답을 도출하였다. 이러한 과정에서 학생 수는 정수로 나와야 한다는 것을 ChatGPT가 인지하고 있어 합격한 학생 수 22.5의 값을 23명으로 반올림한 과정을 통하여 인공지능 언어 모델이 단순히 문장에 제시된 수학적 정보를 추출하여 기계적으로 연산하는 것이 아니라 문제의 맥락까지 파악하여 문제를 해결하려 하고 있음을 알 수 있다. [그림 III-6]에서는 서준이네 반 학생 25명 중에서 매일 아침을 먹는 학생은 16명이라면, 서준이네 반 전체 학생 수에 대한 매일 아침을 먹는 학생 수의 비율을 구하고자 하였다. 이때 문제를 바르게 이해하였으나 백분율로 나타내는 과정에서 오류가 나타났으며 결과적으로 답을 바르게 구하지 못하였다.

<p>You 시험에 응시한 전체 학생 수에 대한 합격한 학생 수의 비율은 3/5입니다. 전체 학생 60명 중에서 합격한 학생은 몇 명인지 수직선을 이용하여 구해 보세요.</p> <p>ChatGPT 합격한 학생 수와 전체 학생 수의 비율이 3/5이므로, 합격한 학생 수를 나타내는 부분을 $3x$, 전체 학생 수를 나타내는 부분을 $5x$로 표현할 수 있습니다. 따라서, $3x + 5x =$ 전체 학생 수 $8x = 60$ (전체 학생 수가 60명이라고 주어졌습니다) $x = 60 / 8$ $x = 7.5$ 이제 합격한 학생 수를 구해봅시다. 합격한 학생 수 = $3x = 3 \times 7.5 = 22.5$ 하지만 학생 수는 정수로 나와야 하므로, 합격한 학생 수를 소수에서 가장 가까운 정수로 반올림합니다. 따라서, 합격한 학생 수는 23명입니다.</p>	<p>You 서준이네 반 학생 25명 중에서 매일 아침을 먹는 학생은 16명입니다. 서준이네 반 전체 학생 수에 대한 매일 아침을 먹는 학생 수의 비율을 백분율로 나타내어 보세요.</p> <p>ChatGPT 매일 아침을 먹는 학생 수의 비율은 다음과 같이 나타낼 수 있습니다: 매일 아침을 먹는 학생 수 : 전체 학생 수 = 16 : 25 이 비율을 백분율로 나타내려면, 비율의 합을 100으로 만들어야 합니다. 비율의 합이 16 + 25 = 41이므로, 비율을 각각 16/41과 25/41로 나누어 줍니다. 매일 아침을 먹는 학생 비율 = $\frac{16}{41} \times 100 \approx 39.02\%$ 따라서, 서준이네 반 학생 중에서 매일 아침을 먹는 학생의 비율은 약 39.02%입니다.</p>
---	--

[그림 III-5] B출판사 6학년 1학기 수학 교과서 비와 비율 7번(한대희 외, 2023a, p.96) 옳게 해결하지 못한 경우

[그림 III-6] J출판사 6학년 1학기 비와 비율 5번 (박성선 외, 2023a, p.97) 옳게 해결하지 못한 경우

인공지능 언어 모델의 문제해결 과정에서 항상 옳은 결과만 도출되는 것은 아니었으며 초등학생을 대상으로 진행한 선행연구(석경희, 백석운, 2004)에서 유형화한 것과 유사하게 각 단계에 따라 문제의 문장에 대한 오해, 전략의 성급한 선택, 잘못 이해된 수학 내용의 사용, 계산 과정의 오류 등 다양한 유형의 오류가 발생하였다. 경우에 따라 한 가지 이상의 오류가 발생하기도 하였다. 이러한 풀이 과정을 통하여 학생이 문제를 푸는 것과 유사하게 여러 가지 유형의 오류가 발생할 수 있으며 그러한 오류와 과정을 확인함으로써 문제를 바르게 해결하지 못한 원인을 파악할 수 있으리라 짐작할 수 있다.

2. ChatGPT의 정답률에 대한 일관성 분석

ChatGPT의 여러 가지 특성 중 하나는 실시간으로 답변을 구성하는 과정에서 나타나는 변동성이다. 이러한 변동성으로 인하여 동일한 질문에도 여러 가지 답변이 가능하며 이는 인공지능 언어 모델의 사용 목적에 따라 장점 또는 단점으로 작용할 수 있다. 연구 대상으로 선정된 각각의 문제를 5회 해결하였을 때 답이 옳게 나타난 횟수를 정리하면 다음 <표 III-2>와 같다.

<표 III-2> ChatGPT의 정답률에 대한 일관성 분석

	답이 옳은 횟수						계
	5회	4회	3회	2회	1회	0회	
비와 비율	12(75%)	2(12.5%)	0(0%)	2(12.5%)	0(0%)	0(0%)	16(100%)
비례식과 비례배분	8(38.10%)	6(28.57%)	3(14.29%)	2(9.52%)	2(9.52%)	0(0%)	21(100%)
	20(54.04%)	8(21.62%)	3(8.11%)	4(10.81%)	2(5.41%)	0(0%)	37(100%)

ChatGPT의 풀이 과정 및 결과를 분석하였을 때 5회에 걸친 조사 동안 5회 모두 옳게 해결한 경우는 비와 비율 단원에서 12회(75%), 비례식과 비례배분 단원에서 8회(38.10%)로 나타났다. 4회 일치하는 경우는 비와 비율 단원에서 2회(12.5%), 비례식과 비례배분 단원에서 6회(28.57%), 3회 일치하는 경우는 각각 0회(0%), 3회(14.29%), 2회 일치하는 경우는 각각 2회(12.5%), 2회(9.52%), 1회 일치하는 경우는 각각 0회(0%), 2회(9.52%), 0회 일치하는 경우는 두 단원 모두 나타나지 않았다.

두 단원의 결과를 전체적으로 살펴보면 5회 모두 일치하는 경우 20회(54.05%), 4회 일치하는 경우 8회(21.62%), 3회 일치하는 경우 3회(8.11%), 2회 일치하는 경우 4회(10.81%), 1회 일치하는 경우 2회(5.41%), 모두 일치하지 않는 경우는 나타나지 않았다. 이때 5회에 걸친 풀이 과정은 각각 독립적으로 진행되었다. 특정한 회차에서 정답 또는 오답이 옳게 나타나는 것이 아니라 각 문제에 대한 풀이가 독립적으로 진행되었으며 선행된 문제의 풀이가 특정 문제의 풀이에 영향을 미치는 경우는 나타나지 않았다.

ChatGPT가 문제를 해결한 결과 중 문제를 한 번도 해결하지 못한 경우는 나타나지 않았으며 다섯 번 중 한 번만 해결에 성공한 경우가 2개 존재하였다. 이는 ChatGPT가 수학 문장제를 해결할 수 있는 확률이 높다는 것을 보여준다. 다만 여러 번 문제를 연속적으로 옳게 해결하거나 옳게 해결하지 못하였더라도 각각의 경우가 다르게 나타났다. 이를 통하여 각각의 경우 독립적으로 문제 해결이 진행됨을 알 수 있다. 풀이 과정이 다양하게 나타남에 따라 동일한 문제의 오답이 여러 개 발생하거나 답이 같더라도 풀이 과정이 다르게 나타나는 등의 경우를 쉽게 찾아볼 수 있다. 이처럼 동일한 문제에 대하여 다른 해결 방법이 옳게 구성되거나 오답이라도 다양한 풀이 과정이 나타나는 것을 통하여 인공지능 언어 모델의 확장성과 변동성을 확인할 수 있으며 이러한 결과를 통하여 다양한 해결 사례의 수집이 가능할 것이라는 짐작이 가능하다.

5회에 걸친 문제 해결 결과가 모두 옳았던 경우, 문제의 답이 모두 일치하였더라도 문제를 해결하는 과정이

다르게 나타나는 경우가 나타났다. 예를 들어 [그림 III-7]의 문장제는 치킨을 구입할 때 할인 쿠폰을 사용한 상황을 통하여 할인쿠폰의 할인율을 구하고자 하였다. 그 결과, ‘할인된 금액=원래 가격-할인 후 가격’이라는 정보를 활용하여 문제를 해결하였다. 반면, 동일한 문제를 해결한 [그림 III-8]에서는 ‘할인율=(1-정상 가격/실제 결제액)×100’이라는 공식을 활용하여 문제를 해결하였다. 이러한 사례는 ChatGPT가 동일한 문제를 한 가지 이상의 방법으로 해결할 수 있다는 것을 보여준다. 다양한 종류의 풀이가 가능한 경우 ChatGPT는 한 가지 방법을 이해하지 못하는 학생에게 여러 가지 방법을 제안해 줄 수 있으며 여러 가지 방법을 고민하는 교사들에게도 여러 가지 방안을 탐색할 수 있도록 유의미한 도움을 줄 수 있으리라는 예상이 가능하다.

<p>You 시우는 할인 쿠폰을 사용하여 18000원짜리 치킨을 14400원에 주문했습니다. 시우가 사용한 쿠폰은 몇 %짜리 할인 쿠폰인지 () 안에 알맞은 수를 써넣으세요.</p> <p>ChatGPT 할인된 가격은 원래 가격에서 할인된 금액을 뺀 가격입니다. 할인된 금액은 원래 가격에서 할인 후 가격을 뺀 것입니다. 할인된 금액 = 원래 가격 - 할인 후 가격 주어진 정보에 따라 계산하면: 할인된 금액 = 18000 원 - 14400 원 = 3600 원 이제 할인율을 구하기 위해 할인된 금액을 원래 가격으로 나누고 100을 곱합니다. 할인율 = $\frac{\text{할인된 금액}}{\text{원래 가격}} \times 100$ 할인율 = $\left(\frac{3600}{18000}\right) \times 100 = 20\%$ 따라서, 시우가 사용한 쿠폰은 20% 할인 쿠폰입니다.</p>	<p>You 시우는 할인 쿠폰을 사용하여 18000원짜리 치킨을 14400원에 주문했습니다. 시우가 사용한 쿠폰은 몇 %짜리 할인 쿠폰인지 () 안에 알맞은 수를 써넣으세요.</p> <p>ChatGPT 할인율을 계산하기 위해 다음과 같은 공식을 사용할 수 있습니다: 할인율 = $\left(1 - \frac{\text{실제 결제액}}{\text{정상 가격}}\right) \times 100$ 여기서, • 실제 결제액은 14400원, • 정상 가격은 18000원입니다. 할인율을 계산해봅시다: 할인율 = $\left(1 - \frac{14400}{18000}\right) \times 100$ 할인율 = $(1 - 0.8) \times 100$ 할인율 = $0.2 \times 100 = 20\%$ 따라서, 시우가 사용한 쿠폰은 20%짜리 할인 쿠폰입니다.</p>
--	--

[그림 III-7] E 출판사 6학년 1학기 수학 교과서 비와 비율 5번(강완 외, 2023, p.99) 옳게 해결한 경우(1)

[그림 III-8] E출판사 6학년 1학기 수학 교과서 비와 비율 5번(강완 외, 2023, p.99) 옳게 해결한 경우(2)

<p>You 우주네 집에 주스 3 1/2 L와 우유 13/5L가 있습니다. 우주네 집에 있는 주스와 우유의 양의 비율 비율이 같은 간단한 자연수의 비로 나타내 보세요.</p> <p>ChatGPT 주스의 양을 $\frac{7}{2}$ L, 우유의 양을 $\frac{13}{5}$ L로 나타내겠습니다. 이때, 주스와 우유의 비를 나타내는 간단한 자연수의 비를 찾기 위해 최대 공약수를 구합니다. 주스와 우유의 비를 나타내는 비율을 $\frac{a}{b}$라고 할 때, 다음과 같은 식을 세울 수 있습니다: $\frac{7}{2} : \frac{13}{5} = \frac{a}{b}$ 이를 통해 최대 공약수를 찾습니다. $a = 7 \times$ 최대 공약수 $b = 2 \times$ 최대 공약수 주스와 우유의 비를 간단한 자연수의 비로 나타내기 위해 최대 공약수를 구합니다. 최대 공약수 $(7, 2) = 1$ 따라서, 최대 공약수는 1이며, 주스와 우유의 간단한 자연수의 비는 다음과 같습니다: $\frac{7}{2} : \frac{13}{5} = \frac{7}{2} \times \frac{5}{1} : \frac{13}{5} \times \frac{1}{1} = \frac{35}{2} : \frac{13}{5}$ 따라서, 주스와 우유의 양의 비를 간단한 자연수의 비로 나타내면 $\frac{35}{2} : \frac{13}{5}$입니다.</p>	<p>You 우주네 집에 주스 3 1/2 L와 우유 13/5L가 있습니다. 우주네 집에 있는 주스와 우유의 양의 비율 비율이 같은 간단한 자연수의 비로 나타내 보세요.</p> <p>ChatGPT 우주네 집에 있는 주스와 우유의 양을 각각 간단한 자연수로 나타내기 위해서는 주스와 우유의 양을 최대공약수로 나누어야 합니다. 주어진 주스와 우유의 양은 다음과 같습니다: 주스: $3\frac{1}{2}L = \frac{7}{2}L$ 우유: $1\frac{3}{5}L = \frac{8}{5}L$ 먼저, 주스와 우유의 양을 최대공약수로 나누어 보겠습니다. 최대공약수는 1이므로, 주스와 우유의 양은 이미 간단한 자연수의 비로 나타낼 수 있습니다. 따라서, 주스와 우유의 양의 비를 간단한 자연수의 비로 나타내면: 주스:우유 = 7:8</p>
---	---

[그림 III-9] B출판사 6학년 2학기 수학 교과서 비례식과 비례배분 5번(한대희 외, 2023b, p.98) 옳게 해결하지 못한 경우(1)

[그림 III-10] B출판사 천재한 6학년 2학기 수학 교과서 비례식과 비례배분 5번(한대희 외, 2023, p.98) 옳게 해결하지 못한 경우(2)

문제의 답이 옳지 않은 경우에도 풀이 과정에 따라 풀이 과정과 답이 다르게 나타났다. [그림 III-9]는 우주네 집에 주스 $3\frac{1}{2}$ L와 우유 $1\frac{3}{5}$ L가 있을 때, 주스와 우유의 양의 비를 비율이 같은 간단한 자연수의 비로 나타내라고 요구하고 있다. [그림 III-9]에서는 문제에서 간단한 자연수의 비로 나타내라고 하였으나 자연수의 비로 나타내지 않고 주스는 $\frac{7}{2}$, 우유는 $\frac{8}{5}$ 라며 분수로 비를 나타내었다. [그림 III-10]에서는 주스는 $\frac{7}{2}$, 우유는 $\frac{8}{5}$ 로 나타낼 수 있으나 이때 분모의 크기가 다른 것을 고려하지 않아 7:8이라는 오답을 도출하였다.

3. ChatGPT의 오류 유형 분석

ChatGPT의 문제 해결 과정을 분석 방법에 따라 답이 옳은 것과 옳지 않은 것으로 각각 구분하여 나타낸 결과는 다음과 같다. 답이 옳은 경우와 답이 옳지 않은 경우를 각각의 풀이 방법에 따라 구분하였다. 문제에 따라 한 가지 이상의 오류가 발생할 수 있어 총 횟수는 190회로 문제를 풀이한 개수인 185회보다 더 많게 나타났다. 오류를 유형별로 분석한 결과는 <표 III-3>와 같다.

<표 III-3> 비와 비율 및 비레식과 비례배분 단원 문항에 대한 ChatGPT의 오류 유형

구분	답이 옳은 경우(T)				답이 옳지 않은 경우(F)				계
	T1	T2	T3	F1	F2	F31	F32	F33	
횟수	131(68.4%)	19(10%)	0(0%)	0(0%)	0(0%)	13(6.84%)	18(9.5%)	9(4.74%)	190(100%)

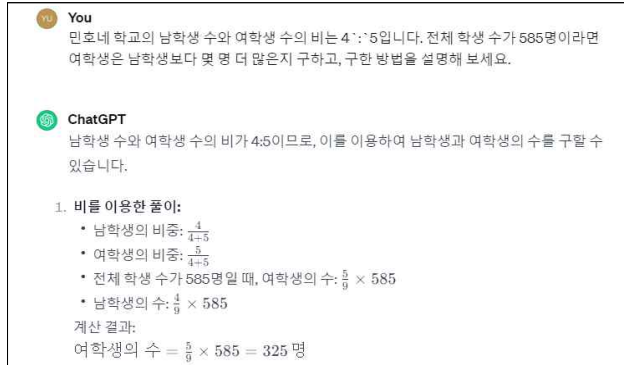
분석 결과, 답이 옳은 150회 중 풀이가 옳은 경우(T1)는 131회(68.4%), 풀이가 옳지만 초등학교 6학년 교육과정을 벗어나는 경우(T2)는 19회(10%)로 나타났으며 답이 옳지만 풀이가 옳지 않은 경우(T3)는 나타나지 않았다. 답이 옳지 않은 35회 중 풀이가 옳은 경우는 0회(0%), 풀이가 옳지만 6학년 교육과정을 벗어나는 경우는 0회(0%)로 나타났다. 답이 옳지 않고 풀이가 옳지 않은 경우 중 문제 이해 오류가 나타난 경우(F31)는 13회(6.84%), 식 세우기 오류가 나타난 경우(F32)는 18회(9.5%), 계산 오류가 나타난 경우(F33)는 9회(4.74%)로 나타났다. 답이 옳으나 풀이가 옳지 않은 경우(T3), 답이 옳지 않지만 풀이 과정이 옳은 경우(F1, F2)의 경우가 각각 0회로 나타나지 않았다는 것은 ChatGPT의 문제 풀이 능력에 대한 근거로 작용할 수 있다.

문장제의 해결 과정에서 발생한 오류 경향의 분석은 학생들의 지도에 필요한 개별 정보를 얻을 수 있는 기회를 제공하고, 오류에 대하여 정확한 피드백을 하는데 좋은 자료가 될 수 있다(이병욱, 안병근, 2008). 이러한 측면에서 ChatGPT의 문제 해결 과정에서 나타나는 오류를 유형별로 살펴보았다.

첫째, 문제 이해 오류이다.

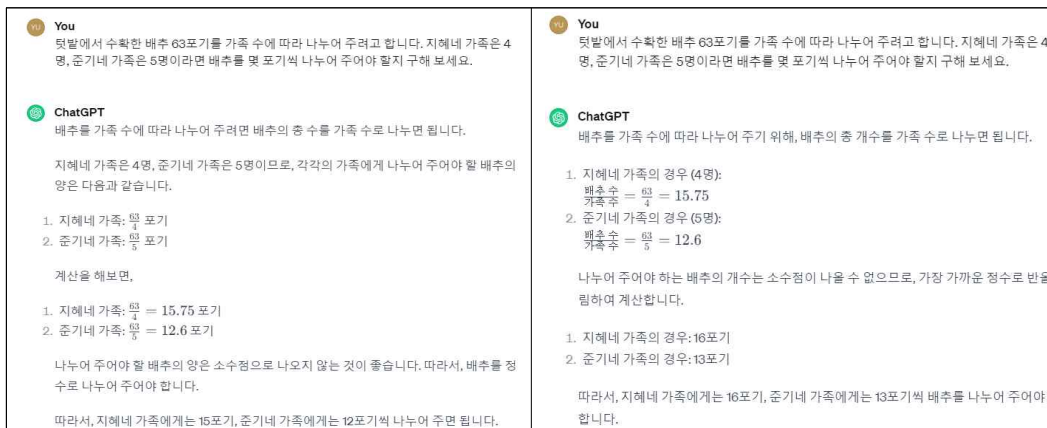
ChatGPT는 인공지능을 활용한 강력한 언어 모델이지만 문장이나 문제를 제대로 이해하지 못하여 문제를 해결하지 못할 수 있다. 언어로 기술된 문제 자체가 모호하거나 어려운 어휘나 문법을 포함하는 경우, 해석 또는 문장제의 질문에 여러 방향으로 해석할 여지가 있는 경우 등이 해당한다. 사용자의 질문이 모호하거나 충분한 세부 정보가 없는 경우, 특정 주제나 문제에 대한 충분한 데이터가 부족하거나 누락된 경우의 가능성도 있다. 이러한 이유 중 하나 또는 여러 이유가 결합하여 모델이 문제를 정확하게 이해하지 못할 수 있다.

예를 들어, [문제 III-11]에서는 민호네 학교의 남학생 수와 여학생 수의 비는 4:5라고 가정한다. 해당 문제는 전체 학생 수가 585명이라면 여학생은 남학생보다 몇 명 더 많은지 구하고 구한 방법을 설명해 보라고 지시하였다. 그러나 인공지능 언어 모델은 남학생과 여학생의 수를 구해야 한다고 문제를 잘못 이해하였다. 그 결과 여학생이 남학생보다 몇 명 더 많은지가 아니라 남학생과 여학생의 수를 각각 구하였다.



[그림 III-11] H출판사 6학년 2학기 수학 교과서 비례식과 비례배분 5번(신항균 외, 2023, p.89) 옳게 해결하지 못한 경우

[그림 III-12]와 [그림 III-13]은 동일한 문제를 해결한 것이다. 해당 문장제를 바르게 해결했을 경우 지혜네 가족은 28개, 준기네 가족은 35개의 배추를 나누어 가지게 된다. 하지만 [그림 III-12]와 [그림 III-13]의 풀이 과정에서 문제를 해결하기 위하여 배추 수를 가족의 수로 나누어야 한다고 잘못 이해하는 오류를 범하고 있다. 그 결과 지혜네 가족이 나누어 가져야 하는 배추의 수와 준기네 가족이 나누어 가져야 하는 배추의 수는 12.6개로 나타난다. 그러나 [그림 III-12]에서는 나누어 주어야 하는 배추 수에 대한 맥락을 충분히 고려하지 못하여 소수점이 포함된 결과를 제시하고 있으나 [그림 III-13]에서는 배추의 수라는 맥락을 고려하여 가장 가까운 정수로 반올림하여 계산하고 있다. 동일한 문제에 대하여 유사한 방향으로 해결하는 두 가지 풀이 과정은 문제 이해의 오류에 따라 동일한 오답이 도출될 수 있었으나 문제의 맥락을 충분히 이해하였는가에 따라 오답이라 할지라도 해결 과정이 다르게 전개될 수 있음을 보여준다.



[그림 III-12] 국정 6학년 2학기 수학 교과서 비례식과 비례배분 6번(교육부, 2020b, p.87) 옳게 해결하지 못한 경우(1)

[그림 III-13] 국정 6학년 2학기 수학 교과서 비례식과 비례배분 6번(교육부, 2020b, p.87) 옳게 해결하지 못한 경우(2)

둘째, 식 세우기 오류이다.

이러한 상황은 수학적 문제를 해결할 때 잘못된 계산이나 해석을 하는 경우를 의미하며 이러한 상황이 발생하는 이유는 다양할 수 있다. 예를 들어, 문제 자체나 수학적 표현이 모호하게 작성되었거나, 여러 해석이 가능한 경우 모델은 그중 하나를 선택할 수 있으며 이때 모델이 선택한 해석이 질문의 의도와 다를 수 있다. 또는, ChatGPT가 복잡한 수학적 계산이나 논리적인 단계를 이해하는 데 어려움을 겪을 수 있으며 긴 문장이나 복잡한 수학적 계산을 완전히 이해하는 데 어려움이 있을 수 있다. 이러한 이유 중 하나 이상이 결합하여 모델이 수학적 문제를 정확하게 해결하지 못할 수 있다. 경우에 따라 한 가지 이상의 이유가 복합적으로 작용할 수 있다. 비록 문제 이해에 성공하였어도 식 세우기 오류가 나타나기도 하였으며 문제를 이해하지 못하여 식 세우기 오류가 나타나는 경우도 있다.

[그림 III-14]에서 ChatGPT는 여행경비를 사람 수에 따라 나누어 내기 위하여 각 가족의 인원수에 따라 비례식을 세울 수 있다고 응답하고 있다. 하지만 정작 식을 세울 때는 비례식을 옳지 않게 세우는 오류를 범하고 있다. [그림 III-15]의 경우에도 유산균 음료를 만드는 데 필요한 우유와 요구르트의 비율을 구해야 한다고 문제를 빠르게 이해하고 있으나 우유 양에 대한 요구르트 양의 비율에 관련된 식을 세우는 과정에서 분모와 분자를 반대로 대입한 오류가 나타났다. 식이 잘못 세워졌기 때문에 빠르게 연산하였다면 15%가 나와야 하지만 666.67%라는 틀린 답을 도출하였다. 이러한 오류를 줄이기 위하여 인공지능 모델의 답변에 대해 추가적인 질문이나 설명을 요청하여 오해를 해소하거나, 문제를 더 자세히 설명하여 더욱 명료한 해결 방안을 유도할 수 있다.

<p>You 경규네 가족과 현서네 가족은 함께 사용한 여행 경비 160000원을 사람 수에 따라 나누어 내려고 합니다. 경규네 가족이 3명이고, 현서네 가족이 5명이라면 여행 경비를 얼마씩 나누어 내야 하는지 구해 보세요.</p> <p>ChatGPT 여행 경비를 사람 수에 따라 나누어 내기 위해서는 각 가족의 인원수에 따라 비례식을 세울 수 있습니다.</p> <p>경규네 가족이 3명이고, 현서네 가족이 5명이라고 할 때, 각 가족이 내야 하는 여행 경비를 A라고 하면 비례식은 다음과 같습니다.</p> $\frac{A}{3} = \frac{160000}{5}$ <p>이 비례식을 풀면:</p> $A = \frac{3 \times 160000}{5}$ <p>계산하면:</p> $A = \frac{480000}{5} = 96000$ <p>따라서, 경규네 가족은 96000원을 내야 하고, 현서네 가족은 160000원에서 96000원을 뺀 나머지인 64000원을 내야 합니다.</p>	<p>You 유산균 음료를 만드는 데 우유 1000 mL와 요구르트 150 mL가 필요합니다. 유산균 음료를 만드는 데 필요한 우유 양에 대한 요구르트 양의 비율을 구하고, 백분율로 나타내어 보세요.</p> <p>ChatGPT 유산균 음료를 만드는 데 필요한 우유와 요구르트의 비율을 구하기 위해선 다음과 같은 계산을 할 수 있습니다.</p> <ol style="list-style-type: none"> 우유 양에 대한 요구르트 양의 비율: 비율 = $\frac{\text{우유양}}{\text{요구르트양}} = \frac{1000}{150} = \frac{20}{3}$ 비율을 백분율로 나타내기: 백분율 = $(\frac{\text{비율}}{1}) \times 100 = (\frac{20}{3} \times 100) \approx 666.67\%$ <p>따라서, 유산균 음료를 만드는 데 필요한 우유와 요구르트의 비율은 20:3이며, 백분율로 나타내면 약 666.67%입니다.</p>
--	---

[그림 III-14] H 출판사 6학년 2학기 수학 교과서 비례식과 비례배분 4번(신항균, 2023, p.89) 옳게 해결하지 못한 경우

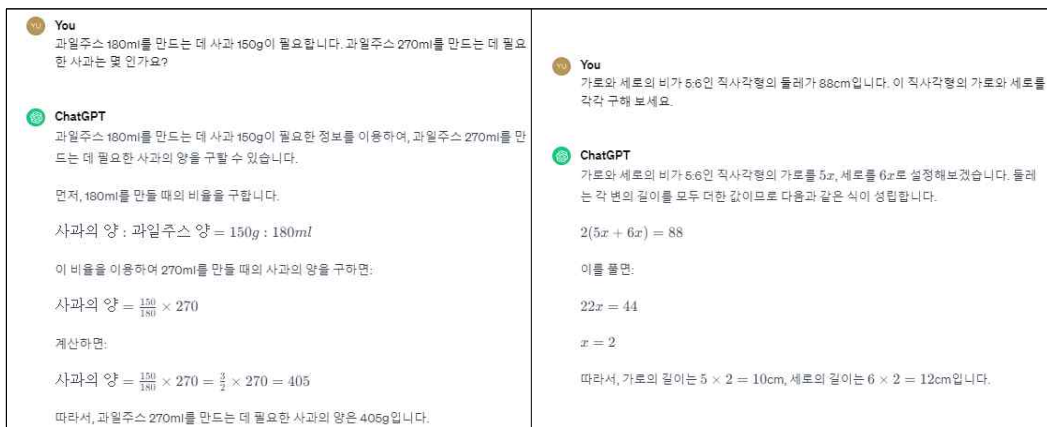
[그림 III-15] D 출판사 6학년 1학기 수학 교과서 비례식과 비례배분 5번(안병곤, 2023, p.99) 옳게 해결하지 못한 경우

셋째, 계산 오류이다.

인공지능 언어 모델은 훈련 데이터에서 학습한 수학적 계산 능력을 기반으로 한다. 그러나 모델이 특정 유형의 계산이나 논리를 이해하지 못할 수 있으며 복잡하거나 긴 계산이 있는 경우 모델의 한계가 드러날 수 있다. 또는, 훈련 데이터에 특정 유형의 계산이나 문제가 부족한 경우, 최신 수학적 개념이나 계산법에 대한 지식이 부족한 경우 해당 유형의 계산에 대한 정확한 결과를 제공할 수 어려울 수 있다. 이러한 상황에서 모델의 답변에

대해 추가 질문을 하거나 입력 데이터를 검토하여 오류를 찾아내고 수정할 수 있다.

ChatGPT의 문제해결 과정 중 식을 세우고 해결하는 과정에서 계산이 바르게 수행되지 못하는 오류가 나타났다. 이는 수립된 식의 계산이 바르게 진행되지 않은 경우를 가리킨다. 식이 바르게 수립되었거나 수립되지 않은 경우 모두 해당 오류가 발생하였다. 예를 들어, [그림 III-16]의 경우 과일주스를 만드는데 사과 150g이 필요한 상황을 제시하였으며 이를 통하여 과일주스 270mL를 만드는 데 필요한 사과의 양을 구하고자 한다. 문제를 해결하기 위하여 사과의 양은 $150/180 \times 270$ 이라는 식을 수립하였으나 식의 해결 과정에서 $150/180$ 을 $3/2$ 로 계산하는 오류가 발생하였다. [그림 III-17]에서는 가로와 세로의 비가 5:6인 직사각형의 둘레를 이용하여 직사각형의 가로와 세로의 비를 구하라고 요구한다. 이를 위하여 “ $2(5x+6x)=88$ ”이라는 식을 수립하였으나 문제의 해결 과정에서 “ $22x=44$ ”라는 계산 실수가 발생하였다. 바르게 식을 풀었다면 “ $22x=88$ ”이 되어야 하며 이러한 실수로 인하여 최종 답까지 영향을 미친 것이다. 이외에도 $45/300$ 을 $1/6$ 으로 계산하는 등 식을 바르게 수립하였더라도 식의 해결 과정에서 계산 오류가 발견되는 경우가 나타났다. 이러한 상황에서 모델의 답변에 대한 추가 질문을 하거나 검산을 요구하고 입력 데이터를 검토하여 오류를 찾아내는 등의 지시가 가능하며 필요에 따라 더 자세한 계산 과정을 요청할 수 있다.



[그림 III-16] C 출판사 6학년 2학기 수학 교과서 비례식과 비례배분 4번(박교식 외, 2023, p.94) 옳게 해결하지 못한 경우

[그림 III-17] J 출판사 6학년 2학기 수학 교과서 비례식과 비례배분 7번(박성선 외, 2023b, p.99) 옳게 해결하지 못한 경우

IV. 결론 및 제언

본 연구는 자연어 처리가 가능한 인공지능 언어 모델인 ChatGPT의 정답률, 정확도, 오류 유형을 분석하여 비와 비율 및 비례식과 비례배분 단원평가 문장제의 해결에 대한 ChatGPT의 성능을 확인하고 수학교육에서의 교수학적 활용 방안을 모색하고자 하였다. 연구 결과로부터 도출한 결론 및 제언은 다음과 같다.

첫째, 인공지능 언어 모델의 문장제 해결 능력을 확인하였으며 문제해결 과정에서 획득한 정보를 교수학습 설계에 활용할 가능성을 확인하였다. 문장제의 정답률은 전체 185회 중 150회(81.08%)에 달하였으며 5회 풀이 중 80% 이상의 정답률을 유지하는 경우는 전체 37회 중 28회(75.68%)로 나타났다. ChatGPT의 오답률이 높은 경우 문제의 유형, 내용 등의 차이는 있었으나 문제에서 요구하는 과정이 복잡하게 나타나는 등 문장제를 이해

하기 어렵다고 여겨지는 요소를 포함하는 경우가 많았다. 이외에도 간단한 계산을 요구하는 경우보다 문제의 맥락을 파악해야 문제를 해결할 수 있는 경우 정답률이 낮게 나타났다. 이러한 결과는 선행된 학습 내용과 여러 개념을 복합적으로 요하는 문제의 해결에 대해 어려움을 더 느낀다는 선행연구(박장희 외, 2012)와 유사한 방향이다. 또한, ChatGPT는 인간이 아니지만 인간이 생성한 데이터를 기반으로 답변을 생성하기에 문제해결 과정 및 정보에 대한 여러 가지 추가 정보를 수집할 수 있으리라는 예상이 가능하다. 다시 말하면, ChatGPT가 언어를 통하여 질문을 이해하고 답변을 구성하는 과정이 인터넷의 방대한 양의 데이터를 통해 훈련된 것을 바탕으로 하기에 문제해결 과정 및 결과를 통하여 학생들의 학습 과정을 유추하는 데 도움이 되는 정보를 얻을 수 있다는 가능성을 엿볼 수 있다. 이러한 정보는 다양한 학생들을 만나는 교사들에게 필요한 것으로 교사들은 다양한 사례를 수집하여 교수학습 과정을 더 정교하게 설계하고 개별 학생에게 적절한 피드백을 고안할 수 있다. 예를 들어, 학생이 문제를 해결하지 못한 경우 도출된 오답의 유형과 해결 과정에 대한 정보를 미리 파악하고 있다면 더 신속하고 적절한 대안의 제시가 가능하다.

둘째, 인공지능 언어 모델의 해결 유형을 파악하여 학생들의 개별화 학습 도구를 위한 정보 수집에 활용할 수 있다. 인공지능 언어 모델의 경우 문제의 정답률(81.08%)이 높을 뿐 아니라 다양한 해결 유형이 나타났다. 문제를 빠르게 해결하지 못한 경우에도 해결 과정에서 문제 이해 오류, 식 세우기 오류, 계산 오류 등의 여러 유형이 나타났다. 이러한 인공지능 언어 모델의 정답률과 오답률, 해결 방법을 분석하여 다양한 풀이의 방향을 제시하거나 비계를 설정할 수 있으리라 기대할 수 있다. AI 모델은 학생들에게 문장제와 계산을 위한 강력한 도구일 수 있다(Remoto, 2023). 인공지능 언어 모델의 특성을 고려하였을 때 문장제의 이해 및 문제 재구성 등에 대한 구체적인 역할을 부여할 수 있으며, 인공지능 언어 모델이 학생들의 실수를 감지하고 답변을 변경할 수도 있다는 것을 고려하면(Remoto, 2023) 다양한 문제해결 방법을 지속적으로 생산할 수 있는 특성이 오히려 장점으로 작용할 수 있다. 더욱이, 대형 언어 모델인 ChatGPT는 방대한 양의 데이터를 학습하며 학습한 데이터의 특성이 ChatGPT의 반응에 영향을 미치므로(Azaria, 2022) 이러한 점을 고려하면 학습자 개인에게 적절하게 개별화된 학습 도구로 역할할 수 있으리라고 예상할 수 있다. 최근 인공지능 언어 모델을 활용한 학습용 챗봇 등의 기술이 발전하고 있으며 단순히 정답을 제시하는 것이 아니라 명령 프롬프트에 원하는 종류의 출력의 예시를 포함하거나 "이것에 대해 단계적으로 생각해 봅시다"와 같은 일반적인 지시 사항을 포함하는 것으로 밝혀졌다(Davis, 2023). 목적에 따라 인공지능 언어 모델에 활용할 수 있는 수학적 개념의 범위를 제한하고 학생들이 직접 생성한 데이터를 입력하여 그를 기반으로 새로운 답변을 생성하도록 요구하는 등 구체적인 지시 사항을 설정하여 다양한 방향으로의 활용을 기대할 수 있다.

셋째, 인공지능 언어 모델의 문제해결 과정을 통하여 문장제의 언어적 표현과 수학적 개념, 맥락 요소의 적절성을 검증할 수 있다. 분석 결과, 6학년 2학기에 해당하는 비례식과 비례배분 단원에서 상대적으로 정답률이 낮게 나타났다. 이는 다른 단원에 비하여 복잡하게 문장을 구성한 해당 단원의 특징이 반영된 결과인 동시에 문장에 내제된 언어적 특성이 문제가 요구하는 수학적 정보와는 별개로 문장제의 해결 여부에 영향을 미칠 수 있다는 것을 보여준다. 교과서의 집필진 또는 수학 수업을 진행하는 교사는 이러한 유형의 문장제에 익숙한 초등 수학 전문가로 구성되어 있다. 이러한 경우 초등 수학 전문가들은 실제로 문제를 해결해야 하는 초등학생의 눈높이를 고려하지 못하여 문제의 난도 조절에 어려움을 겪을 수 있다. 문장제의 해결을 위하여 언어적 표현, 수학적 개념, 맥락 요소 등을 입체적으로 고려해야 하기 때문이다. 이때 인공지능 언어 모델의 풀이를 통하여 실제로 학생들이 문제를 해결하기 전 문제의 완성도나 표현 등을 점검할 수 있다. 인공지능 언어 모델에게 초등 수학교육 과정 및 교과서와 관련된 데이터를 입력하고 해당 범위 내에서 문제를 해결하도록 요구하거나 초등학생의 어휘 수준에서 문제를 이해할 것을 요구하고 이해하기 어려운 단어를 탐색하도록 하는 등 자연어 처리와 관련하여 여러 가지 역할을 수행하도록 지시할 수 있다.

넷째, 인공지능 언어 모델을 이용하여 수학교육 연구를 위한 연구 도구 및 방법으로 활용할 방안을 탐색해야

한다. 인공지능 언어 모델은 학생이 아니기 때문에 개별 학생이 문제를 해결하는 방식과 유사하거나 차이가 있으리라 확신할 수 없다. 그럼에도 불구하고 다양한 답변을 무한정 도출할 수 있으며 문제에서 답으로 이어지는 풀이 과정에서 오답에 도달할 때 어떠한 과정에서 오류가 있었는지 결과를 통하여 그 과정을 유추할 수 있다. 더욱이, 학생들이 실제로 어떠한 사고 과정을 통하여 문제를 해결하거나 해결하지 못하는지 파악하려면 많은 수의 경우를 일일이 조사해야 하지만 이는 현실적으로 어렵다는 점을 고려하면 이러한 인공지능 언어 모델의 가능성이 더욱 중요해진다. ChatGPT는 무한한 답변을 생성할 수 있는 인공지능 언어 모델이며 답변의 다양성은 오히려 다양한 풀이 과정 탐색에 유리하게 작용할 수 있기 때문이다.

ChatGPT라는 인공지능 언어 모델의 도입이 사회 전반에 중요한 영향을 미칠 것이라고 거론되는 시점에서 이것을 교육적으로 어떻게 활용할 것인가에 대한 점검과 고민이 필요하다. 본 연구는 이러한 관점에서 접근하였으며 새로운 기술의 발전인 인공지능 언어 모델을 수학교육에 활용할 수 있는 방안을 모색할 필요를 주장하였다. 해당 기술이 계속 발전하고 있어 앞으로 다양한 기능이 추가될 가능성이 있으며 이로 인하여 수학교육에서 큰 변화가 예상되기에 현 시점에서 이러한 기술과 교육의 접목에 관련된 연구가 더욱 필요하다. 추후 다양한 소재 및 영역, 유형의 문장제를 대상으로 여러 가지 인공지능 언어 모델의 풀이 과정 및 결과를 살펴볼 필요가 있으며 지속적으로 최신 기술을 활용한 연구를 진행할 필요가 있다.

참 고 문 헌

- 강완 · 백석윤 · 전인호 · 이경화 · 김연 · 이미연 외 15인 (2023). 수학 6-2. 대교출판.
- Kang, W., Baek, S., Jeon, I., Lee, K., Kim, Y., ..., Lee, M. (2023). *Mathematics 6-2* Daekyo.
- 강윤지(2023). 수학교육의 교수학적 실천과 연구 측면에서 AI 언어 모델 활용 방안 탐색. 수학교육학연구, **33(4)**, 1023-1040.
- Kang, Y.(2023) Exploring the use of AI language models in terms of pedagogical practice and research in mathematics education. *Journal of Educational Research in Mathematics*, **33(4)**, 1023-1040.
- 교육부 (2020a). 수학 6-1. 비상교육.
- Ministry of Education (2020a). *Mathematics 6-1*. Visang.
- 교육부 (2020b). 수학 6-2. 비상교육.
- Ministry of Education (2020b). *Mathematics 6-2*. Visang.
- 권오남 · 오세준 · 윤정은 · 이경원 · 신병철 · 정원 (2023). ChatGPT 의 수학적 성능 분석: 국가수준 학업성취도 평가 및 대학수학능력시험 수학 문제 풀이를 중심으로. 수학교육 논문집, **37(2)**, 233-256.
- Kwon, O., Oh, S., Yoon, J., Lee, K., Shin, B., & Jung, W.(2023). Analyzing mathematical performances of chatGPT: Focusing on the solution of national assessment of educational achievement and the college scholastic ability test. *Communication of Mathematical Education*, **37(2)**, 233-256.
- 박교식 · 정영옥 · 고정화 · 권석일 · 남진영 · 박진형 외 27인 (2023). 수학 6-2. 두산동아.
- Park, K., Chong, Y., Ko, J., Kwon, S., Nam, J., ..., Prak, J. (2023). *Mathematics 6-2* Bookdonga.
- 박만구 · 강경은 · 김대진 · 김도경 · 김수정 · 김은혜 외 8인 (2023). 수학 6-2. 천재교육.
- Park, M., Kang, K., Kim, D., Lim, D., Kim, S., ..., Kim E. (2023). *Mathematics 6-2* Chunjae.
- 박성선 · 류성립 · 김상미 · 권성룡 · 김남균 · 강호진 외 11인 (2023a). 수학 6-1. YBM.
- Park, S., Ryu, S., Kim, S., Kwon, S., Kim, N., ..., Kang, H. (2023a). *Mathematics 6-1*. YBM.
- 박성선 · 류성립 · 김상미 · 권성룡 · 김남균 · 강호진 외 11인 (2023b). 수학 6-2. YBM.
- Park, S., Ryu, S., Kim, S., Kwon, S., Kim, N., ..., Kang, H. (2023b) *Mathematics 6-2* YBM.

- 박장희 · 유석규 · 이증권(2012). 실생활 문장제의 해결 과정에 나타나는 오류유형 분석. 한국학교수학회논문집, **15(4)**, 699-718.
- Park, J., Ryu, S. & Lee, J.(2012). The analysis of mathematics error type that appears from the process of solving problem related to real life. *Journal of the Korean School Mathematics Society*, **15(4)**, 699-718.
- 손태권(2023). ChatGPT의 수학교육 활용 가능성 탐색: 분수 문제에 관한 학생의 산출물과 예비 교사의 담화 사례를 중심으로. 초등수학교육, **26(2)**, 99-113.
- Son, T. (2023). Exploring the possibility of using ChatGPT in mathematics education: Focusing on student product and pre-service teachers' discourse related fraction problems. *Education of Primary School Mathematics*, **26(2)**, 99-113.
- 석경희 · 백석운(2004). 초등 수학의 문장제 해결 과정에 나타난 오류 분석. 한국초등수학교육학회지, **8(2)**, 147-168.-
- Seok, k. & Paik, S.(2004). An analysis of errors on the process of word problem solving in the elementary mathematics. *Journal of Elementary Mathematics Education in Korea*, **8(2)**, 147-168.
- 안병곤 · 나귀수 · 김민경 · 이광호 · 류현아 · 최지선 외 14인 (2023). 수학 6-2. 두산동아.
- Ahn, B., Na, G., Kim, M., Lee, K., Ryu, H., ..., Choi, J. (2023). Mathematics 6-2. Bookdonga.
- 이병옥 · 안병곤(2008). 수학 문장제의 문장 구조와 해석상의 오류 분석-초등학교 2 학년을 중심으로. 한국초등수학교육학회지, **12(2)**, 185-204.
- Lee, B. O., & Ahn, B. G. (2008). An analysis on sentence structures and interpretation errors in word problems in mathematics-focussing on the 2nd grade elementary students. *Journal of Elementary Mathematics Education in Korea*, **12(2)**, 185-204.
- 장혜원 · 서동엽 · 김민희 · 김선 · 김주숙 · 김차명 외 8인 (2023). 수학 6-2. 미래엔.
- Chang, H., Seo, D., Kim, M., Kim, S., Kim, J., ..., Kim, C. (2023). *Mathematics 6-2* Mirae-n.
- 한대희 · 고은성 · 이수진 · 조형미 · 한상의 · 신희영 외 11인 (2023a). 수학 6-1. 천재교육.
- Han, D., Ko, E., Lee, S., Cho, H., Han, S., ..., Shin, H. (2023a). *Mathematics 6-1*. Chunjae.
- 한대희 · 고은성 · 이수진 · 조형미 · 한상의 · 신희영 외 11인 (2023b). 수학 6-2. 천재교육.
- Han, D., Ko, E., Lee, S., Cho, H., Han, S., ..., Shin, H. (2023b). *Mathematics 6-2*. Chunjae.
- Azaria, A. (2022). ChatGPT usage and limitations. HAL Open Science. hal-03913837
- Davis, E. (2023). Mathematics, word problems, common sense, and artificial intelligence. arXiv preprint arXiv:2301.09723.
- OpenAI (2023a). Improving mathematical reasoning with process supervision. <https://openai.com/research/improving-mathematical-reasoning-with-process-supervision>
- Remoto, J. P. (2023). ChatGPT and other AIs: Personal relief and limitations among mathematics-oriented learners. *Environment and Social Psychology*, **9(1)**. 1-13.
- Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM*, **52**, 1-16.
- Zong, M., & Krishnamachari, B. (2023, June). Solving math word problems concerning systems of equations with gpt-3. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 13, pp. 15972-15979).

A study on the didactical application of ChatGPT for mathematical word problem solving

Kang, Yunji

Seoul Hongyeon Elementary School

E-mail : angie0718@sen.go.kr

Recent interest in the diverse applications of artificial intelligence (AI) language models has highlighted the need to explore didactical uses in mathematics education. AI language models, capable of natural language processing, show promise in solving mathematical word problems. This study tested the capability of ChatGPT, an AI language model, to solve word problems from elementary school textbooks, and analyzed both the solutions and errors made. The results showed that the AI language model achieved an accuracy rate of 81.08%, with errors in problem comprehension, equation formulation, and calculation. Based on this analysis of solution processes and error types, the study suggests implications for the didactical application of AI language models in education.

* 2000 Mathematics Subject Classification : 97U70

* Key words : ChatGPT, word problem, mathematics education, artificial intelligence

† corresponding author