

ChatGPT의 수학교육 활용 가능성 탐색: 분수 문제에 관한 학생의 산출물과 예비교사의 담화 사례를 중심으로

손태권(봉명초등학교, 교사)

본 연구는 수학교육에서 ChatGPT의 활용 가능성을 탐색하기 위해 분수 문제에 대한 학생의 산출물과 예비교사와 학생과의 담화 자료를 사례로 선정하여 ChatGPT를 통해 분석하고 수학교육전문가의 분석 결과와 비교하였다. 학생 자료는 학생의 문제해결 전략과 수학적 사고를 분석했으며 예비교사 자료는 예비교사의 발문이 학생의 반응을 토대로 이루어지는지 그리고 발문이 수학적 사고를 이끌어내는지를 평가 기준으로 선정하고 ChatGPT에게 분석을 요청하였다. ChatGPT의 분석 결과, 학생 자료에 대한 분석 결과는 수학교육전문가의 분석 결과와 유사하게 나타났다. 예비교사 자료는 유사점과 차이점이 함께 나타났다. 이러한 결과를 바탕으로 수학교육에서 ChatGPT의 활용 가능성과 시사점을 도출하였다.

I. 서론

인공지능(Artificial Intelligence: 이하 AI)이 빠르게 발전함에 따라 AI는 교육 분야에서 무엇을 가르치고 어떻게 가르칠지에 관한 교육 방식을 변화시키고 있다(Popenici & Kerr, 2017). 이러한 변화와 함께 교육에서의 AI는 인간 교사나 교사 교육자가 수행하던 노력을 대체하거나 보조하여 시간이나 비용을 절감해줄 것이라 기대되고 있다(OECD, 2021). 특히, 여러 교과 중 수학적 학생들이 배워야 할 수학적 개념이 위계적이고 구조화되어 있다는 점에서 AI를 적용하기에 적합한 교과이며(Home et al., 2019), 이러한 이유에서 여러 연구자들은 다양한 관점에서 수학교육에서의 AI 활용 가능성을 탐색해왔다. 예를 들어, 지능형 교수 시스템을 활용한 개별화 학습(이지혜, 허난, 2020; Bernacki & Walkington, 2018), 반응적 교수 관행을 위한 챗봇

발(Lee & Yeo, 2022), 개별 학습자를 위한 맞춤형 학습 경로, 학습 자료, 교수·학습 전략 제공(예: Baylari & Montazer, 2009; Dwivedi et al., 2018; Soltanpoor et al., 2018)과 같은 연구들은 숙련된 인간 교사가 수행하던 교수 역할을 AI가 부분적으로 대체할 수 있는 가능성을 보여주었다. 그러나 대부분의 연구들은 연구자가 정한 규칙에 따라 AI가 작동하는 규칙 기반(rule-based) 기술을 적용하였다. 이러한 전통적인 AI 기술은 제한적인 상황에서만 학습이 가능하므로 실제 교육성과를 일반화하기는 어려운 실정이었다(Shamir & Levin, 2022).

이러한 상황에서 자연어 처리(natural language process) AI 모델인 ChatGPT가 2022년 10월 30일 발표되었다. ChatGPT는 빅데이터를 기반으로 훈련된 대규모 언어 모델(large language model)로서 입력된 텍스트와 맥락을 기반으로 사람과 유사한 텍스트를 생성한다(Kasneci et al., 2023). ChatGPT는 학생들을 위한 과제를 만들고 산출물에 대한 평가를 수행하며 학술 논문을 작성하는 등 기존의 교사와 교사 교육자가 수행해온 복잡한 작업을 수행하는 능력을 보여주었다(Zhai, 2022). 이러한 ChatGPT의 능력에도 불구하고, ChatGPT의 교육적 활용에 대한 연구자들의 의견은 서로 대립되고 있다. 일부 연구자들은 ChatGPT가 교육의 자동화와 창의성에 획기적인 기여를 하며 에세이와 같은 전통적인 과제와 평가의 종언을 고할 수 있는 강력한 도구라고 평가하였다(Stokel-Walker, 2022; Zhai, 2022). 반면, 다른 연구자들은 ChatGPT가 가진 단점을 부각하며 ChatGPT의 교육적 사용에 대한 우려를 표하고 있다. 예컨대, ChatGPT가 학생이 수행해야 할 과제를 AI가 대신 수행하고 부정확한 정보를 제공할 수 있다는 점, 특정 데이터에 대한 의존으로 인한 잠재적 편견과 차별을 가진 답변을 제시하거나 창의성 및 비판적 사고가 부족하다는 점 등이 잠재적 악

* 접수일(2023년 3월 23일), 심사(수정)일(2023년 4월 13일), 게재확정일(2023년 4월 20일)
* MSC2000분류 : 97U70
* 주제어 : 인공지능, ChatGPT, 수학적 사고, 발문

영향으로 제시되고 있다(예: Atlas, 2023; D'Amico et al. 2023; Mhlanga, 2023). ChatGPT의 교육적 사용에 대한 논쟁이 활발하게 이루어지고 있지만, ChatGPT는 현재 새로운 기술로써 수학교육에 어떻게 활용할지에 대한 실증적 연구는 부족한 실정이다(Rudolph et al., 2023). 따라서 교육 현장에서 ChatGPT가 적절하게 활용되려면 학생, 교사와 같은 교육 주체 측면에서 ChatGPT 사용에 대한 분석과 논의가 선행될 필요가 있다.

이에 본 연구에서는 수학교육에서 학생과 교사 교육을 위한 ChatGPT의 활용 가능성에 초점을 두어 ChatGPT의 수학교육 활용 가능성을 탐색하고자 하였다. 수학교육에서 학생의 사고와 교사의 관행을 분석하기 위해 널리 사용되는 학생의 문제해결 산출물과 예비교사와 학생과의 담화 자료를 사례로 선정하여 ChatGPT를 통해 분석한 후 수학교육전문가의 분석 결과와 비교하였다. 이를 통해 ChatGPT가 수학교육에 제공하는 교육적 잠재력과 제한점을 검토하고 AI가 수학교육에 미치는 영향을 이해하는데 도움을 제공하고자 한다.

II. 이론적 배경: ChatGPT와 수학교육

ChatGPT는 AI가 적용된 최신 언어 모델로서 사용자와 대화를 할 수 있는 챗봇이다(Heaven, 2020). 사용자는 ChatGPT에게 질문을 하거나 요청을 할 수 있으며 ChatGPT는 딥 러닝(deep learning)을 활용하여 사용자가 입력한 텍스트에 대응하는 응답을 생성한다(Rudolph et al., 2023). ChatGPT의 짧은 역사로 인해 ChatGPT를 수학교육에 활용하려는 연구는 아직 매우 미흡한 편이라 할 수 있다. Google Scholar, ERIC, RISS, KCI와 같은 논문 검색 포털에서 찾을 수 있는 ChatGPT와 관련된 수학교육 논문들은 ChatGPT의 수학적 표현 능력 검토(Frieder et al., 2023), 수학 문장제 생성(Shakarian et al., 2023), 대수 문제 해결에서 인간 교사와 ChatGPT의 힌트 제공 효과 비교(Pardos & Bhandari, 2023)와 같이 ChatGPT의 성능을 평가하는 연구가 주를 이루었으며, 그마저도 심사 과정을 거치지 않은 출판 전 공개 논문(preprint)이었다. 본 연구의 목적은 ChatGPT를 교육용 챗봇으로 간주하고 수학교육

에서 활용할 수 있는 가능성을 탐색하는 것이므로 보다 넓은 영역에서 교육용 챗봇이 학생과 교사 교육으로 사용되는 선행 연구들을 중점적으로 살펴보았다.

1. 학생을 위한 교육용 챗봇

수학교육에서 학생을 위한 교육용 챗봇을 개발하거나 그 효과를 검증하는 연구는 점점 늘어가는 추세이다(Ferrell & Frrell, 2020). 예를 들어, Nguyen 외(2019)는 고등학생들에게 함수를 가르치기 위해 지능형 챗봇을 설계하고 학생에게 힌트를 제공하거나 문제를 해결하는 방법을 가르치게 하였다. Cai 외(2021)는 등차수열에서 공차(common difference) 개념을 학생에게 안내하는 MathBot을 개발하고 온라인 및 서면 학습과의 효과를 비교하였다. MathBot은 온라인과 서면 학습에서 효과의 차이가 없었으며, 이러한 결과를 통해 저자는 MathBot이 기존 온라인 학습 자원을 보완하는 유망한 도구라고 제안하였다. Pai 외(2021)는 시간 표현 및 연산을 가르치기 위한 대화 기반의 지능형 교수 시스템을 개발하고, 그 효과성을 검증하기 위해 대만의 5학년 학생 134명을 대상으로 비교 연구를 수행하였다. 그 결과, 대화 기반 지능형 교수 시스템은 인간 교사의 수업과 마찬가지로 5학년 학생들의 수학 학습 능력과 동기를 향상시켰다.

이처럼 수학교육에서 교육용 챗봇의 활용에 관한 연구들은 주로 AI가 교사 역할을 수행하거나 보조할 수 있는 가능성을 탐색하고 있었다. 다만, 대부분의 연구가 정해진 교육 경로를 통해 학생의 답변을 유도하고 있으며, 이는 학생의 역할을 능동적인 학습의 주도자가 아니라 수동적인 지식의 수용자로 한정짓는다. AI가 교사로서의 역할을 올바르게 수행하려면 학생에게 지식을 전달하는 역할뿐만 아니라 학생의 사고를 올바르게 이해할 수 있어야 한다(Bommasani et al., 2021). 본 연구에서는 학생의 문제해결 과정을 사례로 제시하고 ChatGPT가 학생의 문제해결 전략과 수학적 지식을 이해하는지 살펴보았다.

2. 교사를 위한 교육용 챗봇

수학교육에서 AI는 학생용으로 주로 사용되며 교사를 위해 교육용 챗봇을 적용한 연구는 거의 수행되지

않았다(Hwang & Chang, 2021). 다만 몇 가지 사례는 존재하며, 주로 교사의 전문성개발을 위해 교육용 챗봇을 사용하고 있다. 예컨대, Beyer(2022)는 교사의 자료 기반의 수학 교수 관행을 개선하기 위한 모바일용 챗봇을 개발하였다. 교사의 프로파일 정보에 따라 적절한 수학적 내용을 담은 자료를 제안하고 자료 기반의 자기 조절(self-regulation) 교사 학습의 지원 방향을 논의하였다. Lee와 Yeo(2022)는 예비교사의 반응적 교수 관행을 증진시키기 위한 교육용 챗봇을 개발하였다. 개발한 교육용 챗봇은 학생용 챗봇이 교사의 역할을 수행한 것과는 달리 가상 학생이 되어 예비교사와 상호작용하였으며 이를 통해 예비교사들이 가상 학생의 사고를 이해하고 반응적으로 교수할 수 있는 연습 기회를 제공하였다.

교사가 질 높은 수학 수업을 하려면 학생의 필요에 따라 교수 결정을 내릴 수 있어야 하고 학생의 사고와 지식을 이끌어낼 수 있어야 한다(NCTM, 2000). 이를 위해 교사 교육자는 교사가 성장할 수 있도록 연구와 성찰의 기회를 제공해야 한다(Graham & Fennell, 2001). 본 연구에서는 예비교사와 학생의 담화 자료를 ChatGPT에게 제시하고 예비교사의 발문이 학생의 반응에 따라 이루어지는지, 수학적 사고를 유발하는 발문인지 분석하도록 요청하였다. 이를 통해 ChatGPT가 교사를 위한 교육용 챗봇으로 활용 가능한지 탐색해보았다.

III. 연구방법

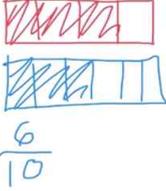
1. 분석 사례

본 연구에서는 사례 연구 분석 방법(Merriam, 1998)

을 사용하였으며, 선정된 사례는 분수의 크기 비교 문제에서 어려움을 보인 초등학교 3학년 슬기의 사례이다. 분수는 학생들이 이해하기에 어려우며(Aliustaoğlu et al., 2018; Kara & Incikabi, 2018), 문제해결과정에서 나타난 학생의 어려움을 탐구하는 과정은 수학적으로 중요한 정보를 얻을 수 있고 이를 통해 학생들에게 능동적이고 탐색적인 참여를 장려할 수 있다(Metcalf, 2017). 분석 자료는 학생용 자료와 예비교사용 자료로 구분된다. 먼저 학생용 자료는 분수의 크기 비교 문제에 대한 슬기의 문제해결 과정 산출물이며 구체적인 내용은 [표 1]과 같다. 이 문제는 Webel과 Conner(2017)가 예비교사의 발문 관행을 분석하기 위해 설계하였으며, 이분모 분수의 덧셈을 배우지 않은 초등학교 3학년 학생들이 $\frac{1}{2}$ 을 사용하여 $\frac{3}{4}$ 와 $\frac{3}{6}$ 를 비교한 후 1보다 큰지 작은지 추론할 수 있는지를 유도하는 문제이다. 본 연구에서는 이 문제를 C시의 초등학교 3학년 학생 84명에게 제시하고 이 중 단위의 크기를 고려하지 않고 자연수의 덧셈과 같이 분자는 분자끼리 분모는 분모끼리 더하여 오답을 도출한 슬기의 사례를 분석 사례로 선정하였다.

예비교사용 자료는 슬기의 오개념을 교정하기 위한 목적으로 수행된 슬기와 예비교사 2인의 담화 자료를 사용하였다. 예비교사 2인은 초등수학교육과 2학년에 재학 중이며 분수의 개념과 연산, 분수에 관한 학생의 문제해결과정에서 나타나는 수학적 사고를 살펴보는 수학교육 방법론을 수강하는 학생들이었다. 담화 자료에 대한 데이터 수집은 2022년 9월 말에 이루어졌다. 두 예비교사는 슬기와의 면담을 각자 수행하였으며, 담화 내용은 녹음한 후 분석을 위해 전사하였다.

[표 1] 분수의 크기 비교 문제와 슬기의 문제해결 과정

분수의 크기 비교 문제	민수는 밀가루 반죽을 만들고 있습니다. 민수는 먼저 그릇에 $\frac{3}{4}$ 컵의 밀가루를 붓고 $\frac{3}{6}$ 컵의 밀가루를 더 부었습니다. 민수가 사용한 밀가루의 총 양은 1컵보다 많은가요? 적은가요?	
슬기의 문제해결 과정 (그림, 설명)		먼저 $\frac{3}{4}$ 이니까 하나, 둘, 셋, 넷, 4개 중에 3개예요. 그리고 이것은 $\frac{3}{6}$ 이니까 하나, 둘, 셋, 넷, 다섯, 여섯, 6개 중에 3개구요. 그리고 이것을 다 더하면 $\frac{6}{10}$ 이 되요. 그래서 민수가 사용한 밀가루는 한 컵보다 작아요.

ChatGPT의 분석에 앞서, 초등수학교육 박사 학위를 소지한 전문가 2명이 토론을 통해 두 자료에 대한 분석을 진행하였다. 학생용 분석 자료는 슬기의 문제 해결 전략과 사고에 대한 분석이 이루어졌으며, 예비교사용 분석 자료는 반응적 교수(Ball, 1993; Ball et al., 2001; Webel & Conner, 2017))에 근거하여 예비교사의 발문이 학생의 반응을 토대로 이루어졌는지 그리고 사용한 발문이 학생의 수학적 사고를 이끌어냈는지를 분석하였다. 여기서 학생의 반응을 토대로 발문이 이루어졌다는 점은 단순히 학생의 반응에 응답하는 것이 아니라 학생의 지식과 사고에 따라 교수 전략을 유연하게 변화시킬 수 있는지를 의미한다(Ball et al., 2001). 예컨대, 학생이 ‘근데 저는 분모를 같게 만드는 법을 아직 못 배웠어요.’ 라고 했는데 예비교사가 ‘하지만 조각의 크기가 다르면 조각을 더할 수 없어’라고 응답했다고 하자. 학생이 분모를 같게 만드는 방법(공통분모)에 대해 배우지 않았다고 했음에도 교사는 다른 교수 전략의 사용이나 학생의 사고를 이해하려는 발문을 사용하는 대신 교사가 가진 교수 의도에 따라 조각의 크기를 같게 만들어야 된다고 이야기하고 있다. 따라서 이 발문은 학생의 반응에 따라 이루어지지 않았으며 학생의 수학적 사고를 이끌어내지도 못했다고 판단할 수 있다.

슬기의 문제해결 전략과 지식에 대한 수학교육전문가들의 분석 결과는 다음과 같다.

슬기는 $\frac{3}{4}$ 과 $\frac{3}{6}$ 을 크기가 같은 2개의 직사각형(whole)으로 표현한 후 $\frac{3}{4}$ 을 4개의 조각(part) 중 3개의 조각이 색칠된 빨간 직사각형, $\frac{3}{6}$ 은 6개의 조각 중 3개의 조각이 색칠된 파란 직사각형으로 적절하게 표현하였다. 그러나 슬기는 조각의 크기가 다른 점을 고려하지 않고 조각의 수를 더하여 계산하였다. 빨간 직사각형이 4개의 조각으로 나뉘어져 있고 파란 직사각형은 6개의 조각으로 나뉘어져 있으므로 분모를 $4+6=10$ 으로 생각했으며, 분자는 빨간 직사각형에서 색칠된 조각의 수가 3개이고 파란 직사각형에서 색칠된 조각의 수가 3개이므로 $3+3=6$ 으로 생각하여 $\frac{6}{10}$ 이라

고 생각하였다. 따라서 $\frac{6}{10}$ 은 1보다 작으므로 민수가 사용한 밀가루가 한 컵보다 작다고 설명하였다. 슬기는 분수의 개념과 전체(whole)와 부분(part)을 이해하고 주어진 분수를 사각형 모델로 표현할 수 있다. 그러나 분수의 덧셈에서 각 조각의 크기가 다른데도 불구하고 개수로만 이해하여 더하고 있으며, 자연수의 덧셈에 편향된 사고를 가지고 분자는 분자끼리 분모는 분모끼리 더하고 있다.

교사용 분석 자료로 사용한 슬기와 예비교사 2명의 담화 자료는 [표 2]와 같다. 예비교사 B가 먼저 슬기와 면담하였으며, 이를 간격을 두고 예비교사 A가 면담을 수행하였다. 예비교사 A, B의 발문이 학생의 반응을 토대로 이루어졌는지 그리고 학생의 수학적 사고를 이끌어냈는지를 분석한 결과는 다음과 같다.

예비교사 A는 슬기의 해결 전략을 탐색하기 위해 분모가 10이라고 답한 이유를 물었다. 그리고 문제해결 과정에서 나타난 그림을 보고 한 조각의 크기(단위 분수 $\frac{1}{4}$, $\frac{1}{6}$)와 색칠된 조각의 크기($\frac{3}{4}$, $\frac{3}{6}$)를 비교하게 하였다. 그리고 하나의 전체에 색칠된 조각들을 넣었을 때 전체가 1보다 큰지 예상해보게 하였다. 그러나 슬기가 이 질문에 적절하게 대답하지 못하자 예비교사 A는 발문의 방향을 유연하게 전환하였다. $\frac{3}{4}$ 과 $\frac{3}{6}$ 의 크기를 직관적으로 가늠할 수 있는 $\frac{1}{2}$ (reference number)과 비교하게 하고 이를 통해 슬기가 $\frac{3}{4}$ 과 $\frac{3}{6}$ 의 합이 1보다 크다는 것을 깨닫게 하였다. 예비교사 A가 사용한 발문들은 학생이 사용한 전략과 사고 과정을 적절하게 이끌어냈으며 슬기의 현재 지식과 반응을 기반으로 이루어졌다. 따라서 예비교사 A는 학생의 반응을 토대로 수학적 사고를 이끌어내는 발문을 사용하는 교사이다.

예비교사 B는 먼저 슬기가 분모가 다른 분수를 더하는 방법과 공통분모의 개념을 알고 있는지 확인하였다. 그리고 공통분모를 사용하도록 직접 가르치려하거나(예: 분수를 더할 때는 공통분모를 사용해야 해) 학생의 사고를 교사의 의도에 따라 유도하였다(예: 빨간 색과 파란색 조각의 수를 더 많이 나누어 같은 크기로

[표 2] 슬기와 예비교사 A, B와의 담화

예비교사 A	예비교사 B
<p>예비교사 A: 답의 분모가 왜 10이라고 생각하니?</p> <p>슬기: 빨간색 조각이 4개있고 파란색 조각이 6개 있으니 다 합치면 10이에요.</p> <p>예비교사 A: 사각형 하나를 4로 나눈 크기와 6으로 나눈 크기를 비교할 수 있어?</p> <p>슬기: 빨간색 사각형의 조각 하나의 크기가 파란색 사각형의 조각 하나의 크기보다 커요.</p> <p>예비교사 A: 아 그렇구나. 그럼 빨간색 사각형의 색칠된 부분과 파란색 사각형의 색칠된 부분의 크기를 비교할 수 있어?</p> <p>슬기: 빨간색 사각형의 색칠된 부분이 파란색 사각형의 색칠된 부분보다 커 보여요.</p> <p>예비교사 A: 그럼 하나의 사각형에 빨간색과 파란색으로 색칠된 조각들을 모두 넣으면 1보다 크니?</p> <p>슬기: 아뇨.. 두 분수의 합이 $\frac{6}{10}$ 이니 1보다 작아요.</p> <p>예비교사 A: 빨간색으로 칠해진 조각을 전체의 절반과 비교해볼까?</p> <p>슬기: 음... 빨간색으로 색칠된 것이 반보다 큰 것 같아요.</p> <p>예비교사 A: 파란색으로 색칠된 조각의 크기를 절반과 비교해볼래?</p> <p>슬기: 파란색 조각은 $\frac{3}{6}$ 이예요. $\frac{3}{6}$ 은 절반과 같아요.</p> <p>예비교사 A: 빨간색 조각이 절반보다 크고 파란색 부분이 절반이라고 했지? 그럼 빨간색 조각들과 파란색 조각들을 합치면 1보다 클까 작을까?</p> <p>슬기: 아하! 내가 틀린 것 같아요! $\frac{3}{4}$ 는 반 컵보다 크고 $\frac{3}{6}$ 은 반 컵이니까 이 둘을 더하면 1컵보다 커야 돼요.</p>	<p>예비교사 B: 분모가 다른 분수를 더하는 방법을 기억하니?</p> <p>슬기: 분수와 그 크기를 비교하는 방법을 배웠어요. 아직 분수를 더하는 방법을 배우지 않았어요.</p> <p>예비교사 B: 공통분모가 뭘까?</p> <p>슬기: 뭘지 모르겠어요.</p> <p>예비교사 B: 분수를 더할 때는 공통분모를 사용해야 해.</p> <p>슬기: 근데 저는 분모를 같게 만드는 법을 아직 못 배웠어요.</p> <p>예비교사 B: 하지만 조각의 크기가 다르면 조각을 더할 수 없어.</p> <p>슬기: 네.</p> <p>예비교사 B: 슬기가 그린 그림에서 빨간색과 파란색 조각들의 크기가 같니?</p> <p>슬기: 빨간색 사각형은 4등분, 파란색 사각형은 같은 크기로 6등분해서 빨간색 사각형 한 조각이 파란색 사각형 한 조각보다 커요.</p> <p>예비교사 B: 일단 너는 공통분모가 뭘지 배워야해. 공통분모는 두 분모를 같은 크기로 만드는 거야.</p> <p>슬기: 무슨 말인지 잘 모르겠어요.</p> <p>예비교사 B: 그럼 빨간색과 파란색 조각의 수를 더 많이 나누어 같은 크기로 만들어 봐.</p> <p>슬기: 어떻게 같은 크기로 만들어요?</p> <p>예비교사 B: 그니까 공통분모를 배워야 문제를 푼다 니깐.</p>

만들어 봐). 학생이 공통분모와 조각의 크기를 같게 만드는 방법을 모른다고 답했는데도 불구하고 예비교사 B는 학생의 반응을 무시하고 교사가 의도하는 특정 지식(공통분모 사용)을 가르치려고 노력하였다. 이 담화에서 사용된 교사의 발문은 학생의 반응을 고려하지 않고 이루어졌으며, 사용한 발문들은 학생이 수학적

개념을 회상하게 하거나 직접적으로 가르치고 학생의 사고 방향을 의도적으로 유도하고 있다. 또한 공통분모를 사용하는 절차적 지식만을 강조하여 학생의 수학적 사고를 이끌어내는데 어려움을 보였다. 따라서 예비교사 B는 학생의 반응을 고려하지 않고 교사의 의도에 따라 학생의 수학적 사고를 이끌어가는 교사이다.

2. 분석 방법

분석에 사용된 ChatGPT는 2023년 3월 15일에 출시된 ChatGPT Plus(ver 4.0)를 사용하였다. ChatGPT는 사용하는 언어 종류에 따라 지원하는 데이터의 양과 질의 차이가 있으며, 영어가 아닌 다른 언어의 경우 명료성이 현저하게 떨어진다(Seghier, 2023). 이에 ChatGPT 내에서 자동 언어 번역과 전환을 지원하는 확장프로그램을 통해 텍스트를 입력하였으며, 언어 번역 프로그램의 특성상 오역된 부분(예: responsiveness를 민감도나 응답성으로 번역, fractional을 소수로 번역)이나 문맥이 어색한 부분은 영어 원문을 검토하여 수정(예: 응답성은 반응성으로 통일) 후 기술하였다.

분석은 학생용 자료와 예비교사용 자료를 구분하여 실시하였다. 학생용 자료는 슬기에게 제시한 분수의 크기 비교 문제와 슬기의 문제해결과정으로서 이 자료를 ChatGPT에 입력하고 슬기의 문제해결 전략과 수학적 지식(아는 것과 모르는 것)을 ChatGPT가 평가할 수 있는지 살펴보았다. 분수의 크기 비교 문제와 슬기의 문제해결 과정을 텍스트로 입력하고 슬기의 문제해결 전략을 구체적으로 설명하도록 ChatGPT에게 요청하였다. 슬기가 문제 해결을 위해 그린 그림은 텍스트로 설명하였다(슬기는 빨간색 직사각형과 파란색 직사각형을 똑같은 크기로 그렸습니다. 빨간색 직사각형을 4등분 한 뒤 3개를 빨간색으로 색칠하고 파란색 직사각형을 6등분 한 뒤 3개를 파란색으로 색칠했습니다.

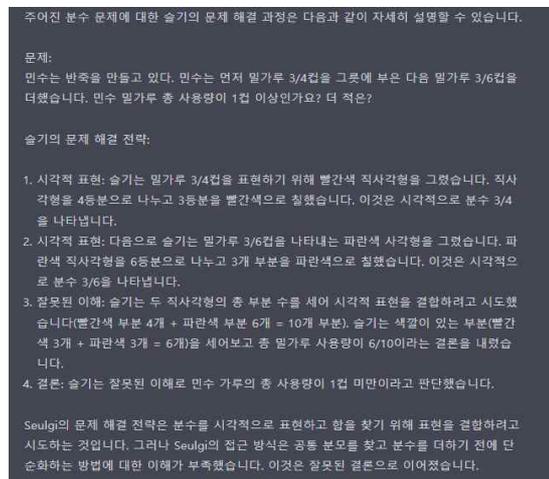
그리고 $\frac{6}{10}$ 이라고 정답을 썼습니다). 교사용 자료는 슬기와 두 예비교사의 담화 자료를 사용하였으며, 두 개의 담화 내용을 각각 ChatGPT에 입력하고 예비교사들의 발문이 학생의 반응에 따라 이루어졌는지, 발문이 수학적 사고를 이끌어내는데 적절했는지를 기준으로 두 예비교사의 발문을 평가하도록 요청하였다. 이러한 질문을 통해 도출된 ChatGPT의 응답을 살펴보고 수학교육전문가들이 분석한 결과와 비교하였다. ChatGPT 3.0의 경우, 부정확하고 무의미한 답변을 작성하거나 정확하지 않은 출처를 제공할 수 있고 텍스트의 구성성과 명료성에 따라 응답이 달라질 수 있다는 제한점이 있었다(OpenAI, 2022). 이에 어떤 기준으로 예비교사들의 발문을 평가했는지 그 평가 기준의 출처가 어디인지에 대해 질문하였으며, 분석 사례에

대한 평가를 요청할 때 여러 어감의 텍스트를 입력하고 이 중 질문 의도에 맞는 가장 적절한 답변을 연구 결과에 기술하였다.

IV. 연구 결과

1. 학생용 자료 분석 결과

슬기의 문제해결 전략에 대한 ChatGPT의 답변은 [그림 1]과 같다.

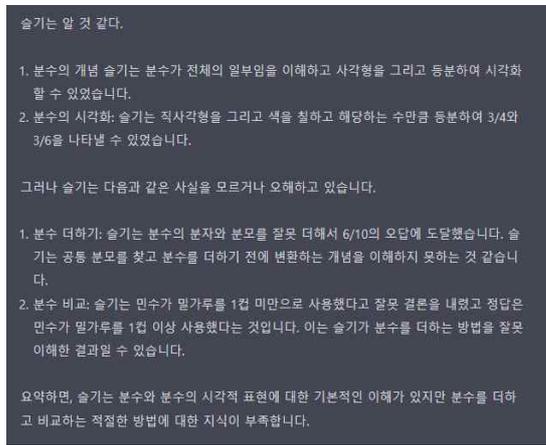


[그림 1] 슬기의 문제해결 전략에 대한 ChatGPT의 응답

ChatGPT는 슬기의 문제해결 전략에 대해 $\frac{3}{4}$ 컵과 $\frac{3}{6}$ 컵의 밀가루를 시각적 표현으로 적절하게 나타냈다고 평가하였다. 또한 두 직사각형의 총 부분 수를 세어(4등분된 빨간색 부분의 수+6등분된 파란색 부분의 수=10개) 분모를 나타내고, 색칠된 조각의 수를 세어(빨간색으로 색칠된 부분의 수 3개+파란색으로 색칠된 부분의 수 3개=6개)를 분자로 나타내어 $\frac{6}{10}$ 이라고 답한 것이 슬기의 오개념이라고 분석하였다. 슬기의 문제해결 전략은 분수를 시각적으로 표현하여 합을 찾기 위한 표현의 결합이며 공통분모를 찾고 약분하는 방법에 대한 이해가 부족하다고 지적하였다. ChatGPT는

입력한 분수 크기 문제, 그림에 대한 설명, 슬기의 문제해결 과정의 맥락을 정확하게 이해하고 있었다. 또한 응답 결과는 주어진 문제를 직사각형 모델로 적절하게 표현하였으나 조각의 크기가 다른데도 불구하고 개수로만 이해하여 더했다고 한 수학교육전문가의 분석 결과와 유사하게 나타났다.

다음으로 슬기의 수학적 지식에 대한 ChatGPT의 답변은 [그림 2]와 같다.



[그림 2] 슬기의 수학적 지식에 대한 ChatGPT의 응답

ChatGPT는 슬기가 분수의 개념과 분수의 시각화를 알고 있다고 평가하였다. 구체적으로 살펴보면, 분수가 전체의 일부임을 이해하고 분수를 사각형으로 시각화할 수 있으므로 분수의 개념을 이해한다고 하였으며, 직사각형으로 주어진 두 분수 $\frac{3}{4}$ 과 $\frac{3}{6}$ 을 표현할 수 있으므로 분수를 시각화 할 수 있다고 분석하였다. 슬기가 모르거나 잘못 이해한 것은 분수 덧셈과 분수 비교라고 응답하였다. 분수의 분자와 분모를 잘못 더하여 $\frac{6}{10}$ 이라는 오답을 도출했으며 이는 분수를 더할 때 공통분모를 사용하는 개념을 이해하지 못하고 있다고 하였다. 또한 분수 크기 비교 문제의 정답으로 밀가루를 1컵 미만으로 사용했다고 잘못 결론 낸 점에 대해 분수 비교를 이해하지 못한다고 분석하였다. 수학교육전문가는 슬기가 분수의 개념과 전체와 부분을 이해하고 주어진 분수를 사각형 모델로 표현할 수 있

으나 분자와 분모를 자연수처럼 더하여 오답을 도출했다고 분석하였으며, 이는 ChatGPT의 응답 결과와 유사하다.

2. 교사용 자료 분석 결과

두 예비교사의 발문에 대한 ChatGPT의 평가 결과는 [그림 3]과 같다.

ChatGPT는 예비교사 A가 슬기의 응답을 바탕으로 문제해결 과정을 안내하고 분수와 그 속성을 비교, 분석, 재고하도록 유도하였으므로 학생의 반응을 토대로 수학적 사고를 이끌어낸 발문을 사용했다고 평가하였다. 구체적으로 살펴보면, 분모가 10이라고 답한 이유를 묻는 발문은 슬기의 기존 이해를 파악하고 추가적인 설명을 요청하는 발문이라고 하였다. 이어지는 발문인 사각형의 크기를 4와 6으로 나눈 조각의 크기를 묻는 발문(두 사각형의 단위분수 크기)은 슬기의 응답을 기반으로 제시되었으며 분수 사이의 크기 차이를 인식하도록 돕는다고 분석하였다. 또한 색칠된 조각의 크기($\frac{3}{4}$ 과 $\frac{3}{6}$)를 비교하는 발문이 슬기가 색칠된 조각의 크기를 분석하고 조각들 사이의 관계를 더 깊이 이해하는 발문이라고 하였다. 빨간색과 파란색 조각을 하나의 사각형으로 합쳤을 때 1보다 크지를 묻는 발문은 슬기가 이해한 개념을 비판적으로 생각하도록 요구하였다고 분석하였으며 색칠된 조각의 크기를 절반($\frac{1}{2}$)과 비교하는 발문은 슬기의 초기 이해를 재고하고 올바른 결론에 도달할 수 있는 발문이라고 평가했다.

ChatGPT는 예비교사 B의 발문이 슬기의 사전 지식을 인정하고 오개념을 해결하려는 시도를 보이므로 어느 정도의 반응성을 보인다고 평가했다. 그러나 제시한 발문과 설명은 슬기가 공통분모의 개념을 이해하는데 도움이 될 만큼 명확하거나 상세하지 못하므로 수학적 사고를 효과적으로 이끌어내는데 부족하다고 분석하였다. 구체적으로 살펴보면, 분모가 다른 분수를 더하는 방법을 기억하는가를 묻는 발문이 슬기의 사전 지식을 고려했으나 제시된 분수 크기 문제를 직접적으로 다루지 않는다고 하였다. 또한 슬기가 공통분모를 배운 적이 없다고 응답하여 공통분모에 대한 설명을 제시했으나 그 설명이 명확하거나 상세하지 않으므로

<p>두 가지 평가 기준(반응성 및 수학적 사고 유도 여부)을 사용하여 예비 교사 A의 질문에 대한 보다 자세한 평가를 제공하려면:</p> <p>민감도:</p> <ol style="list-style-type: none"> 예비교사 A는 먼저 슬기에게 답의 분모가 10이라고 생각하는 이유를 묻습니다. 이 질문은 슬기의 기존 이해를 인정하고 추가 설명을 유도합니다. 슬기가 4와 6으로 나눈 사각형의 크기를 비교하자 예비교사 A는 슬기의 반응을 인지하고 슬기에게 색칠한 부분의 크기를 비교하라고 한다. 이것은 Seolgi의 입력에 대한 주의력을 보여주고 이를 기반으로 합니다. 예비 교사 A는 슬기의 답변에 계속해서 빨간색과 파란색 조각을 하나의 사각형으로 합쳤을 때의 합을 고려하도록 요청하여 슬기가 개념에 대해 비판적으로 생각하도록 격려합니다. <p>수학적 사고 유도:</p> <ol style="list-style-type: none"> 예비교사 A는 슬기에게 4로 나눈 제곱의 크기와 6으로 나눈 크기를 비교하도록 하여 슬기가 분수 사이의 크기 차이를 인식하도록 돕습니다. 예비 교사 A는 슬기에게 빨간색과 파란색 사각형의 색칠된 부분을 분석하도록 격려하여 분수가 서로 어떻게 관련되어 있는지 더 깊이 이해하도록 합니다. 예비 교사 A는 슬기에게 전체의 절반에 대해 빨간색과 파란색 부분을 비교하도록 요청함으로써 슬기가 초기 오해를 재고하고 보다 정확한 결론에 도달할 수 있도록 합니다. <p>전반적으로 예비 교사 A의 질문은 슬기의 답변을 바탕으로 문제 해결 과정을 안내함으로써 강력한 응답성을 보여줍니다. 또한 이러한 질문은 Seolgi가 분수와 그 속성에 대한 이해를 비교, 분석 및 재고하도록 유도하여 수학적 사고를 성공적으로 이끌어냅니다. 이 접근 방식은 학생에게 더 정확한 결론과 풍부한 학습 경험을 제공합니다.</p>	<p>두 가지 평가 기준(반응성 및 수학적 사고 유도 여부)을 사용하여 예비 교사 B의 질문을 전체적으로 평가하려면:</p> <p>민감도:</p> <ol style="list-style-type: none"> 예비교사 B는 슬기가 분모가 다른 분수를 더하는 방법을 기억하는지 묻는 것으로 시작합니다. 이 질문은 슬기의 사전 지식을 고려한 것이지만 현재의 문제를 직접적으로 다루는 것은 아닙니다. 슬기가 공통분모를 모른다고 하자 예비교사 B가 설명을 해준다. 그러나 슬기의 혼란을 해소하기에는 설명이 부족할 수 있다. 예비 교사 B는 슬기에게 빨간색과 파란색 조각의 크기를 묻고 학생이 문제에 대해 생각하도록 격려합니다. 그러나 후속 질문과 설명은 슬기의 혼란을 적절하게 다루거나 공통 분모의 개념을 이해하는 데 도움이 되지 않습니다. <p>수학적 사고 유도:</p> <ol style="list-style-type: none"> 예비교사 B는 슬기에게 공통 분모에 대해 물어보면서 분모가 다른 분수를 더하는 핵심 개념을 소개합니다. 그러나 슬기가 그 중요성을 이해하기에는 설명이 명확하지 않습니다. 빨간색과 파란색 조각이 같은 크기인지 슬기에게 물어보면 공통 분모 개념과의 연결이 효과적으로 확립되지는 않았지만 학생이 문제를 분석하도록 유도합니다. 예비교사 B는 공통분모에 대한 학습의 필요성을 반복적으로 강조하지만 슬기가 개념을 파악하고 문제에 적용할 수 있도록 충분한 안내나 설명을 제공하지 않습니다. <p>전반적으로 예비교사 B의 질문은 슬기의 사전 지식을 인정하고 그들의 혼란을 해결하려고 시도함으로써 어느 정도 반응성을 보여줍니다. 그러나 제공된 질문과 설명은 슬기가 공통 분모의 개념을 이해하는 데 도움이 될 만큼 명확하거나 상세하지 않습니다. 결과적으로 슬기는 개념과 이를 당면한 문제에 적용하는 방법에 대해 확신하지 못하기 때문에 문제가 수학적 사고를 효과적으로 이끌어내는 데 부족합니다.</p>
--	---

예비교사 A

예비교사 B

[그림 3] 두 명의 예비교사에 대한 ChatGPT 평가 결과

학생의 반응을 고려하고 수학적 사고를 이끌어내는데 부족한 발문이라고 평가하였다. 또한 빨간색과 파란색 조각의 크기를 묻는 발문은 학생이 문제에 대해 생각하도록 격려하고 슬기에게 문제를 분석하도록 요구하는 발문이었으나 공통분모와의 연결성은 떨어진다고 하였다. 이어지는 발문들은 슬기의 오개념을 다루거나 공통분모의 개념을 이해하는데 도움이 되지 않는다고 하였다. 또한 예비교사 B가 공통분모의 필요성을 반복적으로 강조하지만 슬기가 개념을 파악하고 문제에 적용할 만큼 충분한 안내와 설명을 제공하지 않았다고 분석하였다.

이상의 ChatGPT가 수행한 예비교사 2인의 분석 결과는 수학교육전문가의 분석 결과와 유사점과 차이점을 보였다.

첫째, 전반적인 평가 결과에서 유사점과 차이점을 보였다. ChatGPT는 수학교육 전문가들과 유사하게 예비교사 A가 학생의 반응을 토대로 수학적 사고를 이끌어내는 발문을 구사했다고 분석했으며, 예비교사 B가 수학적 사고를 효과적으로 이끌어내는데 부족하다고 평가하였다. 그러나 예비교사 B가 학생의 반응을 토대로 발문을 사용했는지를 평가한 결과는 차이를 보

였다. 수학교육전문가는 예비교사 B가 학생의 반응을 고려하지 않은 발문을 구사한다고 하였으나 ChatGPT는 예비교사 B가 어느 정도의 반응성을 보인다고 평가하였다. 이러한 차이가 나타난 이유를 살펴보면, ChatGPT가 예비교사 B의 ‘분모가 다른 분수를 더하는 방법을 기억하니?’ 라는 발문이 학생의 사전 지식을 인정하는 발문이며, 여러 발문을 통해 학생이 가진 오개념을 해결하려고 시도한 점을 반응성이 있다고 보고 있었기 때문이다. 그러나 이 발문은 담화의 시작 발문이고 단순히 수학적 개념을 회상하게 하는 발문이므로 학생의 반응과는 무관하다. 또한 학생의 오개념을 해결하려고 여러 발문을 시도한 점을 학생의 반응에 따라 발문을 구사했다고 보기는 어렵다.

둘째, 개별 발문에 대한 해석이 유사하다. 예를 들어, ‘답의 분모가 왜 10이라고 생각하니?’라는 예비교사 A의 발문에 대해 수학교육전문가는 슬기의 해결 전략을 탐색하는 발문이라고 평가했으며, 이와 유사하게 ChatGPT는 슬기의 기존 이해를 인정하고 추가적인 설명을 요구하는 발문이라고 평가하였다. 또한 빨간색과 파란색의 색칠된 조각들의 크기를 $\frac{1}{2}$ 과 비교

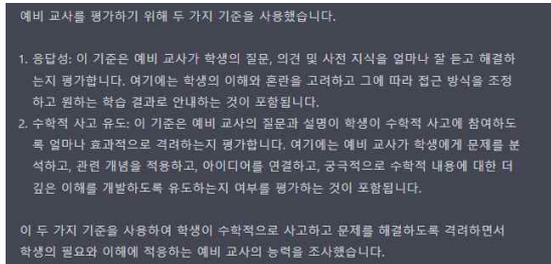
하는 예비교사 A의 발문에 대하여 수학교육전문가는 $\frac{3}{4}$ 과 $\frac{3}{6}$ 의 크기를 직관적으로 가늠할 수 있으므로 슬기가 오개념을 깨닫게 하는 발문이라고 평가하였다. 이와 유사하게 ChatGPT도 이 발문이 슬기의 초기 오해를 재고하게 하고 정확한 결론에 이를 수 있는 발문이라고 평가하였다.

셋째, 평가 기준의 명확성에서 차이를 보였다. 두 가지 기준(학생의 반응에 따라 발문했는지, 발문이 수학적 사고를 이끌어내는데 적절했는지)에 따라 예비교사 2인의 발문의 평가를 ChatGPT에게 요청했으나 일부 응답은 평가 기준을 벗어나 있다. 예를 들어, ChatGPT가 예비교사 B의 반응성에 대해 분석한 일부 결과를 살펴보면 반응성과는 무관한 평가 결과가 일부 나타났다(예: 슬기의 혼란을 해소하기에는 설명이 부족합니다). 이러한 응답은 교사의 발문이 학생에게 기대되는 효과에 가까우며 학생의 반응에 따라 발문을 제시했는지를 판단하는 것과는 무관하다. 또한 예비교사 B에 대한 응답에서 수학적 사고를 이끌어내는 발문인지를 평가했을 때, ChatGPT는 설명의 명확성과 충분한 안내가 부족하므로 수학적 사고를 이끌어내는데 실패했다고 분석하였다. 그러나 교사가 의도적으로 학생의 사고 방향을 변화시키거나(taking over) 직접적으로 개념을 설명하는(directive teaching) 발문은 학생의 수학적 사고를 이끌어내는데 도움을 제공하기 어렵다 (Webel & Conner, 2017).

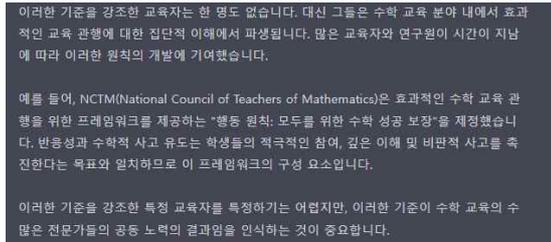
ChatGPT에게 어떤 기준을 가지고 예비교사들을 그림 3과 같이 평가했는지 그리고 그 기준의 출처에 대해 질문하였으며, 이에 대한 응답은 [그림 4, 5]와 같다.

ChatGPT는 예비교사들이 학생의 반응을 토대로 수학적 사고를 이끌어냈는지에 대한 평가 기준으로 반응성과 수학적 사고 유도도의 두 가지 기준을 제시했으며, 이는 연구자가 ChatGPT에게 요청한 두 가지 평가 기준과도 일치했다. 다만 두 평가 기준에 대한 용어를 변화시키는 모습을 보였다. 학생의 반응에 따라 발문했는지를 반응성이라고 정의하고 발문이 수학적 사고를 이끌어냈는지는 수학적 사고 유도도로 정의하고 있었다. ChatGPT는 반응성을 학생의 질문, 의견, 사전 지식을 잘 듣고 해결하는지 그리고 학생의 이해와 혼란을 고려하고 접근 방식을 조정하고 원하는 학습 결과로 안내하는 과정을 포함한다고 진술했다. 또한 수

학적 사고 유도는 예비교사의 질문과 설명이 얼마나 효과적으로 학생들의 수학적 사고에 참여하도록 격려하는지 평가하는 것이며, 문제 분석, 개념 적용, 아이디어 연결을 통해 수학적으로 더 깊은 이해를 개발하도록 유도하는지 여부를 평가한다고 하였다. 진술한 반응성과 수학적 사고 유도도의 정의는 수학교육에서 학생의 수학적 사고 개발의 기회를 제공하고 학생의 사고 특성을 평가, 유도, 참여, 사용하는 반응적 교수(responsive teaching)의 개념(Ball, 1993; Ball et al., 2001)과 유사하였다.



[그림 4] 평가 기준에 대한 ChatGPT 응답



[그림 5] 평가 기준의 출처에 대한 ChatGPT 응답

평가 기준의 출처를 묻는 질문에 대해서 ChatGPT는 한 명의 교육자가 아니라 집단적인 이해에 따라 파생된 원칙이며 NCTM(2014)의 ‘Principles to Actions: Ensuring Mathematical Success for All’의 분석틀의 구성 요소라고 진술했다. 구체적인 출처를 묻자 ChatGPT는 [그림 6]과 같이 응답하였다.

ChatGPT는 NCTM(2014)의 ‘Principles to Actions: Ensuring Mathematical Success for All’에서 강조한 8가지 수학 교수법 중 ‘의미 있는 수학적 담론 촉진’, ‘목적이 있는 발문 제기’의 두 가지 평가 기준과 일치한다고 응답하였다. 특히, ‘의미 있는 수학적 담론 촉

진'은 학생의 아이디어에 적절하게 반응하는 반응성, '목적에 있는 발문 제기'는 학생이 비판적으로 사고하고, 추론하고, 수학적 개념과 관련된 문제에 참여하도록 격려하는 수학적 사고 유도과 관련된다'고 응답하였다. 이러한 결과는 ChatGPT의 응답이 자체적인 기준을 근거로 도출된 것이 아니라 구체적인 이론적 근거를 바탕으로 제시되었음을 시사한다.

NCTM(National Council of Teachers of Mathematics)의 행동 원칙: 모두를 위한 수학 성공 보장에서 프레임워크는 8가지 수학 교수법에 중점을 둡니다. 그들은:

1. 학습에 집중할 수 있도록 수학 목표를 설정합니다.
2. 주론과 문제 해결을 촉진하는 작업을 구현합니다.
3. 수학적 표현을 사용하고 연결합니다.
4. 의미 있는 수학적 대화를 촉진합니다.
5. 의도적인 질문을 던집니다.
6. 개념 이해를 통해 절차적 유창성을 구축합니다.
7. 수학 학습에서 생산적인 투쟁을 지원합니다.
8. 학생 사고의 증거를 도출하고 사용합니다.

이전에 논의된 반응성 및 수학적 사고 유도에 대한 평가 기준은 이러한 교육 관행 중 몇 가지, 특히 관행 4(의미 있는 수학적 담론 촉진) 및 5(목적에 있는 질문 제기)와 일치합니다. 교사가 학생의 아이디어에 적절하게 반응하고 생각의 교환을 장려해야 하므로 교수 맥락에서 반응성은 의미 있는 대화로 연결될 수 있습니다. 수학적 사고 유도는 학생들이 비판적으로 사고하고, 추론하고, 수학적 개념과 관련된 문제 해결에 참여하도록 격려하는 의도적인 질문을 제기하는 것과 관련이 있습니다.

[그림 6] 구체적인 평가 기준 출처에 대한 ChatGPT의 응답

V. 논의

본 연구의 목적은 학생과 교사 교육용으로서 수학 교육에서의 ChatGPT 활용 가능성을 탐색하는 것이다. 이를 위해 분수 크기 비교 문제에 대한 슬기의 사례를 분석 사례로 선정하고 ChatGPT가 학생의 문제해결 전략과 수학적 지식, 예비교사의 발문을 평가 기준에 근거하여 분석할 수 있는지 살펴보았다. 또한 ChatGPT의 응답과 수학교육전문가와의 분석 결과를 비교하였다. 연구 결과를 바탕으로 수학교육에서 ChatGPT의 활용 가능성에 대한 논의를 기술하면 다음과 같다.

첫째, ChatGPT는 학생의 문제해결 전략과 수학적 지식을 탐색하는 능력을 보였다. ChatGPT는 분수 크기 비교 문제, 슬기의 그림에 대한 텍스트, 슬기의 문제해결 과정에 대한 설명의 맥락을 분석할 수 있었으며, 이를 토대로 슬기의 문제해결 전략과 수학적 지식

을 평가하였다. 이러한 응답 결과는 수학교육전문가의 분석 결과와 유사하게 나타났다. 특히, 조각의 크기 차이를 고려하지 않고 조각의 개수로 분수의 덧셈을 수행한 슬기의 오개념을 정확하게 지적한 것은 ChatGPT가 학생 사고를 분석하는 능력을 보여주는 흥미로운 예이다. 기존 학생용 챗봇에 관한 선행연구(예: Cai et al., 2021; Nguyen et al., 2019; Pai et al., 2021)들은 학생의 사고를 파악하지 못하고 교과 지식을 전달하거나 안내하는데 그친 반면, ChatGPT는 학생에게 지식을 전달하는 역할을 넘어 학생 사고 분석 결과를 바탕으로 학생의 개별적 특성을 이해할 수 있는 활용 가능성을 보였다. 이러한 결과는 ChatGPT가 학생의 지식과 사고를 파악하고 효과적인 교수·학습을 위한 정보를 제공할 수 있는 교사의 보조 도구로 활용될 수 있는 가능성을 보여준다.

둘째, ChatGPT는 평가 기준을 토대로 교사의 발문을 분석할 수 있었다. ChatGPT는 학생의 반응에 따라 발문이 이루어지는지, 발문이 수학적 사고를 이끌어내는지를 평가 기준으로 삼아 예비교사의 발문을 적절하게 분석하였다. 전체적인 분석 결과에서 ChatGPT는 수학교육전문가의 분석 결과와 유사하게 예비교사 A가 강한 반응성과 수학적 사고를 성공적으로 이끌어냈다고 분석했으며 예비교사 B는 어느 정도의 반응성과 수학적 사고를 이끌어내기에 부족했다고 평가했다. 또한 개별 발문이 두 평가 기준을 바탕으로 어떤 의미를 가지는지 분석하는 능력을 보였다(예: 답의 분모가 왜 10이라고 생각하니? 라는 발문을 슬기의 기존 이해를 인정하고 추가적인 설명을 요구하는 발문이라 평가). 그러나 평가 기준을 적용하는데 있어 일부 한계도 드러났다. 반응성을 기준으로 개별 발문을 분석했을 때, 반응성과 무관한 분석 결과를 도출했으며, 이로 인해 예비교사 B의 반응성에서 수학교육전문가와 차이 보이기도 하였다. 또한 수학적 사고 유도에 관한 응답에서도 수학적 사고를 이끌어내는 발문 기준으로 적합하지 않는 응답 결과를 도출하였다. 이러한 한계에도 불구하고 ChatGPT의 응답은 수학교육전문가와 상당히 유사한 결과를 보였으며 교사의 관행을 분석할 수 있는 잠재력을 보여주고 있다. 교사 교육자는 학생과의 대화를 통해 도출되는 교사의 관행을 분석하고 그 과정에서 전문성 개발을 지원하는 역할을 수행해야 한다(NCTM, 2000). 연구 결과에서 나타난 바와 같이,

ChatGPT는 교사 교육자가 교사의 관행을 분석하고 이해하기 위한 도구로 활용될 수 있는 가능성이 있음을 시사한다.

셋째, ChatGPT는 이론적 근거를 바탕으로 평가 기준을 설정하고 있었다. ChatGPT는 예비교사의 발문을 분석한 평가 기준을 반응성과 수학적 사고 유도라고 정의하고 응답하였으며 두 가지 평가 기준에 대한 정의는 반응적 교수에 관한 정의와 유사하게 나타났다. 또한 ChatGPT는 설정한 평가 기준에 대한 출처로 NCTM(2014)의 'Principles to Actions: Ensuring Mathematical Success for All'에서 강조한 8가지 수학교수법이라 응답했으며, 그 중 '의미 있는 수학적 담론 촉진'과 '목적어 있는 발문 제기'가 두 가지 평가 기준인 반응성과 수학적 사고 유도와 관련된 요소라고 응답하였다. 이처럼 ChatGPT는 수학적 이론을 근거로 평가 기준을 제시하고 있었으며, ChatGPT 3.0에서 지적되어 왔던 정확하지 않은 출처(OpenAI, 2022)의 문제가 ChatGPT 4.0에서는 개선된 것으로 보인다. 그러나 연구 결과에서 나타난 바와 같이 ChatGPT는 연구자가 의도한 평가 기준을 자체적인 기준에 따라 용어를 변화시켜 사용하고 있으며, 수학적 이론을 바탕으로 평가 기준을 설정했음에도 학생의 반응성과 수학적 사고를 정확하게 이해하지 못하는 모습을 보였다(예: 정확한 설명이 부족한 점을 반응성이 부족하다고 평가한 점). 이는 ChatGPT의 교육적 활용에 있어 숙련된 교사와 교사 교육자의 검토가 추가적으로 필요함을 시사한다.

VI. 결론 및 제언

이상의 결과와 논의를 바탕으로 결론과 시사점을 제시하면 다음과 같다.

첫째, ChatGPT는 수학교육에서 학생의 사고와 교사의 관행을 분석할 수 있는 도구로서의 활용 가능성이 있다. ChatGPT는 지금까지 수학교육에서 숙련된 교사와 교사 교육자가 수행해온 학생과 교사의 수학적 지식과 아이디어에 대한 분석을 보조할 수 있는 가능성을 보여주었다. 다만, ChatGPT가 제공한 정보를 확신해서는 안 된다. 본 연구의 결과에서 나타나듯이 ChatGPT가 제시하는 정보는 부정확할 수 있으며, 수

학적 이론을 분석 과정에 적절하게 적용할 수 있는지 여부 또한 불확실하다. 따라서 수학교육에서 ChatGPT를 효과적으로 활용하려면 수학적 이론의 출처와 용어 사용, 응답한 정보에 대한 충분한 검토가 병행해서 이루어져야 할 것이다. 또한 AI 기술의 교육적 활용에 대한 검토가 이루어지기 전에 ChatGPT의 교육적 사용이 보편화될 가능성을 살펴볼 필요가 있다. 이미 여러 교육청에서 ChatGPT의 교육 활용에 대한 교사 연수가 실시되거나 교육용 안내서가 배부되었으며(예: 박현아, 2023; 윤근호, 2023; 한형진, 2023), 이는 ChatGPT가 가진 단점에 대한 충분한 검토 없이 학교 현장에서 보편적으로 사용될 수 있음을 보여준다. 따라서 ChatGPT가 가진 잠재적 이점과 단점을 고려하면서 교육적 사용에서 발생하는 문제를 수정하는 데 중점을 둔 검토 방식을 고려해볼 수 있다.

둘째, 수학교육에서 AI를 활용한 실증적인 연구가 활성화될 필요가 있다. 수학교육에서 AI를 적용하기 위한 연구들이 지속적으로 이루어지고 있지만 여전히 수학교육에서 AI를 적용한 실증적인 연구는 거의 수행되지 않았다(Zhai, 2022). 본 연구의 결과는 일부 사례의 분석을 통해 나타난 결과이므로 AI가 인간 교사와 교사 교육자를 어떤 부분까지 보조할 수 있는지 확신할 수 없다. 실증적인 연구를 통해 수학교육에서 어떤 부분이 AI로 대체될 수 있고 어떤 부분이 그렇지 않은지를 확인해볼 필요가 있으며, 이러한 차이를 이해함으로써 수학교육에서 가르치고 배우는 방식이 어떻게 변화하는지 이해하는데 기여할 수 있을 것이다. 예컨대, AI 도구를 사용하여 메타 평가를 수행하고 개별적인 피드백을 받는 것이 향후 수학교육 목표의 일부가 될 수도 있다.

본 연구의 제한점과 후속 연구를 위한 제언은 다음과 같다. 첫째, 본 연구는 분수 크기 비교 문제를 해결하는 학생과 예비교사 2인의 담화 자료를 사례로 선정하고 ChatGPT의 응답 데이터를 분석하였다. 사례 연구의 특성상 본 연구의 결과를 수학교육의 전 영역으로 확장하기는 어렵다(Merriam, 1998). 따라서 후속 연구에서는 수학과와 여러 주제와 학생과 교사의 사고를 드러내는 여러 산출물을 활용하여 본 연구의 결과를 확장할 필요가 있다. 둘째, ChatGPT는 인터넷에서 얻은 방대한 양의 텍스트 데이터를 학습한 언어 모델이다(Mhlanga, 2023). 따라서 학생과 교사의 사고를 실

제로 추론하기보다는 학습한 데이터를 토대로 응답을 도출한다. 따라서 본 연구의 결과에서 나타난 ChatGPT의 응답은 편향된 데이터에 기반한 결과일 수 있으므로 교육적 활용에서는 신중한 검토가 필요하다. 이러한 문제를 성공적으로 해결하고 AI 기술이 수학교육에서 학생과 교사에게 유익하게 사용될 수 있음을 보장하려면 지속적인 연구와 모니터링이 수행되어야 할 것이다. 셋째, 입력하는 사용자의 텍스트에 따라 ChatGPT의 응답이 달라지므로 유의할 필요가 있다. 예컨대, ChatGPT에게 예비교사 A의 발문에 대한 평가를 요청할 때 '예비교사 A의 발문이 학생의 반응을 토대로 수학적 사고를 이끌어내는지 평가해줘'라고 질문하면 학생의 반응보다는 수학적 사고에만 초점이 맞추어져 응답이 제시되는 경향이 나타났다. 오히려 '두 가지 평가 기준(학생의 반응에 따라 발문이 이루어졌는지, 발문이 학생의 수학적 사고를 이끌어내는지)을 가지고 예비교사 A의 발문을 평가해줘'라고 질문했을 때 반응성과 수학적 사고를 평가 기준으로 정확하게 응답하는 결과를 얻을 수 있었다. 향후 수학교육에서 ChatGPT 활용 연구를 수행할 때는 연구자가 요구하는 평가 기준과 질문을 AI가 이해할 수 있도록 구체적이고 명확하게 입력할 필요가 있다.

참 고 문 헌

- 박현아(2023). 전남도교육청, 챗GPT 활용 교육용 안내서 배부. 광주드림, 2023년 4월 12일. <https://www.gjdream.com/news/articleView.html?idxno=626079>
- 윤근호(2023). 대전교육정보원, 교사 대상 ChatGPT 연수 실시. 충남일보, 2023년 4월 17일. <https://www.chungnamilbo.co.kr/news/articleView.html?idxno=710475>
- 이지혜, 허난(2020). 인공지능을 활용한 맞춤형 수학 학습 프로그램 개발. *East Asian Mathematical Journal*, 36(2), 273-289.
- 한형진(2023). 새로운 물결 'ChatGPT'...제주 교사 연수에서도 인기. 제주의 소리, 2023년 3월 13일. <http://www.jejusori.net/news/articleView.html?idxno=412820>
- Aliustaoğlu, F., Tuna, A., & Biber, A. Ç. (2018). The misconceptions of sixth grade secondary school students on fractions. *International Electronic Journal of Elementary Education*, 10(5), 591-599.
- Atlas, S. (2023). *ChatGPT for higher education and professional development: A guide to conversational AI*. Independently Published.
- Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal*, 94, 373-397.
- Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook research on teaching* (4th ed., pp. 433 - 456). American Educational Research Association.
- Baylari, A., & Montazer, G. A. (2009). Design a personalized e-learning system based on item response theory and artificial neural network approach. *Expert Systems with Applications*, 36(4), 8013-8021.
- Bernacki, M. L., & Walkington, C. (2018). The role of situational interest in personalized learning. *Journal of Educational Psychology*, 110(6), 864-881.
- Beyer, S. (2022, April). Developing a chatbot for mathematics teachers to support digital innovation of subject-matter teaching and learning. *In society for information technology & teacher education international confèrence* (pp. 1344-1348). Association for the Advancement of Computing in Education (AACE).
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258. <https://doi.org/10.48550/arXiv.2108.07258>
- Cai, W., Grossman, J., Lin, Z. J., Sheng, H., Wei, J. T. Z., Williams, J. J., & Goel, S. (2021). Bandit algorithms to personalize educational chatbots. *Machine Learning*, 110(9), 2389-2418.

- Dwivedi, P., Kant, V., & Bharadwaj, K. K. (2018). Learning path recommendation based on modified variable length genetic algorithm. *Education and Information Technologies, 23*(2), 819-836.
- Ferrell, O. C., & Ferrell, L. (2020). Technology challenges and opportunities facing marketing education. *Marketing Education Review, 30*(1), 3-14.
- Frieder, S., Pinchetti, L., Griffiths, R. R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., ... Berner, J. (2023). Mathematical capabilities of chatgpt. arXiv preprint arXiv:2301.13867. <https://doi.org/10.48550/arXiv.2301.13867>
- Graham, K. J., & Fennell, F. (2001). Principles and standards for school mathematics and teacher education: Preparing and empowering teachers. *School Science and Mathematics, 101*(6), 319-327.
- Heaven, W. D. (2020, July 20). OpenAI's new language generator GPT-3 is shockingly good—and completely mindless. MIT Technology Review, Retrieved from <https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education*. Center for Curriculum Redesign.
- Hwang, G. J., & Chang, C. Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments, 1-14*.
- Kara, F., & Incikabi, L. (2018). Sixth grade students' skills of using multiple representations in addition and subtraction operations in fractions. *International Electronic Journal of Elementary Education, 10*(4), 463-474.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274.
- Lee, D., & Yeo, S. (2022). Developing an AI-based chatbot for practicing responsive teaching in mathematics. *Computers & Education, 191*, 104646.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. Jossey-Bass.
- Metcalf, J. (2017). Learning from errors. *Annual Review of Psychology, 68*, 465-489.
- Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. SSRN. <http://dx.doi.org/10.2139/ssrn.4354422>
- National Council for Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. NCTM.
- National Council of Teachers of Mathematics (2014). *Principles to actions: Ensuring mathematical success for all*. NCTM.
- Nguyen, H. D., Pham, V. T., Tran, D. A., & Le, T. T. (2019, October). Intelligent tutoring chatbot for solving mathematical problems in High-school. *In 2019 11th international conference on knowledge and systems engineering (KSE)* (pp. 1-6). IEEE.
- OECD (2021). *AI and the future of skills*. OECD Publishing,
- OpenAI. (2022, November 30). ChatGPT: Optimizing language models for dialogue. Retrieved from <https://openai.com/blog/chatgpt/>
- Pai, K. C., Kuo, B. C., Liao, C. H., & Liu, Y. M. (2021). An application of Chinese dialogue-based intelligent tutoring system in remedial instruction for mathematics learning. *Educational Psychology, 41*(2), 137-152.
- Pardos, Z. A., & Bhandari, S. (2023). Learning gain differences between ChatGPT and human tutor generated algebra hints. arXiv preprint arXiv:2302.06871. <https://doi.org/10.48550/arXiv.2302.06871>
- Popenici, S. A., & Kerr, S. (2017). Exploring the

- impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 1-13.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 1-22.
- Seghier, M. L. (2023). ChatGPT: not all languages are equal. *Nature*, 615(7951), 216-216.
- Shakarian, P., Koyyalamudi, A., Ngu, N., & Mareedu, L. (2023). An independent evaluation of chatgpt on mathematical word problems (MWP). arXiv preprint arXiv:2302.13814. <https://doi.org/10.48550/arXiv.2302.13814>
- Shamir, G., & Levin, I. (2022). Teaching machine learning in elementary school. *International Journal of Child-Computer Interaction*, 31, 100415.
- Soltanpoor, R., Thevathayan, C., & D'Souza, D. (2018). Adaptive remediation for novice programmers through personalized prescriptive quizzes. In *Proceedings of the 23rd annual ACM conference on innovation and technology in computer science education* (pp. 51-56). ACM.
- Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays—should academics worry? *Nature*. <https://doi.org/10.1038/d41586-022-04397-7>
- Webel, C., & Conner, K. A. (2017). Using simulated teaching experiences to perturb preservice teachers' mathematics questioning practices. *Mathematics Teacher Educator*, 6(1), 9-26.
- Zhai, X. (2022). ChatGPT user experience: Implications for education. Available at SSRN 4312418. <http://dx.doi.org/10.2139/ssrn.4312418>

Exploring the possibility of using ChatGPT in Mathematics Education: Focusing on Student Product and Pre-service Teachers' Discourse Related to Fraction Problems

Son, Taekwon

Bongmyong Elementary School

E-mail : sontaekwon7@gmail.com

In this study, I explored the possibility of using ChatGPT math education. For this purpose, students' problem-solving outputs and conversation data between pre-service teachers and a student were selected as an analysis case. A case was analyzed using ChatGPT and compared with the results of mathematics education experts. The results that ChatGPT analyzed students' problem-solving strategies and mathematical thinking skills were similar to those of math education experts. ChatGPT was able to analyze teacher questions with evaluation criteria, and the results were similar to those of math education experts. ChatGPT could also respond with mathematical theory as a source of evaluation criteria. These results demonstrate the potential of ChatGPT to analyze students' thinking and teachers' practice in mathematics education. However, there are limitations in properly applying the evaluation criteria or providing inaccurate information, so the further review of the derived information is required.

* 2000 Mathematics Subject Classification : 97U70

* Key Words : Artificial Intelligence, ChatGPT,
Mathematical Thinking, Questioning