

RESEARCH ARTICLE

ChatGPT as a Scaffolding Tool: Evaluating the Impact on Elementary Students' Mathematical Logic Problem-Solving Skills

Doyeon Ahn¹ · Taekwon Son² · Kwangho Lee^{3*}

¹Cheonancheongdang Elementary School

²Bongmyeong Elementary School

³Korea National University of Education

비계설정 도구로서의 ChatGPT: 초등학생의 수학 논리 문제 해결 능력에 미치는 영향 평가

안도연¹ · 손태권² · 이광호^{3*}

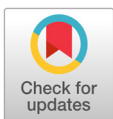
¹천안청당초등학교 · ²봉명초등학교 · ³한국교육대학교

*Corresponding Author: paransol@knue.ac.kr

ABSTRACT

This study explored the effectiveness of personalized scaffolding using OpenAI's large language model, ChatGPT-4.0, with the aim of enhancing elementary students' mathematical logical problem-solving abilities. Based on Vygotsky's socio-cultural learning theory, the influence of AI-based hint provision on students' problem-solving abilities and self-regulated learning was analyzed. The results showed that the hints provided by ChatGPT effectively improved the students' problem-solving levels, particularly utilizing restructuration and verification strategies efficiently. Furthermore, this AI model accurately identified the students' errors and provided appropriate hints considering these errors. However, such effects varied depending on the difficulty of the problem and the students' levels. These findings demonstrate the potential of AI as a tool for personalized educational support and suggest its applicability, especially in the field of mathematics education.

Key words: Mathematical logic, problem-solving, scaffolding, large language model, generative pretrained transformer, ChatGPT, GPT-4



OPEN ACCESS

Brain, Digital, & Learning
2023, Vol. 13, No. 2, 183-196.

<https://doi.org/10.31216/BDL.20230011>

Received: June 08, 2023

Revised: June 20, 2023

Accepted: June 20, 2023

© 2023. Institute of Brain based Education,
Korea National University of Education



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

모든 학문 분야에서 논리적 이론의 전개는 매우 중요하며, 논리의 학문이라고 할 수 있는 수학에서 논리적 타당성은 특히 강조된다(Shin et al., 2014). 수리논리학은 현대 사회에서 사람이 사용하는 언어를 컴퓨터가 이해할 수 있는 기호로 변환하는 중요한 역할을 수행한다(Fitting, 1990). 따라서 수리논리 문제를 다루고 학습하는 것은 비판적이고 논리적인 사고력을 기르는데 도움을 제공하지만, 초등학생들이 엄밀한 수리논리를 학습하는 것은 문제의 난이도로 인해 어려움이 따르곤 한다(Shin et al., 2014).

이러한 어려움은 최근 인공지능의 발달로 인해 새로운 해결 가능성이 제시되고 있다. 교육에서의 인공지능 활용은 Vygotsky (1978)의 사회 구성주의 이론에 근거한 비계설정 이론을 반영하며, 학습자들이 혼자서는 이해하지 못하는 개념을 학습하거나 문제를 해결할 때 비계에 해당하는 문항이나 힌트, 예시 자료의 형태로 도움을 제공할 때 가장 많이 배울 수 있다고 가정한다(Beal et al., 2010). 이로 인해 교육에서 인공지능은 주로 학생들의 개별적인 교육적 요구를 파악하여 개별화된 학습 방안을 지원하는 보조 교사로 활용되며 여러 연구에서 그 효과성이 입증되고 있다(Dwivedi et al., 2018; Pardos & Bhandari, 2023). 특히, 최근 인공지능 기반의 대규모 언어 모델(Large Language Model; 이하 LLM)인 ChatGPT가 발표되었다. ChatGPT는 교사와 교사 교육자가 수행하던 복잡한 교수 작업을 수행하는 능력을 보여주었을 뿐만 아니라(Zhai, 2022), 교사의 역할을 보조하는 비계설정 기능을 제공하는 인공지능 튜터로서도 다양하게 활용되고 있다(e.g., Shakarian et al., 2023; Pardos & Bhandari, 2023). 그러나 ChatGPT는 최신 기술이므로 수학교육에서 ChatGPT를 활용하는 방법에 대한 실증적인 연구는 부족한 실정이다(Rudolph et al., 2023). 이에 본 연구에서는 현재 대표적인 LLM인 ChatGPT 4.0을 활용하여 인공지능 모델이 초등학생들의 수학적 논리 문제 해결 능력을 향상시키는 비계설정 도구로 사용될 수 있는지를 확인하고 수리 논리 학습에 대한 시사점을 제공하고자 하였다. 이에 따라 본 연구에서 설정한 연구문제는 다음과 같다.

연구문제1. ChatGPT가 제공하는 수리논리문제의 힌트는 초등학생들의 수리논리 문제 해결을 돕는데 효과가 있는가?

연구문제2. 학생들의 수리논리 문제 해결을 돕기 위해 ChatGPT가 제공하는 힌트에서 나타난 비계 설정 전략은 어떠한가?

Theoretical Background

Mathematical Logic

수리논리학(mathematical logic)은 논리학에서 사용하는 명제들을 수학적 기호로 표현하는 학문이다. 이는 일상 언어와 같은 자연언어의 복잡성과 오류 가능성을 줄이고, 명제를 효과적이며 간결하게 다룰 수 있도록 도입된 현대 논리학의 이론이다(Kleene, 1943). 수리논리학의 하위 분야로는 집합론, 증명이론, 재귀이론, 모델이론 등이 있다. 특히 명제논리(propositional logic)와 술어논리(predicate logic)는 사람이 사용하는 언어를 컴퓨터가 인식할 수 있는 명확한 기호로 변환하며, 이는 컴퓨터 과학 및 인공지능의 발달에 기여하고 있다(Fitting, 1990).

수리논리 개념은 초등학생들이 학습하기 어렵지만, 수학 교육에 있어 매우 중요한 역할을 한다. Paul & Elder (2006)는 논리적 사고와 추론 능력이 비판적 사고 능력을 개발하는데 중요하며, 수학을뿐만 아니라 삶의 여러 분야에서 합리적인 결정을 내리는 데 도움이 된다고 주장하였다. 또한 Polya (1945)는 수리논리 학습이 수학의 모든 영역에 걸쳐 논리적인 문제 해결에 중요한 역할을 한다고 강조했다. 디지털 중심의 현대 사회에서는 논리적 수학의 중요성이 점점 증가하고 있어, 미래 사회를 대비하는 데에도 중요한 역할을 하고 있다(National Research Council, 2012). 이처럼 수리논리 학습이 중요함에도 불구하고 엄밀한 논리적 사고를 초등학교 학생들에게 요구하는 것은 흥미를 쉽게 잃게 하므로 수리논리학을 가르치는 것은 쉽지 않다. 이에 Shin et al. (2014)은 수리논리학을 초등학생에게 가르치는 효과적인 방법으로 수리퍼즐을 제안하였으며, 수리퍼즐이 초등학생의 논리적 사고력을 향상시키고 흥미를 유발하는 효과적인 방법이라고 주장하였다. 본 연구에서는 초등학교 학생들에게 수리퍼즐을 적용하고 학생들의 수리 논리 문제 해결력이 향상되는지 살펴보았다.

Scaffolding in Problem-solving

비계설정은 Vygotsky의 사회적 구성주의 이론에 근거한 개념으로, 학생들을 점진적인 이해와 독립적인 학습 과정으로 이끌기 위해 동료나 교사가 제공하는 지원을 의미한다(Vygotsky, 1978). 이는 학습자가 그들의 가까운 발달 영역에 따라 발전할 수 있도록, 자신보다 더 경험이 많은 동료 학습자나 교사가 도움이나 힌트를 제공하는 것을 포함한다. 수학 학습, 특히 문제 해결 과정에서 비계설정은 개념 이해와 문제 해결 전략을 증진하는데 효과적인 도구 역할을 할 수 있다. 문제 해결 과정에서 적절한 비계설정은 학생들이 자기 조절, 문제 해결 전략, 그리고 반성적 사고와 같은 메타인지적 사고를 향상시키는 데 도움이 된다(Hmelo-Silver, 2004). Bliss et al. (1996)에 따르면, 학생들에게 힌트를 제공하는 것은 비계설정 전략 중 하나로, 특히 1:1 학습 상황에서 가장 효과적이라고 주장하였다. 문제 해결 과정에서 제공되는 적절한 힌트는 학생들이 문제에 접근하는 방법과 전략을 학습하게 하며, 궁극적으로는 유사한 문제를 독립적으로 해결하는 능력을 향상시키는 데 도움이 된다(Rosenshine & Meister, 1992). 이런 비계설정의 중요성에 대한 인식과 함께, 인공지능의 발전에 따라 개인화된 힌트를 제공하는 자동화 시스템에 대한 연구가 활발히 진행되고 있다. Barnes & Stamper (2008)는 논리 증명 학습에서 학습자의 반응에 따른 힌트를 자동으로 생성하는 지능형 튜터를 개발하였고, 이후 연구에서는 논리 학습에서 지능형 튜터가 힌트를 제공했을 때 학습자의 성적이 유의하게 향상되었으며 문제 해결을 포기하는 확률이 줄어들었다는 것을 입증하였다(Stamper et al., 2013). 이에 따라 본 연구에서도 수리 논리 문제를 해결할 때 ChatGPT가 제공하는 비계 설정이 학생들의 수리 문제 해결력 향상에 도움이 되는지 살펴보았다.

Large Language Model (LLM)

LLM은 상당한 양의 텍스트 데이터를 학습하여 자연어 이해와 생성 작업을 수행하는 인공지능 모델이다(Kasneci et al., 2023). 이는 간단한 단어부터 문서 단위의 텍스트까지 이해하고 생성하는 능력을 지니고 있으며, 다양한 언어를 지원한다. OpenAI의 ChatGPT와 Google의 Bard가 대표적인 예이다. LLM은 대규모 텍스트 데이터를 이용해 사전 학습되며, 이렇게 학습된 모델은 기계 번역, 문장 생성, 요약, 질의응답, 텍스트 분류 등의 작업을 수행하는 데 사용될 수 있다(Devlin et al., 2018). 특히 수학 교육에서는 문제 해결 지원, 개별적인 학습 지원, 개념 설명 등에 활용될 수 있다(Radford et al., 2019). 본 연구에서는 LLM의 이러한 특성을 고려하여 대표적인 LLM인 ChatGPT4.0을 사용하여 수리 논리 문제에서 적절한 비계설정이 가능한지 살펴보았다.

Methodology

Participants

연구 참여자는 대전광역시 J초등학교 6학년 1개 학급에서 연구 참여에 동의한 25명의 학생들이다. 수리퍼즐 문제에서 요구되는 논리적 사고 수준을 고려하여 초등학교 6학년으로 연구 참여자를 선정하였으며, 25명의 참여자 중 국어와 수학 교과에서 학습 부진을 보이는 1명을 제외한 24명(남12명, 12명)을 최종적으로 선정하였다.

Instrument

수리 논리 과제는 Shin et al. (2014)에서 초등학생용으로 제시한 논리 퍼즐 두 문제를 한국어로 번역하여 사용하였다. 두 명의 초등학교 경력교사가 번역된 내용에서 어색한 문맥을 수정하였다. Fig. 1은 학생들에게 제시한 수리 논리 과제이다.

<문제 1>	<문제 2>
청록마을에 네 가족이 이사를 오게 되었습니다. A, B, C, D 가족은 각각 지붕의 색을 다르게 칠해서 집을 구분하기로 하였습니다. 다음 진술이 모두 참이라고 할 때, A가족의 지붕 색은 어떤 색일까요? 지붕 색은 노란색, 붉은색, 초록색, 파랑색입니다. B가족의 지붕 색은 노란색이 아닙니다. C가족의 지붕 색은 파란색입니다. D가족의 지붕 색은 노란색도 초록색도 아닙니다.	마음초등학교의 5학년은 1반부터 5반까지 있습니다. 이 중 1반 학생들은 축구를 굉장히 좋아해서 기회만 있으면 다른 반과 축구시합을 하였습니다. 다음 진술이 모두 참이라고 할 때, 1반 학생들과 축구시합을 한 반은 어느 반인지 모두 찾아보세요. 1반은 2반과 축구시합을 하지 않았다. 1반은 3반과 축구시합을 했거나 그렇지 않으면 5반과 축구시합을 했다. 1반은 4반과 축구시합을 하지 않았다면 2반과 축구시합을 했다. 1반이 3반과 축구시합을 했다면 4반과 축구시합을 하지 않았다.

Fig. 1. Mathematical logic tasks

Data Collection

연구문제1을 해결하기 위해서 ChatGPT에게 힌트를 받기 전·후의 학생들의 풀이 데이터를 수집하였다. 학생들은 힌트 없이 수리 논리 과제를 먼저 풀고, 이후 ChatGPT가 제시한 힌트를 보고 다시 풀이를 작성하였다. 학생들은 활동지를 통해 주어진 과제를 해결하였으며 풀이 과정은 약 15분 정도 소요되었다. 학생들의 풀이 활동지는 ChatGPT가 인식할 수 있도록 모두 전사하였다. Fig.2은 ChatGPT-4.0에게 제시한 과제와 함께 ‘나는 초등학생입니다. 내가 이 문제를 잘 풀 수 있도록 답을 제외하고 쉬운 언어로 힌트를 주세요.’라고 요구했을 때 제시된 힌트이다. 힌트는 ChatGPT의 번역 확장 프로그램인 프롬프트 지니로 번역된 한국어로 제시하였다.

면 다른 반과 축구시합을 하였습니다. 다음 진술이 모두 참이라고 할 때, 1반 학생들과 축구시합을 한 반은 어느 반인지 모두 찾아보세요.

1반은 2반과 축구시합을 하지 않았다.

1반은 3반과 축구시합을 했거나 그렇지 않으면 5반과 축구시합을 했다.

1반은 4반과 축구시합을 하지 않았다면 2반과 축구시합을 했다.

1반이 3반과 축구시합을 했다면 4반과 축구시합을 하지 않았다.



이 문제를 푸는 방법은 각 진술을 주의 깊게 살펴보고, 가능한 상황을 모두 고려하는 것입니다. 각 진술에서 나온 경우의 수를 조합하여 축구시합을 했을 수 있는 반들을 찾아봅시다.

첫 번째 힌트는 진술 1번입니다. 여기서 우리는 1반이 2반과는 축구시합을 하지 않았다는 것을 알 수 있습니다.

두 번째 힌트는 진술 2번에서 얻을 수 있습니다. 여기서는 1반이 3반 또는 5반과 시합을 했다고 말하고 있습니다. 하지만 어느 반과 시합을 했는지는 아직 확실하지 않습니다.

세 번째 힌트는 진술 3번입니다. 여기서는 1반이 4반과 시합을 하지 않았다면, 2반과 시합을 했다는 것입니다. 하지만 우리는 이미 1반이 2반과 시합을 하지 않았다는 것을 알고 있습니다. 그러면 이 상황에서 어떤 결론을 내릴 수 있을까요?

마지막 힌트는 진술 4번에서 얻을 수 있습니다. 이는 1반이 3반과 시합을 했다면, 4반과는 시합을 하지 않았다는 것입니다. 이 정보가 무엇을 의미하는지 생각해 보세요.

각 진술에서 얻은 정보를 종합해서, 어떤 반과 1반이 축구 시합을 했는지 결정해 보세요.

Fig. 2. The hint for Task2 provided by ChatGPT

연구문제2를 해결하기 위하여 힌트를 보기 전 학생들의 풀이를 ChatGPT-4.0에 입력하고, ChatGPT가 제공하는 힌트를 수집하였다. 학생들이 작성한 풀이 내용 중 논리적 오류 없는 풀이, 정답을 제시한 풀이, 아무 풀이도 작성하지 않은 응답을 제외하고 총 19개의 풀이를 ChatGPT에 입력하고 19개의 힌트를 수집하였다.

Analysis Methods

수리 논리 과제에 대한 학생의 풀이를 평가하기 위해 6학년 지도 경험이 있는 교직 경력 5년 이상의 경력교사 3명을 평가자로 선정하였다. 이들은 힌트를 보기 전과 후의 풀이가 무작위로 섞인 답안지를 Table 1의 제시된 표의 기준에 따라 1~5점으로 채점하였다. Table 2는 Table 1의 기준에 따라 채점한 답안의 예시이다. 평가자 3명의 점수를 합하여 각 문항의 최종 점수로 정하였다(최고 15점). ChatGPT가 제시한 힌트를 보여주기 전과 후의 문제 해결 결과를 비교하기 위해 대응표본 t검정을 실시하였다.

Table 1. Rubric for score

Criteria	Score
Provides detailed solution process along with the correct answer.	5
Provides an incomplete solution process but presents the correct answer; or presents the accurate solution process but gives an incorrect answer.	4
Contains some inaccuracies but provides correct explanations along with an inaccurate answer.	3
Mostly incorrect explanations, but some logical reasoning is evident, and an incorrect answer is provided.	2
Provides an incorrect answer without any explanation or does not provide an answer.	1

Table 2. Example of scoring according to the scoring rubric

1번 지명의 색을 다 다르게 칠한다해서"	1 ~ 5번 1번은 2번이랑 착기함 x 1번 = 3, 5 1번 = 4, 2 } = 2번, 5번 1번 = 3, 4	5번 5학년이 1번 ~ 5번까지 있을때 1번은 2번과 5번 (착기함) x, 3, 4, 5 1번은 3번과 5번이면 5번과 4번 1번은 4번과 5번이면 5번과 3번 1번이 3과 4번이면 5번과 2번
score 1	score 2	score 3
5, 4 왜냐하면 1번과 2번은 안함 3, 4, 5 1번이 3, 4와 했지만 4번은 3, 5 1번이 4, 2번과 착기함 2번은 안함 1번이 3, 4와 안함 근데 4번만 5 그러서 5, 4	B기죽은 노랑이 아니고, C기죽은 파랑이요, D기죽은 노랑색과 초록색이 있습니다. (빨, 노, 초, 파) C기죽은 파랑색이기에 다른 기죽의 재능색은 파랑 x D기죽은 노랑색과 초록색이 아니기에 남은 색 파랑 0 B기죽은 노랑색이 아니지만 남은 색은 노랑과 초록이다. 그러므로 초록 0 A기죽은 여기서 남은 색 노랑색을 지능색으로 활용하게 된다. 답: 노랑색.	
score 4	score 5	

연구문제2에 해당하는 ChatGPT의 힌트를 평가하기 위해 3명의 평가자(학생응답 평가자와 동일)가 19개의 풀이에 대해 ChatGPT에게 힌트를 요청하고 ChatGPT의 힌트 수준을 평가하였다. 정확한 힌트 요청을 위해 Fig.3과 같이 프롬프트를 입력하고 문제와 풀이를 각각 제시했다.

이후 각 평가자는 ChatGPT가 제시한 힌트의 수준을 Table 3에서 제시한 평가표에 작성하였다. 평가표는 Anghileri (2006)의 비계설정 전략 분류틀과 Lee (2019)의 비계설정 수준 기준표를 참고하여 개발하였다. 정확한 평가를 위해 평가자에게는 Vygotsky의 비계설정이론과 평가기준을 충분히 이해할 수 있도록 사전에 자료와 평가표를 제공하였다.



안녕하세요. 저는 초등학교생입니다. 수리논리문제를 풀어야하는데 도움이 필요합니다. 제가 문제와 저의 풀이를 제시하면 제가 스스로 문제를 해결할 수 있도록 힌트를 주세요. 저의 풀이에서 잘못된 점을 고려해서 힌트를 주세요. 답은 절대로 알려주시면 안됩니다.



안녕하세요! 네, 문제를 제시하시면 도와드리겠습니다. 제가 어떻게 도울 수 있는지 알려주시면 감사하겠습니다. 



<문제>

청록마을에 네 가족이 이사를 오게 되었습니다. A, B, C, D 가족은 각각 지붕의 색을 다르게 칠해서 집을 구분하기로 하였습니다. 다음 진술이 모두 참이라고 할 때, A가족의 집은 지붕은 색은 어떤 색일까요?

1. 지붕 색은 노란색, 붉은색, 초록색, 파랑색 입니다.
2. B가족의 지붕 색은 노란색이 아닙니다.
3. C가족의 지붕 색은 파랑색입니다.
4. D가족의 지붕 색은 노란색도 초록색도 아닙니다.

<풀이>

A가족의 지붕은 붉은색입니다. 왜냐하면 B,C,D 가족들 모두 붉은색 빼고 다른색들은 다 아니라고 했는데 이걸 반대로 하면 A가족은 붉은색이겠죠. 그래서 A가족 빼고 다른 B,C,D 가족들은 붉은색이고, 문제에 나오는 '참'이라는 뜻은 옳다 혹은 거짓이라는 뜻이니까, 붉은 색만 쏙 빼고 말한 B,C,D 가족들은 모두 거짓말을 하고 있는거니까 답은 A가족의 집은 붉은색입니다.

Fig. 3. Prompt to get hints from ChatGPT

Table 3. Rubric for assessing ChatGPT's hints

Item		Rating				
Logical accuracy		5	4	3	2	1
Reflecting student errors		5	4	3	2	1
Consideration of student's level of understanding		5	4	3	2	1
Scaffolding strategy	Reviewing	prompting and probing				Y / N
		interpreting students talk				Y / N
	Restructuring	provision of meaningful contexts				Y / N
		simplifying the problem				Y / N
		rephrasing students talk				Y / N

Results

Impact of Hints

ChatGPT가 제시한 힌트를 보여주기 전과 후의 문제 해결 결과를 비교하기 위해 대응표본 t검정을 실시하였다. 두 문제를 합산한 결과를 비교하였을 때 힌트를 제공한 후에 문제해결 결과가 유의미하게 향상된 것으로 나타났다($p < .05$, Table 4). 각 문제별로 살펴보았을 때, 문제1에서는 힌트가 제공된 후 문제 풀이 점수가 상승했으나 통계적으로 유의미한 차이가 나타나지 않았고, 문제2에서는 문제풀이 점수가 유의미하게 높아진 것으로 나타났다($p < .05$).

Table 4. Paired t-test results of before and after hints

Task		M	SD	t	p
Total	pre	22.9167	5.54755	-2.036	0.048*
	post	26.0000	4.92533		
Task1	pre	13.7083	3.16886	-0.465	0.644
	post	14.0883	2.35753		
Task2	pre	9.2083	4.05376	-2.362	0.022*
	post	11.9167	3.88885		

* $p < .05$

문제1에서는 24명 학생 중 20명이 힌트를 보기 전에 이미 문제를 해결하였다. 이에 따라 평균(M) 차이도 크지 않았으며 표준편차(SD)도 낮게 나타났다. 문제2에서는 힌트 전과 후의 평균(M) 점수는 문제 1에서 보다 낮았지만, 표준편차(SD)는 더 크게 나타났다. 문제2는 문제1과 비교했을 때 난이도가 더 높은 문제로 학생들의 점수 분포도 다양하게 나타났으며 힌트 제공 전과 후에 학생들의 답변 결과도 차이가 컸다. Fig. 4는 24명의 전체 점수 분포를 총 점수가 낮은 학생부터 높은 순서로 나열하여 나타낸 것이다. 힌트를 제공 받기 전 점수가 높거나 낮은 학생들보다 중간 성적의 학생들에게서 힌트 제공 전·후의 점수 차이가 크게 나타남을 알 수 있다.

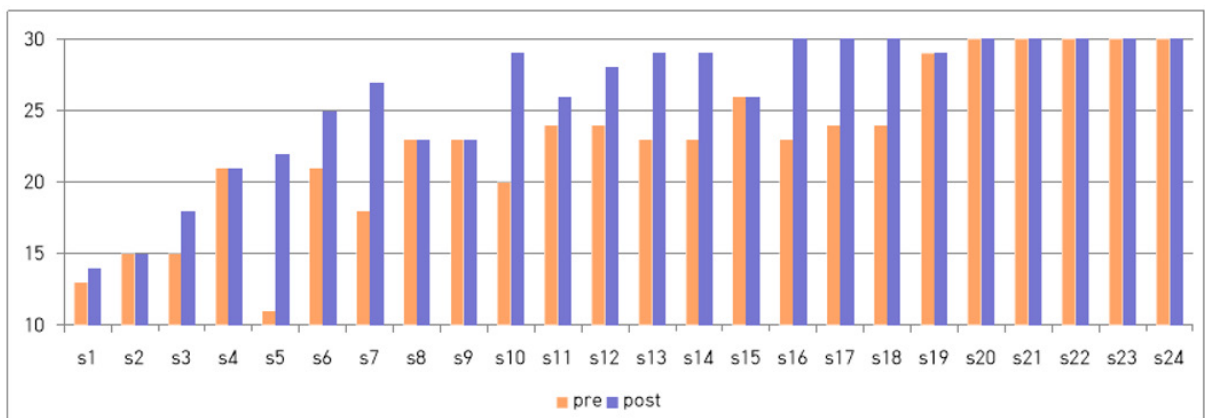


Fig. 4. Result of students' mathematical logic problem solving

Fig. 5는 17번 학생이 작성한 문제2에 대한 힌트 전과 후의 문제풀이 과정이다. 문제2는 주어진 4개의 단서 중 두 개의 단서 사이(2번, 4번)에서 논리적 추론을 할 수 있어야 문제를 해결할 수 있다. 17번 학생은 Fig. 5(a)와 같이 각각의 단서는 이해하였지만 논리적 추론을 하지 못해 오답을 제시하였다. 그러나 힌트를 제공 받은 후인 Fig. 5(b)에서는 올바르게 추론을 하여 옳은 답을 제시하였다. 12, 13, 16, 18번 학생도 이와 유사한 과정을 보였다.

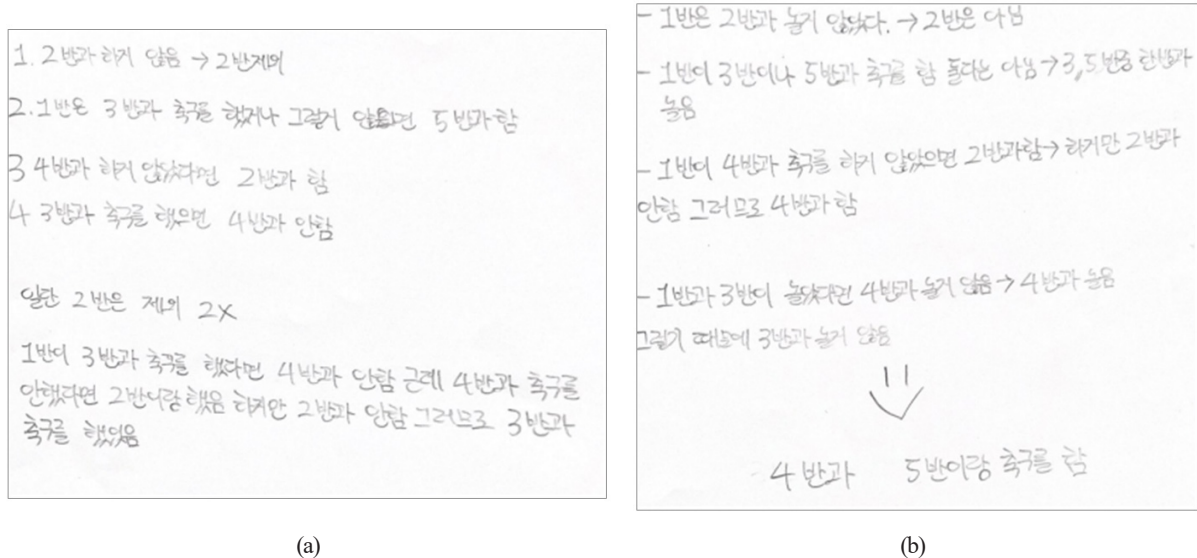


Fig. 5. (a) An answer before the hint (score 3), (b) An answer after the hint (score 5)

Scaffolding Strategies Embedded in the Hints by ChatGPT

오류가 있는 학생들의 풀이를 입력하고 힌트를 요청했을 때, ChatGPT가 제시하는 힌트의 수준은 Table 5와 같다. ‘논리적 정확도’ 측면에서는 총 57개의 답변에서 55개(96.5%)가 논리적 오류를 보이지 않았다. ‘학생의 오류 반영’ 측면에서는 ‘매우 그렇다’가 78.5%로 나타난 반면 ‘전혀 그렇지 않다’의 경우도 7건이 있었다. Fig. 6는 ‘학생 오류 반영’ 항목에서 1점을 받은 예이다. 학생의 풀이는 답은 맞았지만 풀이과정이 생략되어 있었다. 그러나 ChatGPT는 학생의 정답에 오류가 있다고 보았다. ‘학생 이해수준 고려’의 측면에서는 92.9%의 답변이 ‘매우 그렇다’의 평가를 받았다. 낮은 평가를 받은 경우에는 Fig. 7과 같이 ‘내포’, ‘진술’, ‘가정’, ‘모순’ 등 초등학생이 이해하기에는 다소 어려운 한자어를 나열하는 경우가 대부분이었다.

Table 5. General Quality of ChatGPT's Hint

Item	Frequency (%), n=57				
	5	4	3	2	1
Logical accuracy	55(96.5)	2(3.5)	-	-	-
Reflecting student errors	45(78.9)	4(7.0)	-	1(1.8)	7(12.3)
Consideration of student's level of understanding	53(92.9)	2(3.5)	1(1.8)	1(1.8)	-

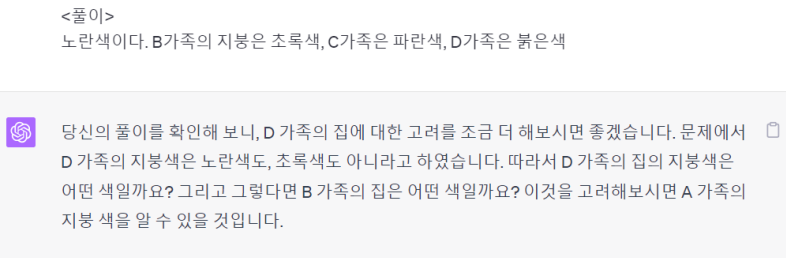


Fig. 6. An example of hint ChatGPT provided

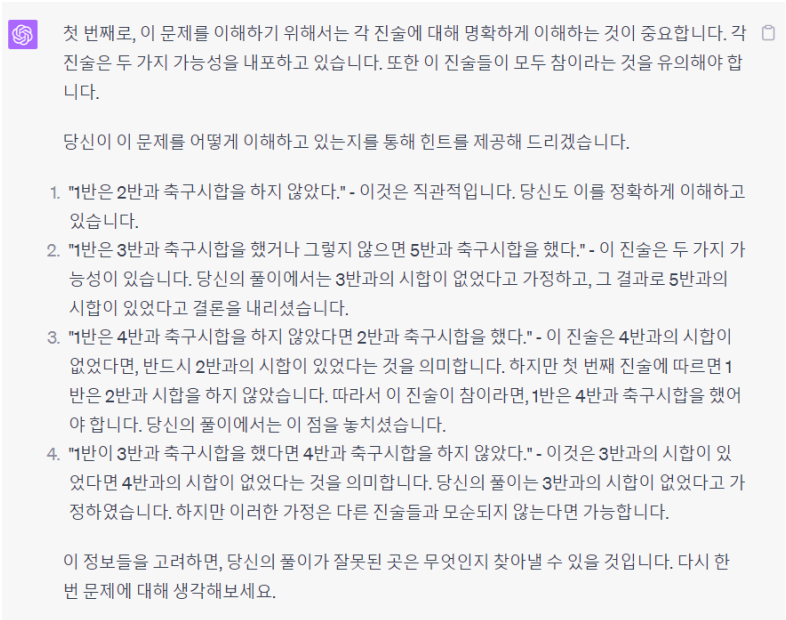


Fig. 7. An example of hint ChatGPT provided

ChatGPT가 힌트를 제시할 때 사용한 비계설정 전략을 살펴보면 Table 6과 같다. 각 전략의 포함 여부를 Y/N로 평가하였기 때문에, 여러 전략이 함께 나타난 경우도 포함되어 있다. 표에서 제시된 백분율(%)은 전체 힌트 중 해당 전략이 나타난 힌트의 비율을 의미한다. Anghileri (2006)의 비계설정 전략 유형에 따라 살펴보면, ‘검토하기’는 전체 힌트의 36.8%로 나타났다. ‘검토하기’ 세부 유형을 살펴보면 ‘독려하고 탐구하는 질문’이 나타난 답변은 전체 답변 중 19.3%, ‘학생의 발언 다시 말하기’는 24.6%로 나타났다. Fig. 8은 ‘독려하고 탐구하는 질문’을 사용한 예이다. 문제는 1반이 함께 축구경기한 반을 찾는 것이지만 그 답을 얻기 위해 알아야 하는 중간 단계인 5반이 함께 축구경기한 반을 생각해보도록 유도하고 있다.

Table 6. Frequency (percentage) of scaffolding strategies in ChatGPT’s hints

Scaffolding strategy (n=57)	Reviewing	21(36.8)	prompting and probing questions	11(19.3)
	Restructuring		interpreting students talk	14(24.6)
		48(84.2)	provision of meaningful contexts	-
			simplifying the problem	8(14.0)
			rephrasing students talk	44(77.2)

<풀이>

1반은 2반과 축구시합을 하지 않았다. -> 2반X

1반이 2반과 축구시합을 하지 않았으므로 4반과 축구시합을 했다. -> 4반O

1반이 4반과 축구시합을 했으므로 3반과 축구시합을 안했다.



아주 잘 하셨습니다. 하지만 조금 더 신중하게 풀어야 할 부분이 있습니다.



문제의 두 번째 진술을 다시 살펴보세요. 이 진술은 '1반은 3반과 축구시합을 했거나 그렇지 않으면 5반과 축구시합을 했다'라고 되어 있습니다. 여러분이 현재 3반과의 축구시합을 하지 않았다고 결론 지었는데, 이는 5반과의 축구시합 여부를 어떻게 변화시키는지 생각해 보세요.

추가적으로, 5반과의 축구시합 여부는 어떻게 되나요? 이것이 확실해진다면 문제의 답을 찾을 수 있습니다.

Fig. 8. An example of hint ChatGPT provided

‘재구조화’ 전략은 84.2%의 비율이었으며 그 중 가장 많이 사용된 유형은 ‘학생의 말을 재구조화하여 말하기’였다(77.2%). ‘문제 단순화하기’는 14%이었다. ‘의미 맥락화’ 전략은 나타나지 않았다.

Discussions

본 연구는 ChatGPT가 제공하는 힌트가 초등학교 6학년 학생들이 수리 논리 과제를 해결할 때 효과적인지 살펴보고 ChatGPT가 제공하는 힌트가 비계설정 수준에서 적절한지 평가하였다. 본 연구의 결과에 따른 논의는 다음과 같다.

첫째, ChatGPT가 제공한 힌트는 초등학생들의 수리 논리 문제 해결에 유의미한 효과가 있었다. 학생들의 다양한 수준의 산출물을 ChatGPT에게 제공했을 때 논리적으로 정확한 힌트를 제시하였으며(96.5%) 힌트 제공 전과 후의 학생들의 문제 풀이 수준은 유의하게 높아졌다. 이러한 결과는 개인 맞춤형 AI 학습도구의 효과에 대해서 연구한 선행 연구들(eg. Son, 2023; Stamper et al., 2013; Winkler et al., 2021)에서도 유사하게 나타났다. 성취결과뿐만 아니라 문제 해결을 중도 포기하는 확률을 낮추고(Stamper et al., 2013), 문제해결과정에서 문제 해결의 단계를 스스로 인지하는 효과(Winkler et al., 2021) 또한 기대해볼 수 있을 것이다.

둘째, ChatGPT는 힌트를 제공할 때 ‘재구조화’ 비계설정 전략을 가장 많이 사용하였다(84.2%). 재구조화 방법 중 ‘학생의 말을 재구조화하여 말하기’(77.2%), ‘문제를 단순화하여 말하기’(24.6%) 순으로 사용하였으나 ‘의미의 맥락화’는 사용하지 않았다. ‘검토하기’ 전략도 3문제 중 1문제에서 나타났다. 문제에 대한 이해를 돕기 위해 ‘학생의 발언 다시 말하기’, ‘독려하고 탐구하는 질문’ 등의 전략을 사용할 수 있었다. 수학 수업에서 비계설정 교사의 비계설정 전략을 분석한 Lee (2019)의 연구에서 교사가 ‘독려하고 탐구하는 질문’을 주된 비계설정 전략으로 사용한 것과 차이가 있었으며, ‘의미의 맥락화’ 전략을 전혀 사용하지 않는 것은 ChatGPT와 유사하게 나타났다.

셋째, ChatGPT가 제공하는 힌트의 수준은 학습자의 오류와 응답 수준에 따라 차이가 있었다. 학생들의 다양한 수준의 산출물을 ChatGPT에게 제공했을 때 학생의 오류를 정확하게 찾아 관련 도움을 제시할 수 있었

으며(78.5%), 학생의 수준을 고려한 응답을 제시하였다(92.9%). 학습자의 오류를 ChatGPT가 이해하고 이론적인 근거에 따라 논리적으로 해석할 수 있다는 점과(Son, 2023), 지능형 개인 학습 지원 도구가 학생들의 문제 해결 단계를 파악하고 알맞은 학습 지원을 제공한다는 점(Winkler et al., 2021)으로 미루어보아, ChatGPT가 개인형 학습 보조 시스템으로 사용될 수 있는 잠재력을 내포하고 있다고 볼 수 있다. 그러나 ChatGPT는 Fig. 5와 같이 학생의 오류를 정확하게 파악하지 못함에도 매우 자연스럽게 잘못된 힌트를 제공할 수 있으므로(Frieder et al., 2023), 교육 현장에서 적용할 때 유의할 필요가 있다.

넷째, 문제의 난이도와 학생의 수준에 따라 힌트의 효과가 다르게 나타났다. 본 연구에서는 학생들에게 너무 쉬운 문제에서는 힌트가 큰 효과를 보지 못했으나 적절한 난이도의 문제에서는 ChatGPT가 제공한 힌트가 더 효과적으로 작용했다. Ferguson (2022)은 인공지능 튜터가 근접발달영역을 고려한 지침을 지속적으로 제공하였을 때, 비교집단과 실험집단의 학습 효과는 유사하지만 인지부하가 줄어들었다고 보고하였다. 이러한 결과는 Vygotsky (1978)의 비계설정이론과 관련하여 개인 맞춤형 지능형 교수 시스템 도입에 학습자의 근접발달영역을 고려할 필요가 있음을 시사한다.

Conclusions

연구결과와 논의를 바탕으로 본 연구에서는 다음과 같은 결론을 도출하였다.

첫째, ChatGPT는 초등학생을 위한 수리논리 학습에서 비계설정 도구로서 활용할 수 있다. 수학퍼즐 문제에서 학생들의 수준에 맞는 힌트를 제공할 수 있었으며, 학생들에게 이를 적용해 본 결과 유의미한 효과를 보였다. ChatGPT가 제공한 힌트는 논리적 정확도가 높았고 학생의 오류를 찾아 관련 도움을 제시할 수 있었다. 또한 ‘학생의 말을 재구조화하여 말하기’, ‘문제 단순화하여 말하기’, ‘학생의 발언 다시 말하기’, ‘독려하고 탐구하는 질문’ 등과 같은 다양한 비계설정 전략을 사용할 수 있었다.

둘째, 교사는 인공지능 언어모델을 수학수업에 활용할 때 사회적 비계설정 전략에 집중할 필요가 있다. 연구 결과 ChatGPT는 수리논리 문제에 대한 힌트에서 ‘검토하기’, ‘재구조화’와 같은 인지적 비계설정 전략을 다양하게 사용할 수 있었다. ‘학급 분위기 조성’, ‘동료 학습’과 같이 인공지능 모델이 사용하기 어려운 전략에 집중할 필요가 있음을 의미한다.

셋째, ChatGPT의 힌트를 수리논리 문제 해결에 적용할 때, 학생의 근접발달영역을 고려한 문제를 활용한다. 수리논리 문제의 난이도를 학생의 근접발달영역에 맞게 설정함으로써, ChatGPT의 힌트 제공이 더욱 효과적일 수 있음을 발견하였다. 또한, 이미 학습에 적용되고 있는 학생 학습 수준 평가 인공지능 모델과의 연계를 통해 학생 개인의 수준에 맞는 맞춤형 힌트를 제공할 수 있을 것으로 기대된다.

넷째, 어린이가 사용할 수 있는 LLM 개발이 필요하다. 효과적인 비계설정을 위해서는 학생과의 실시간 상호작용이 필수적이다. 본 연구에서는 ChatGPT와 같은 인공지능 LLM이 초등학생을 대상으로 한 수리논리 학습의 비계설정에 도움이 된다는 것을 확인하였다. 최근 대중적으로 사용되고 있는 ChatGPT와 Bard는 모두 이용대상이 만 13세 이상으로 초등학생은 사용할 수 없다. 비계설정전략은 1:1학습에서 큰 효과를 발휘한다는 점을 고려하면(Bliss et al., 1996) 현 상황은 매우 제약적이다. 다만, ChatGPT와 같은 생성형 인공지능은 사실이 아닌 것도 매우 논리적으로 설명하는 경향이 있으므로 어린이가 사용할 경우 이에 대한 보완과 경계가 필요하다.

본 연구에서는 인공지능 모델이 초등학생의 수학적 논리 문제 해결 능력을 향상시키는 비계설정 도구로써 사용될 수 있는지 확인해보고자 하였다. 그러나 상황적인 제약으로 인해 학생들에게 동일한 힌트를 제공하는 방식으로 실험을 진행하였기 때문에, 개별 응답에 따른 힌트의 효과도 검증해 볼 필요가 있다. 또한, 유사한 상황에서 교사와 인공지능 모델의 반응을 비교하여 교사가 인공지능을 활용함에 있어 어떤 역할을 맡아야 하는지에 대해 구체적으로 살펴볼 필요가 있다. 이러한 추가 연구는 인공지능을 교육 현장에서 활용하는 데 있어서 중요한 지침을 제공할 수 있을 것으로 기대된다. 더 나아가, 본 연구 결과는 교사와 인공지능 모델이 상호보완적으로 협력하여 학생들의 학습을 지원하는 방안을 모색하는 데에도 기여할 수 있을 것이다.

References

- Anghileri, J. (2006). Scaffolding practices that enhance mathematics learning. *Journal of Mathematics Teacher Education*, 9, 33-52.
- Barnes, T., & Stamper, J. (2008). Toward automatic hint generation for logic proof tutoring using historical student data. In *Intelligent Tutoring Systems: 9th International Conference, ITS 2008, Montreal, Canada, June 23-27, 2008 Proceedings 9* (pp. 373-382). Springer Berlin Heidelberg.
- Beal, C. R., Arroyo, I. M., Cohen, P. R., & Woolf, B. P. (2010). Evaluation of AnimalWatch: An intelligent tutoring system for arithmetic and fractions. *Journal of Interactive Online Learning*, 9(1).
- Bernard, J. B., Nicole, M. G. (2007). *Cognition, Brain, and Consciousness* (1st ed.). New York, NY: Academic Press.
- Bliss, J., Askew, M., & Macrae, S. (1996). Effective teaching and learning: Scaffolding revisited. *Oxford review of Education*, 22, 37-61.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Fitting, M. (1990). *First-Order Logic and Automated Theorem Proving*. Graduate Texts in Computer Science. Springer-Verlag.
- Dwivedi, P., Kant, V., & Bharadwaj, K. K. (2018). Learning path recommendation based on modified variable length genetic algorithm. *Education and Information Technologies*, 23, 819-836.
- Ferguson, C., van den Broek, E. L., & van Oostendorp, H. (2022). AI-Induced Guidance: Preserving the Optimal Zone of Proximal Development. *Computers and Education: Artificial Intelligence*, 3, 100089.
- 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*.
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16, 235-266.
- 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.
- Kleene, Stephen Cole (1943). Recursive predicates and quantifiers. *Transactions of the American Mathematical Society*, 53, 41-73. [Doi:10.2307/1990131](https://doi.org/10.2307/1990131). JSTOR 1990131.
- Lazarus, A. & Buxton. (1981). *The Practice of Multimodal Therapy*. New York: McGraw-Hill Book Company.
- Lee, E. J. (2019). An analysis of characteristics and strategies of scaffolding in mathematics classroom: A case of a middle school teacher. *Journal of Educational Research in Mathematics*, 29, 301-319.
- Lezak, M. D. (1995). *Neuropsychological Assessment* (3rd ed.). New York: Oxford Press.
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting*

Concepts, and Core Ideas. National Academies Press.

- 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>
- Paul, R., & Elder, L. (2006). Critical thinking: The nature of critical and creative thought. *Journal of Developmental Education*, 30, 34-35.
- Polya, G. (1945). *How to Solve It: A New Aspect of Mathematical Method*. Princeton University Press.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1, 9.
- Rosenshine, B., & Meister, C. (1992). The use of scaffolds for teaching higher-level cognitive strategies. *Educational leadership*, 49, 26-33.
- 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, 1-22.
- 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>
- Shin, H. Y., Lee, J. Y., Kang, H. J., & Kim, J. Y. (2014). *Introduction to Mathematics for Elementary School Teachers*. Mathematics Education Series 11. Seoul: Kyowoo-sa.
- Son, T. K. (2023). Exploring the possibility of using ChatGPT in Mathematics Education: Focusing on Student Product and Pre-service Teachers' Discourse Related to Fraction Problems. *Education of Primary School Mathematics*. 26, 99-113.
- Stamper, J., Eagle, M., Barnes, T., & Croy, M. (2013). Experimental evaluation of automatic hint generation for a logic tutor. *International Journal of Artificial Intelligence in Education*, 22, 3-17.
- Vygotsky, L. S. (1978). *Mind in society: The Development of Higher Psychological Processes*. Harvard University Press.
- Zhai, X. (2022). ChatGPT user experience: Implications for education. Available at SSRN 4312418. <http://dx.doi.org/10.2139/ssrn.4312418>

Authors Information

Ahn, Doyeon: Cheonancheongdang Elementary School, Teacher, First Author

Son, Taekwon: Bongmyeong Elementary School, Teacher, Co-Author

Lee, Kwangho: Korea National University of Education, Professor, Corresponding Author