영어어문교육 *제30권* 1호 2024 봄
https://doi.org/10.35828/etak.2024.30.1.41

메타분석에 의한 AI 챗봇 기반 영어교육 프로그램의 교육효과 비교

박은정
(신도시평생학습센터)

Park, Eun-Jeong. (2024). A meta-analysis on the comparative effects of AI chatbot utilization in English education programs. *English Language & Literature Teaching, 30*(1), 41-61.

This study endeavors to assess the impact of AI chatbot-based English education programs on the effectiveness of language education for EFL learners. Additionally, it aims to examine moderator variables that influence the efficacy of these programs. A total of seventy-nine dissertations and academic journals, spanning from 2002 to October 2023, were chosen as the subjects for analysis. The effect size was computed utilizing a random effects model. A meta-regression analysis was conducted to validate moderation effects, and Egger's regression test was employed to assess publication bias. The impact of AI education programs was statistically significant, demonstrating a moderate effect size (0.44; 95% Confidence intervals: 0.34-0.54). The results of comparing educational effects in cognitive and non-cognitive domains revealed a higher effect size in the cognitive domain. The effect sizes based on educational outcomes ranked highest to lowest as follows: achievement, motivation, interest, confidence, attitude, and immersion. This study provides valuable insights into areas that need improvement to enhance the educational effects of utilizing AI chatbots. In future designs of English education programs utilizing AI chatbots, enhancing the pronunciation recognition component to ameliorate negative attitudes towards AI and diversifying conversation patterns to elevate learner immersion are anticipated to improve overall educational effectiveness.

[AI chatbots/educational effectiveness/comparative analysis/meta-analysis/AI, 챗봇/교육효과/비교분석/메타분석/인공지능]

Ⅰ. 서론

코로나 19 이후 대면 수업이 실시 되면서 비대면 수업 기간 동안에 벌어진 학생들 간의 학력격차가 문제가 되고 있다. 상위권 학생은 학업성취도가 높아진 반면 하위권의 학업성취도는 낮아졌기 때문이다(김경근 외, 2022; 정송, 안영은, 2021). 학생들 간의 학습격차가 큰 상황에서 모두에게 똑같은 내용을 전달하는 전통적인 수업방식으로는 학습자에게 적절한 교육을 제공하기 힘들 수 있다. 인공지능(AI) 기반 교육은 학습자가 자기주도학습을 할 수 있도록 개인 맞춤형 콘텐츠를 제공해 줄 수 있다는 점에서 이러한 상황의 해결책이 될 수 있다. 특히 영어를 외국어로서 학습하는 EFL 상황에서 AI 기반 영어교육은 시간에 구애 받지 않고 학습자가 필요할 때 자신의 수준에 맞게 일대일로 지도 받는 것과 같은 효과를 얻을 수 있을 것으로 기대된다.

AI 기반 교육프로그램은 다양하지만 영어교과에서는 주로 AI 챗봇을 활용하거나 자동 쓰기평가(Automated writing evaluation, AWE) 프로그램을 사용하는 경우가 많다. AI 챗봇은 인간이 의사 소통하는 것처럼 AI와 소통할 수 있는 프로그램을 뜻한다. AI 챗봇 관련 국내 연구로는 구글(Google)에서 제공하는 API(Application programming interface)인 다이얼로그플로우(Dialogflow)을 이용해 개인이 제작한 AI 챗봇을 이용한 연구(권해경, 2023; 김세연, 2023), 혹은 국내에서 최초로 국가적 차원에서 개발된 ‘AI 펭톡’을 이용한 연구(성수진, 2022; 우미혜, 김지연, 2023; 차수미, 2023)가 있다. AWE 프로그램은 영작 후 문법적 피드백을 주고 교정 서비스를 제공하는 AI 기반 프로그램을 말한다 (O’Neill & Russell, 2019). AWE 프로그램과 관련해서는 주로 Grammarly를 활용한 연구(손희정, 2023; O’Neill & Russell, 2019)가 이루어지고 있다. 그동안 AI 기반 영어교육프로그램 관련 실험연구가 많이 수행되어 오기는 했으나, AI 펭톡 외에는 주로 개별 연구자에 의해 개발된 프로그램으로 실험이 진행되었기 때문에 AI 기반 영어교육 프로그램의 교육효과를 종합적으로 알아보기는 쉽지 않았다. 교육부(2023)는 2025년에 인공지능 영어교과서를 전국 초·중·고에 보급하기로 확정하였다. 따라서 전국적인 AI 교과서 보급 이전에 지금까지 이루어진 개별적인 AI 기반 영어교육프로그램 관련 선행 연구를 종합하여 실제로 AI 기반 영어교육프로그램이 교육효과가 있는지를 분석해 볼 필요가 있다.

그러나 EFL 학습자를 대상으로 한 AI기반 영어교육프로그램에 관한 누적된 실험연구를 종합적으로 분석한 선행 연구(박은정, 2023a; Lee & Hwang, 2022)는 아직 많지 않다. 더구나 AI 챗봇을 수업에서 사용한 정도와 실험 기간을 조절 변인들 중 하나로 삼은 연구(Lee & Hwang, 2022)와 AI챗봇의 사용시간과 학교 급에 따른 교육효과크기를 비교한 연구(박은정, 2023a)는 있으나 아직 국내에서 AI 챗봇을 수업에 활용 시 교육효과 간에 효과크기를 비교한 연구는 없다. 박은정(2023a)은 초기 연구 설계 당시 교육효과 간의 효과크기를 비교하고자 하였으나 교육효과를 조절 변인으로 보기에는 선행 연구 수가 충분하지 않아 통

계적으로 유의미하지 않았고, 선행 연구 별로 제시한 교육효과가 너무 다양하여 출판된 모든 선행 연구의 교육효과를 하나로 보아 분석했을 때 이질성이 너무 높았으며 출판편향도 있었다고 보고했다. 박은정(2023a)의 연구 당시 보다 발표된 선행 연구가 많아진 현 시점에서, 다양한 AI 기반 영어교육프로그램 중 AI 챗봇을 활용한 영어교육프로그램에 관한 선행 연구의 특정 교육효과만을 분석대상으로 삼아 이질성을 줄인다면 교육효과 간에 효과크기를 비교해 AI 챗봇 기반 영어교육프로그램 개발 시 보완해야 할 점을 알 수 있을 것이라 생각한다. AWE 프로그램을 활용한 선행 연구를 제외하고 AI 챗봇에 관한 선행 연구만을 분석대상으로 한 것은 박은정(2023a)이 밝힌 것처럼 이질성과 출판편향을 줄이기 위해서이기도 하지만 23년 3월 공개된 생성형 AI 챗봇(Generative AI Chatbot)인 ChatGPT 4.0이 교육 패러다임을 바꿀 것으로 기대되기 때문이다. AI 챗봇은 대화를 통해 학습자와 상호작용하며 학습자가 실제 상황에서 영어를 사용하는 것과 유사한 경험을 제공해 줄 수 있다. 기존의 AI 챗봇은 실제 사람과 대화하는 것만큼 창의적이고 다양한 패턴이 가능하지는 않았다. 그 결과 성수진(2022)은 AI 챗봇 활용 수업을 통한 동기유발이 학습지를 활용한 수업과 큰 차이가 없었다고 보고했으며, 이삭(2019)은 대화의 자유도의 제한으로 AI 챗봇 활용 수업이 몰입에 부정적인 영향을 줄 수 있다고 했다. 그러나 ChatGPT 4.0을 활용해 챗봇과의 대화 시 답변 온도조절기능을 통해 대화의 자유도와 창의성을 조절할 수 있게 됨에 따라 이전보다 학습자가 챗봇과의 학습에서 몰입감을 느낄 가능성이 높아졌다. 답변의 창의성이 다소 제한되어 있던 ChatGPT 4.0 공개 이전의 챗봇을 활용한 선행 연구를 종합하여 분석해 본다면 차후 ChatGPT 4.0 기반 AI 챗봇 제작을 통한 수업설계안을 작성하는데 필요한 기초자료를 제공할 수 있을 것으로 기대된다.

따라서 본 연구에서는 2002년부터 2023년 10월까지 최근 20년간 연구된 국내외 석·박사 학위논문 및 학술지를 분석대상으로 삼아 EFL 학습자를 대상으로 AI 기반 영어교육프로그램 중 AI 챗봇을 활용 시 교육효과에 미치는 영향을 메타분석을 통해 종합적으로 알아보고자 한다. 또한 EFL 영어학습자의 교육효과 중 인지적 영역(성취도)과 비인지적 영역(자신감 흥미, 동기, 태도, 몰입)의 효과크기를 비교하고 각 교육효과 간 효과크기의 차이도 알아보고자 한다. 이를 위해 본 연구에서 선정한 연구문제는 다음과 같다.

1. AI 챗봇 기반 영어교육 프로그램이 교육효과에 미치는 전체 평균 효과크기는 어떠한가?
2. AI 챗봇을 기반 영어교육 프로그램의 인지·비인지영역에 따른 효과크기는 어떠한가?
3. AI 챗봇 기반 영어교육 프로그램의 교육효과(성취도, 자신감, 흥미,동기, 태도, 몰입)에 따른 효과크기 간의 차이는 어떠한가?

Ⅱ. 이론적 배경

챗봇은 챗(chat)과 봇(bot)이 합쳐진 말로 인간과 대화하는 것처럼 소통이 가능한 프로그램을 말한다. 이러한 챗봇을 교육에 활용할 경우 인지적 영역 및 비인지적 영역에 있어 교육효과를 기대해 볼 수 있다. 인지적 영역과 비인지적 영역은 교육효과를 높이기 위해 함께 고려되는 경우가 많다. 인지적 영역과 비인지적 영역은 둘 다 교육효과의 성공을 결정하는 중요 지표이기 때문이다 (Bloom, 1977; Gil-Hernández, 2021; Taryana 외, 2024).

인지적 영역은 특히 학업성취도와 높은 연관성이 있다(Kaur & Prajapati, 2022). 학업성취도는 현재의 상태로부터 더 나은 배움을 얻기 위해 학습을 할 때 올라가는 것이다(Heritage & Wylie, 2018). AI 챗봇을 학습에 활용하여 학업성취도가 올랐다는 다수의 선행 연구(김나영, 2018b; 박명호, 김정렬, 2021; 오유진, 백주현, 2022; 최원경, 2020; El Shazly, 2021; Hwang 외, 2022)를 볼 때 AI 챗봇의 주요 교육효과 중 하나가 학업성취도의 향상이라는 것을 알 수 있다. Bloom(1977)은 높은 교육효과를 얻기 위해서 인지적 영역인 학업 성취도를 가장 중요하게 여기기는 하였으나 학업성취도 외에 자신감, 흥미, 동기, 태도, 몰입 및 수업의 질과 같은 비인지적 영역도 함께 고려되어야 한다고 했다. 실제로 AI 챗봇을 활용한 실험 연구 중에는 인지적 영역인 성취도 외에 비인지적 영역을 함께 교육효과로 본 선행 연구가 많았다.

AI 챗봇을 활용한 영어교육 시 성취도와 함께 비인지적 영역 중 하나인 자신감이 향상된다는 선행 연구(김희삼 외, 2021; 윤택남, 2022; 진승희; 2022)가 있다. AI 챗봇 활용 교육 시 자신감은 모국어로서 영어를 사용하지 않는 학습자가 영어로 다른 사람 및 AI와 상호작용하는 데 성공할 것이라는 확신이다(박은정, 2023a). 김소연과 김정렬(2021)의 연구에 따르면, 학생들은 혼자서도 재미있게 학습하고 문장을 말해 볼 기회가 많아서 AI를 활용한 수업이 좋았다고 했다. EFL 환경에서 학습자가 교수자나 다른 학습자와 영어로 충분히 대화하기는 쉽지 않다. 혹은 다른 사람과의 대화 시 문법적으로 틀린 문장을 말할 수 있다는 두려움을 느껴 대화를 시도하는 것에 어려움을 겪기도 한다. 하지만 챗봇과는 언제든지 자유롭게 대화를 할 수 있고, 틀린 문장을 말하더라도 타인의 시선을 의식할 필요가 없어 대화 시 자신감이 향상될 수 있다고 보인다.

자신감과 함께 가장 많이 연구된 교육효과로는 흥미가 있다. AI를 이용한 학습 시 흥미는 영어학습 자체에 대한 내재적 흥미와 AI와 대화한다는 상황 자체에 대한 상황적 흥미를 모두 포함한 개념으로 보아야 한다(박은정, 2023a). AI 챗봇 활용 시 흥미에 영향을 준다는 선행 연구(김나영, 2018a; 장지혜, 김정렬, 2022; 진승희, 2022)에 따르면 흥미 또한 AI활용 영어교육의 주요 효과라는 것을 알 수 있다. 다만, 장지혜와 김정렬(2022)이 게임 형식의 AI 기반 챗봇에 학습자가 흥미를 느끼기도 했으나 같은 공간에서 다른 학습자가 연습하는 모습을 보면서 자극을 받아 더욱 흥미를 느꼈다고 보고한 점을 생각해 본다면, 수업

설계 시 AI 챗봇 자체에 대한 흥미 뿐 아니라 학습자 간의 상호작용이 동기에 주는 영향도 고려해야 한다는 것을 알 수 있다. Flowerday와 Shell(2015)은 90명의 대학생을 대상으로 흥미와 선택이 학습동기에 미치는 영향을 연구한 결과 흥미가 교육효과에 직접적인 영향이 있다고 밝혔다. 또한, 주제에 대한 흥미는 상황적 흥미에 직접적 영향을 주었으며, 상황적 흥미를 매개로 하여 긍정적인 태도를 향상시키고 동기에 영향을 주는 것으로 나타났다고 했다. 이러한 선행연구를 바탕으로 생각해볼 때, 흥미는 태도 및 동기와 밀접한 관련이 있는 교육효과 변인이라는 점을 알 수 있다.

실제로 흥미 외에 태도를 함께 교육효과로 살펴본 선행 연구(서소담, 김정렬, 2023; 차수미, 2023; 한다은, 2020)가 많다. AI 챗봇 활용 교육에서 태도는 AI 챗봇을 활용 후 학생이 수업에 임하는 자세의 변화라고 볼 수 있다. 차수미(2023)는 AI 펭톡을 활용한 수업 시 학습 내용이 영어 교과 시간에도 연계되어 학습자의 흥미 뿐만 아니라 영어수업을 대하는 태도까지도 긍정적으로 변했다고 보고했다. 흥미 외에 동기를 교육효과로 본 선행 연구(성수진, 2022; 신형정, 2002; Han, 2020)도 있다. 흥미와 동기는 비슷한 개념처럼 보이나 약간의 차이가 있다. 흥미는 내재적이며 시간에 따라 변화하기도 하지만 동기는 내재적일 뿐만 아니라 외부의 요인에 의해서도 발생하며 목표를 달성하려는 욕구이기 때문에 흥미보다 오래 유지될 수 있다. 선행 연구(성수진, 2022; Han, 2020)에서도 흥미와 동기가 항상 함께 향상되지는 않은 것으로 확인되어 흥미와 동기가 구별된다는 것을 알 수 있었다. Han(2020)이 AI 챗봇 활용으로 학습자의 흥미와 동기 모두 향상되었다고 보고한 것과는 달리 성수진(2022)은 AI 펭톡 활용 시 학습자의 어휘학습에 대한 흥미가 향상된 것은 맞으나 동기에 있어서는 교과서나 학습지를 활용한 어휘 학습보다 더 긍정적인 영향을 미쳤다고 보기는 어렵다고 하였다. Han(2020)의 연구는 중학생을 대상으로 이루어졌고 성수진(2022)의 연구는 초등학생을 대상으로 이루어졌으며, 각각 다른 AI 챗봇 프로그램으로 연구가 진행되었다. 따라서 이러한 상이한 결과가 연구대상의 연령에 따른 결과인지 프로그램의 차이에 따른 것일지는 알기 어려우나, 흥미와 달리 동기의 경우 AI 챗봇 활용에 따른 교육효과가 일관적으로 향상되지 않을 수 있는 것으로 확인되었다.

몰입 또한 주요한 교육효과 변인 중 하나이다. Reschly와 Christenson(2012)은 동기는 초반의 의도를 나타내고 몰입은 후반의 행동을 표현한다고 했다. 몰입에 대한 정의는 학자마다 다양하다. AI 챗봇을 이용한 선행 연구를 분석대상으로 삼은 본 연구에서는 몰입을 본질적으로 즐겁고, 자의식 상실이 수반되면서 자기 강화되는 상태(Novak & Hoffman, 1997)가 AI와의 상호작용에 의해 촉진되어 AI를 활용하는 동안 연속적으로 느껴지는 것으로 보고자 한다. AI 챗봇 영어교육프로그램에 관한 선행 연구 중 몰입에 관한 연구(서정임, 2021; 윤택남, 이승복, 2021; 이삭, 2019)도 동기처럼 시간이 지날수록 교육효과가 높아지는 양상만을 보이지는 않았다. 4주 동안 초등학생을 대상으로 실시된 연구(서정임, 2021;

윤택남, 이승복, 2021)에서는 몰입에서 유의미한 향상이 보였다. 그러나 이삭(2019)은 하위 수준의 학습자는 학습자의 발화 오류 등으로 인해 챗봇이 음성 인식을 하지 못하거나 정해진 말만 해야 하는 챗봇과의 대화에서 자유도의 제한을 느낄 때 몰입에 부정적인 영향을 받았다고 보고하였다.

기존의 챗봇을 활용한 선행 연구에서 나타난 몰입이나 동기에 있어서의 부정적인 영향은 실제 인간과 소통하는 것과 달리 대화에 있어서 제한된 내용만을 반복해야 하는 데서 오는 지루함이 원인일 가능성이 있다. 이러한 문제는 생성형 AI 챗봇의 발전과 함께 해결될 가능성이 높다. 규칙 기반 모델인 챗봇은 미리 정한 규칙과 입력한 데이터에 따른 대답만을 한다(손태권, 2023). 이에 반해 규칙 기반 챗봇에 머신러닝(machine learning)기술을 더해 사람이 실제 사용하는 언어의 형태로 소통할 수 있는 '자연어 처리(NLP) 기술'이 접목된 AI 기반 챗봇은 대화의 자유도가 더 높은 편이다(박정아, 이향, 2021). 2022년 11월 Open AI에서 개발한 생성형 AI 챗봇 'ChatGPT'가 발표된 이후로 딥러닝을 활용한 생성형 AI 챗봇에 대한 관심이 높아지고 있다. 생성형 AI 챗봇은 기존 챗봇과 달리 따로 업데이트를 해주지 않아도 지속적으로 스스로 학습하여 대화에 필요한 능력을 향상시키는 것이 가능하다(전종희, 2023). 따라서 시간이 흐를수록 대화의 자유도가 높아지고 맥락을 파악하여 실제 인간과 대화하는 것과 같은 흐름을 만들어낼 수 있다. 이로 인해 학습자는 챗봇과의 상호작용을 실제 다른 학습자나 교수자와의 상호작용과 비슷하게 느껴 동기와 몰입감이 높아질 가능성이 높다. 그러나 생성형 AI 챗봇을 활용한 연구는 현재 충분히 이루어지지 않은 상황이다(장혜지, 소효정, 2023).

생성형 AI 챗봇은 처음부터 교육을 위해 개발된 것이 아니기 때문에 교수자가 교수안을 개발하고 학습자에 맞춰 친근한 정도나 대답의 자유도 정도를 설정해 주는 것이 필요하다. 기존 AI 챗봇을 활용한 연구를 종합하여 성취도, 자신감, 흥미, 동기, 태도, 몰입과 같은 교육효과 변인 중 교육효과에 미치는 효과크기를 비교하여 생성형 AI 챗봇을 활용한 교수안에 적용해 본다면 기존 AI 챗봇을 활용한 영어교육에서 부족한 점을 보완할 수 있을 것으로 기대된다.

선행 연구에서 언급한 교육효과로는 성취도, 자신감, 흥미, 동기, 태도, 몰입 외에도 AI에 대한 인식 및 영어로 말하고자 하는 의지 등도 있었으나 사전검사 및 평균과 표준편차가 없는 연구를 제외 후 해당 변인에 관한 연구 수가 거의 남지 않아 전체 교육효과 분석 시 이질성을 줄이기 위해 분석대상 연구에서 제외하였다. 따라서 본 연구에서는 교육효과를 AI 챗봇을 활용한 영어교육프로그램을 통해 성취도, 자신감, 흥미, 동기, 태도, 몰입의 향상된 정도로 정의하고자 한다.

Ⅲ. 연구 방법

1. 메타분석

Plonsky와 Oswald(2015)에 따르면 메타분석은 좁은 의미와 넓은 의미로 나누어 정의될 수 있다. 메타분석의 좁은 정의에 따르면 메타분석은 전체 연구의 효과 크기를 모아 평균과 분산 및 상관 관계(r)나 표준화된 평균 차이(d)를 계산하기 위한 통계적 방법이다. 그러나 메타 분석을 더 넓게 정의하자면 단순한 통계 계산 뿐만 아니라 문헌의 개념적 통합과 메타 분석에 실질적인 의미를 부여하는 결과라고 보아야 한다.

메타분석은 동질적인 연구들을 종합하여 일반화할 수 있는 결과를 얻기 위해서도 실시되지만 각 연구마다 연구대상의 연령대 및 연구대상에게 실시한 수업 차수 및 시간이 모두 달라 이질성이 높은 다양한 선행 연구를 종합하여 분석하기 위해서도 필요하다. 메타분석에서 서로 상이한 집단간의 연구결과는 효과크기(effect size, ES)간 차이 비교를 통해 가능하다. 효과크기란 표준화된 평균 차이 혹은 상관관계이며, 이를 통해 동일한 주제에 대한 다양한 연구결과를 비교하는 것이 가능해진다(황성동, 2016).

메타분석을 활용한다고 해서 상이한 연구들을 모두 종합할 수 있는 것은 아니다. 지나친 이질성을 줄이기 위해 최소한 같은 주제나 프로그램을 적용한 연구 결과 간에 유의미한 통계치를 얻을 수 있는 연구를 선정하여 분석하는 것이 필요하다. Borenstein 외(2017)는 메타분석에서의 이질성에 관해 말할 때 실질적인 의미를 이해하는 것이 필요하다고 하였다. I^2에 의해 이질성이 판단된다는 일반적인 믿음이 있지만 실제로는 그렇지 않다는 것이다. 만약 효과크기가 50이라는 결과가 나왔다 하더라도 다양한 모집단의 효과크기가 40에서 60까지 변하는 지 10에서 90까지 변하는지는 I^2만으로는 판단을 내리기 어렵다. 황성동(2016)도 통계치가 절대적인 것은 아니며 연구자 개인의 판단이 중요하다고 강조하였다.

본 연구에서는 AI 기반 영어교육프로그램 중에서도 챗봇과 관련한 연구만을 분석대상으로 선정하고 교육효과를 성취도, 자신감, 흥미, 동기, 태도, 몰입으로 한정하여 이질성을 최소화하고자 하였다. 또한 메타분석을 통해 다양한 선행연구 전반에 걸친 효과 크기모음의 평균 및 분산과 더불어 상관 관계(r)와 교정된 표준화된 평균효과크기인 Hedge's g를 구하여 AI 챗봇을 활용한 영어교육프로그램의 전반적인 교육효과에 대해 분석해보고자 하였다. 표준화된 평균 차이(d)대신 교정된 표준화된 평균효과크기(g)를 분석에 사용한 이유는 이질성을 최소화하고자 했음에도 분석대상 연구의 연구대상 연령 및 프로그램이 달라 이질성이 높다고 판단되었고, I^2 또한 높았기 때문이다.

2. 분석대상

Plonsky와 Oswald(2015)가 제안한 코딩절차를 참고하여 분석대상 연구를 선정하였다. 본 연구의 메타분석을 위한 데이터는 구글 스콜라(Google Scholar), 한국학술정보(KISS), 교보문고 스콜라(Scholar), 네이버 학술정보(Naver Academic), 한국학술지 인용색인(KCI), 학술논문검색서비스(e-article), 학술연구정보서비스(RISS) 데이터베이스에서 ‘AI’, ‘챗봇’, ‘교육효과’, ‘영어’, ‘EFL’을 조합하여 2023년 10월 20일에 검색하였다.

온라인 검색을 통하여 확보한 검색자료는 2002년부터 2023년 10월까지 출판된 총 2525개의 학위논문 및 학술지논문과 참고문헌을 통해 추출한 연구 7편을 포함하여 총 2532개이다. 2532개의 연구 중 477개의 중복된 논문을 제외하고 남은 2055개의 논문 중 초록과 제목을 확인하여 주제와 관련 없는 1910개의 논문을 추가로 제외하였다. 중복된 연구를 제외할 때 동일한 제목으로 다른 사이트에서 검색된 연구를 제외하였으며, 동일한 내용을 영문과 국문으로 각각 출판된 경우에는 하나의 연구만을 포함시켰다. 동일한 저자가 학위논문으로 발표한 내용을 저널로 다시 발표한 경우 저널만을 분석대상으로 삼았다. 이 과정을 통해 정확한 주제와 실험과정 및 통계 수치를 확인할 수 있는 전문이 있는 연구 145개를 선정하였다. 이 과정에서 원래는 사전-사후검사를 제시한 단일 실험연구만을 대상으로 하거나 통제집단-실험집단의 사전-사후가 모두 제시된 연구만을 대상으로 하는 것이 일반적이다. 그러나 아직 선행 연구가 충분히 많지 않아 분석에 필요한 연구 수를 충족하기 위해 사전-사후검사를 제시한 단일 연구와 더불어 통제집단-실험집단 연구 중 실험집단의 사전-사후 검사 결과를 함께 분석에 활용하게 되었다. 145개의 연구 중 양적 실험연구가 아닌 40개의 논문을 제외하였고, 본 연구에서 교육효과로 본 성취도, 자신감, 흥미, 동기, 태도, 몰입 외의 교육효과를 다룬 연구 12개를 추가로 제외하였다. 성취도의 경우 영어를 외국어로서 학습하는 학생의 영어 말하기, 쓰기, 듣기, 읽기, 어휘에 관한 연구만을 분석 대상으로 삼았다. 동일한 연구 대상에게서 성취도의 5가지 영역 중 2가지 이상의 영역에서 연구를 진행한 경우 전체 평균과 표준편차를 구해 하나의 성취도로 보았다. 영어 말하기의 경우 발음, 유창성 등을 각각의 교육효과로 보고 전체적인 교육효과를 서술하지 않은 선행 연구의 경우 말하기 전체 영역을 하나의 교육효과로 본 다른 연구와의 균형을 맞추기 위해 전체 평균과 표준편차를 따로 계산하여 하나의 성취도 연구로 파악했다. 교육효과를 성취도, 자신감, 흥미, 태도, 동기, 몰입으로 한정한 것은 이질성을 줄이기 위함이다. 박은정(2023a)의 연구에서 교육효과를 특정하지 않은 채로 모든 AI 챗봇을 활용한 선행 연구를 메타분석을 통해 분석하였을 시 이질성이 너무 높고 출판편향이 있는 것으로 나타나 통계수치에 대해 신뢰성이 떨어졌다고 나타났기 때문이다. 하나의 연구에서 성취도, 자신감, 흥미, 태도, 동기, 몰입 중 2개 이상을 교육효과로 본 경우 각각을 개별 연구로 보아 코딩하였으며 본 연구에서 교육효과로

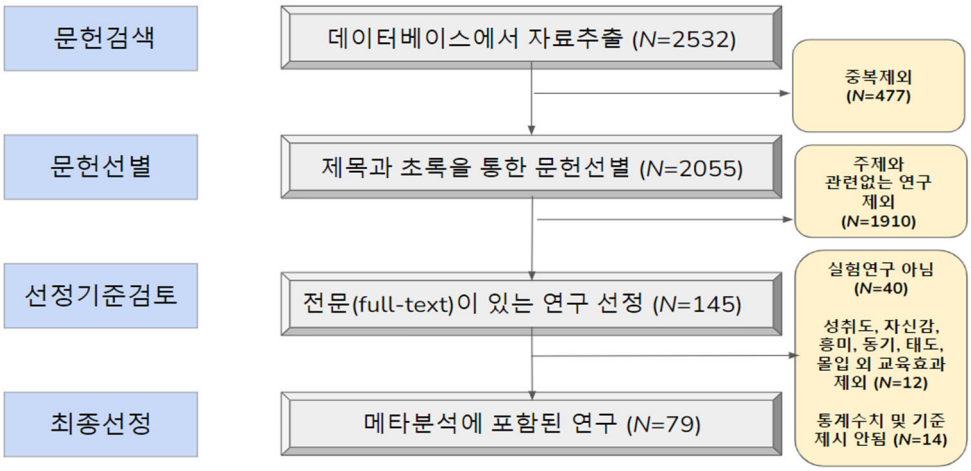
보지 않은 항목은 연구 수에 포함하지 않았다. 그 후 사전검사 및 평균, 표준편차를 제시하지 않은 연구와 만점의 기준을 제시하지 않아 다른 연구와의 효과 비교가 어렵다고 판단된 연구 14개 또한 제외하였다. 이 과정을 통해 최종적으로 메타분석에 포함된 연구는 총 79편이다. 분석대상 연구의 특성은 표 1로 정리하였다.

표 1
메타분석 대상논문

	연구 출처	연구 수	비율
출처	학위논문	29	(36.71)
	학술지	50	(63.29)
합계		79	(100%)

자료 선정 과정은 그림 1로 정리하였다.

그림 1
Prisma Flowchart



3. 자료코딩

교육효과는 크게 인지적 영역과 정의적 영역으로 나누어 생각해 볼 수 있다. 학업성취에 좋은 효과를 얻기 위해서는 정의적 특성 또한 고려되어야 한다는 선행 연구(Bloom, 1977)에 따르면 인지적 영역에 해당하는 학업성취도를 올리기 위해서는 정의적 영역 또한 함께 고려되어야 한다. AI 챗봇에 관한 선행 연구를 분석한 결과 인지적 영역에 해당하는 성취도에 관한 연구가 23편으로 가장 많았고 인지적 영역과 더불어 흥미, 몰입, 자신감, 태도, 동기를 함께 다루는 연구

가 함께 많이 이루어지고 있었다. 따라서 박은정(2023a)의 선행 연구 분석을 바탕으로 이질성을 줄이기 위해 교육효과를 성취도 및 성취도와 관련이 있는 자신감, 흥미, 동기, 태도, 몰입으로 보아 AI 챗봇을 활용한 영어교육프로그램의 전체적인 교육효과를 살펴보고자 하였다. 또한 전체 교육효과 중에 인지적 영역과 비인지적 영역 중 어떤 것이 더 큰 효과가 있는지를 알아보고자 하였다. 인지·정의적 영역 대신 인지·비인지 영역으로 조절 변인을 나누게 된 것은 교육효과를 인지적 영역, 정의적 영역, 심동적 영역으로 나누어 보는 연구(김세기, 2020; 노형규, 2012)가 있기 때문이다. 교육효과를 인지·정의·심동적 영역으로 나누어 보게 되면 흥미, 동기, 몰입의 경우 정의적 영역과 심동적 영역 모두에 영향을 주기 때문에 분류 기준이 명확하지 않다는 문제점이 생긴다. 또한 분석대상으로 삼은 선행 연구에서 정의적 영역 안에 흥미를 포함시킨 연구(윤택남, 이승복, 2021; 장지혜, 김정렬, 2022)가 있어 교육효과를 인지·정의적 영역으로 나누어 보기에는 무리가 있다고 판단했다. 따라서 본 연구에서는 성취도를 인지적 영역으로 보고 인지적 영역 외에 자신감, 흥미, 동기, 태도, 몰입은 비인지적 영역으로 보아 효과를 비교하고자 하였다. 또한 총 6개의 교육효과 중에서 프로그램 구성 시 보완해야 할 변인을 알아보기 위해 성취도, 자신감, 흥미, 동기, 태도, 몰입을 각각 조절 변인으로 코딩하여 효과크기를 비교하였다.

4. 분석방법

자료의 분석은 Borenstein 외(2017)의 연구를 참고하여 메타분석 프로그램 중 Comprehensive Meta-Analysis(CMA) 4.0 프로그램을 이용하였다. 표본이 작을 경우에는 과대 추정될 수 있는 Cohen's d대신 교정된 표준화된 평균효과크기(corrected standardized mean difference)인 Hedge's g를 사용하는 것이 좋다(Borenstein 외, 2017; Borenstein 외, 2021). 본 연구에서는 각 조절변인 간 연구 수의 차이가 있어 Hedge's g를 사용하여 95% 신뢰수준(Confidence intervals, CI)을 계산하였다. 각 효과크기의 가중치(weight)는 분산의 역수(inverse of variance)를 이용하였다. 그 후 동질성 검증(test of homogeneity)으로 각 연구의 효과크기 분포를 분석했다. 본 연구의 동질성 검증은 Cochran의 Q-검정을 통해 이루어졌다. 동질성 검증의 결과는 표 2에서 보듯이 $Q = 572.98$ ($p < .05$)이므로 동일한 모집단에서 표본이 추출되었다는 가정의 영가설을 기각하여 개별 연구들이 이질적이라고 판단했다. 따라서 T^2보다 연구 편수의 영향을 덜 받는 I^2를 사용하기로 결정했다. 분석 결과 I^2가 75% 이상인 86.39% ($p < .05$)였으므로 개별 연구들의 효과크기가 매우 이질적이라고 해석하였다.

표 2
동질성 검증 결과

무선효과모형	연구수	I^2	P	$Q(df)$
	79	86.39	.000	572.98(78)

이질성이 높을 경우 연구들이 기능적으로 동일하며 효과크기가 동질성을 가진다고 가정하는 고정효과모형보다는 분석 대상인 연구들을 잠재적 연구들의 무작위 표본으로 가정하여 해당 전체 연구 집단에 대한 추론을 수행하는 데 사용되는 무선효과 모형을 사용하는 것이 타당하다(Borenstein 외, 2021). 이와 같은 이유로 본 연구에서는 연구들의 효과 크기의 이질성을 고려하여 연구 간 분산을 인정하여 가중치가 보다 균형적이고 표준오차가 큰 무선효과모형을 선택하였다. 마지막으로 출판오류를 검증하고자 하였다. 분석대상 연구 수가 적을 경우 큰 효과크기로 계산되는 것을 출판 오류라고 한다. Egger의 회귀분석를 통해 회귀식 초기값의 P 값이 .98로 .05보다 커 통계적으로 유의하지 않음을 알 수 있었다. 그러므로 초기값은 우연히 생긴 결과라는 영가설에 따라 출판오류가 없다고 판단했다.

Ⅳ. 연구 결과 및 논의

1. AI 기반 영어 교육프로그램의 전체 교육효과크기

총 79편의 연구를 메타 ANOVA와 회귀분석을 통해 분석한 결과 전체 효과크기의 값은 .44($p < .05$)이며 95% 신뢰구간에 해당하는 값이 .34~.54으로 나타나 신뢰구간내에 0이 포함되지 않으므로 통계적으로 유의하다고 보았다. 효과 크기가 $+0.10 \leq ES \leq +0.30$인 경우는 낮은 정도의 효과, $+0.40 \leq ES \leq +0.70$인 경우는 중간 정도의 효과, $+0.80 \leq ES \leq +1.00$인 경우 큰 정도의 효과에 해당한다(황성동, 2016). 따라서 AI 챗봇을 영어교육에 활용 시 .44($p < .05$)의 중간 정도의 효과크기를 기대할 수 있다. 이를 정리한 내용은 표 3과 같다.

표 3
전체 평균효과 크기

모형	연구수	평균 효과크기	표준 오차	95% 신뢰구간		Z	$Q(P)$	I^2
				하한값	상한값			
무선 효과 모형	79	0.44	.05	0.34	0.54	8.83	572.98 (.000)	86.39

Z: 효과크기의유의도, Q: 동질성검정통계량

2. 인지·비인지 영역에 따른 효과크기

인지·비인지 영역에 따라 AI 챗봇을 활용한 영어교육 시 교육효과에 차이가 있는지 알아보기 위하여 메타 ANOVA를 실시하였다. 인지적 영역은 성취도로 보았으며 자신감, 흥미, 동기, 태도, 몰입은 비인지적 영역으로 보았다. 표 4를 보면 인지적 영역(ES = 0.70)이 비인지적 영역(ES = 0.36)보다 큰 효과를 보인다는 것을 알 수 있다. 인지적 영역의 효과크기는 0.70으로 중간효과크기에 해당하며 비인지적 영역의 효과크기는 0.36으로 작은 효과크기에 해당한다. 동질성 검증 결과 $Q = 4.25, df = 1, p < .05$이므로 인지·비인지 영역에 따른 차이가 있음을 알 수 있다.

표4
인지·비인지 영역에 따른 효과크기

교육효과	연구수	효과크기 (ES)	표준 오차	95% 신뢰구간	*Q(df)*	*p*
인지적 영역	22	0.70	0.15	0.40~1.00		
비인지적 영역	57	0.36	0.06	0.24~0.48	4.25(1)	0.04
총합	79	0.41	0.06	0.30~0.52		

Taryana 외(2024)는 Howard Gardner의 다중지능이론과 Daniel Goleman의 감성지능 접근 방식에 따른 문헌연구를 한 결과 교육 전략을 향상시키기 위해서는 교육자가 학업성취에 관한 인지적 영역 뿐만 아니라 정서적, 사회적 측면과 같은 비인적 측면을 고려해야 한다고 밝혔다. Gil-Hernández(2021)과 Bloom(1977) 또한 인지적영역과 함께 비인지적 영역이 교육효과를 높이기 위해 함께 고려되어야 한다고 했다. 특히 Bloom(1977)은 학업 성취도에 인지적 영역이 50% 정의적영역과 수업의 질이 각각 25%씩 영향을 준다고 밝혔다. 즉, 교육효과를 높이기 위해서는 비인지적 영역 또한 고려되어야 하지만 인지적 영역이 교육효과에 미치는 영향력이 더 높을 수 있다는 뜻이다. 본 연구 결과, EFL 영어학습자에게 AI 챗봇으로 영어교육을 실시할 경우 인지적영역과 비인적 영역에 모두 긍정적인 영향을 끼쳤으며, 특히 인지적 영역에 있어 더 큰 교육효과를 보이고 있었다. Bloom(1977)이 밝힌 교육효과에 미치는 인지적 영역과 비인지적 영역의 비율을 고려해 본다면, EFL 영어학습자에게 AI 챗봇을 활용한 수업을 실시할 경우 학업 성취도가 높아질 가능성이 있다.

3. 교육효과(성취도, 자신감, 흥미, 동기, 태도, 몰입)따른 효과크기

교육효과 중에서 가장 효과크기가 높았던 것은 표 5에서 보듯이 성취도(ES = 0.91)였고, 뒤이어 동기(ES = 0.87)의 수치가 높았다. 성취도의 효과크기인 0.91과 동기의 효과크기인 0.87은 0.80보다 크므로 큰 효과크기로 볼 수 있다. 동기 다

음으로 교육효과가 높았던 것은 흥미(ES = 0.63)와 자신감(ES = 0.52)이다. 흥미의 효과크기인 0.63과 자신감의 효과크기인 0.52는 0.40보다 크므로 중간효과크기에 해당한다. 그 다음으로는 태도(ES = 0.26), 몰입(ES = -0.86) 순으로 교육 효과가 높았다. 태도의 효과크기 0.26은 0.30보다 작으므로 작은 효과크기에 해당한다. 몰입은 역방향 효과(inverse-effect)로 나타났는데 이는 메타분석에서 일반적이지 않은 현상으로 본 연구의 메타분석의 대상집단의 연령, 프로그램, 교육효과가 모두 달라 이질성이 너무 커 발생한 것으로 보인다. 동질성 검증 결과 $Q = 48.03$, $df = 5$, $p < .05$이므로 교육효과에 따른 차이가 있음을 알 수 있다.

표 5
교육효과에 따른 효과크기

교육효과	연구수	효과크기 (ES)	표준오차	95% 신뢰구간	*Q(df)*	*p*
성취도	23	0.91	0.18	0.55~1.27		
자신감	14	0.52	0.09	0.33~0.70		
흥미	20	0.63	0.12	0.40~-0.87		
동기	3	0.87	0.33	0.22~1.53	48.03(5)	0.000
태도	4	0.26	0.11	0.05~0.47		
몰입	15	-0.86	0.22	-1.30~-0.42		
총합	79	0.44	0.06	0.36~0.59		

태도에 있어서 낮은 효과크기가 나타난 것은 AI 챗봇과 적절한 상호작용이 안 된다고 느낀 학습자가 학습에 부정적인 태도를 보이게 된 것 때문일 수 있다. AI 챗봇 활용 시 학습자의 발화 오류 등으로 인해 챗봇이 음성인식을 하지 못하는 경우 AI 챗봇 활용에 대해 부정적인 반응을 보였다는 선행 연구(이삭, 2019; 이지효, 민덕기, 2022; 차수미, 2023)가 이를 뒷받침한다. 음성 인식의 정확성이 너무 높게 설정되어 있을 경우 제대로 영어 발음을 하지 않으면 챗봇이 인식하지 못하는 문제가 생길 수 있다. 이러한 현상은 학습자의 수준이 높지 않은 경우에 주로 나타난다. 반대로 학습자의 다소 부정확한 발음까지 인식될 수 있도록 챗봇을 제작할 경우 유사한 단어를 사용할 수 없어서 대화패턴이 한정되는 문제가 생길 수 있다.

몰입의 경우에는 AI 챗봇과의 대화가 오랜 시간 지속될수록 한정된 대화 패턴으로 인해 학습자가 AI 챗봇을 더 이상 대화상대로서 인식하지 못하고 몰입 상태에서 벗어나기 때문에 부정적인 영향으로 나타난 것으로 보인다. 이삭(2019)은 챗봇이 음성인식을 하지 못하는 경우와 정해진 말만 해야 하는 챗봇과의 대화에서 학습자가 자유도의 제한을 느낄 때 몰입에 부정적인 영향을 준다고 하였다. 챗봇 활용 시 처음에는 신기함을 느껴 몰입에 긍정적인 영향을 주었으나 또래와의 상호작용 없이 챗봇과 계속해서 대화하게 되면 한정된 대화 패턴으로 인해 지루함을 느끼게 되는 것이다. 차수미(2023) 또한 AI 챗봇을 이

용하는 시간이 길어지면서 초반보다 관심도가 떨어지는 학습자가 관찰되었다고 보고했다. 박은정(2023b)은 총 수업시간에 따른 효과크기 차이를 알아보기 위해 EFL학습자를 대상으로 AI 챗봇을 활용해 영어학습을 진행한 선행 연구를 메타분석을 통해 살펴본 결과, 가장 짧은 수업시간이었던 1-100분의 효과크기(ES = 1.30)가 가장 높았고, 가장 긴 수업시간이었던 500분 이상을 수업한 경우의 효과크기(ES = 0.92)가 두번째로 높았다고 보고했다. 총 수업시간이 500분이 넘어가는 경우 교육 효과가 높게 나온 것은 학습의 누적효과 때문으로 볼 수 있으나 가장 짧은 수업 시간이었던 100분 이하의 수업의 효과가 가장 높게 나온 것은 몰입감 때문으로 보인다. Wu와 Yu(2024) 또한 AI 챗봇을 활용한 수업에서 단기간의 수업이 장기간의 수업보다 효과가 좋았다고 밝혔다. 이러한 선행 연구 결과를 바탕으로 생각해볼 때 기존의 AI 챗봇은 다양한 대화패턴을 구사하기 힘들어 총 학습시간이 100분이 넘어가면 학습자가 사람과 대화하는 것과 다르다는 것을 깨닫고 더 이상 대화에 몰입하지 못했을 가능성이 있다.

V. 결론

본 연구는 AI 기반 영어교육 프로그램이 교육효과에 어느 정도 효과를 보이는 지와 교육효과 별 차이를 메타분석을 통해 종합적으로 평가하는 데 목적이 있다. 연구결과에 대한 논의는 다음과 같다.

첫째, 무선효과모형으로 분석대상 79편의 논문을 분석한 결과 AI 챗봇을 활용한 영어교육 프로그램은 EFL 환경의 학습자의 교육효과는 중간정도(ES = 0.44)인 것으로 나타났다. 이것은 영어가 모국어가 아닌 환경에서 부족한 상호작용을 보충하기 위하여 AI를 활용 시 효과가 있다는 것을 보여주는 결과이다. 이는 선행 연구(박은정, 2023a; Lee & Hwang, 2023)와 일치하는 결과이다.

둘째, EFL 영어학습자를 대상으로 한 AI 챗봇 영어교육 프로그램의 인지·비인지영역에 따른 효과크기를 비교한 결과는 인지적 영역(ES = 0.70)이 비인지적영역(ES = 0.36)보다 큰 효과를 보이는 것으로 나타났다.

셋째, EFL 영어학습자가 AI 챗봇 기반 영어교육 프로그램 활용 시 교육효과 간의 크기를 비교해 본 결과 성취도(ES = 0.91), 동기(ES = 0.87), 흥미(ES = 0.63), 자신감(ES = 0.52), 태도(ES = 0.26), 몰입(ES = -0.86) 순으로 교육효과크기가 높은 것으로 나타났다. 즉 AI 챗봇을 영어교육에 활용 시 성취도에 가장 많은 영향을 주고 그 다음으로 동기, 흥미, 자신감, 태도, 몰입 순으로 영향을 준다는 뜻이다. 성취도와 동기의 경우 큰 효과크기에 해당하기 때문에 AI 챗봇을 활용한 영어수업을 할 경우 학습자의 성취도와 동기가 높아질 가능성이 매우 크다. 흥미와 자신감은 중간효과크기에 해당하여 AI 챗봇 활용이 학습자의 흥미와 자신감 또한 높일 수 있다고 판단된다. 그러나 태도의 경우 낮은 효과크기를 보였으며 몰입은 이례적으로 역방향 효과가 나타났다. 태도의 효과크기가 낮은

것과 몰입이 부정적인 효과를 보인 것은 AI 챗봇이 학습자가 하는 말에 따라 적절한 맥락을 유지하면서 이야기가 지속될 수 있도록 상호작용이 가능한 상황이 만들어지지 않았기 때문으로 보인다.

Malloy 외(2013)에 따르면 특정 맥락과 상황에 따라 학습자의 몰입 정도가 달라질 수 있다고 한다. 학습자가 선택할 수 있는 실제적이며 협동적이고 도전적인 과제가 주어질 때 학습자의 몰입이 향상된다는 선행 연구(Parsons 외, 2014)를 생각해보면, 학습자는 실제로 상호작용이 가능하며 매 순간 스스로 선택 가능한 과제가 주어지는 상황에서 몰입이 올라갈 수 있다고 생각된다. 박은정과 김희진(2023)은 K-MOOC수업에서 학습자와 학습콘텐츠 간 상호작용이 몰입에 영향을 미쳤으며, 학습콘텐츠의 하위요인인 마이크로러닝과 게이미피케이션 중 즉각적인 상호작용이 많다고 느낀 게이미피케이션이 몰입감에 더 많은 영향을 주었다고 보고했다. 즉, 학습자의 몰입감을 높이기 위해서는 영어 교육에 활용할 AI 챗봇 개발 시 학습자와의 상호작용에 중점을 두고 학습자에게 선택할 수 있는 과제를 지속적으로 제시하여 실제 사람과 대화하는 것과 같은 실재감을 줄 수 있도록 하는 것 필요하다는 뜻이다. 이지효와 민덕기(2022)는 교감적 언어사용을 하는 AI 챗봇을 제작하여 수업에 활용한 결과 학습자가 AI 챗봇을 사람처럼 여기며 챗봇과 공감했다고 여긴 대화내용을 특히 잘 기억했다고 보고했다. 따라서 단순한 과제 제시 및 피드백에 그치지 않고 AI가 학습자 개개인의 흥미와 상황을 파악하여 공감적 대화까지 가능하도록 설계한다면 전반적인 AI 활용 교육효과가 높아질 것으로 기대된다.

본 연구는 다양한 선행 연구를 종합하여 EFL환경에서 AI 챗봇과의 대화가 학습자의 교육효과에 영향을 준다는 것을 밝혀 추후 AI 챗봇을 활용한 영어교육프로그램의 지속적인 개발에 토대가 되는 자료를 제공한다는 데 의의가 있다. 또한 인지·비인지영역에 따른 효과를 비교하여 제시하였으므로 교수자가 수업 설계 시 AI 챗봇을 활용하여 어떤 효과를 거둘 수 있을 것인지 예측할 수 있는 정보를 제공한다는 데 학문적 의의가 있다고 볼 수 있다. 특히 교육효과(성취도, 자신감, 흥미, 동기, 태도, 몰입)에 따른 효과크기 간의 차이를 밝혀 앞으로 교육 현장에서 AI 챗봇을 활용한 영어교육 프로그램 개발 시 개선해야 할 점을 밝혔다는 데 실제적 의의가 있을 것이다. 특히 사용자의 대화 특성과 목적에 따라 답변의 창의성을 조절할 수 있는 Chat GPT 4.0을 기반으로 한 생성형 AI 챗봇을 활용한 수업을 위한 교수안 개발 시 필요한 기초자료가 될 것으로 기대한다. Chat GPT 4.0은 답변 온도를 0부터 1사이의 값으로 설정할 수 있다. 이 값이 높을수록 창의성이 높은 답변을 할 수 있어 실제 사람과 대화하는 듯한 느낌을 받을 수 있으나 이상하거나 문법적으로 틀린 답변을 가능성도 높아진다. 반면에 답변의 온도를 낮추면 더 확실하고 문법적으로 올바른 답변을 받을 가능성이 높아지나 기존의 챗봇처럼 대화 패턴이 한정되어 학습자의 태도와 몰입을 낮추는 요인이 될 수도 있다. 학습자의 수준과 원하는 교육효과에 따라 답변의 온도를 조절하며 교수자가 AI 챗봇을 제작하는 것이 가능해졌다는 뜻이

다. 따라서 답변의 온도가 낮은 기존의 챗봇을 활용한 선행 연구를 종합하여 분석한 본 연구 결과를 바탕으로 적절하게 창의적인 답변이 가능한 생성형 AI 챗봇 제작을 통해 학습자의 태도와 몰입을 높여 더 높은 교육효과를 얻을 수 있을 것으로 본다.

그러나 본 연구는 이러한 연구의 학문적, 실제적 의의 외에 다음과 같은 제한점도 있다. 첫째, 분석대상이 되는 연구를 검색 시 온라인 검색을 통해 구할 수 있는 자료만을 선정하여 오프라인 상으로만 출간된 연구물이 제외되었고, 연구자의 언어적 능력의 제한으로 영어와 한글로 된 논문만을 선정해야 했으므로 영어나 한글 외 다른 언어로 출간된 EFL 학습자 대상의 AI 챗봇 활용 프로그램 논문을 온라인을 통해 검색하고도 분석대상에 포함시키지 못했다. 둘째, 각 연구가 제시한 만점의 기준이 모두 다르고 전체 평균과 표준편차를 제시하지 않은 연구가 있어 연구자가 수기로 통계치를 다시 계산해야 했다. 특히 성취도의 경우 말하기, 듣기, 읽기, 쓰기, 어휘 중 단일 영역만을 분석 대상으로 하지 못하고 5가지 영역 중 하나 이상에 해당하는 경우 모두 성취도 관련 연구로 보았다는 문제가 있다. 이 과정에서 동일한 연구 대상에게 수행된 하나의 연구에서 다수의 분석대상 연구물이 나오는 것을 막기 위해 두 가지 영역 이상을 각각 평균과 표준편차를 제시한 경우 연구자가 따로 전체 평균과 표준편차를 구해야만 했다. 셋째, 교육효과 간의 효과크기를 알아보기 위해 통계적으로 의미 있는 결과를 얻기 위하여 최대한 많은 연구 편수를 확보하려다 보니 사전-사후 단일연구만을 대상으로 하지 못하고 실험집단과 통제집단이 있었던 실험연구에서 실험집단의 결과만을 분석 대상으로 삼게 되었다.

추후 연구에서는 이러한 점을 보완하기 위하여 충분한 연구 편수가 확보된다면 동일한 척도를 사용한 연구만을 분석대상으로 삼아 메타분석을 돌려 볼 필요가 있다. 또한 교육효과에 있어서도 단일한 효과만을 선정하여 메타분석을 한다면 연령, 총 수업시간 등이 특정 교육효과에 미치는 영향에 대해 좀 더 명확한 정보를 얻을 수 있을 것으로 본다.

참고문헌

교육부. (2023,10.17.). *인공지능(AI)디지털교과서 법적 지위를 얻다: 보도자료* https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&lev=0&statusYN=W&s=moe&m=020402&opType=N&boardSeq=96704

권해경. (2023). *AI 챗봇 활용 사고기능 과업의 초등영어 말하기 수업 적용과 분석*[미출간 석사학위논문]. 청주교육대학교.

김경근, 심재휘, 임혜정. (2022). 코로나19를 전후한 고등학생 수학 성취도 변화: 실태 및 영향요인. *교육과정평가연구, 25*(4), 63-88. https://doi.org/10.29221/jce.2022.25.4.63

김나영. (2018a). 챗봇 활용이 국내 영어 학습자의 어휘 습득에 미치는 영향. *디지털융복합연구, 16*(2), 1-7. https://doi.org/10.14400/JDC.2018.16.2.001

김나영. (2018b). 국내 대학생의 영어 듣기 및 읽기 능력 향상을 위한 챗봇 활용 연구. *디지털융복합연구, 16*(8), 19-26. https://doi.org/10.14400/JDC.2018.16.8.019

김세기. (2020). 코로나 19로 인한 초등학교 체육교과 원격수업 실태 분석. *한국초등체육학회지, 26*(2), 145-158. https://doi.org/10.26844/ksepe.2020.26.2.145

김세연. (2023). *AI 챗봇 활용이 초등학교 6학년 학습자의 영어 말하기 능력 및 정의적 영역에 미치는 영향*[미출간 석사학위논문]. 진주교육대학교.

김소연, 김정렬. (2021). AI 챗봇을 도입한 초등학교 영어 디지털 교과서 활용 수업의 정의적 효과. *학습자중심교과교육연구, 21*(10), 37-49. https://doi.org/10.22251/jlcci.2021.21.10.37

김희삼, 엄한숙, 오형나, 최다인. (2021). 인공지능 영어 말하기 프로그램의 효과성 분석. *교육방법연구, 33*(3), 563-588. https://doi.org/10.17927/tkjems.2021.33.3.563

노형규. (2012). 발달장애인 신체활동 프로그램 유형과 효과에 대한 메타분석. *한국특수체육학회지, 20*(3), 17-28. https://doi.org/10.17006/kjapa.2012.20.3.17

박명호, 김정렬. (2021). AI 챗봇을 활용한 고등학생들의 영어 글쓰기 상호작용 분석. *학습자중심교과교육연구, 21*(9), 1-13. https://doi.org/10.22251/jlcci.2021.21.9.1

박은정. (2023a). AI 챗봇을 활용한 영어교육 프로그램이 언어교육 효과에 미치는 영향에 대한 메타분석. *영어교과교육, 22*(3), 113-132. https://doi.org/10.18649/jkees.2023.22.3.113

박은정. (2023b, 11). *메타분석에 의한 AI챗봇 활용 영어교육 프로그램이 성취도, 자신감, 흥미, 몰입에 미치는 효과검증*[발표논문]. 한국영어어문교육학회 학술대회, 선문대학교 아산캠퍼스.

박은정, 김희진. (2023). 게이미피케이션과 마이크로러닝을 활용한 K-MOOC 수업이 몰입과 학습지속의향에 미치는 영향. *영어교과교육, 22*(2), 65-96. https://doi.org/10.18649/jkees.2023.22.2.65

박정아, 이향. (2021). 한국어 교육용 AI 챗봇 개발을 위한 챗봇 빌더 활용 방안. *외국어로서의 한국어교육, 63*, 51-91 https://doi.org/10.21716/TKFL.63.3

서소담, 김정렬. (2023). AI 챗봇 활용 수업이 초등 학습자의 영어 어휘 능력과 정의적 영역에 미치는 영향. *영어교과교육, 22*(1), 149-170. https://doi.org/10.18649/jkees.2023.22.1.149

서정임. (2021). *AI 챗봇 활용 영어 어휘 학습이 초등학생의 영어 어휘습득과 정의적 영역에 미치는 영향*[미출간석사학위논문]. 사이버한국외국어대학교 https://doi.org/10.5762/KAIS.2022.23.9.312

성수진. (2022). AI펭톡을 활용한 초등학교 5학년 영어 어휘 학습 효과 및 지도 방안. *멀티미디어언어교육 25*(3), 117-145. https://doi.org/10.15702/mall.2022.25.3.117

손태권. (2023). ChatGPT의 수학교육 활용 가능성 탐색: 분수 문제에 관한 학생의 산출물과 예비교사의 담화 사례를 중심으로. *초등수학교육, 26*(2), 99-113. https://doi.org/10.7468/jksmec.2023.26.2.99

손희정. (2023). AI 기반 자동쓰기평가(AWE) 프로그램 활용과 교수자 피드백에 대한 영어 학습자의 인식과 태도. *영어어문교육, 29*(3), 91-112. https://doi.org/ 10.35828/etak.2023.29.3.91

신형정. (2002). *인공지능 채팅프로그램이 고등학생의 영어 쓰기 능력 및 정의적 측면에 미치는 영향 연구*[미출간 석사학위논문]. 이화여자대학교.

오유진, 백주현. (2022). 인공지능 챗봇 'AI 펭톡'이 초등 영어 수업에 미치는 효과 및 활용방안: 말하기 지도 중심으로. *현대영미어문학, 40*(1), 207-235. https://doi.org/10.21084/jmball.2022.02.40.1.207

우미혜, 김지연. (2023). AI펭톡 활용 수업이 초등학생 영어 발음 능력 향상과 정의적 영역에 미치는 효과. *현대영어교육, 24*(1), 26-44. https://doi.org/10.18095/meeso.2023.24.1.26

윤택남. (2022). EBS AI 펭톡을 활용한 초등학습자들의 영어학습 효과. *영상영어교육, 23*(3), 27-38. https://doi.org/10.16875/stem.2022.23.3.27

윤택남, 이승복. (2021). 인공지능 기반 말하기 학습이 초등영어학습자들의 정의적 특성과 학습 만족도에 미치는 영향. *한국콘텐츠학회논문지, 21*(9), 34-41. https://doi.org/10.5392/JKCA.2021.21.09.034

이삭. (2019). 게이미피케이션 기반 AI 챗봇 활용 수업이 초등학생의 영어 말하기 수행 및 정의적 영역에 미치는 영향. *초등영어교육, 25*(3), 75-98. https://doi.org/10.25231/pee.2019.25.3.75

이지효, 민덕기. (2022). 초등학생 대상 AI 챗봇 기반 교감적 영어 말하기 수업 모형 개발 및 적용. *초등영어교육, 28*(3), 133-157. https://doi.org/10.25231/pee.2022.28.3.133

장지혜, 김정렬. (2022). AI펭톡을 활용한 초등영어 말하기 수업의 효과. *초등교과교육연구, 37*, 37-55.

장혜지, 소효정. (2023). ChatGPT의 교육적 활용 관련 연구동향 및 주제 분석. *교과교육학연구, 27*(4), 387-401. https://doi.org/10.24231/rici.2023.27.4.387

전종희. (2023). 대학교육의 챗GPT 대응 현황 및 방안 연구. *문화교류와 다문화교육, 12*(4), 517-548. https://doi.org/10.30974/kaice. 2023. 12.4.21

정송, 안영은. (2021). 코로나 19 전후 학교 내 학력격차 실태 분석: 서울 소재 중학교 학업성취등급 분포를 중심으로. *교육사회학연구, 31*(2), 53-74. https://doi.org/10.32465/ksocio.2021.31.2.003

진승희. (2022). AI 말하기 어플리케이션(앱) 활용 수업이 한국 대학생 학습자의 영어 말하기 능력과 정의적 태도에 미치는 효과. *멀티미디어 언어교육, 25*(2), 232-254. https://doi.org/10.15702/mall.2022.25.2.232

차수미. (2023). AI 펭톡이 초등 영어 학습자들의 어휘력과 정의적 영역에 미치는 효과 및 활용 방안. *초등영어교육, 29*(2), 5-22. https://doi.org/10.25231/pee.2023.29.2.5

최원경. (2020). AI 챗봇을 활용한 초등영어 과정중심 말하기 평가. *초등영어교육, 26*(1), 131-152. https://doi.org/10.25231/pee.2020.26.1.131

한다은. (2020). *AI 챗봇 활용이 한국 EFL 학습자의 말하기 능력 및 정의적 영역에 미치는 영향* [미출간박사학위논문]. 전남대학교.

황성동. (2016). *알기쉬운 메타분석의 이해*(3판). 학지사.

Bloom, B. S. (1977). *Human characteristics and school learning.* McGraw-Hill.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (Second ed.). Wiley.

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Res Synth Methods, 8*(1), 5-18. https://doi.org/10.1002/jrsm.1230

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems, 38*(3), 1-15. https://doi.org/10.1111/exsy.12667

Flowerday, T., & Shell, D. F. (2015). Disentangling the effects of interest and choice on learning, engagement, and attitude. *Learning and Individual Differences, 40*, 134-140. https://doi.org/10.1016/j.lindif.2015.05.003

Gil-Hernández, C. J. (2021). The (unequal) interplay between cognitive and noncognitive skills in early educational attainment. *American Behavioral Scientist, 65*(11), 1577-1598. https://doi.org/10.1177/0002764221996764

Han, D. E. (2020). The effects of voice-based AI chatbots on Korean EFL middle school students′ speaking competence and affective domains. *Asia-pacific Journal of Convergent Research Interchange, 6*(7), 71-80. https://doi.org/10.47116/apjcri.2020.07.07

Heritage, M., & Wylie, C. (2018). Reaping the benefits of assessment for learning: Achievement, identity, and equity. *Mathematics and Interdisciplinary STEM education, 50*(4), 729-741. https://doi.org/10.1007/s11858-018-0943-3

Hwang, W. Y., Guo, B. C., Hoang, A., Chang, C. C., & Wu, N. T. (2022). Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Computer Assisted Language Learning*, 1-27. https://doi.org/10.1080/09588221.2022.2095406

Kaur, G., & Prajapati, P. (2022). Academic achievement in relation to cognitive ability among secondary school students. *scholarly research journal for humanity science and English language, 10*(52), 13143-13151. https://doi.org/10.21922/srjhsel.v10i52.11520

Lee, J-Y., & Hwang, Y. (2022). A meta-analysis of the effects of using AI chat bot in Korean EFL education. *Studies in English Language & Literature, 48*(1), 213-243. https://doi.org/10.21559/aellk.2022.48.1.011

Malloy, J. A., Parsons, S. A., & Parsons, A. W. (2013). Methods for evaluating literacy engagement as a fluid construct. *62nd Yearbook of the Literacy Research Association*, 124-139.

Novak, T. P., & Hoffman, D. L. (1997). Measuring the flow experience among web users. *Interval Research Corporation, 31*(1), 1-35.

O′Neill, R., & Russell, A. (2019). Grammarly: help or hindrance? academic learning advisors' perceptions of an online grammar checker. *Journal of Academic Language & Learning, 13*(1), A88-A107.

Parsons, S. A., Nuland, L. R., & Parsons, A. W. (2014). The ABCs of student engagement. *Phi Delta Kappan, 95*(8), 23-27.

Plonsky, L., & Oswald, F. L. (2015). Meta-analyzing second language research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 106–128). https://doi.org/10.4324/9781315870908

Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 3-19). https://doi.org/10.1007/978-1-4614-2018-7_1

Taryana, T., Safar, M., Rukiyanto, B. A., & Prayitno, M. A. (2024). Contemporary educational perspective: Howard Gardner and Daniel Goleman's approach in cultivating adaptability and resilience. *International Journal of Education, Language, Literature, Arts, Culture, and Social Humanities, 2*(1), 69-80. https://doi.org/10.59024/ijellacush.v2i1.710

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology, 55*(1), 10-33. https://doi.org/10.1111/bjet.13334

예시언어(Examples in): English
적용가능 언어(Applicable Languages): English
적용가능 수준(Applicable Levels): Tertiary

박은정(강사)
신도시평생학습센터
32244 충남 홍성군 홍성읍 충서로 1200
Email: epark0108@gmail.com

Received: January 16, 2024
Reviewed: March 1, 2024
Revised version received: March 21, 2024