

간호사 국가시험에 대한 ChatGPT의 수행능력 평가

장선미
동신대학교 간호학과

Performance Evaluation of ChatGPT on the National Nurse Examinations

Seon Mi Jang
Department of Nursing, Dongshin University

요약 본 연구는 ChatGPT의 간호교육에 활용 가능성을 확인하고자 간호학 분야에서의 ChatGPT의 지식 수준을 측정하였다. 지식수준 측정 방법으로 간호사 국가시험에 대한 ChatGPT의 수행능력을 평가하였다. 수행능력 평가는 2019년부터 2023년까지 5년 동안의 간호사 국가시험에 대한 ChatGPT의 정답률을 확인하는 서술적 연구를 수행하였다. 간호사 국가시험의 합격 기준은 총점 60%이고, 과목별 과락 기준은 40% 미만이다. 총 5년 동안의 간호사 국가시험에 대한 ChatGPT의 평균 정답률은 49.5%로 간호학생의 정답률인 79.3%보다 낮았다. 총점 기준으로 2023년에 53.6%로 가장 높은 정답률을 보였고, 과목별로는 정신간호학이 62.9%로 가장 높은 정답률을 보였다. 5년간 총 40개의 과목 중 29개 과목(72.5%)이 과락 기준 이상의 정답률을 보였다. 연도별 ChatGPT의 정답률을 살펴보면, 2021년까지 정답률이 50% 미만이었으나, 2022년부터는 50% 이상으로 상승하였다. 본 연구의 결과, ChatGPT는 지난 5년 동안 간호사 국가시험에 합격하지 못하였으나 합격 수준에 가까운 수행능력을 보여주어 ChatGPT의 간호교육에 활용 가능성을 보여주었다. 앞으로 ChatGPT 같은 생성형 AI를 활용하는 다양한 연구가 간호교육에서 진행되기를 기대한다.

Abstract This study assessed the knowledge level of ChatGPT in the nursing field to confirm the possibility of using ChatGPT for nursing education. The national nursing exam performance of ChatGPT was assessed by conducting a descriptive study over the five-year period from 2019 to 2023 to determine its knowledge level and accuracy. The standard for a pass is an average total score of 60%, and a score of less than 40% in any of the forty subjects constitutes a fail. Over the five-year study period, ChatGPT had an average total score of 49.5%, which was considerably lower than the 79.3% achieved by nursing students. In 2023, the highest correct answer rate based on total scores was 53.6%, and psychiatric nursing led with a score of 62.9%. Out of 40 subjects, 29 (72.5%) surpassed the passing standard. The correct answer rate of ChatGPT was below 50% until 2021 but rose above 50% in 2022. This study shows that ChatGPT failed to pass the national nursing exam over the past five years but had a performance close to the pass level, which suggested its potential use for nursing education. Future studies utilizing generative AI programs, such as ChatGPT, in nursing education are anticipated.

Keywords : Artificial Intelligence, Natural Language Processing, Nursing Licensure, Nursing Education

*Corresponding Author : Seon Mi Jang(Dongshin Univ.)

email: jsm2076@gmail.com

Received May 9, 2024

Accepted August 2, 2024

Revised June 11, 2024

Published August 31, 2024

1. 서론

1.1 연구 필요성

ChatGPT는 OpenAI가 2022년 개발한 대화형 인공지능(artificial intelligence, AI) 챗봇으로 광범위하게 수집한 데이터를 기반으로 사전 학습되어 주어진 질문에 문장으로 생성된 답을 제시한다[1]. ChatGPT는 대화형 대규모 언어 모델(large language model, LLM)을 기반으로 입력된 텍스트와 이미지에 대해 몇 초 내에 텍스트로 결과물을 보여주고 인간과 유사한 성능을 보여준다[2]. 생성형 AI는 이미 다양한 분야에서 활용되고 있고, 간호를 포함한 보건의료 분야도 생성형 AI를 적용한 연구가 시작되었다[3-7]. 특히, ChatGPT는 보건의료의 교육과 연구 분야에서 교수자, 학습자, 연구자 모두에게 영향을 줄 가능성이 제시되었다[8].

교수자 측면에서 ChatGPT는 효과적인 교수 전략, 학습 선호도, 평가 결과에 대한 자료를 분석하여 교육 과정에 대한 자료를 제공하고, 교육 과정 및 개인 지도 계획, 교수법, 교육 자료 제공, 평가 계획 및 평가 기준 등을 제공한다[8]. 학습자 측면에서는 ChatGPT 활용 시 개인 맞춤형 학습, 즉각적인 피드백, 자기주도학습, 임상술기와 임상 추론 개발, 의사소통 기술 향상 등의 장점이 있다[8]. 연구자는 ChatGPT의 자료를 빠르게 추출하고 분석하는 기능을 활용하여 연구 과정을 간소화하여 시간을 효율적으로 활용할 수 있고, ChatGPT의 연구계획서, 논문 작성, 등의 기능을 활용하여 논문의 질을 높일 수 있다[8]. 또한 ChatGPT는 교육과 연구뿐만 아니라 임상 실무 영역에서도 다양하게 활용될 가능성이 제시되었다[3,8].

간호교육에서 ChatGPT의 역할에 대해 긍정적이어서 빅데이터 교육을 받은 ChatGPT는 간호교육에서 인간 교수자보다 다양한 지식을 갖고 있을 가능성이 제시되면서 ChatGPT의 간호교육 분야에서의 역할이 기대되고 있다[4]. ChatGPT는 학습 과정을 촉진하기 위한 대화형 도구의 역할을 하여[4], 간호교육에서도 협업 활동을 증가시킨다[9]. 그러나 ChatGPT는 편향된 결과물 출력 가능성, 정보의 정확성 및 신뢰성 문제, 과잉 의존, 학생들의 비판적 사고 및 의사소통 능력에 대한 잠재적 영향 등의 제한점도 존재한다[3,4,6,10]. 또한, 개인정보보호, 저작권, 표절 같은 윤리적 문제도 ChatGPT의 활용을 위해 고려해야 할 문제이다[8].

ChatGPT의 교육 간호교육에서의 활용 가능성이 제시되고 있지만, 간호교육에서 ChatGPT 적용은 초기 단

계이다. 국내에서는 2023년 ChatGPT의 기초간호교육 활용 가능성을 평가한 연구[11]와 생성형 AI를 활용한 사례 기반 간호교육 프로그램 개발한 연구[12]가 등장하였다. ChatGPT를 간호교육에 활용하기 위해서는 실증적 연구가 필요한데 특히 ChatGPT의 간호학 분야에 지식수준 등을 확인하기 위한 ChatGPT의 수행능력 평가 연구가 필요하다. 이를 위해 선행문헌에서도 의료인 국가시험에 대한 ChatGPT의 수행능력을 분석하였다. Huang의 연구[13]에서는 간호교육의 미래를 예측하고 준비하기 위하여 ChatGPT의 간호교육에 대한 적용을 검증하고자 대만 간호사 국가시험에 대한 ChatGPT의 정답률을 통해 ChatGPT의 수행능력을 검증하였고, Taira 등의 연구[14]는 비영어권 간호사 국가시험에 대한 ChatGPT의 수행능력을 검증하기 위하여 일본 간호사 국가시험에 대한 ChatGPT의 정답률을 분석하였다. 또한, ChatGPT의 의학적 지식역량과 환자치료에 대한 적용을 평가하기 위하여 미국 의사 면허시험에 대한 ChatGPT의 정답률을 통해 수행능력을 평가하였다[15].

이렇듯 선행문헌에서 간호학 교육에 AI 활용 가능성을 확인하기 위하여 간호사 면허시험[13,14]에 대한 ChatGPT의 수행능력을 분석하였으나 아직 국내에서는 이를 확인한 연구가 없다. 빠르게 발전하는 인공지능 기술에 발맞추어 간호교육의 미래를 적극적으로 준비하기 위하여 ChatGPT의 간호교육에 활용 가능성을 확인하는 것이 필수적이다. 따라서 본 연구에서는 ChatGPT의 간호교육에 활용 가능성을 확인하기 위하여 간호학 지식의 평가 도구인 간호사 국가시험을 사용하여 간호학에 대한 ChatGPT의 수행능력을 평가해 보고자 한다.

1.2 연구 목적

본 연구에서는 간호학에 대한 ChatGPT의 수행능력을 평가해 봄으로써 ChatGPT의 간호교육에서의 적용 가능성을 확인해 보고자 한다.

- ChatGPT의 수행능력을 확인하기 위하여 간호사 국가시험에서 ChatGPT 획득한 점수(정답률)를 통해 간호학 지식수준을 확인한다.
- ChatGPT의 수행능력을 확인하기 위하여 ChatGPT와 간호학생의 간호사 국가시험에 대한 점수(정답률)를 비교한다.
- ChatGPT의 수행능력 변화를 보기 위하여 ChatGPT의 간호사 국가시험에 대한 매년 정답률을 분석하였다.

2. 연구 방법

2.1 연구설계

본 연구는 한국의 간호사 국가시험에 대한 ChatGPT의 수행능력을 평가하기 위한 서술적 조사연구이다.

2.2 연구 자료

분석된 자료는 한국보건직업인국가시험원 홈페이지에서 제공되는 간호사 국가시험 기출 문제 및 정답 자료 [16]이다. 간호사 국가시험은 간호학 관련 지식을 측정하는 공식적인 평가 도구로 ChatGPT의 수행능력 분석에 활용되었다.

간호사 국가시험은 간호사 면허를 취득하기 위한 필수 시험으로 간호학을 전공하고 졸업한 자를 대상으로 간호학 관련 지식을 측정하는 평가 도구이다. 간호사 국가시험은 연 1회 시행되고 성인간호학을 포함하여 총 8개 과목으로 구성된다. 시험은 총 295문항의 사지선다형 객관식 문항으로 구성되고 1문항당 1점이 부여되어 총점 295점이다. 간호사 국가시험의 합격 기준은 총점 60% 이상이고, 각 과목당 과락 기준은 40% 미만이다. 시험 문항의 유형은 지식수준별과 자료유형별로 구분되고, 지식수준별 문항은 암기형, 해석형, 해결형으로 구성되고, 자료유형별 문항은 텍스트형과 자료제시형으로 구성된다.

2.3 연구 도구

간호사 국가시험에 대한 ChatGPT의 수행능력 평가는 GPT-3.5 모델을 사용하였다.

2.3.1 입력 source

본 연구에서 활용된 자료는 2019년부터 2023년까지 총 5회의 간호사 국가시험 문제와 정답이 포함되었다. 1회 시험 당 총 295문항은 성인간호학 70문항, 모성간호학 35문항, 아동간호학 35문항, 지역사회간호학 35문항, 정신간호학 35문항, 간호관리학 35문항, 기본간호학 30문항, 보건의약관계법규 20문항으로 구성된다. 5회 시험의 총문항수는 1,475문항이고, 그림, 표 등의 자료제시형 문항 수는 매년 다르다. 본 연구에서 활용한 GPT-3.5는 이미지 인식을 지원하지 않는 대화형 언어 AI이기 때문에 1,475문항을 모두 검토하여 그림, 표 등의 문항을 선별하였다. 표를 제시한 문항 중 텍스트형으로 변형이 가능한 경우 텍스트형 문항으로 변경하였고, 텍스트형으로 변형할 수 없는 표와 그림을 포함하는 문항은 본 연구에

서 제거하였다. 입력 소스는 하나의 질문과 하나의 객관식 답변으로 구성되었다.

2.3.2 프롬프트(prompt)

ChatGPT와 같은 LLM 기반의 생성형 AI는 텍스트를 입력받고 이에 대한 응답을 생성하는데, 이때 입력하는 텍스트 또는 질문이 프롬프트(prompt) 이다[18]. Textbox 1과 같이 간호사 국가시험의 각 문항을 그대로 입력하여 프롬프트를 구성하였다. 프롬프트 입력 순서는 빠른 연도부터 시작하여 빠른 문항부터 입력하였다. 이에 대한 응답으로 문항 풀이와 정답이 생성되거나 정답만 생성되었다. 각 문항에 대한 입력과 출력은 각 1회만 시행하여 정답률을 산출하였다.

```
Prompt : multiple choice (Simple-choice) questions
Question: (Questionnaire contents)
① Option 1
② Option 2
③ Option 3
④ Option 4
```

Textbox 1. Prompt for questions

2.4 자료 분석

간호사 국가시험에 대한 ChatGPT의 수행능력을 확인하기 위하여 ChatGPT의 간호사 국가시험에 대한 정답률을 총점 및 과목별 점수로 분석하였다.

간호학생의 수행능력은 간호사 국가시험에 응시한 간호학생의 평균 점수를 의미한다. 간호사 국가시험에 응시한 간호학생의 점수는 한국보건직업인국가시험원 홈페이지에서 제공하는 국가시험 분석 결과의 자료[17]로 매년 간호사 국가시험의 총점, 과목별 총점, 등의 정보가 제공한다.

간호사 국가시험에 대한 ChatGPT와 간호학생의 수행능력을 비교하기 위하여 ChatGPT의 정답률과 간호학생의 정답률을 연도별 총점 및 각 과목 정답률을 비교하였다. ChatGPT의 수행능력 변화를 확인하기 위하여 ChatGPT의 매년 정답률을 비교하였다.

3. 연구 결과

3.1 분석된 자료의 특성

최근 5년 간호사 국가시험의 총 1,475개 문항 중 자

료제시형 14개 문항을 삭제하여 본 연구에서 분석된 총 문항수는 1,461개이다(Table 1).

Table 1. Questions Included and Excluded in the Analysis from 2019 to 2023

Subjects (n)	Number of questions (included/excluded)				
	2019	2020	2021	2022	2023
Adult nursing (70)	68/2	70/0	69/1	68/2	68/2
Maternal nursing (35)	35/0	35/0	35/0	35/0	35/0
Child nursing (35)	35/0	35/0	35/0	35/0	35/0
Community nursing (35)	35/0	34/1	33/2	35/0	34/1
Psychiatric nursing (35)	35/0	35/0	35/0	35/0	35/0
Nursing management (35)	35/0	34/1	35/0	35/0	34/1
Fundamental nursing (30)	30/0	30/0	29/1	30/0	30/0
Health & medical laws (20)	20/0	20/0	20/0	20/0	20/0
Total (295)	293/2	293/2	291/4	293/2	291/4
Total 1461/14					

3.2 ChatGPT의 수행능력

ChatGPT의 간호사 국가시험의 정답률은 총점 기준 45.4%~53.6%의 분포를 보였고, ChatGPT의 5년 평균 정답률은 49.5%이다. 분석된 5년 동안의 총점 정답률을 살펴보면 2023년에 53.6%로 가장 높은 정답률을 보였고, 2021년에 45.4%로 가장 낮은 정답률을 보였다. 과

목별 정답률을 살펴보면, 정신간호학이 62.9%로 가장 높았고, 보건의약관계법규가 37.0%로 가장 낮았다. 5년 동안의 과목별 정답률에서 2020년 정신간호학 정답률이 77.1%로 가장 높았고, 2019년 아동간호학 정답률이 28.6%로 가장 낮았다. 또한 5년간 총 40개 과목 중 29개 과목이 과락 기준인 40% 이상의 정답률을 보였다 (Table 2).

3.3 ChatGPT와 간호학생의 수행능력 비교

5년 동안 간호사 국가시험에 대한 간호학생의 평균 정답률은 총점 기준 77.5%~82.4%의 분포를 나타냈고, 정답률이 가장 높은 해는 2023년이었고, 가장 낮은 해는 2020년이였다. 5년 동안 각 과목에서 간호학생의 정답률이 가장 높은 과목은 2023년 정신간호학 91.7%이었고, 2020년 아동간호학이 67.7%로 가장 낮았다(Table 3).

간호사 국가시험에 대한 ChatGPT와 간호학생의 수행능력을 정답률로 비교한 결과, ChatGPT의 5년 평균 정답률은 49.5%로 간호학생의 정답률보다 낮았다. 모든 연도에서 간호학생의 정답률 평균이 ChatGPT의 정답률보다 높았고, 또한 5년 동안의 모든 과목에서 간호학생의 정답률이 ChatGPT의 정답률보다 높았다. 수행능력

Table 2. Percent of Correct Answers by ChatGPT

Subjects	Correct answer/Total questions (%)					
	2019	2020	2021	2022	2023	Total
Adult nursing	37/68 (54.4)	39/70 (55.7)	26/69 (37.7)	38/68 (55.9)	30/68 (44.1)	170/343 (49.6)
Maternal nursing	24/35 (68.6)	13/35 (37.1)	21/35 (60.0)	15/35 (42.9)	18/35 (51.4)	91/175 (52.0)
Child nursing	14/35 (40.0)	10/35 (28.6)	18/35 (51.4)	14/35 (40.0)	20/35 (57.1)	76/175 (43.4)
Community nursing	12/35 (34.3)	18/34 (52.9)	17/33 (51.5)	18/35 (51.4)	24/34 (70.6)	89/171 (52.0)
Psychiatric nursing	18/35 (51.4)	27/35 (77.1)	16/35 (45.7)	23/35 (65.7)	26/35 (74.3)	110/175 (62.9)
Nursing management	20/35 (57.1)	17/34 (50.0)	16/35 (45.7)	21/35 (60.0)	16/34 (47.1)	90/173 (52.0)
Fundamental nursing	9/30 (30.0)	10/30 (33.3)	11/29 (37.9)	16/30 (53.3)	13/30 (43.3)	59/149 (39.6)
Health and medical laws	7/20 (35.0)	7/20 (35.0)	7/20 (35.0)	7/20 (35.0)	9/20 (45.0)	37/100 (37.0)
Total	141/293 (48.1)	141/293 (48.1)	132/291 (45.4)	152/293 (51.9)	156/291 (53.6)	722/1461 (49.5)

Table 3. Comparing the Performance between ChatGPT and Nursing Students

Subjects	Percent of correct answers (%)													
	2019		2020		2021		2022		2023		Total			
	ChatG PT	NS	ChatG PT	NS	ChatG PT	NS	ChatG PT	NS	ChatG PT	NS	ChatG PT	NS		
Adult nursing	54.4	73.9	55.7	79.3	37.7	74.9	55.9	74.0	44.1	75.1	49.6	75.4		
Maternal nursing	68.6	75.7	37.1	79.1	60.0	78.0	42.9	82.3	51.4	75.1	52.0	78.1		
Child nursing	40.0	74.9	28.6	67.7	51.4	88.3	40.0	80.9	57.1	83.1	43.4	79.0		
Community nursing	34.3	80.9	52.9	74.6	51.5	74.0	51.4	68.0	70.6	85.4	52.0	76.6		
Psychiatric nursing	51.4	78.9	77.1	85.4	45.7	85.7	65.7	85.4	74.3	91.7	62.9	85.4		
Nursing management	57.1	83.5	50.0	72.6	45.7	68.9	60.0	83.1	47.1	77.7	52.0	77.2		
Fundamental nursing	30.0	80.3	33.3	79.7	37.9	77.7	53.3	87.9	43.3	91.7	39.6	83.4		
Health and medical laws	35.0	86.0	35.0	81.5	35.0	78.0	35.0	73.5	45.0	79.0	37.0	79.6		
Average	48.1	79.3	48.1	77.5	45.4	78.2	51.9	79.4	53.6	82.4	49.5	79.3		

NS=nursing students

의 차이를 연도별로 분석한 결과, ChatGPT와 학생의 평균 정답률의 차이는 2022년 27.5% 포인트로 가장 적고, 2021년 32.8% 포인트로 가장 컸다. 수행능력의 차이를 교과목별로 살펴보면, 기본간호학에서 가장 큰 차이(43.8% 포인트)를 보였고, 정신간호학에서 가장 적은 차이(22.5% 포인트)를 보였다. 분석된 5년 동안의 40개 교과목에서 차이를 살펴보면, 2023년 기본간호학이 48.4% 포인트로 ChatGPT와 간호학생 간의 가장 큰 수행능력의 차이를 보였고, 2019년 모성간호학에서 가장 작은 7.1% 포인트 차이가 났다.

3.4 ChatGPT 수행능력 변화

ChatGPT의 간호사 국가시험에 대한 연도별 수행능력을 살펴보면, ChatGPT의 정답률은 2021년까지 50% 미만이었으나 2022년부터 50% 이상으로 상승하였다. ChatGPT와 간호학생의 정답률 차이의 추이를 살펴보면, 2021년까지 최대 32% 포인트의 차이가 있었으나 2022년부터는 29% 포인트 이하로 감소하였다(Fig. 1).

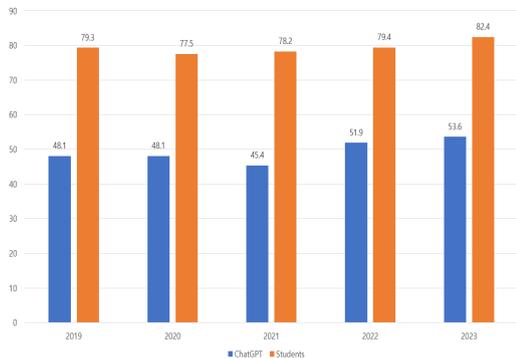


Fig. 1. Comparing the performance between ChatGPT and nursing students

4. 논의 및 결론

본 연구에서는 간호학에 대한 ChatGPT의 수행능력을 확인하기 위하여 간호사 국가시험에 대한 ChatGPT의 정답률을 분석하였다.

본 연구에서 간호사 국가시험에 대한 ChatGPT의 수행능력을 정답률로 본 결과, 총점 기준으로 약 45%~54%의 정답률을 보였다. 이는 간호사 국가시험의 합격 기준인 60%를 충족하지 못하여 ChatGPT는 지난 5년 동안 간호사 국가시험에 합격하지 못했지만, 합격 수

에 가까운 수행능력을 보여주면서 ChatGPT가 간호에 대한 지식을 어느 정도 갖추고 있다는 것을 보여주었다. 연구 결과를 선행연구와 비교 시, 대만과 일본의 간호사 국가시험에 대한 ChatGPT 수행능력을 각각 분석한 선행문헌인 Huang의 연구 결과[13] 54%~64%와 Taira 등의 연구 결과[14]인 60%~80%보다 낮은 수행능력 수준을 보여주었다. 기초간호학에 대한 ChatGPT의 수행능력을 분석한 연구[11]에서 ChatGPT를 한국어로 수행 시 영어에 비하여 ChatGPT의 정답률이 낮다고 하였는데, 한국어로 된 질문을 영어로 입력 시 정답률이 약 73%에서 100%로 증가하였다. 따라서 본 연구의 결과에서 다른 나라에서 수행한 연구 결과보다 낮은 ChatGPT의 수행능력을 보여주는 것은 입력된 언어의 차이로 인한 것으로 해석된다. 추후 한국어로 된 질문을 영어로 번역하여 질문을 하면 ChatGPT의 간호사 국가시험에 대한 정확한 수행능력 수준을 확인할 수 있을 것으로 보인다. 또한, Kim 등[11] 연구에서 기초간호학에 대한 ChatGPT-4와 ChatGPT-3.5의 수행능력 비교 시, ChatGPT-4의 정답률이 100%로 나타나 ChatGPT-3.5의 62.7%보다 높은 수행 수준을 보였다. 무료 버전인 GPT-3.5는 표와 그림에 대한 분석에 제한이 있으므로 간호사 국가시험에 대한 ChatGPT의 정확한 수행수준을 확인하기 위하여 ChatGPT-4.0을 활용한 연구를 시도해 볼 필요가 있다.

ChatGPT의 수행능력 수준은 총점 기준으로 간호사 국가시험의 합격 기준인 60%를 충족하지 못하였지만, ChatGPT의 총점 정답률은 2021년 이후 매년 증가하는 양상을 보여 2023년의 정답률은 약 54%로 합격 기준과 약 6% 포인트 차이를 보였다. 또한, 과목의 과락 기준으로 살펴보면 5년 동안 40개 과목 중 72.5%(29개)의 과목이 과락 기준인 40% 이상의 성취 수준을 보였고, 과락된 과목의 개수도 2019년 3개에서 2023년 0개로 감소하였다. 이를 통해 ChatGPT의 수행능력이 안정적이고 학습을 통해 향상되고 있음을 보여주었다. 아직 ChatGPT가 보여주는 정보가 완전히 신뢰하기는 어렵지만, 향후 몇 년 이내에 ChatGPT의 수행능력이 국가시험의 합격 기준을 넘길 것으로 예측된다. 교육 분야에서 교수설계 등에 ChatGPT가 이미 활용되고 있고[19], 간호 교육에서도 학습자와 교수자에게 긍정적 영향을 기대하며 ChatGPT 활용 가능성이 제시되었다[11,12]. 따라서 ChatGPT가 간호교육에서 다양하게 활용될 도구로써 인지하는 것이 필요하고, 교육의 도구로써 현명하게 이용하기 위해서는 이용 방안에 관한 연구가 필요하다. 즉,

ChatGPT의 안전한 기능과 사용을 위해 간호교육 분야에서의 적극적 활용과 실증적 연구가 필요하고, 관련 지침 등의 필요성을 인식해야 한다.

ChatGPT와 간호학생의 간호사 국가시험에 대한 정답률 비교 시, 간호학생의 정답률이 더 높았다. 모든 연도에서 간호학생의 간호사 국가시험에 대한 평균 정답률이 ChatGPT의 정답률보다 높았고, 또한 5년 동안의 모든 과목에서 간호학생의 정답률이 ChatGPT의 정답률보다 높았다. ChatGPT의 평균 정답률이 49.5%이고, 간호학생은 79.3%로 29.8% 포인트의 차이를 보였다. 이는 의학 분야에서 ChatGPT의 수행능력을 연구한 Huh의 연구[20] 결과와 유사하여 Huh의 연구에서도 기생충학 문제에 대해 ChatGPT는 60.8%의 정답률을 보였지만 의대생은 90.8%의 정답률을 보였다. 이를 통해 ChatGPT가 간호에 대한 전문 지식을 어느 정도 갖추고 있으나 아직은 4년 동안 간호학을 학습한 간호학과 졸업생의 수준보다 수행능력이 낮다는 것을 보여준다. 그러나 추후 ChatGPT의 간호학에 대한 학습의 기회가 증가한다면 정답률이 향상될 것으로 해석된다. 따라서 지속적인 연구로 AI의 수행능력 향상을 확인하여 추후 간호교육의 다양한 부분에 활용이 가능할 것으로 기대된다.

본 연구 결과에서 간호사 국가시험 연도가 증가할수록 ChatGPT의 정답률이 상승하고, ChatGPT와 간호학생의 정답률 차이가 감소하고 있다는 것에 관심을 두어야 한다. 이는 간호사 국가시험에 대한 ChatGPT의 학습 경험이 증가하면서 ChatGPT의 간호학에 대한 학습을 통해 지식이 향상되어 정답률이 향상되고 있다는 것을 보여준다. ChatGPT는 간호사 국가시험의 문제에 대한 몇 초 이내에 빠른 답변을 내놓는다. 또한 AI의 학습 속도는 인간과 비교할 수 없고 예측할 수도 없을 만큼 빨라서 추후 반복되는 간호학 관련 학습의 기회가 증가하면서 ChatGPT의 수행능력은 급속히 상승할 것으로 기대된다. 또한 현재 AI의 발전 속도를 고려하면 향후 수행능력은 향상될 것으로 판단된다. ChatGPT는 교육뿐만 아니라 연구, 임상실무 영역에서도 다양하게 활용될 가능성이 제시되면서[3], 교수자뿐만 아니라 학습자, 연구자, 실무자도 사용하기 시작하였다. 그러나 ChatGPT의 제한점으로 제시된 편향된 답변 가능성, 정보의 정확성, 저작권 문제 등을 ChatGPT의 활용 시 반드시 고려해야 할 문제이다[3,4,6,8,10]. 따라서 개인정보보호 및 저작권에 대한 윤리적 교육 등과 함께 ChatGPT가 제공하는 정보의 정확성 판단과 비판적 사고를 높이는 방법을 모색해야 할 것이다.

본 연구에서 ChatGPT의 간호학에 대한 지식수준을 확인하고 간호교육에 활용 가능성을 확인하기 위하여 간호사 국가시험에 대한 수행능력을 분석하였다. 분석 결과 ChatGPT의 수행능력은 간호학에 대한 전문적인 지식을 어느 정도 갖고 있으나 4년 간호학 교육을 마친 간호학생의 수행수준보다는 낮았다. 그러나 추후 ChatGPT의 간호학에 대한 학습의 기회가 증가한다면 수행능력이 향상될 것으로 기대된다.

본 연구를 통해 다음과 같이 제언하고자 한다. 추후 연구에서는 ChatGPT의 수행능력을 세부적으로 알아보기 위하여 간호사 국가시험의 과목별로 ChatGPT의 수행능력을 확인해 볼 필요가 있다. 또한 다양한 간호학 시험에서의 ChatGPT 수행능력을 확인해 볼 필요가 있다. 마지막으로 다양한 간호교육 프로그램에 ChatGPT를 적용하여 교육적 효과와 제한점을 평가해 보기를 기대한다.

References

- [1] Daum Encyclopedia. ChatGPT [Internet]. Daum encyclopedia, [cited 2024 March 1]. Available from: <https://100.daum.net/encyclopedia/view/47XXXXXb2696>
- [2] OpenAI. Research GPT-4 [Internet]. OpenAI, [cited 2024 April 1]. Available from: <https://openai.com/research/gpt-4>
- [3] M. Sallam, "ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns", *Healthcare*, Vol.11, No.6, pp.887, Mar. 2023. DOI: <https://doi.org/10.3390/healthcare11060887>
- [4] H. Miao, H. Ahn, "Impact of ChatGPT on interdisciplinary nursing education and research", *Asian Pacific Island Nursing Journal*, Vol.7, No.1, pp.e48136, Apr. 2023. DOI: <https://doi.org/10.2196/48136>
- [5] T. Talan, Y. Kalinkara, "The role of artificial intelligence in higher education: ChatGPT assessment for anatomy course", *International Journal of Management Information Systems and Computer Science*, Vol.7, No.1, pp.33-40, Jun. 2023. DOI: <https://doi.org/10.33461/uybisbbd.1244777>
- [6] M. Sallam, NA. Salim, M. Barakat, "Al-Tammemi A. ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations", *Narra Journal*, Vol.3, No.1, pp.e103, Apr. 2023. DOI: <https://doi.org/10.52225/narra.v3i1.103>
- [7] K. Ali, N. Barhom, F. Tamimi, M. Duggal, "ChatGPT—A double-edged sword for healthcare education? Implications for assessments of dental students", *European Journal of Dental Education*, Vol.28, No.1,

- pp.206-211, Aug. 2024.
DOI: <https://doi.org/10.1111/eje.12937>
- [8] S. Shorey, C. N. Mattar, T. L. B. Pereira, M. Choolani, "A scoping review of ChatGPT's role in healthcare education and research", *Nurse Education Today*, Vol.135, No.106121, pp.1-11, Apr. 2024.
DOI: <https://doi.org/10.1016/j.nedt.2024.106121>
- [9] G. Nalli, D. Amendola, S. Smith, "Artificial intelligence to improve learning outcomes through online collaborative activities", *Proceedings of the European Conference on e-Learning*, Brighton, UK, pp.475-479, October 27-28 2022.
- [10] J. Qadir, "Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. *Proceedings of the 2023 IEEE Global Engineering Education Conference, EDUCON*, Kuwait, Kuwait, pp.1-9, October 30 2023.
DOI: <https://doi.org/10.1109/EDUCON54358.2023.10125121>
- [11] S. Kim, J. Kim, M. J. Choi, S. H. Jeong, "Evaluation of the applicability of ChatGPT in biological nursing science education", *Journal of Korean Biological Nursing Science*, Vol.25, No.3, pp.183-204, 2023.
DOI: <https://doi.org/10.7586/jkbns.23.0013>
- [12] J. Ahn, H.O. Park, "Development of a case-based nursing education program using generative artificial intelligence", *The Journal of Korean Academic Society of Nursing Education*, Vol.29, No.3, pp.234-246, Aug. 2023.
DOI: <https://doi.org/10.5977/ikasne.2023.29.3.234>
- [13] H. Huang, "Performance of ChatGPT on registered nurse license exam in Taiwan: a descriptive study", *Healthcare*, Vol.11, No.21, pp.2855, Oct. 2023.
DOI: <https://doi.org/10.3390/healthcare11212855>
- [14] K. Taira, T. Itaya, A. Hanada, "Performance of the large language model ChatGPT on the National Nurse Examinations in Japan: evaluation study", *JMIR Nursing*, Vol.6, No.e47305, pp.1-8, Jun. 2023.
DOI: <https://doi.org/10.2196/47305>
- [15] A. Gilson, C. W. Safranek, T. Huang, V. Socrates, L. Chi, R. A. Taylor, D. Chartash, "How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment", *JMIR Medical Education*, Vol.9, No.1, pp.e45312, Aug. 2023.
DOI: <https://doi.org/10.2196/45312>
- [16] Korea Health Personnel Licensing Examination Institute. National nursing exam questions [Internet]. Korea Health Personnel Licensing Examination Institute, 2019~2023 [cited 2024 Mar 1]. Available from: https://www.kuksiwon.or.kr/CollectOfQuestions/brd/m_116/list.do
- [17] Korea Health Personnel Licensing Examination Institute. National nursing exam analysis [Internet]. Korea Health Personnel Licensing Examination Institute; 2019~2023 [cited 2024 Mar 1]. Available from: https://www.kuksiwon.or.kr/analysis/brd/m_91/list.do
- [18] TTA Information and Communication Terminology Dictionary. prompt engineering [Internet]. Daum encyclopedia; [cited 2024 Mar 1]. Available from: <https://100.daum.net/encyclopedia/view/55XXXXX98533>
- [19] S. Kang, H. Heo, "Development and application of generative AI-based instructional design platform", *The Journal of Korean Association of Computer Education*, Vol.26, No.6, pp.143-153, Oct. 2023.
DOI: <https://doi.org/10.32431/kace.2023.26.6.012>
- [20] S. Huh, "Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination? A descriptive study", *Journal Educational Evaluation for Health Professions*, Vol.20, No.1, pp.1-5, Jan. 2023.
DOI: <https://doi.org/10.3352/jeehp.2023.20.1>

장 선 미(Seon Mi Jang)

[정회원]



- 2013년 2월 : 서울대학교 간호학과 (간호학석사)
- 2020년 2월 : 서울대학교 간호학과 (간호학박사)
- 2021년 3월 ~ 현재 : 동신대학교 간호학과 교수

〈관심분야〉

간호정보학, 간호역량, 간호교육