Can ChatGPT Physiotherapy Field Pass the Professional Practice Exam?

Fatma ERDEO*

Necmettin Erbakan University, Nezahat Keleşoğlu Health Science Faculty,
Department of Physical Therapy and Rehabilitation, Konya, Türkiye

Article Info	ABSTRACT
	The development of advanced language models has revolutionized the way humans interact with computers, allowing machines to better understand and respond to complex linguistic input. As AI systems become increasingly capable, they offer the potential to improve the quality of life for more people.Many people say that the capabilities of ChatGPT1, a new chatbot model (Generative Pre-trained Transformer) from OpenAI, are impressive, and that the model can even pass university exams, and in particular, ChatGPT can also pass Multiple Choice Question (MCQ)-based exams for higher education. This study focuses on how ChatGPT programs affect academic success by demonstrating latent and advanced reasoning abilities by addressing the significant challenge posed by online exams. This study planned to examine the current research evidence on ChatGPT's performance on MCQ-based exams and its success on professional practice exams in physiotherapy education. ChatGPT3 free version 1.2023.256 and ChatGPT4 paid version were used. The final exam questions of the fall 2022 semester were used as the basis. ChatGPT3 passed with 61.01, while ChatGPT4 failed with 52.45. 58 out of 98 students scored 60 or above in this exam. 40 people scored below 60. MCQ-based assessments for professional qualifications should be conducted under secure conditions with limited access to similar tools like ChatCPT. The study highlights the need for robust online exam security measures, such as improved proctoring systems and more complex multi-modal exam questions, to reduce potential academic misconduct caused by AI technologies. This does not mean that AI tools like ChatGPT should not be used in education. On the contrary, artificial intelligence, which can be easily acquired by everyone, can be used to increase the quality of education in order to ensure sustainability in education in parallel with the developing technology. Additionally, AI should be used to address inequality in education.

ChatGPT Fizyoterapi Alanında Mesleki Uygulama Sınavını Geçebilir mi?

Makale Bilgisi	ÖZET
	Birçok kişi, OpenAI'nin yeni bir sohbet robotu modeli olan (Generative Pre-trained Transformer) ChatGPT1'in yeteneklerinin etkileyici olduğunu, modelin üniversite sınavlarında bile başarılı olabileceğini, özellikle ChatGPT'nin yüksek öğrenimde meslek edinmek için Çoktan Seçmeli Soru (ÇSS) tabanlı sınavları da geçebileceği söylenmektedir. Bu çalışma ChatGPT programların çevrimiçi sınavların bütünlüğü üzerinde oluşturduğu önemli zorluğu ele alarak, bu modellerin gizli ve gelişmiş muhakeme yeteneklerini göstererek akademik başarıyı nasıl etkilediğine odaklanmaktadır. Bu çalışmada, ChatGPT'nin ÇSS tabanlı sınavlardaki performansı ve fizyoterapi eğitiminde mesleki uygulama sınavındaki başarısı hakkındaki güncel araştırma kanıtlarının incelenmesi planlandı. ChatGPT3 ücretsiz sürümü 1.2023.256 ve ChatGPT4'ün ücretli sürümü kullanıldı. 2022 güz dönemi final sınav soruları baz alındı. ChatGPT3 61.01 ile başarılı olurken, ChatGPT4 52.45 ile başarısız oldu. Bu sınavda 98 öğrencimizden 58'i 60 veya üzeri puan aldı. 40 kişi 60 puanın altında puan aldı. Mesleki yeterlilik için ÇSS tabanlı değerlendirmeler, ChatCPT gibi benzer araçlara sınırlı erişimin olduğu güvenli koşullar altında yapılmalıdır. Çalışma, yapay zekâ teknolojilerinin yol açtığı olası akademik suistimalleri azaltmak için gelişmiş gözetmenlik sistemleri ve daha karmaşık çok modlu sınav soruları gibi sağlam çevrimiçi sınav güvenlik önlemlerine olan ihtiyacı da vurgulamaktadır. Bu, ChatGPT gibi yapay zekâ araçlarının eğitimde kullanılmaması gerektiği anlamına gelmez. Aksine, gelişen teknolojiye paralel olarak eğitimde sürdürülebilirliğin sağlanması amacıyla herkes tarafından kolayca edinilebilen yapay zekalar, eğitimde kaliteyi artırmak için kullanılabilir. Ayrıca, yapay zekâ eğitimde eşitsizliği sağlamak amacıyla kullanılmalıdır.

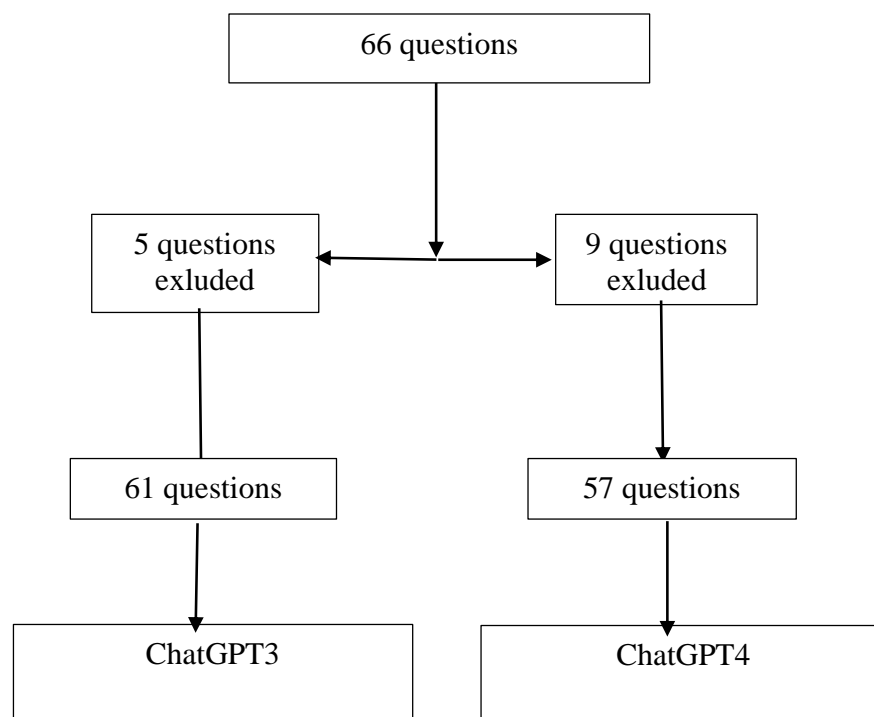*** Corresponding Author:** Fatma ERDEO, *ferdeo@erbakan.edu.tr*

INTRODUCTION

A quality university should provide safe, reliable and valid assessment in the award of professional qualifications, especially in the field of health. With the pandemic process, this evaluation process has completely turned from classical to multiple choice questions (MCQs). In the MCQ, the student is asked to choose the best answer among 4 or 5 options. Since MCQs do not contain comments, they are objective compared to other types of evaluations. If prepared appropriately, it can evaluate higher-level learning (Newton&Philip.,2023). ChatGPT (OpenAI, 2023; San Francisco, CA) is a large language model designed to function as a 'chatbot'. The version of ChatGPT running GPT-4 became available on a subscription basis on March 14, 2023 (OpenAI. 2023). Farazoli et al. A study he conducted indicates that ChatGPT will significantly disrupt education systems around the world (Farazouli et al. 2023). Hammer suggested that ChatGPT could beat university MCQ-based exams in law, medicine, and business (Hammer, Alex. 2023). This raises concerns about ChatGPT's AI-supported cheating, especially in online exams. For this reason, access to artificial intelligence is prohibited in some higher education institutions (ChatGPT). As ChatGPT continues to gain great traction and is increasingly used by students, there is a need to understand its impact on education and take immediate action against potential threats. In our study, professional MCQ exam questions applied to students receiving four-year physiotherapy education were asked to ChatGPT3 and ChatGPT4.

METHOD

This is an observational study. Ethical approval was not obtained because the artificial intelligence robot ChatGPT3 and ChatGPT4 versions were used in this study. ChatGPT's free version 1.2023.256 was used for the study. It was based on the 2022 fall semester final exam questions. The question content consisted of professional questions such as anatomy, neurology, orthopedics, treatment, prosthesis-orthosis. Afterwards, questions were asked out loud. And the answer was received in written form (Figure 1).

Figure 1

Exam Question Selection of Artificial Intelligence ChatGPT

RESULTS

While ChatGPT3 succeeded with 61.01, ChatGPT4 failed with 52.45. 60 students out of 98 passed the exam. 30 students failed the exam. The average of students is 62.11. The student received a minimum of 41 and a maximum of 86 points (Table 1).

Table 1

ChatGPT's response rate to questions

	Correct Responses		Incorrect Responses		Grade	z	p
	N	%	N	%			
ChatGPT3	36	61.01	25	42.37	61.01	-.784[b]	.433
ChatGPT4	32	52.45	25	25	52.45		

N: Number of Correct Answers, z: Wilcoxon Signed Rank test, p<.005

DISCUSSION

The aim of this study is to evaluate the success of physiotherapy students with ChatGPT3 and ChatGPT4 and to review the risks and precautions that may occur for MCQ. The ChatGPT3 version performed close to the students. He answered about half of the questions correctly. ChatGPT4 showed lower results in physiotherapy MCQ compared to ChatGPT3. A review of the literature showed that GPT3 failed the exam and had poor performance when compared to average student scores. ChatGPT4, on the other hand, performed significantly better, passing most tests (Newton & Xiromeriti, 2023). It was observed that ChatGPT4 was less successful in the physiotherapy professional practice MCQ, especially in case questions, and ChatGPT4 had more difficulty in understanding the questions. Talan and Kalınkara, (2023) showed in their study that ChatGPT was more successful than undergraduate students in anatomy questions (Talan, T., & Kalınkara, Y. (2023). Gilson et al. (2022) demonstrated that ChatGPT demonstrated comparable performance to medical students on medical licensing exams (Gilson et. al. 2023). Similar to our result, Huh (2023) found that ChatGPT's overall performance on the parasitology exam was lower than that of medical students in Korea (Huh. 2023).

Although the results with ChatGPT vary depending on the field, it is predicted that these results will be similar with deep learning in the very near future. While with ChatGPT3 you can ask a voice question and get a written answer, in ChatGPT4 it is a disadvantage that you cannot complete long questions out loud. ChatGPT4 was more efficient in terms of copying and pasting and giving the answer in written form. Although there are difficulties, ChatGPT's ability to produce highly realistic texts poses a potential risk to the integrity of online exams and measures need to be taken to prevent this (Qureshi, B. 2023). When asking questions at this time, it is important to ask loud and clear questions in order to get correct answers. However, it is certain that this will make students difficult in the exam. It is likely that this will be overcome in the near future. With its wide range of applications in various fields, questions arise as to whether ChatGPT can also be a valuable asset in the physiotherapy profession. But despite its impressive capabilities, ChatGPT falls short of passing the professional practice exam. Physiotherapy is a specialized field that requires extensive knowledge, critical thinking skills and practical experience. Physical therapists are trained to evaluate, diagnose and provide hands-on treatment to patients with musculoskeletal, neurological and cardiopulmonary disorders. They are also responsible for developing personalized treatment plans and monitoring patients' progress throughout their rehabilitation process. While ChatGPT can process huge amounts of information and produce coherent responses, it lacks the practical experience and expertise required to perform physiotherapy tasks effectively. The professional practice examination serves as a comprehensive assessment of the physical therapist's knowledge, clinical judgment, and practical skills. It tests the ability to make an accurate diagnosis, design appropriate treatment plans, and demonstrate interpersonal communication

skills. These are areas where ChatGPT falls short. Because he lacks the ability to make physical assessments, interpret visual cues, or engage in empathetic patient interactions. The limitation of ChatGPT is that questions are asked in Turkish and answers are received in Turkish. More successful results could have been obtained if the questions were asked in English.

CONCLUSION

As a result, although ChatGPT has tremendous linguistic talent and practical skills, its lack of humanity is evident. Although artificial intelligence technologies can play a supporting role in physiotherapy, they cannot replace the expertise and skills of physiotherapists. Passing the professional practice examination in physiotherapy requires a combination of theoretical knowledge, practical skills and the ability to establish therapeutic relationships with patients; Not all of these are available in ChatGPT. As a result, it is important to recognize the limitations of AI systems and continue to prioritize the human element in healthcare.

SUGGESTIONS

According to Edward Fredkin, an authority in the field of artificial intelligence, "There are three great events in history. The first is the formation of the universe. The second is the beginning of life. The third is the emergence of artificial intelligence." According to this statement, it is not possible to predict where the limits of artificial intelligence will go. In order for this technology, which is advancing at a dizzying pace, to add power to education from different perspectives and to be used to overcome the problems encountered in the teaching process, studies should be carried out to develop education in this field.

Statement of Research and Publication Ethics

This research has been written in accordance with scientific research and publication ethics rules.

Ethics Committee Approval

Since no human or animal subjects were used in this study, no application was made to the ethics committee.

Author Contributions

Research Design (CRediT 1) Author 1 (%100)

Data Collection (CRediT 2) Author 1 (%100)

Research - Data Analysis - Validation (CRediT 3-4-6-11) Author 1 (%100)

Writing the Article (CRediT 12-13) Author 1 (%100)

Revision and Improvement of the Text (CRediT 14) Author 1 (%100)

Finance

There is no financial support.

Conflict of Interest

There is no any declaration of interest.

Sustainable Development Goals (SDG)

Sustainable Development Goals: 4 Quality Education

Sustainable Development Goals: 10 Reducing Inequalities

REFERENCES

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education, 49*(3), 363-375.

Edwards, A. A., Steacy, L. M., Siegelman, N., Rigobon, V. M., Kearns, D. M., Rueckl, J. G., & Compton, D. L. (2022). Unpacking the unique relationship between set for variability and word reading development: Examining word- and child-level predictors of performance. *Journal of Educational Psychology, 114*(6), 1242–1256. https://doi.org/10.1037/edu0000696

Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Medical Education, 9*, e45312. https://doi.org/10.2196/45312

Hammer, A. (2023). The rise of the machines? ChatGPT CAN pass US Medical Licensing Exam and the Bar, experts warn-after the AI chatbot received B grade on Wharton MBA paper. Daily Mail.

Huh S. (2023). Issues in the 3rd year of the COVID-19 pandemic, including computer-based testing, study design, ChatGPT, journal metrics, and appreciation to reviewers. *Journal of educational evaluation for health professions, 20*, 5. https://doi.org/10.3352/jeehp.2023.20.5

Newton, P. M. (2024). Guidelines for creating online MCQ-based exams to evaluate higher order learning and reduce academic misconduct. In Second Handbook of Academic Integrity (pp. 269-285). Cham: Springer Nature Switzerland.

Newton, P. M., & Xiromeriti, M. (2023). ChatGPT performance on MCQ exams in higher education. A pragmatic scoping review. EdArXiv. February, 21.

OpenAI. (2023). 'New GPT-3 Capabilities: Edit & Insert'. 2023. https://openai.com/blog/gpt-3-edit

Dibble, M. (2023). Schools ban ChatGPT amid fears of artificial intelligence-assisted cheating. VOA News.

Qureshi, B. (2023). Exploring the use of chatgpt as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges. arXiv preprint arXiv:2304.11214.

Talan, T., & Kalınkara, Y. (2023). The role of artificial intelligence in higher education: ChatGPT assessment for anatomy course. *International Journal of Management Information Systems and Computer Science, 7*(1), 33-40.