

Using Generative Artificial Intelligence in Exams: A Research on KPSS with ChatGPT

Ali Öter¹ Betül Ersöz² Halil İbrahim Bülbül³ Şeref Sağıroğlu⁴

Article History:

Received 03.01.2024

Received in revised form
26.09.2024

Accepted

Available online 01.10.2024

ChatGPT is a language model developed by OpenAI that uses generative artificial intelligence (GenAI) algorithms for natural language processing (NLP) tasks. It is a popular Chatbot due to its impressive ability to respond consistently to users' various prompts. In this study, the Public Personnel Selection Examination (KPSS) questions, organized by ÖSYM (Turkish Student Selection and Placement Center) every year, were asked to ChatGPT, and its performance was evaluated. The evaluation was made based on fundamental text questions asked in the last three years (2021-2023) in Biology, History, Geography, Turkish Language Literature, and Social Studies in KPSS. This study also aims to investigate the accuracy and reliability of the questions asked to ChatGPT in the field of KPSS educational sciences. According to the research results, ChatGPT's KPSS knowledge and interpretation skills are significantly lower than the university graduate candidates who took the exam. Some suggestions are included for better Chatbot performance, more effective prompt engineering and ethical use of these tools.

© IJERE. All rights reserved

Keywords: ChatGPT, GenAI, KPSS, LLM, Prompt Engineering

INTRODUCTION

Generative artificial intelligence (GenAI) is a type of artificial intelligence that is responsible for creating original content, using deep learning algorithms to generate content such as images, videos, music, and more. ChatGPT is an important language model developed by OpenAI and uses generative artificial intelligence algorithms (Guo et al., 2021; Khanzode & Sarode, 2020). It is an AI algorithm that generates consistent and contextually appropriate responses by training on an extensive library of text data using Natural Language Processing (NLP) natural language input. It is regarded as a text-writing of effective text generation model (Liu et al., 2023). It is one of the most effective at performing NLP (Awasthi et al., 2021). Furthermore, ChatGPT has made great strides in all aspects of human contact, including Chatbots, customer support, and content creation. The potential use of ChatGPT has attracted attention recently due to the rapid increase in the use of artificial intelligence models in professional environments.

Chatbots are an essential part of AI technology and are widely used in various fields such as healthcare, banking, education, and marketing. ChatGPT has been developed using AI methods to improve efficiency in multiple aspects of the healthcare system, such as personalized medicine, medical imaging analysis, diagnostics, and predictive models (Hardy & Harvey, 2020; King, 2023). In the banking sector, Chatbots also help customers deliver customer service and support, account information and transactions, investment advice, loan and loan applications, and other similar topics faster and more efficiently, as well as increasing the efficiency of operations (Sarbabidya & Saha, 2020). The educational guide published by OpenAI has different approaches to using ChatGPT as a tutor to develop curricula based on teacher input, to generate practical explanations, examples, and analogs, to help students learn by teaching, to guide them to understand concepts, and to ask them. It also guides questions and encourages their answers.

Furthermore, ChatGPT is employed in several worldwide educational tests. These tests measure ChatGPT's effectiveness (Biswas, 2023; Sok & Heng, 2023). In the United States, medical licensure exams have used ChatGPT. The results have indicated a significant improvement in natural language, especially regarding the capacity to respond to medical questions (Gilson et al., 2023). The successful completion of ChatGPT in the Certified Public Accountant (CPA) and Uniform Bar Examination (UBE) tests (Bommarito et al., 2023) and (Bommarito II & Katz, 2022) demonstrates its capacity to produce precise text responses to challenging input standards. When the literature is examined, it is seen that ChatGPT offers a wide range of work thanks to its automation and improvement potential. It distinguishes strengths and weaknesses in a particular task and recommends student work (Grassini, 2023). ChatGPT is a new paradigm in education, which includes creating learning assessments, improving pedagogical practice, offering virtual personal lessons, creating an

Kahramanmaraş Sıtkı İmam University, aliooter@ksu.edu.tr, orcid.org/0000-0002-9546-0602
Gazi University, betul.ersoz@gazi.edu.tr, orcid.org/0000-0001-6221-1530
Gazi University, bhali@gazi.edu.tr, orcid.org/0000-0002-6525-7232
Gazi University, ss@gazi.edu.tr, orcid.org/0000-0003-0805-5818

outline, and brainstorming. However, there are also risks related to academic integrity issues, unfair learning assessment, misinformation, and over-reliance on artificial intelligence (Sok & Heng, 2023).

KPSS placements are calculated according to the fields and scores of four-year undergraduate graduates in public procurement. Candidates taking the KPSS mostly use the information produced by Chatbot in their studies during the preparation period for the exam (Kocaoğlu & Özmen, 2019). For this reason, it is necessary to determine how accurately the recently based on GenAI, ChatGPT can solve KPSS-related questions. ChatGPT's ability to answer questions prepared in Turkish should be evaluated. The study evaluates the accuracy and reliability of undergraduate teacher candidates' answers to ChatGPT's KPSS questions in five fundamental science fields. Additionally, suggestions for users were made by questioning ChatGPT's responses to the prompts.

METHOD

This research examined the questions asked in the five fundamental tests in 2021, 2022, and 2023 from 17 fundamental fields in the KPSS held yearly by the Measurement, Selection, and Placement Center (ÖSYM). The study used ChatGPT to analyze topics in five core tests: Biology, History, Geography, Social Studies, and Turkish Language and Literature. Among the questions asked as prompts, questions containing figures, graphs, and equations were not included in the study. A total of 210 questions, 42 from each field, were included in the survey in 2021. In the next two years (2022-2023), fewer questions were evaluated, as 10% of the questions in the five main areas were published. As a prompt, we asked ChatGPT 32 questions in 2022 and 37 in 2023. The distribution of questions according to main areas is given in detail in Table 1.

Table 1. Questions and answers for KPSS in 2021, 2022, and 2023

Lessons / KPSS	2021	2022	2023	Total
Biology	42	7	8	57
Geography	42	5	6	53
History	42	4	8	54
Turkish language and literature	42	8	8	58
Social studies	42	8	7	57
Total question	210	32	37	279

In this study, ChatGPT shows that it can answer questions without changing the structure of the KPSS question format. Since the inquiries are routinely disclosed to the public every year, they are considered a public data set, and no institutional permission is required. To comprehensively evaluate the accuracy of the ChatGPT output, the data were standardized in the input formats of KPSS question sets by performing the following operations.

- To develop ChatGPT's Turkish language library, the questions were entered in Turkish as a prompt.
- The main text of the questions is arranged in a single line, and the answer options are arranged in 5 lines.
- All the questions were asked on 01.01.2024.

FINDINGS

ChatGPT's answers to the KPSS2021 questions published as open source are evaluated in this section. Figure 1 shows the correct answer distribution of the 210 questions asked in 5 areas in KPSS-2021 and the graphical representation of the correct answer distribution given by ChatGPT to these 210 questions. While ChatGPT showed the best performance in the fundamental field of Geography with 29 correct answers out of 42 questions, it showed the lowest performance in Social sciences with 15 correct answers out of 42.

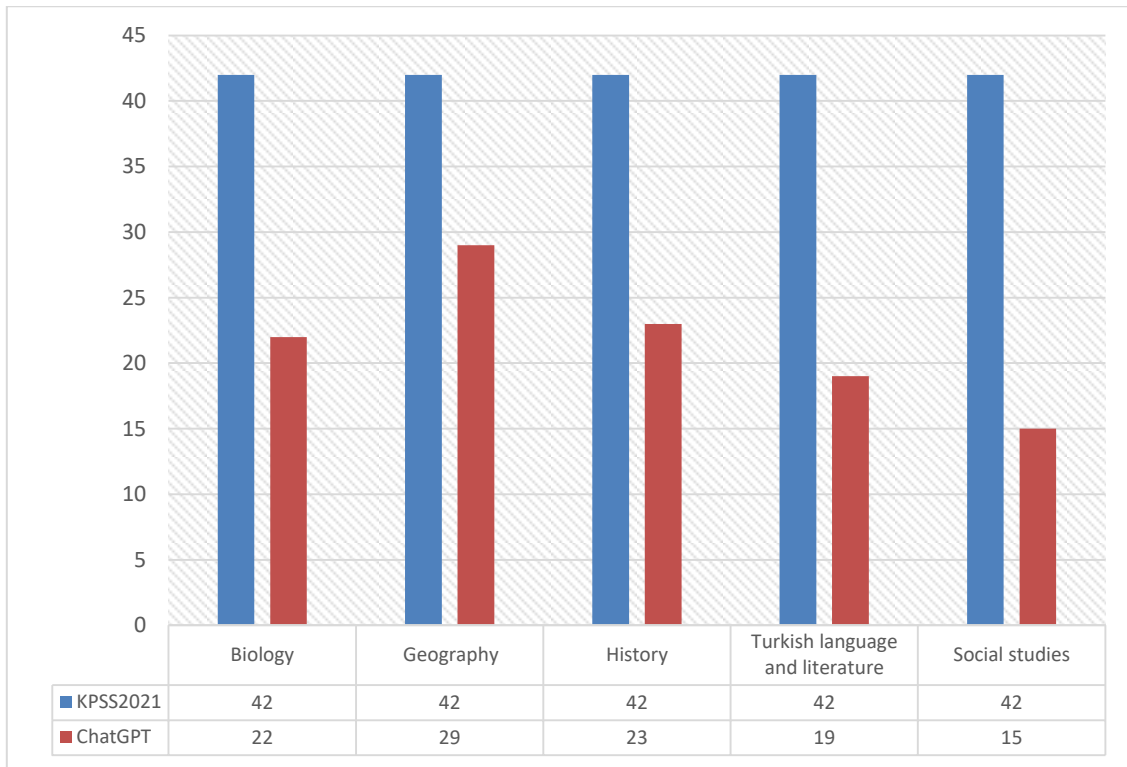


Figure 1. Correct answer distribution of 2021 KPSS questions according to fields and comparison of ChatGPT's correct answers.

According to Table 2, in KPSS2021, where all questions were published, and 42 questions from each test were evaluated, 108 questions (Biology 22; Geography 29, History 23, Turkish Language Literature 19 and Social Studies 15) out of 210 questions written as a prompt to ChatGPT were answered correctly. In this case, ChatGPT has a 51% success rate in the KPSS2021.

Table 2. Distributions of correct answers to KPSS questions achieved from ChatGPT

Lessons / KPSS	2021	2022	2023	Total
Biology	22	4	4	30
Geography	29	4	1	34
History	23	3	1	27
Turkish language and literature	19	2	5	26
Social studies	15	3	4	20
Total correct answers	108	16	15	137
Total Questions	210	32	37	279

Since 10% of the questions were published in the exams held in 2022 and 2023, 32 questions for 2022 and 37 questions for 2023 were asked to ChatGPT as a prompt. In the KPSS2022 exam, a total of 16 questions were answered correctly with ChatGPT, 4 from Biology, 4 from Geography, 3 from History, 2 from Turkish Language Literature, and 3 from Social Studies. In this case, ChatGPT showed a 50% success in KPSS2022. In the KPSS2023 exam, ChatGPT answered 15 questions, 4 from Biology, 1 from Geography, 1 from History, 5 from Turkish Language Literature, and 4 from Social Studies. ChatGPT showed 40% success in KPSS2023. The correct answer distribution of 69 questions published by KPSS in five fundamental tests in 2022 and 2023 and the correct answers given by ChatGPT are shown graphically in Figure 2.

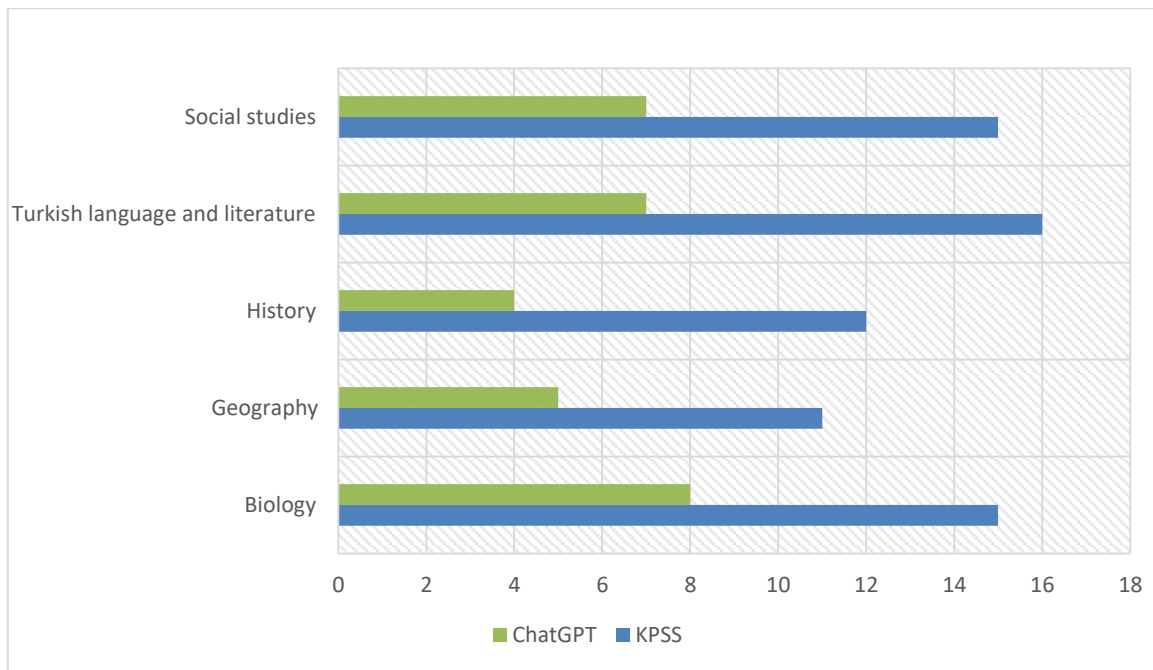


Figure 2. Comparison of correct answers for the questions of the OSYM exams in 2022 and 2023.

Although just 10% of the KPSS questions, rather than all, are published by OSYM, some of these questions involve numbers, and only a few questions are analyzed; moreover, it is possible to assess ChatGPT's replying capacity. Following is a list of the questions from the KPSS in 2022 and 2023 that ChatGPT correctly answered, illustrative of the current circumstance.

- An accuracy rate of 57.14% and 50% was attained in the fundamental biology test, where seven questions in 2022 and 8 questions in 2023 were analyzed.
- The fundamental geography test was assessed with five questions in 2022 and six in 2023. ChatGPT obtained 80% and 16.67% accuracy rates for each test, with four and one correct answer.
- In the history fundamental test, 4 questions in 2022 and 8 questions in 2023 were evaluated, and 3 and 1 correct answer, 75% and 12.50% accuracy rates were obtained by ChatGPT, respectively.
- In the Turkish Language and Literature fundamental test, eight questions were evaluated in 2022 and 2023, and 2 and 5 correct answers, 25% and 62.50% accuracy rates, respectively, were obtained by ChatGPT.
- Finally, in the Social Studies fundamental test, eight questions in 2022 and 7 questions in 2023 were evaluated, and 3 and 4 correct answers, 37.50% and 57.14% accuracy rates were obtained by ChatGPT, respectively.

As a result, in this study, in which 69 questions were used and the correct answer rate of ChatGPT was evaluated, it was more successful in KPSS questions in 2022 than in 2023 in the fundamental Geography, History, and Biology tests. In the fundamental Turkish Language and Literature and Social Studies test, the success of 2023 is higher than the answers given to the questions asked in KPSS in 2022.

RESULT

The use of ChatGPT in the field of education is increasing day by day. For this reason, ChatGPT's Turkish libraries and open data integration must be implemented. The research showed that in the questions asked as a prompt to ChatGPT in the last 3 years, the success rate was 51% in KPSS2021, 50% in KPSS2022, and 40% in KPSS2023. KPSS was offered as an open source in 2021, and it was observed that the success rate increased as more questions were asked. Additionally, when the way of asking questions is changed, success rates increase even more. It is thought that it may be more successful if the questions are asked outside the ÖSYM format and improved with prompt engineering techniques. ChatGPT's low success rate in answering

questions will enable better results by increasing the number of libraries in the electronic environment, improving algorithms, and prompt writing skills.

Some advice for users about Prompt engineering:

- The more detailed and descriptive your prompts are; the better results will be achieved.
- A single word, such as a keyword in the prompts, does not give good results. Therefore, you can provide some background information to set the context of the conversation.
- After giving the topic, you can ask it to outline it. You can print the text on the subject containing the concepts you want in detail, tone, and length.
- You can assign this task to a business analyst, teacher, or academician in a prompt you have written about a particular subject.
- You can rewrite the prompts in different ways. If it misinterprets or gives wrong information, it will try to correct it yourself when you point this out.
- Chatbots, generative artificial intelligence (GENAI) tools, are still in their infancy. Therefore, it will be able to provide better results by improving its infrastructure and library day by day.

DISCUSSION

GenAI tools are increasing day by day. It is an AI tool that generates content in various formats, from pictures to music, from text to coding. The development of ChatGPT, which uses GenAI algorithms, has undoubtedly been a significant milestone in artificial intelligence and NLP. ChatGPT, which is also a great language model, responds to users' prompts more effectively day by day with its new versions. Continuous improvement of ChatGPT will also likely impact the future of AI. This Chatbot in education offers personalized learning materials and the ability to answer many exam-related questions. Therefore, ChatGPT effectively improves students' overall learning experience using advanced NLP capabilities, providing more generative and exciting responses. Additionally, ChatGPT's ability to generate explanations and summaries can also be used to help students understand complex learning materials more efficiently. However, for any of these applications to be helpful, ChatGPT must perform tests of medical judgment and knowledge in a manner comparable to that of human experts. Only then can users trust the solutions offered by ChatGPT.

It is a problem of GenAI tools, and using Chatbots such as ChatGPT in education can cause ethical violations. When using this Chatbot in education, users need to evaluate their tasks, such as; thesis, articles, homework, and exams within the framework of ethical principles. Ethical problems can be prevented by users expressing their thoughts within the scope of ethical principles and using their interpretation skills instead of directly using the answers provided by Chatbots in education. Ethical violations can be prevented if users comply with the ethical guidelines published for Chatbot use and continue their work within this scope.

Declarations

Conflict of Interest

No potential conflicts of interest were disclosed by the author(s) concerning this article's research, authorship, or publication.

Contribution Rates of Authors to the Article

These authors contributed equally to this work.

Acknowledgements

This study is supported by the Gazi University AI R&D Center.

REFERENCES

- Awasthi, I., Gupta, K., Bhogal, P. S., Anand, S. S., & Soni, P. K. (2021). Natural language processing (NLP) based text summarization-a survey. 2021 6th International Conference on Inventive Computation Technologies (ICICT). 10.24926/iip.v13i2.4839. <https://doi.org/10.1109/ICICT50816.2021.9358703>
- Biswas, S. (2023). Role of Chat GPT in Education. Available at SSRN 4369981. (Web page: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4369981)
- Bommarito II, M., & Katz, D. M. (2022). GPT takes the bar exam. *arXiv preprint arXiv:2212.14402*. <https://doi.org/10.48550/arXiv.2212.14402>
- Bommarito, J., Bommarito, M., Katz, D. M., & Katz, J. (2023). Gpt as knowledge worker: A zero-shot evaluation of (ai) cpa capabilities. *arXiv preprint arXiv:2301.04408*. <https://doi.org/10.48550/arXiv.2301.04408>
- 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? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9(1), e45312. <https://doi.org/10.2196/45312>
- Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, 13(7), 692. <https://doi.org/10.3390/educsci13070692>
- Guo, K., Yang, Z., Yu, C.-H., & Buehler, M. J. (2021). Artificial intelligence and machine learning in design of mechanical materials. *Materials Horizons*, 8(4), 1153-1172. <https://doi.org/10.1039/D0MH01451F>
- Hardy, M., & Harvey, H. (2020). Artificial intelligence in diagnostic imaging: impact on the radiography profession. *The British journal of radiology*, 93(1108), 20190840. <https://doi.org/10.1259/bjr.20190840>
- Khanzode, K. C. A., & Sarode, R. D. (2020). Advantages and disadvantages of artificial intelligence and machine learning: A literature review. *International Journal of Library & Information Science (IJLIS)*, 9(1), 3. <https://doi.org/10.17605/OSF.IO/GV5T4>
- King, M. R. (2023). The future of AI in medicine: a perspective from a Chatbot. *Annals of Biomedical Engineering*, 51(2), 291-295. <https://doi.org/10.1007/s10439-022-03121-w>
- Kocaoğlu, M., & Özmen, Z. N. (2019). Türkiye'deki kamu yönetimi eğitimi ve personel istihdam etme usullerinin kamu personeli seçme sınavı (KPSS) üzerinden değerlendirilmesi. *Erciyes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 33(46), 241-266. (Web page: <https://dergipark.org.tr/tr/pub/erusosbilder/issue/46591/527487>)
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1-35. <https://doi.org/10.48550/arXiv.2107.13586>
- Sarbabidya, S., & Saha, T. (2020). Role of chatbot in customer service: A study from the perspectives of the banking industry of Bangladesh. *International review of business research papers*, 16(1), 231-248. (Web page: <https://zantworldpress.com/product/role-of-chatbot-in-customer-service-a-study-from-the-perspectives-of-the-banking-industry-of-bangladesh/>)
- Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. Available at SSRN 4378735. <https://doi.org/10.62037/cjer.2023.03.01.06>