



ORIGINAL ARTICLE

Exploring the potential of ChatGPT to create multiple-choice question exams

Cristian N. Rivera-Rosas*, J.R. Tadeo Calleja-López, Enrique Ruibal-Tavares, Arturo Villanueva-Neri, Cinthya M. Flores-Felix, Sergio Trujillo-López

Department of Medicine and Health Sciences, University of Sonora, Hermosillo, Sonora, México

Received 29 December 2023; accepted 3 April 2024

KEYWORDS

Artificial intelligence;
Medical education;
Multiple-choice
questions

Abstract

Introduction: This study aims to describe our experience using ChatGPT to create exam multiple-choice questions (MCQs) and describe the students' opinions.

Material and methods: We used ChatGPT to create 55 MCQ. Furthermore, we conducted a questionnaire to evaluate students' perceptions.

Results: 89% of the students considered the questions concise and comprehensible; 91% mentioned language was clear; 76% described it as "simple".

Conclusion: ChatGPT has the potential to support teachers in generating MCQ and students' perception of the test of syntactic comprehension of the exam was positive.

© 2024 The Author(s). Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

PALABRAS CLAVE

Inteligencia artificial;
educación médica;
exámenes de opción
múltiple

Explorando el potencial de ChatGPT para crear exámenes de opción múltiple

Resumen

Introducción: El objetivo de este estudio es describir nuestra experiencia usando ChatGPT para crear un examen de opción múltiple y describir la opinión de los estudiantes.

Material y métodos: Usamos ChatGPT para crear 55 preguntas. Posteriormente, realizamos un cuestionario para evaluar la percepción del estudiante.

Resultados: 89% de los estudiantes consideraron las preguntas concisas y comprensibles; 91% que el lenguaje era claro y 76% «sencillo».

Conclusión: ChatGPT tiene el potencial de apoyar a docentes en la generación de exámenes de opción múltiple y la percepción de los estudiantes sobre la comprensión sintáctica del examen fue positiva.

* Corresponding author at: Cristian N. Rivera-Rosas, Department of Medicine and Health Sciences, University of Sonora, Blvd. Luis D. Colosio SN, col. Centro, Hermosillo CP 83000, Sonora, México.

E-mail address: md.cristian.rivera@gmail.com (C.N. Rivera-Rosas).

Introduction

The demands of fulfilling teaching responsibilities, along with increasing workloads in healthcare, result in detrimental effects on the labor of healthcare personnel who attempt to carry out these multiple tasks simultaneously, including the augmented risk of burnout syndrome, stress, and reduced job performance.^{1,2} Therefore, with the increasing workload, it is crucial to acknowledge the clear need to implement or develop tools that can assist healthcare personnel in managing said responsibilities. This is where AI-powered tools in medical education, such as ChatGPT, come into play, as they can be a helpful resource to mitigate some of the negative effects of workload in healthcare academics.

Previous studies have shown promising outcomes of large language models (LLM) like ChatGPT in the creation of exercises and quizzes,³ as well as generating high-quality MCQs that are comparable to those created by expert teachers,⁴ which are activities that can demand a lot of time from professors. However, it is still important to explore the student's perception of the questions generated by these tools in real settings. The current study aims to describe our experience using ChatGPT-3.5 to create MCQ exams and the students' opinions at a university in Mexico.

Methods

We used ChatGPT-3.5 to create an MCQ exam of 50 questions for the anatomy and kinesiology course in one group of 33 students of the physical culture and sports degree program. The exam topics were cellular biology and muscular and skeletal systems. Three different commands were used to request ChatGPT-3.5 to create the questions (Table 1). Since the prompts were originally written in Spanish, we relied on ChatGPT-3.5 to translate both the prompts and the MCQs into English. A total of 55 questions were created (Table 2). Furthermore, the professor designed the exam's final version, and no changes were made to the questions created

by ChatGPT-3.5. Two external professors reviewed the exam before its application. In accordance with the Delphi process, we developed a satisfaction questionnaire employing a Likert scale to evaluate the assessment's redaction, clarity, coherence, and simplicity. Additionally, we inquired about the pretest preparation of the alumni (Table 3), subsequently we administered it using a Google Form. The results of both the exam and the perception survey were analyzed with Google Sheets and Google Forms. None of the students knew that ChatGPT-3.5 was used in exam creation.

Results

From the 55 questions created, the professor reviewed and selected 50. The remaining questions were eliminated for being either redundant or outside the focus of the course. Regarding the students' performance, the interval of correct answers was between 17 and 44 points out of 50 possible and a mean score of 26/50. Concerning the quality of the test measured using the Likert scale, 89% of the students considered the questions to be concise and the writing comprehensible; 91% mentioned that the language was clear, while on the other hand, 76% of the students described the language as "simple". Finally, 40% of participants thought they studied enough, and 24% referred to the questions as difficult. For each command, the time needed for ChatGPT-3.5 to create the questions was approximately 10–15 s, whereas the time invested by the teacher for question selection and test final design was close to 60 min.

Discussion

Our results are consistent with those of previous studies, in which ChatGPT demonstrated the capability to support teachers in generating MCQ.⁵ Also, a probable capability to decrease the time and effort required for exam preparation by the professor was observed. This could contribute to decreasing the workload and thereby positively impacting

Table 1 Topics included in the exam and prompts given to ChatGPT-3.5 to generate MCQs.

| Topic name | Prompts used to MCQs generation |
|------------------|--|
| Cellular biology | Imagine you are a professor from the course of anatomy and kinesiology of the degree in physical culture and sports at a university in Mexico. Create 15 MCQ about cellular biology, each question should have 4 options and display me the correct answer. |
| Muscular system | Imagine you are a professor from the course of anatomy and kinesiology of the degree in physical culture and sports at a university in Mexico. Create 20 MCQ about the muscular system, each question should have 4 options and display me the correct answer. |
| Skeletal system | Imagine you are a professor from the course of anatomy and kinesiology of the degree in physical culture and sports at a university in Mexico. Create 20 MCQ about the skeletal system, each question should have 4 options and display me the correct answer. |

MCQ, multiple choice questions.

Table 2 MCQs generated by ChatGPT-3.5.

| Topic | Questions |
|------------------|---|
| Cellular biology | Which of the following organelles is known as the “control center” of the cell? (a) Ribosomes (b) Nucleus (c) Lysosomes (d) Mitochondria Correct answer: (b) Nucleus |
| | What organelle is responsible for protein synthesis in the cell? (a) Lysosomes (b) Rough endoplasmic reticulum (c) Golgi apparatus (d) Vacuoles Correct answer: (b) Rough endoplasmic reticulum |
| | Which of the following statements about eukaryotic cells is true? (a) They have no defined nucleus. (b) Their genetic material is found in the cytoplasm. (c) They have a nuclear membrane. (d) They lack mitochondria. Correct answer: (c) They have a nuclear membrane |
| | What is the main function of ribosomes in the cell? (a) Storing energy (b) Performing photosynthesis (c) Synthesis of lipids (d) Synthesis of proteins Correct answer: (d) Synthesis of proteins |
| | ^a Which of the following structures is common in plant cells but not in animal cells? (a) Lysosomes (b) Cell wall (c) Centrioles (d) Smooth endoplasmic reticulum Correct answer: (b) Cell wall |
| | What organelle is known as the “powerhouse” of the cell? (a) Nucleus (b) Lysosomes (c) Mitochondria (d) Golgi apparatus Correct answer: (c) Mitochondria |
| | Which of the following organelles is involved in the degradation of cellular waste and recycling of materials? (a) Nucleus (b) Ribosomes (c) Lysosomes (d) Mitochondria Correct answer: (c) Lysosomes |
| | What is the main function of the Golgi apparatus in the cell? (a) Protein synthesis (b) Protein transport (c) Water storage (d) Nutrient digestion Correct answer: (b) Protein transport |
| | Which of the following cell structures is responsible for regulating the passage of substances into and out of the cell? (a) Nucleus (b) Cell membrane (c) Ribosomes (d) Rough endoplasmic reticulum Correct answer: (b) Cell membrane |
| | What function does the cytoplasm serve in the cell? (a) Storing genetic material (b) Carrying out protein synthesis |

Table 2 (continued)

| Topic | Questions |
|-----------------|--|
| | <p>(c) Providing structural support</p> <p>(d) Controlling the passage of substances into the nucleus</p> <p>Correct answer: (c) Providing structural support</p> <p>What type of cell lacks a nucleus and other membranous organelles?</p> <p>(a) Eukaryotic cell</p> <p>(b) Plant cell</p> <p>(c) Animal cell</p> <p>(d) Prokaryotic cell</p> <p>Correct answer: (d) Prokaryotic cell</p> <p>^aWhat organelle is responsible for photosynthesis in plant cells?</p> <p>(a) Mitochondria</p> <p>(b) Rough endoplasmic reticulum</p> <p>(c) Chloroplasts</p> <p>(d) Lysosomes</p> <p>Correct answer: (c) Chloroplasts</p> <p>What is the main function of peroxisomes in the cell?</p> <p>(a) Protein synthesis</p> <p>(b) Lipid metabolism</p> <p>(c) Substance transport</p> <p>(d) Energy storage</p> <p>Correct answer: (b) Lipid metabolism</p> <p>^aWhat organelle is responsible for intracellular digestion and degradation of macromolecules?</p> <p>(a) Mitochondria</p> <p>(b) Smooth endoplasmic reticulum</p> <p>(c) Lysosomes</p> <p>(d) Ribosomes</p> <p>Correct answer: (c) Lysosomes</p> <p>What cell structure regulates vesicle trafficking and protein transport in the cell?</p> <p>(a) Golgi apparatus</p> <p>(b) Nucleus</p> <p>(c) Ribosomes</p> <p>(d) Lysosomes</p> <p>Correct answer: (a) Golgi apparatus</p> <p>What is the largest muscle in the human body?</p> <p>(a) Biceps</p> <p>(b) Quadriceps</p> <p>(c) Deltoid</p> <p>(d) Trapezius</p> <p>Correct answer: (b) Quadriceps</p> <p>Which of the following muscles is responsible for elbow flexion?</p> <p>(a) Brachialis</p> <p>(b) Triceps brachii</p> <p>(c) Deltoid</p> <p>(d) Pectoralis major</p> <p>Correct answer: (a) Brachialis</p> <p>What muscle is located in the back of the thigh and is important for hip extension?</p> <p>(a) Rectus abdominis</p> <p>(b) Gluteus maximus</p> <p>(c) Quadriceps</p> <p>(d) Trapezius</p> <p>Correct answer: (b) Gluteus maximus</p> <p>Which of the following muscles is located in the shoulder region and is responsible for arm abduction?</p> <p>(a) Deltoid</p> <p>(b) Pectoralis major</p> <p>(c) Trapezius</p> <p>(d) Rectus abdominis</p> <p>Correct answer: (a) Deltoid</p> |
| Muscular system | |

Table 2 (continued)

| Topic | Questions |
|-------|--|
| | <p>^aWhat muscle is located in the front of the thigh and is important for hip flexion?</p> <p>(a) Tibialis anterior (b) Biceps femoris (c) Quadriceps (d) Gastrocnemius Correct answer: (c) Quadriceps</p> <p>Which of the following muscles is responsible for wrist extension?</p> <p>(a) Triceps brachii (b) Brachialis (c) Deltoid (d) Pectoralis major Correct answer: (a) Triceps brachii</p> <p>What muscle is located in the front of the neck and is responsible for head flexion?</p> <p>(a) Trapezius (b) Sternocleidomastoid (c) Deltoid (d) Pectoralis major Correct answer: (b) Sternocleidomastoid</p> <p>Which of the following muscles is a calf muscle and is used for plantar flexion of the foot?</p> <p>(a) Tibialis anterior (b) Biceps femoris (c) Quadriceps (d) Gastrocnemius Correct answer: (d) Gastrocnemius</p> <p>What muscle is located in the back of the arm and is responsible for elbow extension?</p> <p>(a) Brachialis (b) Triceps brachii (c) Deltoid (d) Pectoralis major Correct answer: (b) Triceps brachii</p> <p>Which of the following muscles is located in the chest and is responsible for arm flexion?</p> <p>(a) Deltoid (b) Trapezius (c) Pectoralis major (d) Rectus abdominis Correct answer: (c) Pectoralis major</p> <p>What muscle is located in the back of the leg and is responsible for knee flexion?</p> <p>(a) Rectus abdominis (b) Gluteus maximus (c) Quadriceps (d) Hamstrings Correct answer: (d) Hamstrings</p> <p>Which of the following muscles is located in the abdominal region and is important for trunk stabilization?</p> <p>(a) Brachialis (b) Triceps brachii (c) Deltoid (d) Rectus abdominis Correct answer: (d) Rectus abdominis</p> <p>What muscle is located in the back of the shoulder and is responsible for external rotation of the arm?</p> <p>(a) Deltoid (b) Pectoralis major (c) Trapezius (d) Infraspinatus Correct answer: (d) Infraspinatus</p> <p>Which of the following muscles is located in the upper arm and is responsible for forearm supination?</p> <p>(a) Pronator teres (b) Biceps brachii (c) Triceps brachii</p> |

Table 2 (continued)

| Topic | Questions |
|-----------------|---|
| Skeletal system | (d) Deltoid Correct answer: (b) Biceps brachii What muscle is located in the groin region and is responsible for hip flexion? |
| | (a) Gluteus maximus (b) Rectus abdominis (c) Sartorius (d) Trapezius Correct answer: (c) Sartorius |
| | Which of the following bones is part of the vertebral column? |
| | (a) Femur (b) Tibia (c) Vertebra (d) Humerus Correct answer: (c) Vertebra |
| | What is the longest bone in the human body? |
| | (a) Radius (b) Ulna (c) Femur (d) Tibia Correct answer: (c) Femur |
| | Which of the following bones forms the back part of the skull? |
| | (a) Mandible (b) Scapula (c) Occipital (d) Maxilla Correct answer: (c) Occipital |
| | Which of the following bones forms the front part of the rib cage? |
| | (a) Sternum (b) Humerus (c) Femur (d) Radius Correct answer: (a) Sternum |
| | What bone is known as the "arm bone" and is located between the shoulder and the elbow? |
| | (a) Humerus (b) Radius (c) Ulna (d) Clavicle Correct answer: (a) Humerus |
| | ^a Which of the following bones forms the shoulder joint? |
| | (a) Femur (b) Radius (c) Scapula (d) Tibia Correct answer: (c) Scapula |
| | What bone forms the lower part of the vertebral column? |
| | (a) Cervical vertebra (b) Lumbar vertebra (c) Thoracic vertebra (d) Sacrum Correct answer: (d) Sacrum |
| | Which of the following bones is part of the hand and is located in the wrist? |
| | (a) Femur (b) Radius (c) Scapula (d) Carpals Correct answer: (d) Carpals |
| | What bone forms the top of the skull and is known as the "cranial vault"? |
| | (a) Temporal |

Table 2 (continued)

| Topic | Questions |
|-------|---|
| | (b) Parietal (c) Occipital (d) Maxilla Correct answer: (b) Parietal Which of the following bones is part of the rib cage and connects to the ribs? (a) Humerus (b) Vertebra (c) Sternum (d) Femur Correct answer: (c) Sternum Which of the following bones forms the upper part of the vertebral column and the neck? (a) Cervical vertebra (b) Lumbar vertebra (c) Thoracic vertebra (d) Sacrum Correct answer: (a) Cervical vertebra Which of the following bones forms the lower part of the leg and is known as the "shinbone"? (a) Femur (b) Radius (c) Tibia (d) Ulna Correct answer: (c) Tibia What bone is known as the "calf bone" and is located alongside the tibia in the leg? (a) Femur (b) Radius (c) Tibia (d) Fibula Correct answer: (d) Fibula Which of the following bones is part of the hand and forms the fingers? (a) Humerus (b) Radius (c) Ulna (d) Phalanges Correct answer: (d) Phalanges Which of the following bones is located at the back of the foot and forms the heel? (a) Calcaneus (b) Astragalus (c) Tibia (d) Humerus Correct answer: (a) Calcaneus |

^a Questions that were eliminated

their performance as educators and healthcare workers. Nonetheless, even though we found benefits using this AI in education, our sample is too small to confirm its true utility, and more studies are needed to determine if a significant difference is observed between exams done by professors and ChatGPT and the professors' perceptions in the usage of AI in academia and their potential advantages related to diminishing workload.

Overall, according to the satisfaction survey in the present study, a generally favorable perception was observed concerning syntactic comprehension of the exam. Despite the discussion about the potential utilities of ChatGPT in education and the known capability of ChatGPT to generate text, understand context, and be versatile.⁶ Its acceptance in the classrooms as a tool for students for

studying or as a supporting tool for their teachers in the development of class material or assessments continues to be debated. Although, according to Tangadulrat et al.,⁷ students perceive that the use of ChatGPT has a positive effect on their education. Both the acceptance by students and the documented benefits and capabilities of this LLM could give more confidence and motivation to healthcare academics in the usefulness of this AI as a novel support tool for classes and therefore serve as a beneficial aid to alleviate the workload related to academia tasks. Finally, this positive attitude by students and professors could also facilitate the integration of AI-related subjects into the health curricula of universities.

However, it's important to note that professors remain responsible for verifying the validity of the questions as

Table 3 Satisfaction questionnaire used for students' perception about the exam.

| Question made | Rating (1–5) |
|---|----------------------|
| Overall, I consider that the exam was well-written. | 1. Strongly disagree |
| The language used in the exam was clear. | 2. Disagree |
| The language used in the exam was confusing. | 3. Neutral |
| The language used in the exam was concise. | 4. Agree |
| The language used in the exam was coherent. | 5. Strongly agree |
| The language used in the exam was simple. | |
| The questions of the exam were difficult. | |
| I consider that I studied enough for this exam. | |

ChatGPT can create duplicated questions or “AI hallucinations”.^{5,8} Also, in further studies, the quality of the questions created by this LLM should be measured with a validated instrument, as well as the accuracy of the prompts used by educators for they will impact the quality, complexity, and ambiguity of the generated questions,⁹ as seen in the present study.

Conclusion

Our results describe promising results that ChatGPT can efficiently generate quality evaluation instruments that are understandable and clear to students, saving time and effort for the evaluator. This suggests that LLMs can be a potential resource to enhance education on both sides while at the same time positively impacting healthcare workers' performance by reducing their workload. However, studies with more rigorous methodology and larger sample sizes are needed to know the true scope of such a tool.

Ethics disclosure

The authors declare that no experiments were performed on humans or animals for this study. The authors declare that no patient data appear in this article. The authors declare that no patient data appear in this article.

Funding

This research was funded by the Department of Medicine and Health Sciences from the University of Sonora, Hermosillo.

Declaration of competing interest

The authors declare no conflicts of interest.

References

1. Nassar AK, Reid S, Kahnamoui K, Tuma F, Waheed A, McConnell M. Burnout among academic clinicians as it correlates with workload and demographic variables. *Behav Sci (Basel)*. 2020;10(6):94. Available in: <https://doi.org/10.3390/bs10060094>.
2. Nassar AK, Waheed A, Tuma F. Academic clinicians' workload challenges and burnout analysis. *Cureus*. 2019 Nov 8;11(11):e6108. <https://doi.org/10.7759/cureus.6108>.
3. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - reshaping medical education and clinical management. *Pak J Med Sci*. 2023;39(2):605–7. <https://doi.org/10.12669/pjms.39.2.7653>.
4. Cheung BHH, Lau GKK, Wong GTC, Lee EYP, Kulkarni D, Seow CS, et al. ChatGPT versus human in generating medical graduate exam multiple choice questions-a multinational prospective study (Hong Kong S.A.R., Singapore, Ireland, and the United Kingdom). *PLoS One*. 2023;8(8):e0290691. <https://doi.org/10.1371/journal.pone.0290691>.
5. Khilnani AK. Potential of large language model (ChatGPT) in constructing multiple choice questions. *GAIMS J Med Sci*. 2023, 3;3(2):1–3. Available in: <https://gjms.gaims.ac.in/index.php/gjms/article/view/71>.
6. Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*. 2023;3:121–54. <https://doi.org/10.1016/j.iotcps.2023.04.003>.
7. Tangadulrat P, Sono S, Tangtrakulwanich B. Using ChatGPT for clinical practice and medical education: cross-sectional survey of medical students' and physicians' perceptions. *JMIR Med Educ*. 2023 Dec 22;9:e50658. <https://doi.org/10.2196/50658>.
8. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023, 3;11(6):887. <https://doi.org/10.3390/healthcare11060887>.
9. Giray L. Prompt engineering with ChatGPT: a guide for academic writers. *Ann Biomed Eng*. 2023 Dec;51(12):2629–33. <https://doi.org/10.1007/s10439-023-03272-4>.