The Journal of Turkish Dental Research
Türk Diş Hekimliği Araştırma Dergisi

e-ISSN: 2822-4310 | Yıl / Year: 2025 | Cilt – Volume: 4 | Sayı – Issue: 1

Can ChatGPT-3.5 Be Used to Answer Questions Related to Pediatric Dentistry?

ChatGPT-3.5 Çocuk Diş Hekimliği Alanı ile İlgili Soruları Yanıtlamak İçin Kullanılabilir mi?

ChatGPT-3.5 and Pediatric Dentistry

Şeyma MUSTULOĞLU

Doç. Dr., Department of Paediatric Dentistry, Faculty of Dentistry, Mersin University, Mersin, Türkiye
dtseymaozturk@gmail.com
ORCID: 0000-0001-9796-4348

Author contributions : Conceptualisation: Seyma Mustuloglu; Methodology: Seyma Mustuloglu; Formal analysis and investigation: Seyma Mustuloglu; Writing-original draft preparation: Seyma Mustuloglu; Writing-review and editing: Seyma Mustuloglu; Supervision: Seyma Mustuloglus.

Acknowledgements The authors wish to thank all participants for their invaluable contribution to this study.

Makale Bilgisi / Article Information
Makale Türü / Article Types: Araştırma Makalesi / Research Article
Geliş Tarihi / Received: 09-07-2024
Kabul Tarihi / Accepted: 12-03-2025
Sayfa / Pages: 12-18

Sorumlu Yazar / Corresponding Author: Şeyma MUSTULOĞLU

DOI: https://doi.org/10.58711/turkishjdentres.vi.1513154

Can ChatGPT-3.5 Be Used to Answer Questions Related to Pediatric Dentistry?

ChatGPT-3.5 Çocuk Diş Hekimliği Alanı ile İlgili Soruları Yanıtlamak İçin Kullanılabilir mi?

ABSTRACT

Aim: This study aimed to evaluate and compare the quality of chatGPT-3.5's responses to frequently asked and undergraduate education-related questions provided by chatGPT-3.5 in the field of pediatric dentistry.

Materials and Methods: Frequently asked questions by patients about procedures in pediatric dentistry and questions for educational purposes were generated to be asked on ChatGPT-3.5. The questions were divided into two categories consisting of 14 Patient Questions and 18 Educational Questions. Specific topics covered: Oral hygiene and preventive dentistry, patient management, dental trauma, restorative dentistry and pulp therapy, orthodontics, dental pathology and dental radiology. ChatGPT-3.5 answered these questions on April 15, 2024. Responses were scored on a 1 to 5 scale by pediatric dentists using a modified global quality scale. All statistical calculations were performed using IBM SPSS Statistics version 25.0.

Results: Responses were scored by 25 pediatric dentists. The mean score for responses to Patient Questions was 4.21±0.53, while the mean score for responses to Education Questions was 3.44±0.71. The difference in scores between the two question categories and the subgroups of the categories was found to be statistically significant ($p<0.001$).

Conclusion: GPT-3.5 provided more reliable information on frequently asked questions by patients in pediatric dentistry than on undergraduate education-related questions. Beyond being used for clinicians, ChatGPT may be more appropriate for patients to obtain general information in the field of pediatric dentistry

Anahtar Kelimeler: Artificial intelligence; Chatbot; Chatgpt; Pediatric dentistry.

ÖZET

Amaç: Bu çalışmada çocuk diş hekimliği alanı ile ilgili sık sorulan ve lisans eğitimi ile ilgili sorulara chatGPT-3.5'in sunduğu yanıtların kalitesinin değerlendirilmesi ve karşılaştırılması amaçlandı.

Gereç ve Yöntem: Hastaların çocuk diş hekimliğindeki işlemlerle ilgili sıklıkla sorduğu sorular ve eğitim amaçlı sorular ChatGPT-3.5 üzerinden sorulmak üzere oluşturulmuştur. Sorular 14 Hasta Sorusu ve 18 Eğitim Sorusu olmak üzere iki kategoriye ayrıldı. Belirlenen spesifik konular: Ağız hijyeni ve koruyucu diş hekimliği, hasta yönetimi, diş travması, restoratif diş hekimliği ve pulpa tedavisi, ortodonti, diş patolojisi ve diş radyolojisi idi. ChatGPT-3.5, bu soruları 15 Nisan 2024'te yanıtladı. Yanıtlar, çocuk diş hekimleri tarafından modifiye edilmiş bir küresel kalite ölçeği kullanılarak 1 ila 5 arasında bir ölçekte puanlandı. Tüm istatistiksel hesaplamalar IBM SPSS Statistics sürüm 25.0 kullanılarak gerçekleştirildi.

Bulgular: Yanıtlar 25 pediatrik diş hekimi tarafından puanlandı. Hasta Sorularına verilen yanıtların ortalama puanı 4,21±0,53 iken Eğitim Sorularına verilen yanıtların ortalama puanı 3,44±0,71 idi. İki soru kategorisi ve kategorilerin alt grupları arasındaki puan farkı istatistiksel olarak anlamlı bulunmuştur ($p<0,001$).

Sonuç: GPT-3.5, çocuk diş hekimliği hastalarında sıklıkla sorulan sorular hakkında lisans eğitimi ile ilgili sorulara göre daha güvenilir bilgi sağlamıştır. ChatGPT, klinisyenlerin kullanımı dışında, hastaların çocuk diş hekimliği alanında genel bilgi edinmeleri için daha uygun olabilir.

Keywords: Yapay zeka; Sohbet robotu; Chatgpt; Çocuk diş hekimliği.

DOI: https://doi.org/10.58711/turkishjdentres.vi.1513154

Introduction

The use of an artificial intelligence (AI) program called ChatGPT-3.5, which creates text in response to written instructions over the web via OpenAI (San Francisco, CA, USA), has become widespread in recent years. Using natural language processing (NLP), ChatGPT-3.5 understands and interprets people's spoken language and predicts subsequent items from previous natural language texts to produce responses that resemble human responses. However, it has been noticed that there is a risk of giving wrong answers even to basic problems and that it has some limitations.[1]

AI-based chatbots have great potential to provide relevant information to patients and to be used in health education. However, research on the application of ChatGPT-3.5 in dentistry is limited. So far, its performance in answering dental questions related to endodontics[2], surgery[3], public health4, and pediatric dentistry[5] has been evaluated. However, comprehensive studies on pediatric dentistry have not been conducted yet.

Answering parents' questions correctly in pediatric dentistry is of great importance in improving the oral health of children. Since it is not always possible for parents to consult pediatric dentists, revealing the advantages and disadvantages of artificial intelligence applications in pediatric dentistry can make important contributions to this issue. Therefore, this study aims to measure the frequently asked questions of patients and the quality of educational information about pediatric dentistry provided by ChatGPT-3.5 and evaluate the usability of the information generated, and the performance of chatGBT. The null hypothesis (H0) of this study is that there is no difference in the quality of responses to frequently asked questions in pediatric dentistry and undergraduate education questions provided by ChatGPT-3.5. The H1 hypothesis of this study is that ChatGPT-3.5 provides higher quality answers to frequently asked questions in pediatric dentistry than it does to undergraduate education questions.

Materials and Methods

2.1. Study design and participants

This cross-sectional study was reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.6 Informed consent was obtained from all participants. Ethics committee approval was not required as no patient data was used, and clinicians self-enrolled.

Common questions that patients typically ask about Pediatric Dentistry, as well as questions that could be utilized for training purposes, were identified to be posed to the artificial intelligence-based chatbot (ChatGPT-3.5). To create frequently asked pediatric dentistry questions, searches were conducted on Google ("Frequently Asked Questions About Pediatric Dentistry"). The search yielded 39 questions. A pediatric dentist with five years of experience then evaluated the questions for similarity and scientific aspects. As a result, six very recurring (some questions regarding dental caries and causative factors were similar) questions and three non-scientific questions (e.g., cost). were excluded from the study. Thirty two questions covering oral hygiene and preventive dentistry, patient management, dental trauma, restorative dentistry, orthodontics, dental pathology and dental radiology were obtained. These questions were categorized as Patient Questions and Training Questions (Table 1). Professional textbooks were used by the researcher (Ş.M.) to determine the questions about education and determine the correct answers to all questions. Questions arose covering topics such as oral hygiene and preventive dentistry, patient management, dental trauma and pulp therapy, restorative dentistry, orthodontics, dental pathology and dental radiology (Table 1).

ChatGPT-3.5 was accessed on April 15, 2024, with a new account opened for this study. The questions were manually written into a new chat window for each question. The questions asked, and responses provided by ChatGPT-3.5 were recorded in a Google Form to create an evaluation survey. Descriptive information about pediatric dentists, such as gender, age, professional experience and experiences with artificial intelligence, were also added to the survey. Finally, two questions were added to the survey about whether ChatGPT-3.5 could be used for patient information and whether it would be useful in undergraduate dentistry education. Experienced pediatric dentists were contacted via e-mail to evaluate the survey. To evaluate the survey, 70

experienced pediatric dentists with contact information were contacted via e-mail. 25 (%35) pediatric dentists responded to the e-mail. Pediatric dentists were asked to score the questions and answers in the survey between 1 (The answer is correct, and the content is comprehensive) and 5 (The answer and the entire content are incorrect or irrelevant.) using the modified version of the Global Quality Score (GQS).[3] It is a 5-point scale (1–5) that measures the reliability, quality, and usefulness of responses. A score of 4 or 5 indicates high quality, a score of 3 indicates moderate quality, and a score of 1 or 2 indicates low quality.[3]

2.1. Statistical Analysis

All statistical calculations were performed using IBM SPSS Statistics version 25.0. The Shapiro-Wilk test and histograms were used to test the normality of data. Continuous data were presented as mean and standard deviation, while categorical data were presented as frequency and ratio. The chi-square or Fisher's exact test was used to evaluate the relationship for categorical variables. To compare the means of normally distributed data, a paired-samples t-test was used for dependent groups and an independent samples test was used for independent groups. A p-value of less than 0.05 was considered statistically significant.

3. Results

A set consisting of 14 questions frequently asked by patients and 18 questions within the scope of undergraduate dentistry education was asked to ChatGPT-3.5.

A total of 25 pediatric dentists agreed to participate in the study. Sixteen of the participants were female, and the mean age was 32.52. While 19 (76%) of the participants had no experience with artificial intelligence, 6 (24%) had experience. Descriptive characteristics of pediatric dentists are presented in Table 2.

Table 3 evaluates the mean scores of answers to patient questions and answers to education-related questions according to the modified version of the GQS. While the mean score of the responses to patient questions was 4.21±0.53, the mean score of the responses to questions about education was 3.44±0.71. The difference between the two groups was statistically significant ($p<0.001$).

Table 4 shows the results of evaluating the difference between the answers given according to categories. Accordingly, the mean scores of ChatGPT's answers to patient questions were significantly higher in all subgroups ($p<0.001$).

Table 5 shows the ChatGPT-3.5 evaluation results of pediatric dentists. After the scoring, 10 (40%) of the participants stated that they would definitely use ChatGPT-3.5, while no participants said that they would absolutely not use it. In addition, while more than half of the participants thought that ChatGPT-3.5 would be useful if it was further developed in pediatric dentistry education, no participants thought that it was absolutely not useful.

4. Discussion

Today, artificial intelligence chat robots are used by individuals as easy-to-access, powerful information sources. The use of these chatbots in healthcare has the potential to increase patient satisfaction and reduce healthcare costs.[7] However, it is extremely important to critically evaluate the information provided in the field of health and identify the inherent benefits, threats and shortcomings of these data sources.8 Using chatbots to answer frequently asked questions of patients in pediatric dentistry and questions that may be asked within the scope

Table I. Categories of patient questions and education questions regarding pediatric dentistry

	Patient questions	Education questions
Oral hygiene and preventive dentistry	5	2
Patient management	3	3
Dental trauma	1	2
Restorative dentistry and pulp therapy	3	6
Orthodontics	2	2
Dental pathology and dental radiology	0	3
Total	14	18

DOI: https://doi.org/10.58711/turkishjdentres.vi.1513154

Table II. Demographic characteristics of the pediatric dentists who made the evaluation

		Age, years			Professional experience years			Artificial intelligence experience		
	N(%)	Mean	Standard deviation	p-value	Mean	Standard Deviation	p-value	Yes	No	p-value
Female	16 (64)	32.25	3.17	0,605*	6.63	3.36	0.605*	5	11	0.364**
Male	9(36)	33	3.87		7.44	4.39		1	8	
Total	25 (100)	32.52	3.38		6.92	3.69		6 (24)	19 (76)	

* Independent Samples Test. ** Fisher's Exact Test.

Table III. Evaluation of the difference between the answers given to patient questions and the answers given to education-related questions based on the modified version of the GQS

	N	Mean	Standard deviation	p-value
Answers to Patient Questions	25	4.21	0.53	.000
Answers to Educational Questions	25	3.44	0.71	

*paired-samples t-test

Table IV. Evaluating the difference between the answers to patient questions and the answers to education-related questions according to categories

	Answers to patient questions		Answers to educational questions		
	Mean	Standard deviation	Mean	Standard deviation	p-value
Answers about oral hygiene and preventive dentistry	4.21	0.51	3.56	0.83	.000
Answers regarding patient management	4.3	0.63	3.48	0.85	.000
Answers about dental trauma	4,44	0, 82	3.46	0.94	.000
Answers about restorative dentistry and pulp therapy	4, 17	0, 75	3.64	0.64	.000
Answers about orthodontics	4,06	0,71	3.24	0.84	.000
Answers about dental pathology and dental radiology			3.05	0.80	

*paired-samples t-test

Table V. Pediatric dentists' evaluation results of ChatGPT-3.5

Would you consider using ChatGPT-3.5 to inform the patient?	N (%)
I will absolutely use	10 (40)
I can use it if it is developed further	10 (40)
I am undecided	5 (20)
I will absolutely not use it	0 (0)
Is the use of ChatGPT-3.5 useful in undergraduate education in pediatric dentistry?	
Absolutely useful	8 (32)
It could be useful if developed further.	14 (56)
I am undecided	3 (12)
Absolutely not helpful	0 (0)

of graduate education can provide a valuable resource for parents and students seeking information about dental conditions and potential care and treatment outcomes. This study aimed to evaluate the quality of ChatGPT-3.5's answers, an artificial intelligence-supported application, to questions in the field of pediatric dentistry. Therefore, the results of this study provide valuable information about the potential of artificial intelligence in improving the accuracy and efficiency of dental care.

A number of previous studies have evaluated the accuracy of the paid (4.0) and free versions (3.5) of ChatGPT when answering medical questions. In a study by Lee et al.[9] on hyperlipidemia, ChatGPT-3.5 answered 69.33% of the questions correctly, while ChatGPT-4.0 answered 74.67% of the questions. Lechien et al.10 reported that the accuracy of ChatGPT-3.5 and ChatGPT-4.0 in referencing articles published in the field of otolaryngology ranged from 47% to 60% and 73% to 87%, respectively. Additionally, Hirosawa et al.[11] reported that ChatGPT-3.5 creates differential diagnoses with high diagnostic accuracy in clinical cases with common complaints. Massey et al.[12] reported that orthopedic residents were able to answer more questions correctly on the Orthopedic Evaluation Exams than on ChatGPT-3.5 and GPT-4. Since ChatGPT version 4.0, which was released in March 2023, has advanced features but requires a monthly fee, the previous version (3.5), which is accessible to every segment of society, was used in this study.

There are a limited number of studies evaluating the performance of ChatGPT in dentistry. Suarez et al.[13] found a high level of consistency (85%) in ChatGPT's answers to endodontics questions. Mohammad-Rahimi et al.[2] reported that GPT-3.5 provides more reliable information on endodontics-related topics than Google Bard and Bing. Balel[3] reported that ChatGPT has significant potential as a patient information tool in oral and maxillofacial surgery but may not be safe to use in education. Rokhshad et al.[5] reported on the accuracy and consistency of chatbots to clinicians for answering pediatric dentistry questions that ChatGPT4 had the highest accuracy among chatbots, with an accuracy rate of 77.8% ± 5.1%, but significantly lower accuracy than pediatric dentists (mean ± SD: 96.7% ± 4.3%). Özden et al.[14] revealed that ChatGPT 3.5 correctly answered 57.5% of questions regarding dental trauma intervention. In this study, it was observed that ChatGPT3.5 gave more accurate answers to the frequently asked questions of patients than to the questions asked about undergraduate education the H1 hypothesis of the study was accepted. Based on this, ChatGPT3.5 needs to be developed in order to be used as a useful resource in undergraduate education. Particularly with the development of GPT models based on academic databases, it can make a significant contribution to student education by providing access to higher quality and more reliable information in the field of pediatric dentistry. Rokhshad et al.5 found the ChatGPT3.5 response consistency to be questionable (Cronbach's alpha coefficient=0.69). Therefore, there is a need to further examine and improve the consistency of ChatGPT 3.5's answers at different times.

Chatbots can also be used for educational purposes, as they can improve dental students' clinical skills and assist with various aspects of patient care and communication.[15,16,17] However, ChatGPT has limitations, such as it can only answer text-based questions and does not allow questions based on images to be processed. Additionally, short essay questions have limitations, such as not providing detailed information about clinical intervention and follow-up visits.[17] In this study, ChatGPT3.5 provided significantly lower quality answers to patients' frequently asked questions in all education-related subgroups. Moreover, ChatGPT3.5 gave the highest quality answers regarding education to questions in the field of restorative dentistry and pulp therapy, while it gave the lowest quality answers in the field of dental pathology and dental radiology. Similarly, Mago et al.[17] reported that ChatGPT3.5 only emphasizes the basic characteristic features of oral and maxillofacial pathologies and their radiographic findings and is not detail-oriented. Gilson et al.[18] reported that ChatGPT3.5 was initially trained on a corpus with data processing from 2021 or earlier, so it does not have updated sources. Additionally, these sources are not always verifiable. It may not be a useful tool for dental health professionals regarding oral and maxillofacial pathologies and their radiographic features.

DOI: https://doi.org/10.58711/turkishjdentres.vi.1513154

4.1. Limitations

This study has some limitations. First, the consistency of ChatGPT3 was not evaluated by asking questions at different times. Further studies are needed to investigate this important point. Secondly, only the answers given to open-ended questions were assessed. Yes/no, true/false type questions were not asked. However, open-ended questions may be more appropriate for revealing incomplete or incorrect answers. The last limitation is the use of a single version of artificial intelligence. Therefore, similar, multi-institutional studies need to be conducted on a larger scale and with a wider range of questions.

5. Conclusions

Within the confines of this study, ChatGPT may have significant potential as a patient information tool in pediatric dentistry, but caution should be exercised in its use in education. When using this technology, parents and students must be made aware of its inherent limitations and potential risks in terms of the accuracy and reliability of the information provided. In order for pediatric dentists to recommend ChatGPT as a valuable resource for patient and undergraduate education, scientifically validated, freely accessible datasets that pediatric dentists agree on need to be created.

References

1. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial Intelligence-Based Chatbots for Promoting Health Behavioral Changes: Systematic Review. J Med Internet Res. 2023;25:e40789.
2. Mohammad-Rahimi H, Ourang SA, Pourhoseingholi MA, Dianat O, Dummer PMH, Nosrat A. Validity and reliability of artificial intelligence chatbots as public sources of information on endodontics. Int Endod J. 2024;57(3):305-314.
3. Balel Y. Can ChatGPT-3.5 be used in oral and maxillofacial surgery?. J Stomatol Oral Maxillofac Surg. 2023;124(5):101471.
4. Tiwari A, Kumar A, Jain S, et al. Implications of ChatGPT-3.5 in Public Health Dentistry: A Systematic Review. Cureus. 2023;15(6):e40367.
5. Rokhshad R, Zhang P, Mohammad-Rahimi H, Pitchika V, Entezari N, Schwendicke F. Accuracy and consistency of chatbots versus clinicians for answering pediatric dentistry questions: A pilot study. J Dent. 2024;144:104938.
6. Xu L, Sanders L, Li K, Chow JCL. Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review. JMIR Cancer. 2021;7(4):e27850.
7. Dave T, Athaluri SA, Singh S. ChatGPT-3.5 in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell. 2023;6:1169595.
8. Lee TJ, Rao AK, Campbell DJ, Radfar N, Dayal M, Khrais A. Evaluating ChatGPT-3.5 and ChatGPT-4.0 Responses on Hyperlipidemia for Patient Education. Cureus. 2024;16(5):e61067.
9. Lechien JR, Briganti G, Vaira LA. Accuracy of ChatGPT-3.5 and -4 in providing scientific references in otolaryngology-head and neck surgery. Eur Arch Otorhinolaryngol. 2024;281(4):2159-2165.
10. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic Accuracy of Differential-Diagnosis Lists Generated by Generative Pretrained Transformer 3 Chatbot for Clinical Vignettes with Common Chief Complaints: A Pilot Study. Int J Environ Res Public Health. 2023;20(4):3378.
11. Massey PA, Montgomery C, Zhang AS. Comparison of ChatGPT-3.5, ChatGPT-4, and Orthopaedic Resident Performance on Orthopaedic Assessment Examinations. J Am Acad Orthop Surg. 2023;31(23):1173-1179.
12. Suárez A, Díaz-Flores García V, Algar J, Gómez Sánchez M, Llorente de Pedro M, Freire Y. Unveiling the ChatGPT phenomenon: Evaluating the consistency and accuracy of endodontic question answers. Int Endod J. 2024;57(1):108-113.
13. Ozden I, Gokyar M, Ozden ME, Sazak Ovecoglu H. Assessment of artificial intelligence applications in responding to dental trauma. Dent Traumatol. 2024 Dec;40(6):722-729. Epub 2024 May 14. PMID: 38742754.
14. Schwendicke F, Chaurasia A, Wiegand T, et al. Artificial intelligence for oral and dental healthcare: Core education curriculum. J Dent. 2023;128:104363.
15. Li R, Kumar A, Chen JH. How Chatbots and Large Language Model Artificial Intelligence Systems Will Reshape Modern Medicine: Fountain of Creativity or Pandora's Box?. JAMA Intern Med. 2023;183(6):596-597.
16. Ali K, Barhom N, Tamimi F, Duggal M. ChatGPT-A double-edged sword for healthcare education? Implications for assessments of dental students. Eur J Dent Educ. 2024;28(1):206-211.
17. Mago J, Sharma M. The Potential Usefulness of ChatGPT in Oral and Maxillofacial Radiology. Cureus. 2023;15(7):e42133.
18. 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.

DOI: https://doi.org/10.58711/turkishjdentres.vi.1513154