folia medica

Folia Medica 67(4)
DOI: 10.3897/folmed.67.e154338

Original Article

Evaluation of large language models in pediatric dentistry: a Bloom's taxonomy-based analysis

Atanu Mukhopadhyay[1], Santanu Mukhopadhyay[2], Raju Biswas[3]

1 *Amazon.com New York, New York, United States of America*
2 *Department of Dentistry, Malda Medical College and Hospital, Malda, India*
3 *Darjeeling District Hospital, Siliguri, India*

Corresponding author: Santanu Mukhopadhyay, Department of Dentistry, Malda Medical College and Hospital, Malda, West Bengal, India; Email: msantanu25@gmail.com; Tel.: 919732058536

Received: 30 March 2025 ♦ **Accepted:** 8 May 2025 ♦ **Published:** 14 August 2025

Citation: Mukhopadhyay A, Mukhopadhyay S, Biswas R. Evaluation of large language models in pediatric dentistry: a Bloom's taxonomy-based analysis. Folia Med (Plovdiv) 2025;67(4):e154338. doi: 10.3897/folmed.67.e154338.

Abstract

Aim: This study aimed to evaluate the performance of three large language models (LLMs)—ChatGPT-4.0, Claude 3.5 Sonnet, and DeepSeek R1—in answering multiple-choice questions (MCQs) related to pediatric dentistry. Accuracy and justification quality were analyzed using Bloom's taxonomy.

Materials and methods: A total of 90 MCQs were developed based on the American Academy of Pediatric Dentistry (AAPD) guidelines, ensuring cognitive diversity across Bloom's taxonomy levels. The models were assessed for answer accuracy and required to provide justifications, which were scored using a structured 4-point rubric by two independent pediatric dentistry experts. Statistical analyses, including Kruskal-Wallis tests and one-way ANOVA, were used to compare performance.

Results: DeepSeek R1 demonstrated the highest accuracy (92.2%), followed by Claude 3.5 sonnet (86.6%) and ChatGPT-4.0 (72.2%). Significant differences in accuracy were observed at the "Understanding" level (p=0.009). Justification quality also varied significantly among models, with DeepSeek R1 outperforming the others (p<0.001). Inter-rater reliability was high (ICC: 0.615–0.848), showing the reliability of the evaluation.

Conclusion: The study shows variations in LLM performance, with DeepSeek R1 excelling overall. It holds promise for pediatric dentistry education and AI decision-making, but further improvements are needed for better reasoning and clinical use.

Keywords

artificial intelligence, large language models, pediatric dentistry

Introduction

Artificial intelligence (AI) has emerged as a transformative force across various domains, including healthcare and dentistry. In pediatric dentistry, AI technologies have shown significant promise in enhancing diagnostic accuracy, streamlining treatment planning, and improving patient behavior management.[1-5] For instance, AI-powered tools have been utilized to detect dental caries, developmental anomalies, growth assessment, chronological age, and provide personalized preventive care strategies tailored to individual risk factors.[6-10] These advancements help address traditional challenges faced by pediatric dentists, such as managing anxious or uncooperative children during dental visits.

Building on these advancements, large language models (LLMs), a sophisticated subset of AI, represent a re-

markable progression in natural language processing and decision-making capabilities. Trained on vast datasets from diverse sources such as books, articles, and clinical guidelines, LLMs are designed to understand context and generate human-like text. These models operate using neural networks with billions or even trillions of parameters, enabling them to perform tasks such as answering questions, composing text, translating languages, and engaging in contextual conversations.[11-16]

In healthcare and dentistry, LLMs are gaining attention for their potential to assist clinicians by generating evidence-based responses to complex questions, offering clinical recommendations, and providing justifications for decisions. However, while LLMs like ChatGPT have shown promise in general medical applications, their reliability and applicability in pediatric dentistry remain underexplored.

Evaluating the performance of LLMs in pediatric dentistry requires a structured approach aligned with established clinical knowledge frameworks. Bloom's taxonomy serves as an effective tool for this purpose, categorizing cognitive skills into levels ranging from basic recall ("Remembering") to complex problem-solving ("Creating").[17,18]

AIM

This study aims to assess the accuracy and justification quality of three LLMs—ChatGPT-4.0, Claude 3.5 Sonnet, and DeepSeek R1—in answering multiple-choice questions (MCQs) based on the American Academy of Pediatric Dentistry (AAPD) guidelines. By analyzing their performance across Bloom's taxonomy levels, this article seeks to identify strengths, limitations, and areas for improvement in applying LLMs to pediatric dentistry.

Materials and methods

Study design

This quantitative observational study aimed to evaluate the performance of three Large Language Models (LLMs) in answering multiple-choice questions (MCQs) related to pediatric dentistry. The study aimed to evaluate both the accuracy of the LLMs' responses and the quality of their justifications. No Institutional Review Board (IRB) approval was required for this study, as it focused on the evaluation of artificial intelligence (AI) models and did not involve human or animal data. Nonetheless, the reporting complies with the principles of the Declaration of Helsinki.

Question development

A total of 90 MCQs were systematically developed based on the American Academy of Pediatric Dentistry (AAPD) guidelines (see **Appendix**). Each question was categorized according to modified Bloom's taxonomy to ensure comprehensive cognitive coverage across six levels: Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating. The correct answers and justifications were pre-validated by the two pediatric dentistry experts. The choice of Bloom's taxonomy was made to ensure a structured assessment of cognitive skills, allowing a balanced evaluation of LLM performance across different levels of complexity.

Model evaluation

Three LLMs (referred to as ChatGPT- 4.0, Claude 3.5 Sonnet, and DeepSeek R1) were evaluated in this study. These models were selected based on their widespread use in clinical and educational settings, as well as their demonstrated potential for medical knowledge processing. Each model was provided with the 90 MCQs and tasked with selecting the most appropriate answer. Additionally, the models were required to generate a brief justification for each selected response.

Justification scoring

The justifications provided by each model were independently assessed by two experienced pediatric dentists using a structured 4-point rubric:

- 0: No justification or incorrect reasoning
- 1: Minimal justification with limited relevance
- 2: Partially correct justification with some gaps
- 3: Clear and well-supported justification
- 4: Comprehensive and clinically relevant justification

Statistical analysis

Descriptive statistics were calculated to summarize the performance of each model. The normality of justification scores and accuracy rates was assessed using the Shapiro-Wilk test, confirming non-normal distribution. Consequently, the Kruskal-Wallis test was employed to compare performance across models.

Inter-rater reliability was evaluated using the Intraclass Correlation Coefficient (ICC) for rater agreement. Performance differences across modified Bloom's taxonomy levels were analyzed using mean scores and non-parametric statistical comparisons, with significance set at $p<0.05$.

Results

Model accuracy

Table 1 presents the accuracy of each model in selecting correct answers. DeepSeek R1 achieved the highest accuracy (83/90, 92.2%), followed by Claude 3.5 Sonnet (78/90, 86.6%), and ChatGPT-4.0 (65/90, 72.2%) **(Fig. 1)**. A chi-square test revealed a statistically significant difference in accuracy among the three models ($\chi^2=14.064$, $p=0.0008$).

Table 1. Accuracy of large language models

Metric	ChatGPT-4.0	Claude 3.5 Sonnet	DeepSeek R1	χ^2 (*p*-value)
Correct responses (out of 90)	65	78	83	χ^2=14.064, *p*=0.0008
Accuracy (%)	72.2%	86.6%	92.2%	

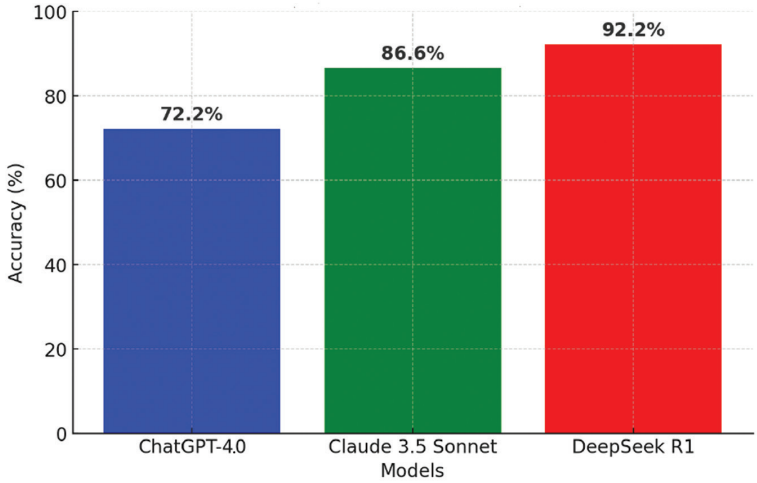

Figure 1. Accuracy of large language models in answering multiple choice questions.

Pairwise comparisons showed a significant difference between ChatGPT-4.0 and DeepSeek R1 (p=0.008), which remained significant after Bonferroni correction (α=0.0167). However, no significant differences were found between ChatGPT-4.0 and Claude 3.5 Sonnet (p=0.064) or between Claude 3.5 Sonnet and DeepSeek R1 (p=0.606).

Performance across modified Bloom's taxonomy levels

Table 2 and **Fig. 2** show model accuracy across Bloom's taxonomy levels. Significant differences were observed only at the "Understanding" level (χ^2=9.27, p=0.009), where DeepSeek R1 achieved the highest accuracy (97.1%), followed by Claude 3.5 Sonnet (91.2%) and ChatGPT-4.0 (70.6%). For other levels, no significant differences were found. Given the small sample sizes in some categories (e.g., Creating, Evaluating), Fisher's exact test was used to minimize the risk of type I errors.

The Kruskal-Wallis H-test indicated significant differences in mean justification scores among models (χ^2=14.11, p<0.001). Dunn's post-hoc tests confirmed significant differences between ChatGPT-4.0 and DeepSeek R1 (p=0.0002) and between Claude 3.5 Sonnet and DeepSeek R1 (p=0.0099), but not between ChatGPT-4.0 and Claude 3.5 Sonnet.

Justification quality

Table 3 presents the mean justification scores across modified Bloom's taxonomy levels, rated by two independent assessors using a structured rubric. ChatGPT-4.0 exhibited the highest mean score for the "Remembering" level (Rater 1: 2.80±0.75, Rater 2: 2.77±0.71), while DeepSeek R1 had the highest mean scores for "Understanding" (Rater 1: 2.76±0.60, Rater 2: 3.09±0.66) and "Creating" (Rater 1:

Table 2. Accuracy of models according to Bloom's level

Revised Bloom's Level	ChatGPT-4.0		Claude 3.5 Sonnet		DeepSeek R1		χ^2 (*p*-value)
	Correct	%	Correct	%	Correct	%	
Remembering	17/20	85.0%	16/20	80.0%	18/20	90.0%	χ^2=0.67 (p=0.72)
Understanding	24/34	70.6%	31/34	91.2%	33/34	97.1%	χ^2=9.27 (***p*=0.009**)
Applying	7/15	46.7%	12/15	80.0%	12/15	80.0%	χ^2=4.71 (*p=0.095*)
Analyzing	8/11	72.7%	11/11	100.0%	11/11	100.0%	χ^2=4.49 (*p=0.10*)
Evaluating	7/7	100.0%	6/7	85.7%	7/7	100.0%	χ^2=1.04 (p=0.59)
Creating	2/3	66.7%	2/3	66.7%	2/3	66.7%	χ^2=0.00 (p=1.00)
Total	65/90	72.2%	78/90	86.6%	83/90	92.2%	χ^2=14.064 (p=0.0008)

Folia Medica

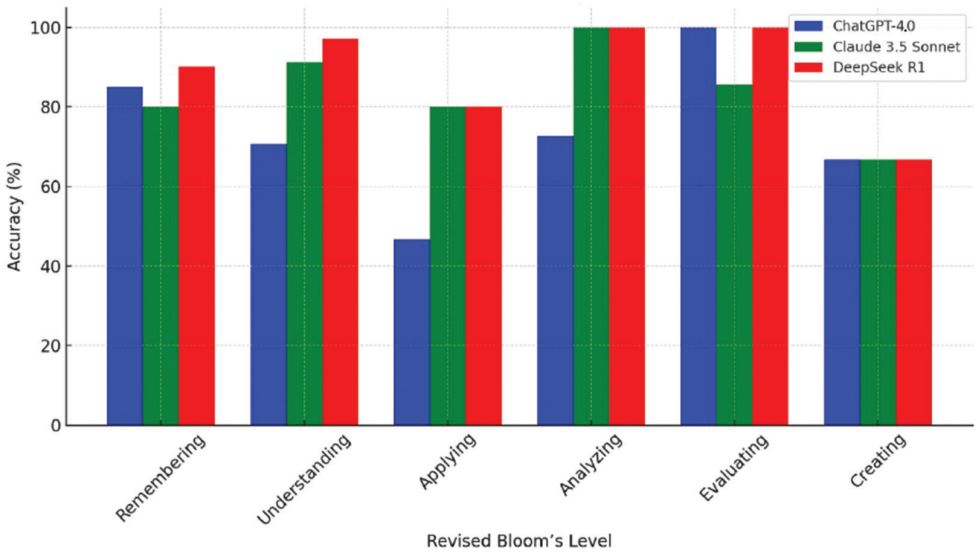

Figure 2. Accuracy of models according to the revised Bloom's level.

Table 3. LLM performance (mean score) across modified Bloom's taxonomy levels

Revised Bloom's Level	ChatGPT-4.0		Claude 3.5 Sonnet		DeepSeek R1	
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
Remembering	2.80±0.75	2.77±0.71	2.92±0.43	2.85±0.40	2.17±0.94	2.50±0.97
Understanding	3.04±0.68	2.95±0.73	2.95±0.63	2.82±0.62	2.76±0.60	3.09±0.66
Applying	3.07±0.94	3.13±0.95	2.63±1.15	2.60±1.52	2.53±1.00	2.70±1.25
Analyzing	3.04±0.47	3.14±0.50	2.59±1.02	2.50±1.07	2.18±1-32	2.72±1.35
Evaluating	2.86±0.38	2.78±0.39	2.78±0.26	3.00±0.50	3.00±0.50	2.85±0.69
Creating	2.67±1.53	2.67±1.53	1.83±1.75	2.16±2.02	2.83±1.04	3.83±2.89
Total Mean± SD	2.97±0.73	2.94±0.75	2.79±0.80	2.74±0.81	2.48±0.85	2.73±0.96

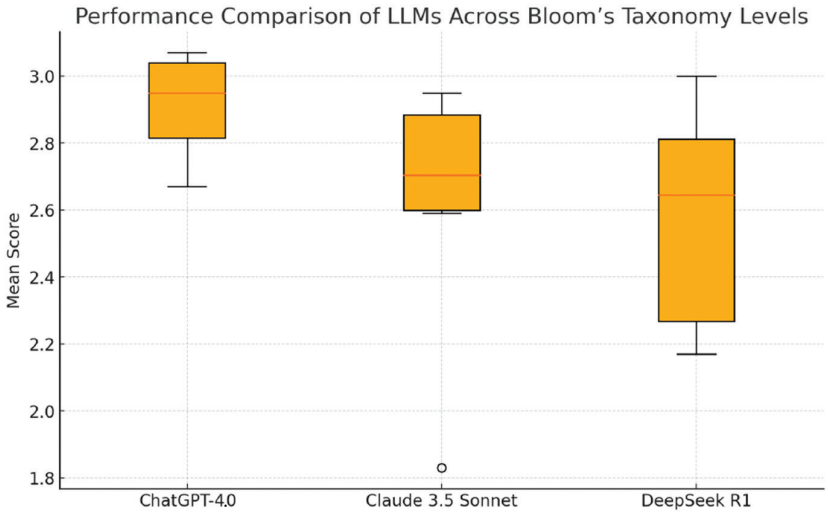

Figure 3. Boxplot comparing the performance of the three models across Bloom's taxonomy levels.

2.83±1.04, Rater 2: 3.83±2.89) **(Fig. 3)**. The ICC values for inter-rater reliability were: ChatGPT-4.0 (0.848, excellent agreement), Claude 3.5 Sonnet (0.783, good agreement), and DeepSeek R1 (0.615, moderate agreement).

The differences in mean justification scores were confirmed using non-parametric analyses. The Kruskal-Wal-

lis H-test indicated significant differences among models (χ^2=14.11, p<0.001). Dunn's post-hoc tests confirmed significant differences between ChatGPT-4.0 and DeepSeek R1 (p=0.0002) and between Claude 3.5 Sonnet and DeepSeek R1 (p=0.0099), but not between ChatGPT-4.0 and Claude 3.5 Sonnet.

Discussion

This study evaluated the performance of three large language models (LLMs)—ChatGPT-4.0, Claude 3.5 Sonnet, and DeepSeek R1—in answering multiple-choice questions (MCQs) in pediatric dentistry based on the American Academy of Pediatric Dentistry (AAPD) guidelines. The findings revealed significant differences in accuracy and justification quality among the models, with DeepSeek R1 demonstrating the best overall performance, particularly at the "Understanding" level of Bloom's taxonomy.

DeepSeek R1 achieved the highest accuracy (92.2%), outperforming Claude 3.5 Sonnet (86.6%) and ChatGPT-4.0 (72.2%). These results are consistent with studies by Giannakopoulos et al.[19] and Tussie and Starosta[20], who found that more advanced models like ChatGPT-4.0 perform better in answering dental questions. However, our study also highlighted that while DeepSeek R1 excelled in comprehension, it struggled with more complex tasks like applying knowledge and creating solutions. This pattern aligns with research by Künzle and Paris[21], who observed that LLMs perform well on straightforward questions but struggle with clinical reasoning.

The quality of justifications provided by the models varied. DeepSeek R1 had moderate inter-rater reliability (ICC=0.615), whereas ChatGPT-4.0 showed higher consistency (ICC=0.848). This variability in DeepSeek R1's evaluations may stem from its unique justification style, which excelled at higher Bloom's taxonomy levels (e.g., Understanding and Creating) but may have been less consistent or clear to raters. This finding aligns with Giannakopoulos et al., who reported that advanced LLMs provide detailed but sometimes inconsistent explanations, and Makrygiannakis et al.[22], who noted that LLMs occasionally give vague or irrelevant justifications. To address this variability, future studies should consider refining the justification rubric to reduce subjectivity or providing rater training to enhance agreement, particularly for models like DeepSeek R1 with complex reasoning patterns.

Previous studies provide additional insight into these results. For instance, Abu Arqub et al.[11] found that ChatGPT's accuracy in answering orthodontic questions was insufficient, highlighting variability in LLM performance across dental specialties. In contrast, Buldur and Sezer[12] showed that ChatGPT provided reliable information about fluoride use, demonstrating its potential for patient education. Maltarollo et al.[13] emphasized that while ChatGPT can be a helpful tool for pediatric dentists, it should be used with caution. Additionally, Dermata et al.[14] found that ChatGPT-4.0 was the most reliable LLM for pediatric dentistry but emphasized that AI should support, not replace, professional expertise.

The significant differences in performance at the "Understanding" level (p=0.009) align with findings from Nguyen et al.[23], who showed that LLMs perform well in comprehension tasks but struggle with higher-order reasoning. Similarly, Künzle and Paris[21] reported that LLMs performed well on straightforward topics like direct restorations and caries management but faced challenges with complex subjects such as endodontic procedures. These findings suggest that while LLMs can effectively recall and interpret information, their ability to apply knowledge in clinical scenarios remains limited.

The study also highlights the importance of model training data and structured prompts. Research by Jones et al.[24] indicates that LLMs trained on domain-specific data tend to perform better. Similarly, ChatGPT-4.0 showed improved accuracy when given structured prompts emphasizing key clinical concepts. This reinforces the need for optimized prompting strategies to maximize AI's effectiveness in dentistry.

Further supporting these findings, Ahmed et al.[25] conducted a comparative analysis of AI-generated dental caries MCQs, evaluating ChatGPT and Google Bard (now Gemini) in terms of accuracy, complexity, and item writing flaws. Their study found no significant differences in question quality between the models; however, Bard-generated questions exhibited higher cognitive levels, whereas ChatGPT showed more formatting errors. This aligns with our findings, as LLM-generated responses in our study also varied in quality and cognitive depth. Additionally, Ahmed et al.[25] emphasized that while AI can generate educational content efficiently, careful human review remains essential to ensure alignment with learning objectives. This further highlights the need for robust validation methods when integrating AI into dental education and assessment.

However, this study has several limitations. The use of MCQs may not fully reflect real-world clinical decision-making, as noted by Tussie and Starosta[20] and Künzle and Paris[21]. While MCQs effectively assess factual knowledge and basic comprehension, they often fail to capture the subtle aspects of clinical reasoning, differential diagnosis, and patient management. Additionally, MCQs provide limited insight into an AI's ability to integrate information across multiple sources, a critical skill in clinical practice. Small sample sizes for higher cognitive levels also limit the generalizability of the results. Additionally, LLM-generated responses occasionally lacked references or contained inaccuracies, a concern raised in studies by Giannakopoulos et al.[19] and Makrygiannakis et al.[22] These limitations highlight the need for careful evaluation before integrating LLMs into clinical practice.

Despite these challenges, LLMs have promising applications in pediatric dentistry. They can support diagnosis by aligning responses with AAPD guidelines, assist in education by helping students and professionals study clinical

Folia Medica

concepts, and even integrate with virtual and augmented reality to manage dental anxiety in children. However, reliance on AI without human oversight could lead to inconsistencies, emphasizing the need for careful implementation. Chatzopoulos et al.[26] warned that improper use of LLMs in clinical settings could negatively impact patient care. Additionally, Bayraktar Nahir[27] pointed out that while ChatGPT-generated content is informative, its readability needs improvement for pediatric patients and their parents.

ChatGPT-4.0, developed by OpenAI, is known for its balance of speed and efficiency, delivering high accuracy across various domains. However, it occasionally struggles with consistency in its justifications, as evidenced by its moderate inter-rater reliability (ICC=0.848). Claude 3.5 Sonnet, created by Anthropic, prioritizes safety and interpretability, often providing detailed and cautious responses. While it performs well in comprehension tasks, its accuracy (86.6%) is slightly lower than that of DeepSeek R1. DeepSeek R1, the newest model evaluated, excels in comprehension and contextual reasoning, achieving the highest accuracy (92.2%) among the three models. However, it faces challenges in more complex tasks such as applying knowledge and creating solutions. This model's moderate inter-rater reliability (ICC=0.615) indicates variability in the quality of its justifications. Each model brings unique strengths, but also exhibits areas where improvement is needed, particularly in handling complex clinical scenarios and ensuring consistent and accurate justifications.

To improve LLM applications in pediatric dentistry, several steps should be considered. Expanding datasets to include case-based questions will help assess LLMs' ability to handle complex clinical decisions. Testing these models in real-world settings, such as clinical workflows, will provide deeper insights into their usefulness. Collaborating with professional organizations to develop domain-specific training datasets can further improve accuracy. Making LLM-generated content more accessible and readable will enhance patient education. Finally, integrating AI into clinical decision-support systems tailored for pediatric dentistry can maximize benefits while minimizing risks.

ChatGPT-4.0, developed by OpenAI, is a widely used LLM that balances speed and efficiency while providing high accuracy in various domains. Claude 3.5 Sonnet, created by Anthropic, focuses on safety and interpretability, often providing detailed, cautious responses. DeepSeek R1, a newer model, has demonstrated strong performance in comprehension and contextual reasoning, making it particularly suitable for specialized applications like dentistry.

Conclusion

The study offers valuable insights into the performance of large language models (LLMs) in pediatric dentistry, emphasizing both the strengths and limitations of each model. Future research should aim to broaden the evaluation scope to encompass more complex clinical scenarios. Additionally, efforts should be made to enhance the training data and prompting strategies, which will improve the accuracy and reliability of LLMs in this field.

Use of Artificial Intelligence in manuscript preparation

During the preparation of this work, the authors utilized ChatGPT-4.0 (OpenAI.com) to check English grammar, spelling, and references, as well as to create bar graphs. After employing this tool, the authors reviewed and edited the content as needed, taking full responsibility for the published article's final content.

Data availability

Data of study is available from the corresponding author on reasonable request.

Declarations funding

The authors did not receive any financial support for their research, writing, or publishing this article.

Authors contributions

Conceptualization: A.M.; data curation, formal analysis: S.M., and R.B.; project administration: A.M.; validation: S.M., and R.B.; methodology, investigation supervision: A.M., S.M., and R.B.; resources: A.M., S.M.; software: A.M.; visualization: A.M.; writing-original draft: S.M.; writing-review & editing: A.M., S.M., and R.B.

Conflicts of interest

All authors declare that they have no conflicts of interest.

Institutional review board permission

No Institutional Review Board (IRB) approval was required for this study, as it focused on the evaluation of artificial intelligence (AI) models.

Acknowledgements

The authors have no support to report.

Appendix

Ninety multiple-choice questions based on the American Academy of Pediatric Dentistry guidelines.

1. What is the primary goal of pulp therapy in pediatric dentistry?

a) To extract the affected tooth

b) To maintain the integrity and health of the teeth and their supporting tissues

c) To replace the tooth with a prosthetic

d) To perform orthodontic treatment

2. Which of the following best describes the purpose of a protective liner in deep cavity preparations?

a) To replace the tooth structure

b) To act as a barrier between the restorative material and the pulp

c) To extract the tooth

d) To perform orthodontic treatment

3. A 7-year-old patient presents with a deep carious lesion in a primary molar but no signs of pulpitis. Which pulp therapy procedure is most appropriate?

a) Pulpectomy

b) Indirect pulp treatment (IPT)

c) Direct pulp cap

d) Extraction

4. Why is mineral trioxide aggregate (MTA) preferred over calcium hydroxide for direct pulp capping in permanent teeth?

a) MTA is less expensive

b) MTA causes tooth discoloration

c) MTA results in more predictable dentin bridging and pulp health

d) MTA is easier to apply

5. A 10-year-old patient has a necrotic immature permanent tooth with an open apex. Which treatment option is most appropriate to induce root end closure?

a) Pulpotomy

b) Apexification

c) Indirect pulp treatment

d) Extraction

6. Which of the following treatment plans is most appropriate for a 5-year-old patient with a primary molar diagnosed with reversible pulpitis and deep caries?

a) Immediate extraction of the affected tooth to prevent further complications

b) Placement of a stainless-steel crown after removing the carious tissue and performing a pulpotomy

c) Application of a fluoride varnish and scheduling a follow-up appointment in six months

d) Removal of all carious tissue, placement of a calcium hydroxide dressing, and temporary restoration with a plan for re-evaluation in six weeks

7. What is the recommended concentration of sodium hypochlorite for irrigation during pulp therapy?

a) 1-5%

b) 10%

c) 20%

d) 30%

8. What is the primary difference between apexogenesis and apexification?

a) Apexogenesis is for necrotic pulp, while apexification is for vital pulp

b) Apexogenesis promotes continued root development, while apexification induces root end closure

c) Apexogenesis is used in primary teeth, while apexification is used in permanent teeth

d) Apexogenesis requires extraction, while apexification does not

9. A 6-year-old patient has a primary tooth with a small mechanical pulp exposure. Which procedure is most appropriate?

a) Pulpotomy

b) Direct pulp cap

c) Pulpectomy

d) Extraction

10. Why is rubber dam isolation considered the gold standard for pulp therapy?

a) It is cost-effective

b) It minimizes bacterial contamination and protects soft and hard tissues

c) It is easier to use than cotton rolls

d) It is required for all dental procedures

11. A 9-year-old patient has a traumatic pulp exposure in an immature permanent tooth. Which procedure is most appropriate to preserve pulp vitality and promote root development?

a) Pulpotomy

b) Direct pulp cap

c) Pulpectomy

d) Extraction

12. Develop a step-by-step protocol for performing a pulpotomy on a primary molar with reversible pulpitis. Include materials and rationale for each step.

a) Diagnosis, Anesthesia, Isolation, Access, Coronal Pulp Removal, Hemostasis, Medicament Application, Restoration, Follow-Up

b) Diagnosis, Isolation, Anesthesia, Coronal Pulp Removal, Access, Hemostasis, Medicament Application, Restoration, Follow-Up

c) Diagnosis, Anesthesia, Coronal Pulp Removal, Access, Isolation, Hemostasis, Medicament Application, Restoration, Follow-Up

d) Diagnosis, Anesthesia, Isolation, Access, Coronal Pulp Removal, Hemostasis, Medicament Application, Restoration, Follow-Up

13. How does moderate sedation differ from deep sedation?

a) Moderate sedation allows purposeful response to verbal commands; deep sedation does not.

b) Deep sedation requires no monitoring.

c) Moderate sedation eliminates protective reflexes.

d) Deep sedation is only used in hospitals.

14. A child is uncooperative despite no history of dental fear. What factor might explain this?

Folia Medica

a) Parental resilience
b) Developmental delay or inadequate coping skills
c) Effective use of tell-show-do
d) High pain tolerance

15. Which factor is most critical when scheduling appointments for anxious children?
a) Time of day and patient tolerance level
b) Dentist's convenience
c) Cost of treatment
d) Availability of toys in the waiting area

16. Which documentation is essential after using protective stabilization?
a) A photo of the procedure
b) Duration of stabilization and patient response
c) Parent's occupation
d) Number of staff present

17. What is the most important component of informed consent for behavior guidance?
a) Discussing risks, benefits, and alternatives with the parent
b) Obtaining a signature without explanation
c) Using technical jargon to ensure clarity
d) Skipping documentation to save time

18. Which method is most effective for assessing pain in a nonverbal child?
a) Self-reporting
b) Observing behavioral changes (e.g., crying, movement)
c) Ignoring signs of distress
d) Relying solely on parental intuition

19. A child with a history of trauma exhibits aggression during treatment. What approach should the dentist take?
a) Use protective stabilization immediately
b) Employ trauma-informed care and gradual desensitization
c) Cancel all future appointments
d) Prescribe general anesthesia for all procedures

20. How would you address a non-English-speaking family's needs during behavior guidance?
a) Proceed without translation services
b) Use gestures exclusively
c) Provide a trained interpreter and culturally sensitive communication
d) Avoid discussing treatment options

21. What is the primary purpose of a knee-to-knee examination in managing traumatic dental injuries in children?
a) Administer local anesthesia
b) Facilitate radiographic imaging
c) Examine a young child effectively
d) Perform emergency extractions

22. Which of the following is NOT recommended for assessing pulp vitality in primary teeth?
a) Tooth mobility evaluation
b) Pulp sensibility tests
c) Observing crown discoloration
d) Radiographic examination

23. What is the ALARA principle in pediatric radiography?
a) Avoiding all radiographs
b) Minimizing radiation exposure as much as possible
c) Using CBCT for all traumatic dental injuries
d) Prioritizing parental preferences

24. Why should avulsed primary teeth NOT be replanted?
a) They regenerate naturally
b) Risk of aspiration and damage to permanent tooth germs
c) Primary teeth lack roots
d) Replantation is painful

25. What is the rationale for documenting parental awareness of potential permanent tooth complications?
a) To increase treatment costs
b) To ensure informed consent and future monitoring
c) To discourage follow-up visits
d) To prioritize primary tooth retention

26. A 4-year-old presents with an intruded primary incisor. What is the FIRST management step?
a) Immediate extraction
b) Observation for spontaneous re-eruption
c) Root canal treatment
d) Splinting the tooth

27. Which homecare instruction is critical for soft tissue injury healing?
a) Avoiding oral hygiene for one week
b) Using alcohol-based mouthwash
c) Cleaning with a soft brush and alcohol-free chlorhexidine
d) Restricting fluid intake

28. How does the management of root fractures in primary teeth differ from permanent teeth?
a) Primary teeth require immediate extraction
b) Root fractures in primary teeth are always splinted
c) Observation is prioritized unless occlusion is affected
d) Permanent teeth are never treated conservatively

29. Why is a structured trauma history form beneficial?
a) Reduces appointment time
b) Improves accuracy of diagnosis and documentation
c) Eliminates the need for radiographs
d) Focuses solely on soft tissue injuries

30. Which factor MOST justifies deferring extraction of a displaced primary tooth?
a) Parental preference
b) Risk of damaging the permanent tooth germ during extraction
c) Cost of sedation
d) Lack of dental materials

31. What is the MOST critical reason to avoid antibiotics for uncomplicated luxation injuries?
a) No evidence supporting their use
b) Parental resistance
c) Risk of tooth discoloration

d) Antibiotic shortages

32. A child with a lateral luxation injury resists examination. What is the BEST approach?

a) Use protective stabilization immediately

b) Refer to a child-oriented dental team with behavioral management expertise

c) Prescribe general anesthesia

d) Delay treatment indefinitely

33. How would you manage a parent concerned about permanent tooth damage after a primary tooth avulsion?

a) Reassure them that no follow-up is needed

b) Schedule regular monitoring and educate about potential complications

c) Extract adjacent primary teeth preventively

d) Recommend CBCT immediately

34. Which injury is MOST associated with permanent tooth developmental anomalies?

a) Crown fracture

b) Intrusion

c) Enamel crack

d) Concussion

35. When is a tetanus booster indicated for a traumatic dental injury?

a) Always required

b) Only for avulsions

c) If environmental contamination is suspected

d) For all soft tissue injuries

36. What is the PRIMARY goal of using alcohol-free chlorhexidine post-injury?

a) Reduce plaque and bacterial load

b) Whiten teeth

c) Promote tooth remineralization

d) Eliminate the need for brushing

37. Which diagnostic tool is RARELY indicated for dental trauma in young children?

a) Periapical radiographs

b) CBCT

c) Clinical photographs

d) Mobility testing

38. What should clinicians prioritize when managing a child's dental anxiety post-dental trauma?

a) Immediate extraction of all injured teeth

b) Avoiding sedation at all costs

c) Minimizing traumatic experiences and using behavioral guidance

d) Delegating care to general dentists

39. Which complication is linked to delayed healing of gingival injuries?

a) Improved occlusion

b) Increased risk of infection and poor oral hygiene

c) Faster eruption of permanent teeth

d) Reduced dental anxiety

40. What is a key component of the International Association of Dental Traumatology Core Outcome Set (COS) for Traumatic dental injuries?

a) Parental satisfaction surveys

b) Standardized generic and injury-specific outcomes

c) Financial cost analysis

d) Aesthetic rankings

41. What is the primary purpose of the Frankl Scale in pediatric dentistry?

a) Diagnose dental caries

b) Assess pain levels during procedures

c) Rate observed patient behaviors during appointments

d) Measure parental satisfaction

42. Which of the following is a contraindication for nitrous oxide/oxygen inhalation?

a) Mild situational anxiety

b) Chronic obstructive pulmonary disease (COPD)

c) Autism spectrum disorder

d) Limited English proficiency

43. What is the first step in the "tell-show-do" technique?

a) Perform the procedure

b) Demonstrate the procedure using non-threatening language

c) Explain the procedure in age-appropriate terms

d) Ask the patient to practice the procedure

44. Which technique involves diverting a patient's attention using audiovisual aids?

a) Voice control

b) Distraction

c) Desensitization

d) Protective stabilization

45. Why is understanding cultural factors critical in behavior guidance?

a) To reduce the cost of dental care

b) To improve communication and foster trust with families

c) To shorten appointment times

d) To eliminate the need for interpreters

46. What is the primary goal of systematic desensitization?

a) To punish disruptive behavior

b) To gradually expose patients to fear-inducing stimuli and reduce anxiety

c) To increase gag reflex sensitivity

d) To replace local anesthesia

47. Why is parental presence/absence used during dental procedures?

a) To reduce the dentist's workload

b) To establish dentist-child roles and improve compliance

c) To eliminate the need for informed consent

d) To prioritize parental preferences over treatment

48. What role do nonclinical staff play in behavior guidance?

a) Administer sedation

b) Set the tone for appointments through welcoming communication

c) Perform dental procedures

d) Diagnose dental anxiety

49. A child with autism spectrum disorder is overwhelmed by sensory stimuli. Which technique is most appropriate?

a) Voice control

b) Sensory-adapted dental environment (SADE)

c) Protective stabilization

d) General anesthesia

50. A child begins crying during a procedure. What should the dentist do first?

a) Proceed quickly to finish the treatment

b) Use protective stabilization

c) Pause and assess pain/discomfort

d) Ask the parent to leave the operatory

51. Which technique is recommended for a child with a hypersensitive gag reflex?

a) Nitrous oxide/oxygen inhalation

b) Memory restructuring

c) Positive reinforcement

d) Ask-tell-ask

52. When is protective stabilization contraindicated?

a) For a cooperative patient requiring urgent care

b) For a non-emergent procedure in a healthy patient

c) For a sedated patient

d) For a patient with special healthcare needs

53. The Canadian cohort study linked maternal fluoride intake to lower IQ in boys, but the New Zealand study found no association. What key difference explains this discrepancy?

a) Fluoride exposure levels and confounding variables

b) Study sample sizes

c) Age of participants at testing

d) Use of fluoride supplements

54. What is the purpose of the Hall Technique for stainless steel crowns (SSCs)?

a) To whiten teeth

b) To cement SSCs without caries removal or tooth preparation

c) To replace orthodontic bands

d) To improve gingival health

55. 5-year-old with high caries risk and poor cooperation needs a Class II restoration. Which material is most appropriate?

a) Composite resin

b) Resin-modified glass ionomer cement (RMGIC)

c) Dental amalgam

d) Compomer

56. According to the AAPD, which criterion is essential for deciding when to restore a carious lesion?

a) Tooth color

b) Visual detection of enamel cavitation

c) Patient age

d) Parental preference

57. Why are preformed metal crowns preferred over amalgam for primary teeth with extensive caries?

a) Greater longevity and lower failure rates

b) Reduced cost

c) Improved esthetics

d) Faster placement

58. Why do palatal and gingival cysts of the newborn not require treatment?

a) They are painful

b) They are filled with keratin

c) They typically disappear during the first three months of life

d) They result from trapped epithelial remnants

59. How do Epstein pearls differ from Bohn nodules in terms of location?

a) Epstein pearls are found on the buccal and lingual aspects of the ridge, while Bohn nodules are found in the median palatal raphe

b) Epstein pearls are remnants of salivary gland epithelium, while Bohn nodules are trapped epithelial remnants

c) Epstein pearls are found in the median palatal raphe, while Bohn nodules are found on the buccal and lingual aspects of the ridge

d) Epstein pearls and Bohn nodules are both found on the crests of the dental ridges

60. Why might composite restorations fail more frequently than amalgam in high- caries-risk patients?

a) Higher susceptibility to recurrent caries

b) Poor esthetics

c) Inadequate fluoride release

d) Difficulty in placement

61. What is the primary microbial cause of dental caries?

a) *Streptococcus mutans*

b) *Lactobacillus acidophilus*

c) *Porphyromonas gingivalis*

d) *Candida albicans*

62. Which component of teeth is first affected by demineralization in caries formation?

a) Dentin

b) Cementum

c) Enamel

d) Pulp

63. What is the term for the sticky biofilm that contributes to tooth decay?

a) Tartar

b) Calculus

c) Dental plaque

d) Gingiva

64. Which dietary factor is most strongly associated with caries development?

a) Protein

b) Fermentable carbohydrates

c) Fat

d) Fiber

65. How does fluoride primarily protect against dental caries?

a) Kills oral bacteria

b) Enhances remineralization of enamel

c) Whitens teeth

d) Reduces saliva production

66. Why are pit-and-fissure surfaces of molars particularly prone to caries?

a) They lack enamel
b) They trap food and bacteria
c) They are exposed to less saliva
d) They are harder to brush

67. What role does saliva play in caries prevention?
a) Neutralizes acids and aids remineralization
b) Increases plaque formation
c) Promotes bacterial growth
d) Destroys enamel

68. What does the term "early childhood caries" refer to?
a) Caries in permanent teeth
b) Severe decay in primary teeth, often due to prolonged bottle-feeding
c) Cavities caused by trauma
d) Genetic enamel defects

69. A patient consumes sugary snacks multiple times a day. Which advice is most effective to reduce caries risk?
a) Brush twice daily
b) Limit snacking frequency
c) Use mouthwash
d) Avoid dairy products

70. Which intervention would best target Streptococcus mutans in high-risk patients?
a) Fluoride varnish
b) Chlorhexidine mouthwash
c) Dental floss
d) Calcium supplements

71. A patient with dry mouth (xerostomia) is at higher caries risk. What is the best preventive measure?
a) Increase sugar intake
b) Use saliva substitutes or sugar-free gum
c) Avoid fluoride toothpaste
d) Reduce brushing frequency

72. Which clinical sign indicates active caries progression?
a) White spot lesion
b) Cavitation with softened dentin
c) Brown staining
d) Enamel hypoplasia

73. How do demineralization and remineralization differ in caries development?
a) Demineralization breaks down enamel; remineralization repairs it
b) Both processes destroy enamel
c) Remineralization causes cavities
d) Demineralization is irreversible

74. Which factor is the primary difference between rampant caries and arrested caries?
a) Active decay vs. inactive, hardened lesions
b) Location on the tooth
c) Bacterial species involved
d) Pain level

75. Why are root surfaces more vulnerable to caries in older adults?
a) Gingival recession exposes cementum
b) Increased saliva flow
c) Thicker enamel
d) Reduced sugar intake

76. Which statement best links diet and caries risk?
a) Frequent acidic drinks lower oral pH, promoting demineralization
b) Protein-rich diets increase plaque
c) Fat inhibits bacterial growth
d) Fiber prevents enamel erosion

77. Which preventive strategy is most effective for populations with limited access to fluoride?
a) Brushing with water
b) Community water fluoridation
c) Avoiding dental checkups
d) Increasing sugar consumption

78. A patient with recurrent caries claims to brush twice daily. What is the most likely oversight?
a) Inadequate interdental cleaning (e.g., flossing)
b) Using fluoride toothpaste
c) Drinking fluoridated water
d) Avoiding sugary snacks

79. Which treatment is least appropriate for early enamel caries?
a) Fluoride therapy
b) Root canal treatment
c) Improved oral hygiene
d) Dietary counseling

80. Why is restoring a cavitated lesion with a filling insufficient for long-term caries control?
a) Fillings weaken teeth
b) It doesn't address the underlying causes (e.g., diet, hygiene)
c) Fillings promote bacterial growth
d) All fillings eventually fail

81. What is the current recommended fluoride level in U.S. community water systems?
a) 0.3–0.6 ppm F
b) 0.7 ppm F
c) 1.2 ppm F
d) 5.0 ppm F

82. As of 2018, what percentage of the U.S. population on community water systems had access to fluoridated water?
a) 35%
b) 60%
c) 73%
d) 90%

83. According to the Cochrane review, water fluoridation led to what percentage increase in caries-free children with permanent dentition?
a) 14%
b) 23%
c) 26%
d) 35%

84. Children aged _____ months are most susceptible to dental fluorosis affecting permanent incisors.
a) 0–6
b) 15–30
c) 36–48

Folia Medica

d) 60–72

85. What amount of fluoridated toothpaste is recommended for children under three years old?

a) Pea-sized

b) Smear or rice-sized

c) Two pumps

d) Full brush coverage

86. Why was the fluoride concentration in U.S. drinking water revised to 0.7 ppm F in 2015?

a) To increase enamel hardness

b) To balance caries prevention with reduced fluorosis risk

c) To align with international standards

d) To reduce production costs

87. What is the primary purpose of silver diamine fluoride (SDF)?

a) Whiten teeth

b) Arrest cavitated caries lesions

c) Replace dental fillings

d) Reduce tooth sensitivity

88. Why are fluoride supplements cautiously recommended for children under six?

a) They cause tooth discoloration

b) Multiple dietary fluoride sources increase fluorosis risk

c) They are ineffective in caries prevention

d) They require parental supervision

89. A child lives in an area with 0.4 ppm F in water and is at high caries risk. What fluoride supplement dose is recommended for a 4-year-old?

a) 0 mg

b) 0.25 mg

c) 0.50 mg

d) 1.00 mg

90. A study found maternal fluoride intake of 1 mg/day during pregnancy was associated with lower IQ in boys. What is the key limitation of this study?

a) Did not adjust for socioeconomic status or maternal IQ

b) Used outdated fluoride measurement tools

c) Excluded girls from the analysis

d) Focused only on permanent teeth

References

1. Kaya E, Gunec HG, Aydin KC, et al. A deep learning approach to permanent tooth germ detection on pediatric panoramic radiographs. Imaging Sci Dent 2022; 52(3):275–81. doi: 10.5624/isd.20220050
2. Bağ İ, Bilgir E, Bayrakdar İŞ, et al. An artificial intelligence study: automatic description of anatomic landmarks on panoramic radiographs in the pediatric population. BMC Oral Health 2023; 23(1):764. doi: 10.1186/s12903-023-03532-8
3. Naeimi SM, Darvish S, Salman BN, et al. Artificial intelligence in adult and pediatric dentistry: a narrative review. Bioengineering (Basel) 2024; 11(5):431. doi: 10.3390/bioengineering11050431
4. Alessa N. Application of artificial intelligence in pediatric dentistry: a literature review. J Pharm Bioallied Sci 2024; (Suppl 3):S1938–S40. doi: 10.4103/jpbs.jpbs_74_24
5. Acharya S, Godhi BS, Saxena V, et al. Role of artificial intelligence in behavior management of pediatric dental patients—a mini review. J Clin Pediatr Dent 2024; 48(3):24–30.
6. Liu L, Wu W, Zhang SY, et al. Dental caries prediction based on a survey of the oral health epidemiology among the geriatric residents of Liaoning, China. Biomed Res Int 2020; 2020:5348730. doi: 10.1155/2020/5348730
7. Lee S, Kim D, Jeong HG. Detecting 17 fine-grained dental anomalies from panoramic dental radiography using artificial intelligence. Sci Rep 2022; 12(1):5172. doi: 10.1038/s41598-022-09083-2
8. Kök H, Acilar AM, İzgi MS. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. Prog Orthod 2019; 20 (1):41. doi: 10.1186/s40510-019-0295-8
9. Farhadian M, Salemi F, Saati S, et al. Dental age estimation using the pulp-to-tooth ratio in canines by neural networks. Imaging Sci Dent 2019; 49 (1):19–26. doi: 10.5624/isd.2019.49.1.19
10. Wang Y, Hays RD, Marcus M, et al. Developing children's oral health assessment toolkits using machine learning algorithm. JDR Clin Trans Res 2020; 5(3):233–43. doi: 10.1177/2380084419885612
11. Abu Arqub S, Al-Moghrabi D, Allareddy V, et al. Content analysis of AI-generated (ChatGPT) responses concerning orthodontic clear aligners. Angle Orthod 2024; 94(3):263–72. doi: 10.2319/071123-484.1
12. Buldur M, Sezer B. Can artificial intelligence effectively respond to frequently asked questions about fluoride usage and effects? A qualitative study on ChatGPT. Fluoride 2023; 56(3):201–16.
13. Maltarollo TFH, Shitsuka C, Sivieri-Araújo G. ChatGPT and its use in pediatric dentistry. Eur Arch Paediatr Dent 2023; 24(5):675–6.
14. Dermata A, Arhakis A, Makrygiannakis MA, et al. Evaluating the evidence-based potential of six large language models in pediatric dentistry: a comparative study on generative artificial intelligence. Eur Arch Paediatr Dent 2025. doi: 10.1007/s40368-025-01012-x
15. Yurdakurban E, Topsakal KG, Duran GS. A comparative analysis of AI-based chatbots: Assessing data quality in orthognathic surgery-related patient information. J Stomatol Oral Maxillofac Surg 2024; 125(5):101757. doi: 10.1016/j.jormas.2023.101757
16. Kusaka S, Akitomo T, Hamada M, et al. Usefulness of generative artificial intelligence (AI) tools in pediatric dentistry. Diagnostics (Basel) 2024; 14(24):2818. doi: 10.3390/diagnostics14242818
17. Anderson LW, Krathwohl DR. A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York (NY): Longmans; 2001.
18. Adams NE. Bloom's taxonomy of cognitive learning objectives. J Med Libr Assoc 2015; 103(3):152–3. doi: 10.3163/1536-5050.103.3.010
19. Giannakopoulos K, Kavadella A, Salim AA, et al. Evaluation of the performance of generative AI large language models ChatGPT, Google Bard, and Microsoft Bing Chat in supporting evidence-based dentistry: comparative mixed methods study. J Med Internet Res 2023; 25:e51580. doi: 10.2196/51580
20. Tussie C, Starosta A. Comparing the dental knowledge of large language models. Br Dent J 2024. doi: 10.1038/s41415-024-8015-2
21. Künzle P, Paris S. Performance of large language artificial intelligence models on solving restorative dentistry and endodontics student assessments. Clin Oral Investig 2024; 28(11):575. doi: 10.1007/s00784-024-05968-w
22. Makrygiannakis MA, Giannakopoulos K, Kaklamanos EG. Evidence-based potential of generative artificial intelligence large language

models in orthodontics: a comparative study of ChatGPT, Google Bard, and Microsoft Bing. Eur J Orthod 2024; cjae017. doi: 10.1093/ejo/cjae017

23. Nguyen HC, Dang HP, Nguyen TL, et al. Accuracy of latest large language models in answering multiple-choice questions in dentistry: a comparative study. PLoS One 2025; 20(1):e0317423. doi: 10.1371/journal.pone.0317423
24. Jones EP. Use of large language model (LLM) to enhance content and structure of a school of dentistry LibGuide. J Med Libr Assoc 2025; 113(1):96–7. doi: 10.5195/jmla.2025.2084
25. Ahmed WM, Azhari AA, Alfaraj A, et al. The quality of AI-generated dental caries multiple-choice questions: a comparative analysis of ChatGPT and Google Bard language models. Heliyon 2024; 10(7):e28198. doi: 10.1016/j.heliyon
26. Chatzopoulos GS, Koidou VP, Tsalikis L, et al. Large language models in periodontology: assessing their performance in clinically relevant questions. J Prosthet Dent 2024; S0022-3913(24)00714-5. doi: 10.1016/j.prosdent.2024.10.020
27. Bayraktar Nahir C. Can ChatGPT be a guide in pediatric dentistry? BMC Oral Health 2025; 25(1):9. doi: 10.1186/s12903-024-05393-1