

The Use of ChatGPT for Creating a Child Abuse Pediatrics Curriculum

Aysha Malawaraarachchi, BS¹, Dana Kaplan, MD, FAAP², Alicia Pekarsky, MD, FAAP³, Kristin Razawich, PNP³, Heather Ross, MD, FAAP³, Elizabeth Simmerman, MD³, Ann S. Botash, MD, FAAP³



UPSTATE
MEDICAL UNIVERSITY

¹SUNY Upstate Medical University, Norton College of Medicine; ²Northwell Health, Department of Pediatrics; ³SUNY Upstate Medical University, Department of Pediatrics

Introduction

- Use of the artificial intelligence open conversation service ChatGPT in medical education as a guide to creating learning objectives and assessment materials has been studied in fields such as biochemistry, pharmacology, and clinical practice.¹⁻³
- We aim to explore the use of ChatGPT in creating child abuse pediatrics board-style learning objectives and corresponding clinical vignette-type multiple choice questions using the American Board of Pediatrics (ABP) child abuse pediatrics content outline consisting of 16 overarching domains with corresponding topics.⁴
- Learning objectives and corresponding multiple-choice questions (MCQs) were generated by ChatGPT and reviewed by expert faculty in the field of child abuse pediatrics to assess for accuracy and utility in medical education.

Methods

Phase 1: ChatGPT Learning Objectives

Domains (16) and associated topics (85) per the ABP content outline for child abuse pediatric board examination informed input into ChatGPT along with this phrase: "In the context of child abuse, please create a learning objective for this topic..." ChatGPT generated multiple objectives per topic.

Phase 3: Collaborative Review

Reviewers asynchronously determined agreement with revised objectives, suggesting further edits. Reviewers self-timed their work for each objective. Final objectives were reviewed, discussed, and edited synchronously.

Phase 2: First Review

Six reviewers were assigned two domains each. Reviewers chose two objectives for each topic. Time for review and editing was self-tracked. Accuracy scale (0-5) where 0 = no changes, 1 = grammatical, 2 = format, 3 = accuracy, 4 = two or more of 1-3, 5 = does not work

Phase 4: ChatGPT MCQs

Phrase entered into ChatGPT: "Please provide a clinical vignette type of multiple-choice question that will assess the learner's ability to accomplish the objective..." Reviewers synchronously discussed appropriateness of generated MCQs and learning objectives.

Example from Domain 1/Topic 1: Head and Spine Trauma/Skull Fractures

Phase 1: ChatGPT generated 8 learning objectives for this topic. Example: "Identify the key indicators and signs of skull fractures in children, distinguishing them from accidental injuries or medical conditions."

**Phase 2:
Asynchronously
Revised Objective
(1 of 2).**
Distinguish skull fractures from suture lines in an infant
Total Time: 1:36.29
Avg. Accuracy: 3
(two or more issues)

**Phase 3:
Collaborative
Review.**
Accepted objective with no changes.

Results

Phase 4: ChatGPT generated MCQ

A 6-month-old infant is brought to the emergency department due to a fall from a bed. The parents are concerned about a head injury. On examination, you carefully assess the infant's skull. Which of the following findings is most indicative of a skull fracture rather than a normal suture line?
A. Widening of the anterior fontanelle
B. Palpable bony ridges along the sagittal suture
C. Depressed area with crepitus on palpation
D. Soft and non-tender swelling along the coronal suture

Revised MCQ (Generated through more specific input to ChatGPT with reviewer revisions: A 3-month-old infant is brought to the emergency department after a minor fall from a height of approximately 2 feet. The infant is crying but consolable, and physical examination reveals a small bump on the head with no other apparent injuries. The parents are concerned about a fracture. A skull X-ray is obtained in the ED. The radiologist notes a linear lucency on the skull X-ray. Which of the following features most accurately distinguishes a skull fracture from a suture line in this infant?
A. Suture lines are more radiolucent than fractures.
B. Suture lines have serrated or zigzag edges, whereas fractures are typically straight.
C. Fractures are accompanied by significant soft tissue swelling, but suture lines are not.
D. Fractures show diastasis (separation of bone), while suture lines do not.

- ChatGPT generated 625 total objectives for 85 topics. Results from the first review phase of all 85 topics showed an average time of review per topic of 70.64 seconds with an average accuracy of 1.78 per objective. The total time spent was 1.88 hours.
- In the first review, Domain 1 (7 topics with resulting 62 objectives) had an average accuracy of 0.63 per objective and average time of review of 71.75 seconds per topic. The total time spent by the reviewer assigned to Domain 1 was approximately 7.17 minutes.
- Asynchronous collaborative review for Domain 1 resulted in changes to 8/11 objectives with an average review time of 36.68 seconds per reviewer for each topic. The total time spent was approximately 31.79 minutes by all six reviewers. Collaborative group discussions were not timed.

Discussion

- Initial asynchronous review of generated objectives demonstrated high accuracy and objectives underwent minimal changes.
- Collaborative synchronous review and discussion resulted in revisions to content and grammar for the majority of Domain 1 objectives. Revisions impacted:
 - Clarity (grammatical changes)
 - Bloom's verbs (too basic)
 - Content gaps (too simplistic)
 - Removal of repetitive language and repeat objectives amongst topics. Example above demonstrated frequent use of accident versus abuse scenario.
- MCQs often failed to incorporate key learning points, requiring re-revision of objectives. Responses required more nuance, including prioritization of knowledge.
- Some MCQs responses by ChatGPT were viewed as incorrect; reviewers recognized that lack of use of clinical images limit question content.
- Limitations:
 - Differences amongst the experts regarding training (four child abuse pediatrician (CAP) board certified/eligible faculty members, one CAP fellow, one PNP) may have influenced revisions and time required.
 - Asynchronous review may have influenced reviewer opinions, as revised objectives were visible to those who reviewed later.
- Future plans include collaborative development of more specific prompts to accompany objectives and generation of corresponding MCQs through ChatGPT.

References

1. Al-Worafi YM, Goh KW, Hermansyah A, Tan CS, Ming LC. The Use of ChatGPT for Education Modules on Integrated Pharmacotherapy of Infectious Disease: Educators' Perspectives. *JMIR Med Educ.* 2024;10:e47339. Published 2024 Jan 12. doi:10.2196/47339
2. Han Z, Battaglia F, Udaiyar A, Fooks A, Terlecky SR. An explorative assessment of ChatGPT as an aid in medical education: Use it with caution. *Med Teach.* Published online October 20, 2023. doi:10.1080/0142159X.2023.2271159
3. 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;9:e50658. Published 2023 Dec 22. doi:10.2196/50658
4. ABP Content Outline for Child Abuse Pediatrics (March 2024). <https://www.abp.org/sites/public/files/pdf/content-outline-child-abuse-pediatrics.pdf>.

Acknowledgements: ChatGPT, developed by OpenAI; JoAnne Race, MSc, and Trish Booth, MA, who support the Child Abuse Medical Provider (CHAMP) team for development of this project. This research was supported by the New York State Office of Children and Family Services, award C029711.