

Exploring the Role of ChatGPT in Enhancing Multiple-Choice Questions for Pharmacy Education



PHARMACY

Benjamin Shultz, Ph.D. and Jeffrey Mucksavage, Pharm.D.
University of Illinois Chicago College of Pharmacy (UIC COP)



KEY TAKEAWAYS

- ChatGPT aids item revision and provides useful templates for item construction. It shows promise in item revision but requires expert oversight.

OBJECTIVES

- Evaluate the effectiveness of ChatGPT in revising multiple-choice exam questions in a therapeutics course
- Highlight the strengths and weaknesses of ChatGPT as a revision tool for pharmacy education

METHODS

- Zero-shot GPT-4 prompts¹
- New session each item²
- Six criteria for item generation
 - Year of Study
 - Bloom's Taxonomy Level
 - Target Knowledge Area
 - Preferred Question Format
 - Possible Answers
 - Additional Instructions
- Items sent to original writers (n=11)
- Documented feedback
- Manually revised as needed

RESULTS

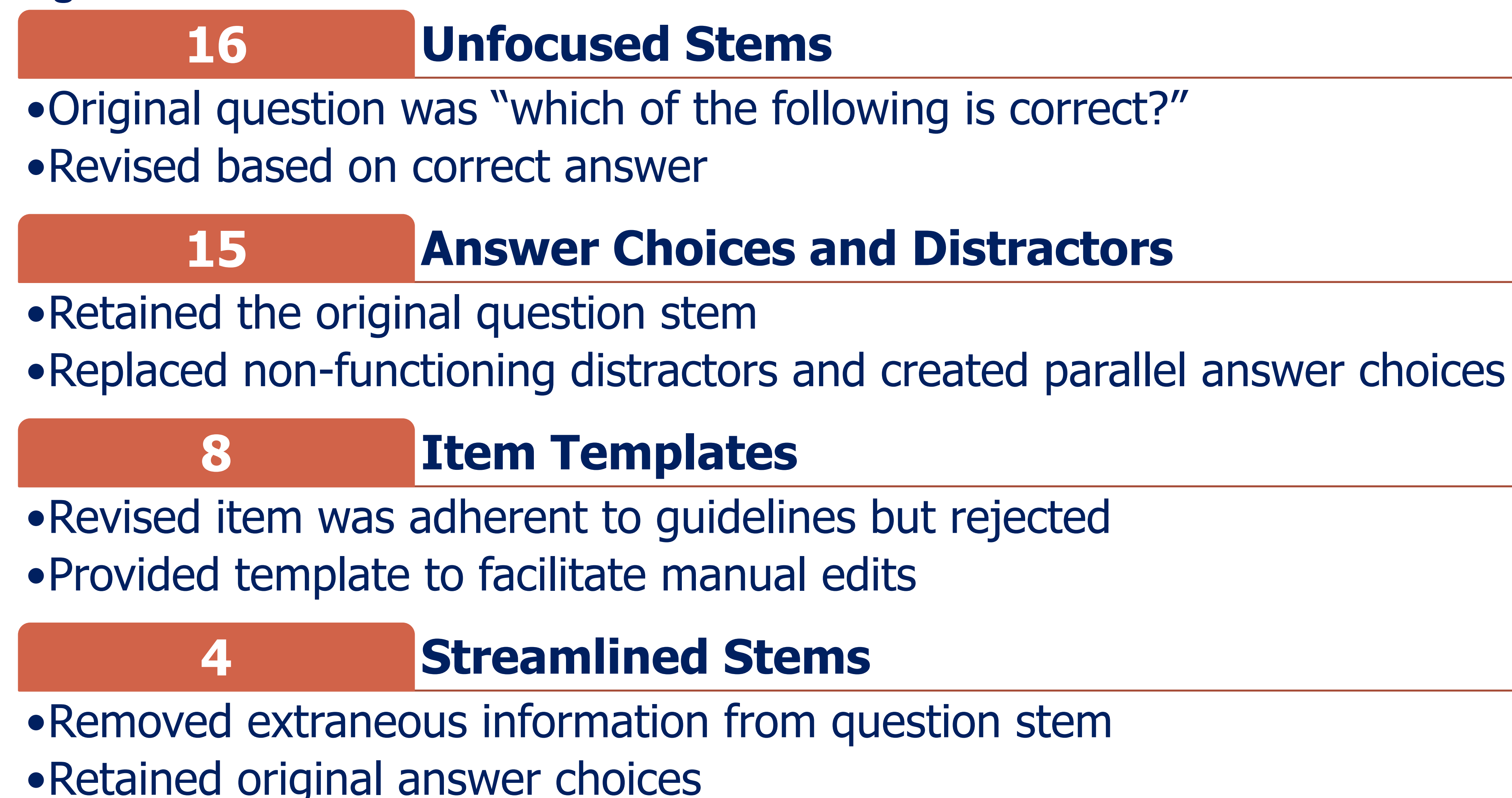
Figure 1: Item Bank Revisions



Table 1: Examples of Revised Items

Original	GPT-4 Revised
Which of the following statements is correct regarding the treatment of hyponatremia?	What is the recommended maximum increase of serum sodium in the treatment of hyponatremia within a 24-hour period?
Which of the following is true when converting between the calcimimetics?	What is the appropriate time frame to switch between etelcalcetide and cinacalcet?

Figure 2: Revisions to Items



FINDINGS

Enhanced Clarity: ChatGPT effectively revised unfocused question stems and aligned them with intended learning objectives

SME Review Essential: AI-generated items require vetting by subject matter experts to ensure accuracy, relevance, and alignment with the instructor's intent.

Limitations with Distractors: ChatGPT often struggled to generate plausible distractors, sometimes producing irrelevant or incorrect options.

Effective Rewording: The AI excelled at rewording text passages, addressing common item flaws like parallel distractors, and providing useful templates even when outputs were inaccurate.

Faculty Development Aid: ChatGPT's strengths in item revision can supplement faculty training programs, highlighting best practices in item construction.

REFERENCES

1. Götz FM, Stieger S, Reips UD. Let the algorithm speak: How to use neural networks for automatic item generation in psychological scale development. *Psychol Methods*. 2023.
2. Reynolds L, McDonell K. Prompt programming for large language models: Beyond the few-shot paradigm. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021.