

33 Do More Qbank Questions Earn You a Higher ITE Score?

Ashley London, Kestrel Reopelle, Timothy Friedmann, Szymon Kutyla, Delaney Bates

Background: A recent decline in American Board of Emergency Medicine Qualifying Exam (ABEM QE) pass rates has driven postgraduate educators to reassess their curricula and reliability of preparation resources. Residency programs nationwide have used the ABEM-administered in-training exam (ITE) scores as an indicator of ABEM QE performance. To facilitate ITE preparation, residency programs often assign practice questions or offer residents access to a test preparation question bank. There is limited evidence to suggest that the number of questions completed correlates with improved ITE scores.

Objectives: This study aimed to determine if there is a correlation between question bank utilization and ITE scores. We hypothesized that completing greater numbers of practice questions will correlate with higher ITE scores.

Methods: This was a retrospective observational cohort study conducted at an urban academic EM residency program. For all residents in 2023-2025, the number of questions completed and the raw ITE scores were collected, de-identified, and collated for the study team. Residents who did not take the ITE were excluded.

Results: 182 resident scores were included in the study; 10 residents excluded. A linear regression analysis of the number of practice questions completed and raw ITE scores demonstrated a weak positive correlation, albeit with a high degree of variation and poor overall fit ($p=0.032$, $R^2=0.0253$) (Figure 1). The number of practice questions completed were broken up into quintiles and an ANOVA analysis was performed, which showed no statistical difference in ITE scores between the five groups ($p=0.0753$) (Figure 2).

Conclusions: The number of practice questions completed prior to ITE is weakly correlated to the score

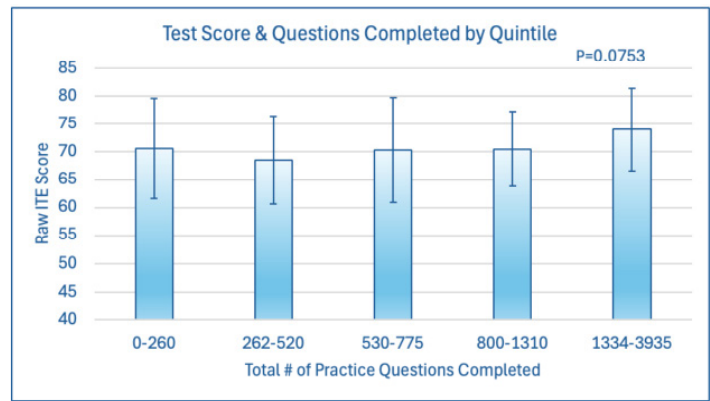


Figure 2: Raw ITE Test Scores and Number of Practice Questions Completed By Quintile. Error bars represent one standard deviation.

residents receive on ITE. The high degree of variation found in the results suggests that other factors are involved in an individual's ITE performance. Volume of questions alone should not be relied upon to improve resident performance.

34 Let's Chat about Large Language Models for Faculty Feedback: A Multi-Institution Feasibility Pilot

Hannah Mishkin, Robyn Hoelle, Danielle Von Nieda, Melody Beltran

Background: Providing faculty feedback (FB) that is anonymous, specific, constructive, and actionable is a core requirement of the Accreditation Council for Graduate Medical Education. Synthesizing qualitative resident comments consistently and without bias can be challenging for Program Directors. Large language models (LLMs) offer a potential tool to summarize resident evaluations, reduce unintended editorial influence, enhance perceived anonymity, and frame feedback using supportive, growth-oriented language.

Objective: To assess faculty perceptions of the learner's feedback summarized by an LLM with regard to clarity, anonymity, and usefulness.

Methods: We developed and piloted a process that used LLMs to generate anonymized summaries of resident evaluations for faculty across two emergency medicine residency programs. Resident comments were de-identified and processed through an iteratively refined prompt that produced a concise summary emphasizing key strengths, one growth opportunity, and a specific actionable educational goal for the coming year. Representative anonymous resident quotes were included to increase specificity, and all outputs were reviewed only for accuracy or professionalism before distribution as part of the annual review process.

Results: Twenty-one faculty completed an anonymous survey evaluating the LLM-generated FB summaries.

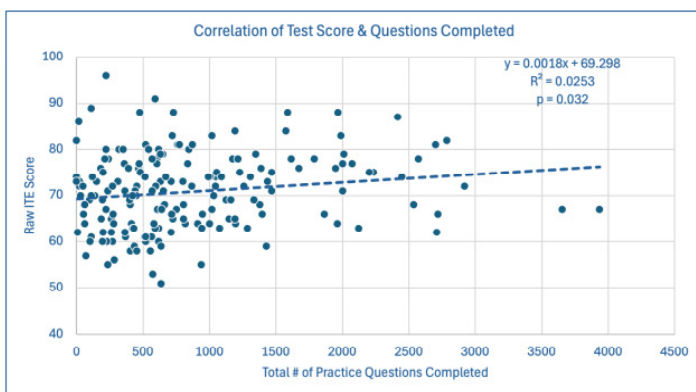


Figure 1: A Correlation of Raw ITE Test Scores and Number of Practice Questions Completed