UC Irvine

Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health

Title

A Comparison of Two Prediction Models for the American Board of Emergency Medicine Qualifying Exam

Permalink

https://escholarship.org/uc/item/0sk33889

Journal

Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health, 27(2.1)

ISSN

1936-900X

Authors

Reopelle, Kestrel
Kutyla, Szymon
Bates, Delaney
et al.

Publication Date

2026-03-26

DOI

10.5811/westjem.64087

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

eScholarship.org Powered by the California Digital Library
University of California

16 A Virtual Reality Approach to Standardizing Mass-Casualty Training across Residency Programs

Cosimo Laterza, Stephen Powell, Michael Berkenbush, Nicholas Cozzi, Matthew Steenberg, Nicholas Kman

Background: Mass-casualty incident (MCI) training is limited by logistical and financial barriers. The ACGME requires education in disaster preparedness, MCI triage, simulation-based training, and systems-based practice. Virtual reality (VR) offers a scalable method that standardizes instruction across programs.

Objectives: To evaluate accuracy, efficiency, and consistency in a VR MCI using SALT triage across geographically distinct EM residencies. We hypothesized that VR would provide uniform assessment across sites and demonstrate reliable performance metrics with expected relationships between accuracy and error patterns.

Methods: We conducted a prospective observational multi-site pilot across four academic EM residency programs in fall 2024. Forty-six residents participated via convenience sampling (n=19,12,9,6), none were excluded. PGY1–3 residents completed a Stop-the-Bleed and SALT refresher prior to the session. Participants then triaged and managed 14 simulated patients in a standardized VR simulation. One trained proctor oversaw all sessions. Simulator log data were reviewed by two statisticians. Descriptive statistics were reported as means(SD). Pearson correlations assessed associations, and one-way ANOVA compared sites.

Results: Mean triage accuracy was 71.8%(11.7) with no site differences(p=0.18). Time to triage was 566s(159) and hemorrhage control time 426s(145), without site differences(p=0.76,p=0.97). Accuracy correlated negatively with total errors(r=−0.80,p<0.001), over-triage(r=−0.50,p<0.001), and under-triage(r=−0.60,p<0.001). Hemorrhage control time showed a trend-level correlation with total errors(r=0.29,p=0.078).

Conclusions: VR MCI training produces consistent performance metrics across EM programs in different states and aligns with ACGME disaster training requirements. Expected accuracy–error correlations support construct validity. Limitations include small sample size, one platform, and no follow-up. These findings support VR MCI simulation as a scalable method to meet disaster education requirements and a practical alternative to resource-intensive MCI drills.

17 A Comparison of Two Prediction Models for the American Board of Emergency Medicine Qualifying Exam

Kestrel Reopelle, Szymon Kutyla, Delaney Bates, Timothy Friedmann

Background: Recently, the American Board of Emergency Medicine Qualifying Exam (ABEM QE) pass rate has declined, forcing post graduate educators to re-evaluate their preparatory curricula as well as the reliability of common preparatory resources and indicators. The ABEM-administered In-Training Exam (ITE) and commercially available question banks (Qbanks) are used for formative assessment by residencies and board eligible EM physicians to guide studying and to predict QE outcomes.

Objectives: This study aimed to determine whether a popular test prep Qbank's QE pass prediction model differs significantly from the ABEM ITE's prediction model. We hypothesized that there would be no difference between the two models for predicted probability of passing the ABEM QE.

Methods: This was a retrospective observational cohort study conducted at an urban academic residency. Qbank dashboard data for EM residents was collected from 2023-2025, correlated to ITE score reports from the same academic years, and de-identified for the study team. Other data collected included resident PGY and number of Qbank questions completed in the given academic year. Residents who did not take the ITE or who did not complete enough Qbank questions to have a reported probability of passing in a given year were excluded.

Results: 177 residents were included in this study and 15 were excluded. The average probability of passing the ABEM QE was 89.2% according to the commercial Qbank's model and 82.4% according to ABEM's model (a statistically significant difference, p<0.001)(figure 1). The significant difference between models persisted for residents with higher Qbank usage (figure 2).

Conclusions: The Qbank's model has a more generous outcome prediction than ABEM's, except at the lowest levels

	PGY1	PGY2	PGY3	PGY4	All Residents
ITE/ABEM reported mean likelihood of passing the ABEM QE:	86.0% chance	77.3% chance	82.4% chance	85.1% chance	82.4% chance
Commercial Qbank reported mean likelihood of passing the ABEM QE:	88.4% chance	90.3% chance	88.5% chance	89.5% chance	89.2% chance
Comparison of above mean values (two-tail paired t-test) :	p=0.314	p<0.001	p=0.003	p=0.01	p<0.001

Figure 1: Comparison of mean probability of passing the ABEM QE by both models, sorted by PGY cohort and total study population

Comparison to mean # of Qbank questions completed:	>1 std dev below	Within 1 std dev below	Within 1 std dev above	>1 std dev above
# of Questions completed over the academic year:	8 - 135 questions	149 - 870 questions	910 - 1616 questions	1674 -3935 questions
ITE/ABEM reported mean likelihood of passing the ABEM QE:	79.4% chance	80.7% chance	85.9% chance	85.6% chance
Commercial Qbank reported mean likelihood of passing the ABEM QE:	86.0% chance	88.11% chance	92.0% chance	90.9% chance
Comparison of above mean values (two tailed paired t-test):	p=0.07	p<0.001	p=0.002	p=0.009

Figure 2: Comparison of mean probability of passing the ABEM QE between both models, organized by number of questions completed in the Qbank

of Qbank usage. Future work should compare these models directly to QE pass rates. This data suggests we should caution trainees to not rely on any single model for formative assessment and study planning.

18 The Resident Effect: Evaluation Inflation for Lower-Performing Students in EM

Xiao Chi Zhang, Alexis Metoyer, Dimitrios Papanagnou, Zachary Kaplan, Chaiya Laoteppitaks, Nina Mingioni, Jiten Patel, Kathleen Cruz, Michael Pasirstein

Background: EM clerkships rely on resident physicians, who serve as approachable, relatable, and readily available near-peer educators for medical students. Their position in the clinical hierarchy can enhance psychological safety and student comfort, but may also raise concerns about their ability to fully address learning objectives or provide fair, competency-based evaluations. Despite the role residents play in the EM clerkship, little is known about how their assessments of medical students compare with those of attending physicians. This gap limits our understanding of the variability of clerkship evaluations and the ability to identify students who may require remediation.

Objective: To compare clinical evaluation scores assigned by resident and attending for medical students completing a 3-week EM clerkship across two academic years, with focus on student performance quartiles.

Methods: We conducted a retrospective analysis of medical student performance data by resident and attending from an urban, multi-site EM clerkship between July 2023 and April 2025. Median evaluation ratings from residents and attendings were compared using paired statistical tests as appropriate. Subgroup analysis examined differences in ratings for the lowest- and highest-performing quartiles of students.

Results: A total of 487 students were analyzed. Resident-based scores were statistically higher than attending-based scores for both academic years, AY23-24 and AY24-25 (88.3% vs. 80.8%, and 86% vs. 80.8%, t-test < 0.05), especially for lower performing (lower 25% quartile) (78.8% vs. 67.5, and 77.6 vs. 66.3%, t-test <0.001). This discrepancy was not observed for upper-quartile students.

Conclusions: Residents were more likely to award higher clinical evaluation scores to lower-performing students in a required EM clerkship. This upward score inflation makes it more challenging to accurately identify students who are struggling or at risk, particularly those in the lower quartile, where early detection is essential for targeted support and remediation. These findings underscore the importance of structured educator training, including the calibration of assessments. If residents are to remain part of clerkship assessment, incorporating these skills into resident-as-teacher curricula is critical.

19 Implementing a Structured Feedback Model to Enhance Medical Student Development during the EM Clerkship

Xiao Chi Zhang, Chaiya Laoteppitaks, Jason Gonzalez, Nina Mingioni, Dimitrios Papanagnou, Samuel Iovine, Joria Le, Michael Pasirstein

Background: EM has become a core clerkship at many medical schools, exposing students to high acuity care in a fast-paced environment. This environment poses a challenge when it comes to providing consistent, actionable feedback to medical students. Formative feedback is critical for guiding students towards the expected clerkship performance goals, but despite its importance, students may not receive timely formative feedback, limiting their opportunities for growth and development.

Objective: To provide a structured post-shift feedback process for students enrolled in a 3-week EM rotation to enhance the quality and frequency of meaningful student feedback.

Methods: We provided a physical feedback card (encased in a retractable badge-reel holder), based on the conceptual framework from the Pendleton feedback, to all third-year medical students rotating through their EM clerkship at a tertiary academic center ED from April to June 2025. Students were provided with verbal instructions on how to solicit feedback using the Pendleton model with their preceptors and were instructed to request a post-shift evaluation using this method after each shift. After each shift, students completed a voluntary likert-scale survey (5=most likely/agree) evaluating the frequency, quality, and satisfaction with their feedback process.

Results: We captured 259 survey data, 102 receiving Pendleton feedback, and 104 non-Pendleton feedback. Students who received Pendleton feedback reported clearer feedback (4.81 vs. 4.61, p < 0.01), more actionable next steps for improvement (4.76 vs. 4.55, p < 0.01), and an improved understanding of their clinical skills (4.75 vs. 4.57, p < 0.01). The Pendleton feedback process was also perceived to be a valuable addition to the rotation (4.52 vs. 4.62, p 0.15).

Conclusions: Students who requested and received structured feedback using the Pendleton method reported increased formative and actionable feedback than traditional feedback. Future research can the applicability of this method outside of the EM rotation to cultivate a supportive and inclusive learning environment where students are empowered to engage in growth-centered dialogue in other academic settings.