Education

CrossMark

FACULTY PREDICTION OF IN-TRAINING EXAMINATION SCORES OF EMERGENCY MEDICINE RESIDENTS: A MULTICENTER STUDY

Amer Z. Aldeen, MD,* Erin N. Quattromani, MD,† Kelly Williamson, MD,‡ Nicholas D. Hartman, MD,§
Natasha B. Wheaton, MD,‖ and Jeremy B. Branzetti, MD¶

*Emergency Medicine Physicians, Ltd., Department of Emergency Medicine, Presence St. Joseph Medical Center, Joliet, Illinois,
†Department of Emergency Medicine, St. Louis University School of Medicine, St. Louis, Missouri, ‡Department of Emergency Medicine,
Advocate Christ Medical Center, Chicago, Illinois, §Department of Emergency Medicine, Wake Forest School of Medicine, Winston-Salem,
North Carolina, ‖Department of Emergency Medicine, University of Iowa Carver College of Medicine, Iowa City, Iowa, and ¶Division of
Emergency Medicine, University of Washington School of Medicine, Seattle, Washington
Reprint Address: Amer Z. Aldeen, MD, Emergency Medicine Physicians, Ltd., Department of Emergency Medicine, Presence St. Joseph
Medical Center, 333 N. Madison Street, Joliet, IL 60435

☐ Abstract—Background: The Emergency Medicine In-Training Examination (EMITE) is one of the few validated instruments for medical knowledge assessment of emergency medicine (EM) residents. The EMITE is administered only once annually, with results available just 2 months before the end of the academic year. An earlier predictor of EMITE scores would be helpful for educators to institute timely remediation plans. A previous single-site study found that only 69% of faculty predictions of EMITE scores were accurate. Objective: The goal of this article was to measure the accuracy with which EM faculty at five residency programs could predict EMITE scores for resident physicians. Methods: We asked EM faculty at five different residency programs to predict the 2014 EMITE scores for all their respective resident physicians. The primary outcome was prediction accuracy, defined as the proportion of predictions within 6% of the actual scores. The secondary outcome was prediction precision, defined as the mean deviation of predictions from the actual scores. We assessed faculty background variables for correlation with the two outcomes. Results: One hundred and eleven faculty participated in the study (response rate 68.9%). Mean prediction accuracy for all faculty was 60.0%. Mean prediction precision was 6.3%. Participants were slightly more accurate at predicting scores of noninterns compared to interns. No faculty background variable correlated with the primary or secondary outcomes. Eight participants predicted scores with high accuracy (>80%). Conclusions: In this multicenter study, EM faculty possessed only moderate accuracy at predicting resident EMITE scores. A very small subset of faculty members is highly accurate. © 2015 Elsevier Inc.

☐ Keywords—residency education; in-training examination; medical knowledge; milestones; competency; assessment

INTRODUCTION

Each February, over 5,000 emergency medicine (EM) residents take the Emergency Medicine In-Training Examination (EMITE), which simulates the American Board of Emergency Medicine (ABEM) qualifying examination. A score of 80% on the EMITE as a postgraduate year 3 (PGY-3) resident predicts a 95% chance of passing the ABEM qualifying examination (1,2). Review of EMITE performance can help residency programs identify a particular resident's strengths and weaknesses in knowledge of core content areas (3). The EMITE is one of the few validated tools for medical knowledge assessment for EM residents (3–6). Other tools, such as direct clinical observation, performance in mock oral examinations, and conference attendance,

do not directly correlate with a resident's medical knowledge(4).

Unfortunately, utilization of the EMITE as a formative assessment tool is limited by its infrequency (taken only once per academic year) and timing (results unavailable until May), thereby delaying the recognition of gaps in a resident's medical knowledge. If EM faculty could accurately predict residents' EMITE scores, then residents at risk for low medical knowledge could be identified earlier each academic year, providing time to institute a remediation plan.

Studies evaluating the ability to predict ITE scores are limited. Jones et al. determined that internal medicine residents do not accurately predict their own ITE scores (7). Hawkins et al. concluded that internal medicine faculty could not accurately predict resident ITE scores on the basis of their clinical assessments (8). A single-site study by Aldeen et al. found that EM faculty possessed only moderate accuracy in predicting ITE scores for their residents (9). It is unknown if these results are generalizable to all EM residency programs. The purpose of this study was to conduct a multicenter trial to evaluate the ability of EM faculty to predict EMITE scores of residents at five different residency programs.

MATERIALS AND METHODS

Study Aim

The goal of this study was to measure the ability of EM faculty physicians to accurately predict the ITE scores of EM resident physicians.

Study Setting

This was a prospective, multicenter trial involving five different EM residency programs. All sites obtained Institutional Review Board approval or exemption. Four sites were PGY 1–3 residencies and one site was a PGY 1–4 residency in its 3rd year of existence.

Study Subjects

Faculty at the five sites were asked to participate via four e-mail solicitations (introduction e-mail plus 3 reminders) sent over 1 week by each respective site investigator. Faculty were eligible to participate if they worked primarily at the main clinical hospital site of the residency program. The following background characteristics were collected on each participant: years of clinical experience after residency, years of clinical experience working with residents, number of monthly clinical hours worked, and status as an educational leader. Educational leader status was defined as residency director (any level), clerkship director (any level), simulation director, or other specific title involving education.

Study Protocol

Data were collected from January 15 to February 1, 2014, about 1 month before the EMITE. Using an online survey instrument (Surveymonkey, Inc., Palo Alto, CA), site investigators asked participants to predict the 2014 EMITE score (0 to 00%) for every resident physician at their respective site. Participants were assigned a unique identifier number known only to the individual participant, the respective site investigator, and the data investigator (the one member of the study team whose site did not participate in the study). Participants were provided the 2013 national median EMITE scores for each PGY class as a reference. After February 1, the survey was closed to data collection and access to the database was restricted to the data investigator to minimize bias. In early May, the ITE scores returned and were reported by each site investigator to the data investigator to analyze.

Outcomes

The primary outcome was prediction accuracy, defined as the proportion of correct predictions (difference between predicted score and actual score $\leq 6\%$) by each faculty member. The range of $\pm 6\%$ was consistent with the protocol used in the prior, single-site study of EMITE prediction (9). Secondary outcomes included prediction precision (mean difference between the predicted scores and actual scores), as well as correlations between background variables and prediction accuracy. Individual site data for residency size, faculty size, and ITE scores were not reported in order to maintain optimal confidentiality for the residency programs.

Data Analysis

Prediction accuracy was found to be relatively normally distributed data, so means were used with 95% confidence intervals. Spearman's coefficient was calculated for correlations between education leader status and prediction accuracy. Statistics were performed using Stata (version 11, College Station, TX).

RESULTS

Over the five sites, 111 (out of 161 eligible) total faculty physicians participated in the study, rendering 3,219 predictions for 147 residents. The aggregate response rate was 68.9% (95% confidence interval [CI] 61.4 to 75.6). See Table 1 for aggregate and site-specific background characteristics. Six total data points were discarded,