Results: 1775 applicants (n=3690 SLOEs) were examined. 1216 SLOEs were from students' home institutions; 2368 were from away rotations. This totaled 3584 included SLOEs. 106 SLOEs were excluded due to inability to identify home or away. Table 1 demonstrates the mean scores, standard deviations, and p-values for home vs away rotation SLOEs. Only C3, anticipated rank list (RL) position (p=0.0017) was statistically significant in favor of higher rank for home SLOEs.

Conclusions: This study demonstrated that most of the mean scores on the SLOE 2.0 were not statistically significant between home vs away institutions. The higher scores on the RL questions on home SLOE 2.0s was the only score signifying statistical significance compared to away SLOE 2.0.

Table 1. Mean and standard deviation for each SLOE 2.0 question for home and away rotations.

Question	HOME Mean (SD)	AWAY Mean (SD)	P-Value
A1 Ability to perform a focused history and physical exam (1-3)	2.75 (0.45)	2.72 (0.48)	0.0705
A2 Ability to generate a differential diagnosis (1-3)	2.55 (0.55)	2.52 (0.55)	0.1222
A3 Ability to formulate a plan(1-3)	2.47 (0.55)	2.42 (0.56)	0.0109
A4 Ability to perform common ED procedure (1-3)	2.39 (0.81)	2.38 (0.79)	0.7221
A5 Ability to recognize and manage basic emergent situations (1-3)	2.61 (0.52)	2.56 (0.55)	0.0087
B1 Compassion, sensitivity, and respect towards patients and team members (1-5)	4.30 (0.74)	4.32 (0.72)	0.4355
B2 Receptivity to feedback and ability to incorporate feedback (1-5)	4.27 (0.73)	4.25 (0.76)	0.4498
B3 Dependability, responsibility, initiative, and work ethic (1-5)	4.35 (0.73)	4.34 (0.77)	0.7080
B4 Punctuality, attendance, and preparation for duty (1-5)	4.32 (0.73)	4.34 (0.76)	0.4498
B5 Timeliness and responsiveness in completing administrative tasks (1-5)	4.21 (0.80)	4.20 (0.78)	0.7187
B6 Interpersonal and communication skills with patients and family members. (1-5)	4.31 (0.71)	4.31 (0.73)	0.9709
B7 Interpersonal and communication skills with faculty, residents and healthcare professionals. (1-5)	4.28 (0.78)	4.25 (0.81)	0.2879
C1 Anticipated Guidance (1-4)	3.20 (0.71)	3.16 (0.73)	0.1171
C3 Rank List (0-4)	2.78 (0.89)	2.68 (0.91)	0.0017

SD, standard deviation

2 ChatGPT Editing Effects on Emergency Medicine Residency Personal Statements

Mark Chesebro, Kristen Whitworth, Matthias Barden, Jesse Kellar, Donna Okoli, Christian Kolacki, Barbara Blasko, Matthew Hysell

Background: Use of artificial intelligence (AI) is increasing historically. Potential effects of AI on the emergency medicine residency application process are unknown.

Objective: To determine if reviewers favor personal statements edited by AI.

Methods: We asked AI system ChatGPT to "Edit a personal statement for an emergency medicine program" of ten application essays used by graduated residents. Faculty from six emergency medicine residencies performed blinded review of both original and edited personal statements. Reviewers first recorded which essay they subjectively favored then scored the essays with an objective rubric. This rubric used an anchored one to four scale for domains of clear focus, organization, grammar, creativity/voice, and vividness of reflection such that an essay could maximally score 20 and minimally 4. Chi-squared testing was used to analyze subjective preference, and paired t-test to examine the essays' objective scoring.

Results: Six reviewers averaging 5 years of experience reviewing applications (ranging from 2 through 11 years) reviewed 10 pairs of essays. Overall, in 4/10 (40%) of essay pairs, reviewers preferred the ChatGPT edited version. In 3/10 (30%) there was no preference, and in 3/10 (30%) reviewers preferred the original version. The ChatGPT version was preferred in reviewers' individual responses in 34/60 (57%) of instances, chi-square p=0.144. Eight out of ten essays received a higher rating using the rubric after ChatGPT editing. The ten essays' rubric scores increased from a mean of 13.0 (SD 1.9) unedited to 14.2 (SD 1.8), p=0.028, CI 0.17 – 2.31, after editing. There was 90% agreement between the subjective and objective analysis of each pair of essays, representing substantial agreement, Cohen's kappa 0.793.

Conclusions: Use of ChatGPT to edit essays did increase their scoring on an objective rubric, however reviewers' subjective review of essays was less impacted by ChatGPT editing.

3 EM Resident Clinical and Communication Performance on Simulated Resuscitations is not Correlated when Stratified by Gender

Bryan Kane, Diane Begany, Matthew Cook, Nicole Elliot, Michael Nguyen

Background: Prior papers evaluated multi-source feedback (MSF) and communication of EM residents managing a high-fidelity simulation (sim) case.

Objective: We seek to determine if, based on gender of the team leader, a correlation exists between clinical performance and consultant communication.

Methods: This IRB approved secondary analysis of enrolled EM residents from a PGY 1-4 program reported gender as male or female. Both sims were toxic ingestions. MSF feedback was generated using a Queens Simulation

Assessment Tool (QSAT) from self-evaluation, a junior resident, an EMS provider, nursing, and two EM faculty. In both sims communication to a toxicologist and intensivist were measured using the 5C's model. The summed QSAT and 5C scores were correlated using Pearson's correlation coefficient with Fisher's z transformation; interpreted as weak (<0.3), moderate (0.3-0.7) and strong (>0.7). Significance was set at 0.05. Positive correlation indicates synchronous movement of scores, negative correlation asynchronous movement.

Results: Table 1 presents 32 ACLS sims. There were moderate positive correlations between all MSF and averaged consultant 5Cs [r=0.412, 95% CI (-0.011, 0.710)] in males, and between average faculty QSAT and intensivist 5C [r=0.589, 95% CI (-0.198, 0.914)] in females. The remaining correlations were weak. 34 residents led the PALS sim (Table 2). Surprisingly, there was a moderate negative correlation between the average attending QSAT score and the Intensivist 5C score in males [r=-0.390, 95% CI (-0.697, 0.038)]. The remaining correlations were weak. All correlations in both sims lacked significance.

Table 1. Correlation of QSAT and 5C's Score in adult simulations stratified by resident gender.

QSAT Metric	5C's Metric	Gender	n	Standard Correlation Coefficient (*r*)[a]	Fisher's z Transformed Coefficient (*zr*) (95% CI)[b]	p-value[c]
Average (All Raters)	**Average (Toxicologist & Intensivist)**	**Male**	**22**	**0.412**	**0.438 (-0.011, 0.710)**	**0.0561**
		Female	**8**	**0.103**	**0.103 (-0.649, 0.753)**	**0.8171**
Average (Faculty Only)	Average (Toxicologist & Intensivist)	Male	22	0.190	0.193 (-0.252, 0.566)	0.4014
		Female	8	-0.248	-0.254 (-0.811, 0.553)	0.5708
Average (Attendings Only)	**Toxicologist Only**	**Male**	**22**	**0.178**	**0.180 (-0.263, 0.558)**	**0.4323**
		Female	**10**	**0.056**	**0.056 (-0.595, 0.662)**	**0.8827**
Average (Attendings Only)	Intensivist Only	Male	22	0.067	0.067 (-0.365, 0.475)	0.7693
		Female	8	0.589	0.676 (-0.198, 0.914)	0.1308

[a]4 assessments were missing either the Tox or Int 5C's score, therefore the average score is also missing, which changes the n depending upon the correlation pairing.
[b]Pearson correlation coefficient
[c]Fisher's z transformed Pearson correlation coefficient
[d]p-value corresponds to the Fisher's z transformed correlation coefficient and 95% CI

Conclusions: In this single site cohort, stratified by team lead resident gender, clinical and communication sim performance do not appear correlated. While there were isolated moderate correlations, they were mixed. This suggests that regardless of gender, clinical performance and communication should be independently evaluated.

Table 2. Correlation of QSAT and 5C's Score in pediatric simulations stratified by resident gender.

QSAT Metric	5C's Metric	Gender	n	Sample Correlation Coefficient (*r*)[a]	Fisher's z Transformed Coefficient (*zr*) (95% CI)[b]	p-value[c]
Average (All Raters)	Average (Toxicologist & Intensivist)	Male	19	0.166	0.168 (-0.311, 0.577)	0.5016
		Female	12	0.006	0.006 (-0.570, 0.578)	0.9852
Average (Attendings Only)	Average (Toxicologist & Intensivist)	Male	19	-0.040	-0.040 (-0.485, 0.422)	0.8736
		Female	12	0.149	0.150 (-0.465, 0.666)	0.6522
Average (Attendings Only)	Toxicologist Only	Male	20	0.281	0.289 (-0.185, 0.643)	0.2341
		Female	12	0.090	0.090 (-0.510, 0.631)	0.7867
Average (Attendings Only)	Intensivist Only	Male	22	-0.390	-0.412 (-0.697, 0.038)	0.0727
		Female	12	0.160	0.161 (-0.456, 0.672)	0.6293

[a]4 assessments were missing either the Tox or Int 5C's score, therefore the average score is also missing, which changes the n depending upon the correlation pairing.
[b]Pearson correlation coefficient
[c]Fisher's z transformed Pearson correlation coefficient
[d]p-value corresponds to the Fisher's z transformed correlation coefficient and 95% CI

4 Describing Preliminary Data on Scoring Using the Standardized Letter of Evaluation (SLOE) 2.0 Format

Aman Pandey, Sharon Bond, Sara Krzyzaniak, Teresa Davis, Cullen Hegarty, Kasia Gore, Thomas Beardsley, Sandra Monteiro, Al'ai Alvarez, Melissa Parsons, Michael Gottlieb, Alexandra Mannix

Background: The Standardized Letter of Evaluation (SLOE) is a very important part of an emergency medicine (EM) bound student's application. The SLOE helps provide objective data on students' performances on EM rotations and helps residency programs screen applicants. The SLOE 2.0 introduced changes to the SLOE and so far there is no data to understand distribution of scores using the SLOE 2.0.

Objective: The objective of this study was to describe the initial distribution of scores on the SLOE 2.0.

Methods: This study was a multi-institution, retrospective cross-sectional study using SLOE 2.0 data from the 2022-2023 application cycle from 5 geographically distinct EM programs across the United States. SLOEs from 4-week EM electives were included and duplicate SLOEs from the 5 institutions were excluded. Also excluded were subspecialty or OSLOEs, SLOEs not written by a faculty group of other qualified person, SLOEs from letter writers that wrote <5 SLOEs last year, or SLOEs with incomplete data. Since Part A and Part C were qualitative questions, they had to be converted to a quantitative point system. We assessed the means, medians, and distribution of scores for