

“Best of the Best” Research Abstracts

1 Increased Use of Generative Artificial Intelligence-Associated Language in Emergency Medicine Residency Personal Statements

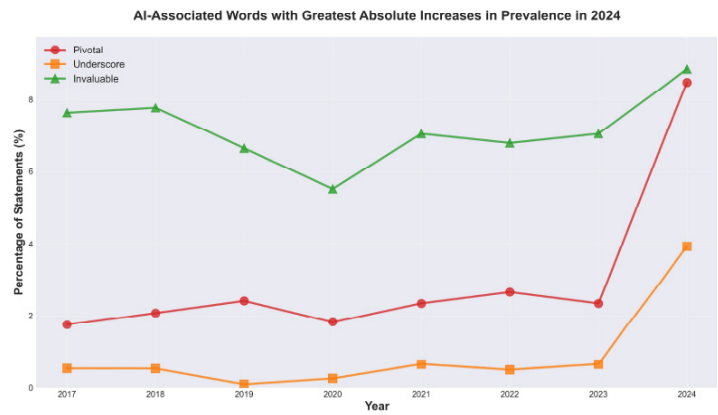
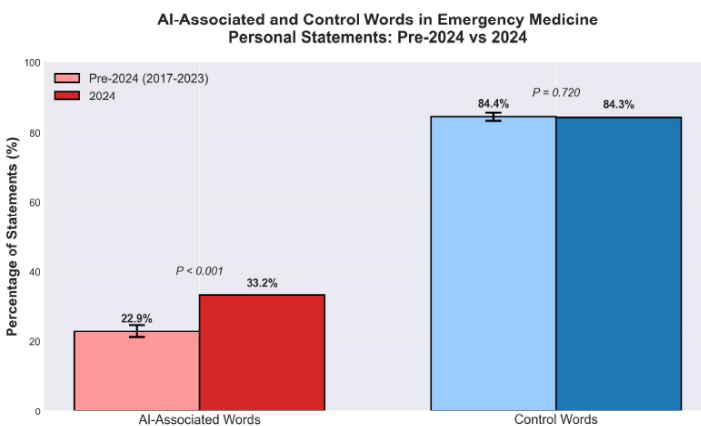
Ryan McKillip, Darrow Traylor, Elise Lovell, Ravi Chacko, Andrea Carlson

Background: Residency leaders increasingly rely on personal statements to select candidates. The availability of artificial intelligence (AI) writing tools raises concerns that personal statements may reflect AI-generated writing rather than authentic applicant voices.

Objective: Assess the prevalence and impact of AI-generated writing in EM residency personal statements submitted for the 2024 application cycle.

Methods: This retrospective study analyzed personal statements submitted to the EM residency of a large academic medical center from 2017 to 2024. The primary outcome was the prevalence of 27 AI-associated target words identified in prior research, or 12 control words, compared between 2024 and 2017–2023 (pre-widespread release of AI writing tools) using one-sample t tests. Secondary outcomes included complexity (Flesch Reading Ease, word count), lexical diversity (type-token ratio), and personalization (first- and third-person pronoun frequency).

Results: A total of 8,617 statements were studied (7,803 pre-2024, 814 in 2024). The proportion of statements with AI-associated words increased significantly from pre-2024 to 2024 (22.9% vs. 33.2%, $P < 0.001$) (Figure 1). Control words were unchanged (84.4% vs. 84.3%, $P = 0.720$). Words with the most significant absolute increases were “pivotal” (2.2% to 8.5%), “underscore” (0.5% to 3.9%), and “invaluable” (6.9% to 8.9%) (Figure 2). Word count decreased (686.5 vs. 674.3 words, $P = 0.005$). Flesch Reading Ease decreased (43.9 vs. 41.9, $P < 0.001$) but remained at the college level. Type-token ratio increased (0.487 vs. 0.500, $P < 0.001$), suggesting greater



lexical diversity. First-person pronouns remained stable (50.7 vs. 51.0, $P = 0.383$), while third-person pronouns increased (8.3 vs. 8.8, $P = 0.035$).

Conclusions: The 10.3% absolute increase in AI-associated word prevalence suggests that approximately 1 in 10 personal statements submitted for the 2024 application cycle contained AI-generated text. Several changes in writing characteristics were observed, and further study is needed to understand the impact on program decision making.

2 Analysis of the eSLOE Score in ResidencyCAS: Comparing Apples to Oranges

Nathan Baggett, Reed Owens-Pochinka, Sarah Schmidt, Graci Gorman, Bradley Hernandez, Cullen Hegarty

Background: The eSLOE is a key feature of EM residency applications. ResidencyCAS (RCAS) calculates an eSLOE score utilizing an applicant’s clerkship grade, estimated guidance and ranking assessments. However, when applicants are evaluated with pass/fail grading, RCAS reassigns the weights for their eSLOE score to disregard grades. This results in applicants being assigned eSLOE scores based on two different formulas. The reliability of this new score has not yet been tested.

Objective: This project aimed to compare our internal eSLOE score with the RCAS eSLOE score and to evaluate the effect of how factors are weighted in the RCAS eSLOE score.

Methods: This was a single-site, cross-sectional study of applicants to our EM residency during the 2025-26 application cycle. Applicants were excluded if they did not receive an eSLOE. We calculated a rating for each eSLOE using our internal scoring system which converts part C of the eSLOE to a numerical score. We calculated the RCAS eSLOE score for all applicants with and without the grades in the formula. We compared the mean and standard deviation (SD) of our internal score and the RCAS eSLOE scores using ANOVA and

Tukey’s post hoc test. RCAS eSLOE scores with and without grades were compared using a T-test. This project was deemed exempt by the IRB.

Results: 917 SLOEs from 454 applicants were analyzed. There was a statistically significant difference in mean eSLOE score between scoring methods ($F(2,1359)=7.93, p<0.001$). Post-hoc analysis found that the mean eSLOE score was significantly different between our internal score and the standard RCAS score including grades ($p<0.001, CI=2.06-9.37$) and the RCAS score with grades excluded ($p=0.004, CI=1.31-8.62$). Grades were unavailable for 63 applicants (13.9%). The RCAS eSLOE score with grades excluded resulted in increased scores for 214 (54.7%) applicants by an average of 6.5 (SD=4.6, $p<0.001$) and in decreased scores for 177 applicants (45.3%) by an average of 5.8 (SD=4.4, $p<0.001$).

Conclusion: The RCAS eSLOE score is calculated differently if the applicant received a pass/fail clerkship grade. Given the absence of standardization in EM clerkship grading, we question the utility of including the clerkship grade when calculating eSLOE scores due to added statistical variability which limits comparisons between applicants.

Table 1. Analysis results.

Scoring Tools		Mean Difference (x1-x2)	Std. Error	Confidence Interval		p-value
(x1)	(x2)			Lower	Upper	
Internal Score	ResidencyCAS Standard Score (including clerkship grade)	5.71*	1.10	p<0.001	8.37	p<0.001
Internal Score	ResidencyCAS Score with Clerkship Grades Excluded	4.98*	1.10	p<0.001	8.62	p<0.001
ResidencyCAS Standard Score (including clerkship grade)	ResidencyCAS Score with Clerkship Grades Excluded	0.75	1.10	p=0.88	4.41	p=0.88

Applicant Categories	n. (%)	Standard ResidencyCAS Score (including clerkship grades)		ResidencyCAS Score with Grades excluded		Average of Differences	SD of Differences	t-test
		M	SD	M	SD			
Applicants with increased eSLOE scores	214 (54.7%)	64.3	23.6	70.8	21.0	6.5	4.6	20.6*
Applicants with decreased eSLOE Scores	180 (46%)	52.4	22.9	46.7	24.4	-5.8	4.4	-17.5*

*Statistically significant with $p<0.001$

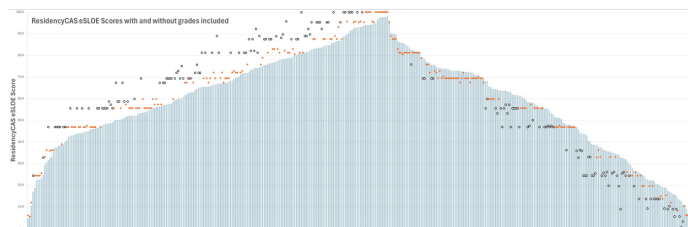


Figure 1. Applicant eSLOE scores with and without grades included. Blue bars represent standard eSLOE scores (weighted 33% each for clerkship grade, guidance, and expected rank list position). Orange circles represent the recalculated eSLOE score with grades excluded (50% weight for guidance and expected rank list position). White circles indicate statistically significant score increases or decreases when grades were excluded.

3 Understanding the Decline in Emergency Medicine Qualifying Examination Pass Rates

Neelou Wecker, Nickolas Srica, Ryan Coughlin, Robert Sobehart, Kaila Pomeranz, Michael Menowsky, Tara Cassidy-Smith

Background: The American Board of Emergency Medicine (ABEM) Qualifying Examination (QE) is a key step in physician certification. In 2024, the first-time pass rate dropped to 82%, the lowest in recent history, raising concerns about resident preparedness and program effectiveness.

Objectives: Identify program-level factors associated with QE performance.

Explore national trends in residency program approaches to exam preparedness.

Methods: This cross-sectional, anonymous survey study targeted U.S. EM residency program directors (PDs). A national working group of EM educators designed the survey, which was distributed via Qualtrics to 280 PDs. Descriptive statistics and chi-square goodness-of-fit tests were used to evaluate response distributions, with $p < 0.05$ considered significant. Qualitative data were analyzed using inductive coding and descriptive analysis.

Results: Out of 280 surveys sent, we received a total of 128 (45.7%) responses. Several variables were collected to assess predictors of ABEM QE failure. The ITE percentile was the strongest individual predictor of ABEM QE failure, with an H-statistic of 35.36 ($p < 0.000004$). Geographic region was associated with a higher incidence of failures ($H = 11.23, p = 0.01$). Programs offering fewer structured educational hours were also associated with higher QE failure rates ($H = 14.85, p = 0.021$). No statistically significant differences in QE failures were observed based on program type (academic vs. community), program length, or reported trends in ITE performance. The qualitative analysis of open-ended responses revealed three major themes: resident study habits, decreased EM competitiveness, and the rigor of undergraduate medical education.

Conclusions: The recent decline in ABEM QE pass rates may reflect broader systemic pressures within EM training rather than a transient testing anomaly. This study highlights the need for ongoing programmatic reflection and national dialogue.

4 Signals of Inclusion: Prevalence and Patterns of DEI Statements on EM Residency Websites

Abagayle Bierowski, Brittany Tian, Sulabh Neupane, Fernando Nombera-Bueno, Erin Hoag, Casey Morrone, Kathleen Cruz, Jiten Patel, Kelly Kehm, Peter Tomaselli

Background: Diversity, equity, and inclusion (DEI) are increasingly recognized as essential components of resident