Title

Comparison of the Standardized Video Interview and Interview Assessments of Professionalism and Interpersonal Communication Skills in Emergency Medicine

Journal

AEM Education and Training, 3(3)

ISSN

2472-5390

Authors

Hopson, Laura R
Dorfsman, Michele L
Branzetti, Jeremy
et al.

Publication Date

2019-07-01

DOI

10.1002/aet2.10346

Peer reviewed

Comparison of the Standardized Video Interview and Interview Assessments of Professionalism and Interpersonal Communication Skills in Emergency Medicine

Laura R. Hopson, MD[1], Michele L. Dorfsman, MD[2], Jeremy Branzetti, MD[3], Michael A. Gisondi, MD[4], Danielle Hart, MD[5], Jaime Jordan, MD[6], James A. Cranford, PhD[7], Sarah R. Williams, MD[8], and Linda Regan, MD[9]

ABSTRACT

Objectives: The Association of American Medical Colleges Standardized Video Interview (SVI) was recently added as a component of emergency medicine (EM) residency applications to provide additional information about interpersonal communication skills (ICS) and knowledge of professionalism (PROF) behaviors. Our objective was to ascertain the correlation between the SVI and residency interviewer assessments of PROF and ICS. Secondary objectives included examination of 1) inter- and intrainstitutional assessments of ICS and PROF, 2) correlation of SVI scores with rank order list (ROL) positions, and 3) the potential influence of gender on interview day assessments.

Methods: We conducted an observational study using prospectively collected data from seven EM residency programs during 2017 and 2018 using a standardized instrument. Correlations between interview day PROF/ICS scores and the SVI were tested. A one-way analysis of variance was used to analyze the association of SVI and ROL position. Gender differences were assessed with independent-groups t-tests.

Results: A total of 1,264 interview-day encounters from 773 unique applicants resulted in 4,854 interviews conducted by 151 interviewers. Both PROF and ICS demonstrated a small positive correlation with the SVI score ($r = 0.16$ and $r = 0.17$, respectively). ROL position was associated with SVI score ($p < 0.001$), with mean SVI scores for top-, middle-, and bottom-third applicants being 20.9, 20.5, and 19.8, respectively. No group differences with gender were identified on assessments of PROF or ICS.

Conclusions: Interview assessments of PROF and ICS have a small, positive correlation with SVI scores. These residency selection tools may be measuring related, but not redundant, applicant characteristics. We did not identify gender differences in interview assessments.

From the [1]Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, MI; [2]Department of Emergency Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA; [3]Ronald O. Perelman Department of Emergency Medicine, New York University School of Medicine, New York, NY; [4]Department of Emergency Medicine, Stanford University, Stanford, CA; [5]Department of Emergency Medicine, University of Minnesota Medical School, St. Paul, MN; [6]Department of Emergency Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA; [7]Department of Psychiatry, University of Michigan, Ann Arbor, MI; [8]Department of Emergency Medicine, Stanford University School of Medicine, Stanford, CA; and the [9]Department of Emergency Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
Received January 4, 2019; revision received March 16, 2019; accepted March 23, 2019.
The authors have no relevant financial information or potential conflicts to disclose.
Supervising Editor: Anne Messman MD, FACEP.
Address for correspondence and reprints: Laura R. Hopson, MD; e-mail: Lhopson@med.umich.edu.
AEM EDUCATION AND TRAINING 2019;3:259–268

doi: 10.1002/aet2.10346

Residency programs invest substantial resources into the high-stakes process of resident selection. These efforts aim to identify individuals with a high likelihood of success during training. Studies of residency applications show mixed results for predicting future performance including medical school grades, United States Medical Licensing Examination (USMLE) performance, and letters of recommendation.[1–3] Success in residency training and beyond likely requires a mixture of cognitive and nontechnical skills. As defined by the Accreditation Council for Graduate Medical Education (ACGME), professionalism (PROF) requires a commitment to carrying out professional responsibilities and an adherence to ethical principles; interpersonal communication skills (ICS) require the effective exchange of information and collaboration with patients, their families, and health professionals.[4] These elements are poorly represented in standard application materials; however, these may be crucial determinants of professional success for future trainees.[5]

Despite the mixed data surrounding their predictive value for training outcomes, conclusions drawn from residency interviews are heavily weighted in the selection process.[1,2] In emergency medicine (EM), assessments of communication skills and interactions with program personnel during interviews are among the most heavily weighted considerations when ranking applicants.[5] There are important reasons for this focus. Deficits in PROF and ICS are exceedingly difficult and time-consuming to remediate during graduate medical education training.[6–9] PROF and ICS impact patient care and have substantial real-world implications beyond the challenges of remediation including risk of future state medical board disciplinary action, medical errors, and malpractice suits.[10–14] There is also a strong correlation between burnout and low PROF, especially in residents and early-career faculty.[15] For all these reasons, PROF and ICS are critically important characteristics to measure in residency applicants.

In response to feedback from program directors about the lack of PROF and ICS data in residency applications, the Association of American Medical Colleges (AAMC) developed the Standardized Video Interview (SVI). Using definitions identical to those of the ACGME, the SVI attempts to provide information about PROF and ICS, through the use of behavioral and situational interview questions, allowing programs to utilize these data when deciding who to invite for an interview.[16] The AAMC SVI became required for all applicants in EM for the 2017 to 2018 residency application season (2018 Match), and expansion to other specialties is under consideration.[17] The SVI consists of six interactions scored by trained raters on a 5-point scale resulting in a numerical score between 6 and 30. The AAMC has ongoing research on the predictive value of the SVI for residency performance. There remain practical questions about how residency programs should utilize the SVI score for applicant invitation and selection purposes as the predictive value of the SVI for future performance is still an area of active research. Additionally, SVI field testing has been met with some controversy, so understanding its potential role in candidate selection is of crucial importance.[18]

Through a multisite study with seven EM residency programs, we studied the correlation of the SVI and interviewer assessments of PROF and ICS conducted during standard residency interviews. The primary objective of this study was to assess the degree of correlation between the SVI and interviewer assessments of PROF and ICS, in an effort to determine whether these assessments contribute similar or different data to the residency selection process. A high correlation (generally defined as $r > 0.4$) would indicate that the SVI and interview assessments are measuring similar candidate attributes, while a low correlation (commonly defined as $r < 0.19$) would indicate that they are measuring different attributes. We hypothesized that we would observe a strong correlation. Secondary objectives included an examination of 1) inter- and intrainstitutional assessments of ICS and PROF, 2) correlation of SVI scores with rank order list (ROL) positions, and 3) finally, due to recent research, we wanted to define the potential influence of gender on interview day assessments for ICS and PROF.[19–21]

METHODS

We conducted an observational study using data collected prospectively from seven EM residency programs and the applicants to these programs during the 2018 Match. The participating programs represent diverse geography and training settings, with representation from both the 36-and the 48-month training formats. Site characteristics are summarized in Table 1. Institutional review boards at each of the participating sites approved the study or deemed it exempt.

All residency applicants who completed an interview at any one of the seven programs during the 2018 Match season and had an SVI score available in the

Table 1
Institutional Characteristics by Site Detailing Program Format, Applicant Interviews, and SVI Prevalence

	Site							Total
	A	B	C	D	E	F	G	
Training venue	Midwest suburban university	West urban university	Mid-Atlantic urban university	Northeast urban university	Midwest urban university	West suburban university	Midwest urban county	
Training duration	48 months	48 months	48 months	48 months	36 months	48 months	36 months	
No. of PGY-1 positions	16	16	12	16	16	15	12	
No. of applicants interviewed	203	132	196	171	177	229	156	
% Female	43.3%	43.2%	40.3%	49.1%	33.3%	38.9%	41.0%	41.1%
Interviews/applicant	5	2–3	3	5	4–5	3–7	3	
Interview duration	15 and 30 min	20 min	20 min	20 min	20 min	10 and 20 min	20 min	
Individual interviews conducted	1,015	360	588	839	787	849	416	4,854
Individual interviews/ interviewer, mean (±SD)	31.7 (±41.8)	10.0 (±10.9)	23.5 (±12.3)	83.3 (±50.4)	52.5 (±53.1)	77.2 (±65.0)	18.9 (±33.4)	
Structured interviews	Yes	No	No	No	No	No	Yes	
No. missing SVI (%)	2 (1.0%)	2 (1.5%)	3 (1.5%)	2 (1.2%)	5 (2.8%)	1 (2.2%)	1 (0.6%)	

SVI = Standardized Video Interview

electronic residency application service (ERAS) were eligible for inclusion. Applicants without a SVI score were excluded from the analysis. Study approval and manuscript review were obtained from the AAMC for use of the SVI data; however, authors retained control over the data and final manuscript.

Two assessment tools were utilized in this study: 1) applicant SVI scores available through ERAS and 2) interviewer assessments of applicant PROF and ICS behaviors including interview discussions and other interactions during on-site interviews. All seven participating sites utilized the same five-point scoring rubric in order to standardize interviewer assessments (Figure 1). Use of a previously validated tool was not possible, as there is no single validated assessment tool widely used. However, many programs are attempting to measure these domains during interviews. Similarly, we did not attempt to completely standardize the student interviews across all seven sites, as we wanted to reflect the usual process conducted by most residency programs. Our assessment scale was developed through an iterative approach that included review of existing assessments from participating programs, the generic SVI scoring algorithm, and group consensus from the study authors.[17] Content validity was established through review of this assessment tool by expert educators who have extensive residency leadership and applicant interview experience as well as advanced medical education training. The assessment items were determined to represent the construct being evaluated, with behavioral characteristics that all program directors believed were critical to assess (Figure 1), and were considered similar constructs of PROF and ICS that are intended to be assessed with the SVI.[17,22] Response process validity was addressed by 1) ensuring that each interviewer completed their applicant ratings prior to discussion of the applicants with other interviewers; 2) attempting to blind interviewers to the applicants' SVI scores during the immediate preinterview file reviews by providing the files as PDF documents, blocking access to the SVI scores through ERAS, or requesting that program leadership (who may still have been able to access SVI scores through ERAS) not look at the SVI scores proximate to interview day; and 3) training interviewers to utilize the anchors to inform their scoring through the use of brief site-specific in-person as well as written methods prior to any interviews being conducted. Internal structure validity evidence was provided by assessing interinstitutional correlations between ICS and PROF scores. All sites assessed ICS and PROF during interview days prior to this study, and no other aspects of the interview day were changed during this process (e.g., number of interviews, timing of interviews, etc).

Data were merged using the AAMC identification number to allow cross-referencing of otherwise blinded

I. Verbal /Communication Skills

Major deficits – 1	Minor deficits – 2	Appropriate – 3	Excellent – 4	Outstanding – 5
Does not meet standards/ Abrasive, unclear, poor focus	*Meets minimum standards/ May fail to be clear or focused at times, requiring redirection*	*Meets expected standards/ Effective in non-complex or stressful situations*	*Above expected standards/ Articulate and effective in both simple and most complex/stressful situations*	*Significantly above expected standards/ Clear, concise, poised with constant success across the spectrum of situations*

II. Professionalism/Maturity

Major deficits – 1	Minor deficits – 2	Appropriate – 3	Excellent – 4	Outstanding – 5
Does not meet standards/ Concerning deficits	*Meets minimum/ May have minor concerns, but can develop*	*Meets expected standards/ No concerns*	*Above expected standards/ Noted to have some advanced qualities*	*Significantly above expected standards/ Wise beyond expected experience*

Figure 1. ICS and PROF scoring tool for interview interactions, which was developed through group consensus after review of the literature. Expected standards would include local-level program judgment as well as performance at Level 1 for the communication and PROF ACGME milestones. ACGME = Accreditation Council for Graduate Medical Education; ICS = interpersonal communication skills; PROF = professionalism.[22]

data across sites. Apart from gender, all other individually identifying candidate information was removed. Collected information about each interviewer included administrative title, years of interview experience, and gender. We utilized data routinely collected as a part of residency applications including the applicant's SVI score (possible range = 6–30), applicant interview day PROF and ICS scores (each scored on a separate five-point scale), and the applicant's final position on the ROL (top third, middle third, lower third, or do not rank [DNR]).

All data were maintained in Excel 2016. Statistical analysis was performed with IBM SPSS Statistics for Windows, Version 25.0. We estimated zero-order Pearson's correlations between interview day assessments of PROF and ICS, each using five-point anchored scales and the SVI score. Subject data for PROF and ICS were aggregated into a single rating by using the mean of all ratings. We conducted a multiple regression analysis of PROF and ICS as predictors of SVI to gain perspective on the degree of overlap between these assessments.

Analysis of the relationship between SVI score and ROL position used a one-way analysis of variance (ANOVA), with ROL position as the independent variable. Follow-up pairwise comparisons were performed with the Tukey HSD test when analyzing SVI scores for ROL position. Using the applicant and the interviewer as the unit of analysis, applicant gender differences, and interviewer assessment differences in PROF, ICS, and SVI scores were assessed with a series of independent-groups t-tests. We did not calculate inter-rater reliability among interviewers, as each interview is a distinct interaction and applicants may perform differently in each interview. For testing our primary objective, the unit of analysis was the applicant (n = 1,264). Tests for secondary objectives used the interviewer (n = 151) or the

interview (n = 4,854) as the unit of analysis where appropriate.

RESULTS

Tables 1 and 2 summarize the descriptive statistics by sites. A total of 1,264 applicants were interviewed by 151 interviewers across the seven participating EM programs. This resulted in a total of 4,854 interviews, across 773 unique applicants. Missing data points were very infrequent. In descending order of frequency, administrative positions included general faculty and fellows (49.7%), residents (16.6%), assistant/associate program directors (15.9%), student clerkship leadership (5.3%), other leadership including deans and chairs (5.3%), program directors (4.6%), and nonphysicians (2.6%). Scores for interview day assessments of ICS and PROF ranged from a low of 1 to a high of 5. However, very few students (less than 0.2%) had a score of 1 on ICS or PROF, and less than 2.1% had a score of 5 on ICS or PROF. Interviewer role was statistically significantly associated with PROF scores ($F(6,143) = 2.2$, $p = 0.04$) and with ICS scores ($F(6,143) = 2.5$, $p = 0.02$). Generally, lower PROF and ICS scores were provided by program directors (mean = 3.6), and higher PROF and communication scores were provided by nonphysicians, residents, and "others" (means = 3.9–4.0). There were strong interinstitutional correlations between ICS and PROF scores, all of which were statistically significant (r range = 0.38 to 0.68, all $p < 0.05$). Full data are provided in Data Supplement S1 (available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10346/full). We calculated interinstitutional correlations using data from students who completed interviews at more than one institution. To achieve 80% power to detect a medium-to-large interinstitutional correlation of 0.40 at an alpha of 0.05, pairwise correlations with a sample size of 40 or higher were examined. Of the 42 possible interinstitutional correlations, 14 met this inclusion criterion. As shown below, the 14 interinstitutional correlations ranged from 0.38 to 0.68, and all were statistically significant at 0.05 (Table 3).

Table 3
Interinstitutional Correlations for ICS and PROF Scores for Applicants Completing Interviews at More Than One Participating Institution Where Adequate Sample Size Existed

	r-value	
Institutions	Between ICS Scores	Between PROF Scores
A and C	0.38	0.48
A and E	0.60	0.62
A and F	0.46	0.45
A and G	0.68	0.42
C and E	0.64	0.55
C and F	0.62	0.54
D and F	0.43	0.50

ICS = interpersonal communication skills; PROF = professionalism.
*All interinstitutional correlations were statistically significant at $p < 0.05$.

Results presented in Table 4 indicate that, for our sample of applicants, SVI scores ranged from 12 to 28 and the mean (±SD) AAMC SVI score was 20.4 (±2.8), which was statistically significantly higher than the mean (±SD) score of 19.1 (±3.1) reported for the AAMC SVI performance data in the 2018 Match.[23] Table 5 also shows a strong positive correlation between interview day assessments of PROF and ICS ($r = 0.81$, $p < 0.001$), which was consistent across all interview sites (r range = 0.71 to 0.89, median r = 0.81). Both PROF and ICS were positively and statistically significantly corrlated with SVI score, but the correlations were small in magnitude (r = 0.16; r values s by site ranged from 0.11 to 0.26, median r = 0.18) and 0.17 (r values by site ranged from 0.05 to 0.29, median r = 0.19), respectively. These patterns of correlation did not vary when examined by institution or by applicant gender (Data Supplement S1). Only two of the participating programs utilized

Table 2
Interviewer Demographics by Site

	Site							
	A	B	C	D	E	F	G	Total
No. of interviewers	32	36	25	10	15	11	22	151
Interviewer years experience, mean (±SD)	5.5 (±5.5)	7.7 (±9.5)	8.3 (±8.2)	5.2 (±8.1)	6.7 (±7.3)	10.4 (±10.2)	9.6 (±9.4)	7.5 (±8.3)
% Female	40.6%	44.4%	40.0%	60.0%	26.7%	36.4%	40.9%	41.1%

Table 4
Summary of SVI, ICS, and PROF Scores by Site and in Aggregate

	Site							Total
	Site A	Site B	Site C	Site D	Site E	Site F	Site G	
No. of applicants interviewed	203	132	196	171	177	229	156	1,264
SVI, mean (±SD)	20.2 (±2.7)	20.6 (±2.9)	20.3 (±2.7)	20.8 (±2.9)	20.2 (±2.5)	20.6 (±2.9)	20.3 (±2.9)	20.4 (±2.8)
ICS, mean (±SD)	3.8 (±0.5)	3.9 (±0.6)	3.8 (±0.5)	3.6 (±0.5)	3.8 (±0.5)	3.5 (±0.7)	3.6 (±0.7)	3.7 (±0.6)
Prof, mean (±SD)	3.8 (±0.5)	3.9 (±0.5)	3.8 (±0.5)	3.7 (±0.5)	3.9 (±0.5)	3.6 (±0.7)	3.6 (±0.6)	3.7 (±0.6)

ICS = interpersonal communication skills; PROF = professionalism; SVI = Standardized Video Interview.

interviews that included predefined behaviorally based interview questions; in the other programs, questions asked were at the discretion of the individual interviewer. When examined by interview method, the two programs utilizing structured interviews did not differ from those with unstructured interviews. Interestingly, despite the small correlation between interview day PROF and ICS scores and SVI scores, there was very little statistical overlap between PROF and ICS scores and AAMC SVI scores when multiple regression analyses were conducted. Taken together, PROF and ICS scores were associated with a statistically significant squared multiple correlation (adjusted $R^2 = 0.028$, $p < 0.001$), indicating that both variables together were associated with 2.8% of the variance in SVI scores. Although statistically significant, these results indicate that over 97% of the variance in SVI scores was not associated with PROF and ICS scores from the interview day assessments.

Results from a one-way ANOVA ($n = 1,264$ applicants) showed that there was a statistically significant association between ROL position and SVI score ($F(2,1176) = 16.5$, $p < 0.001$). Follow-up pairwise comparisons found that mean SVI scores for top-, middle-, and bottom-third and DNR applicants were 20.9, 20.5, 19.8, and 19.8, respectively. All means were statistically significantly different from each ($p < 0.05$), except for the bottom-third and DNR groups ($p = 0.90$).

Table 5
Correlations Between SVI, ICS, and PROF Scores ($n = 1,264$ Applicants)

Variable	1. SVI	2. ICS	3. PROF	Mean	SD
1. SVI	—			20.4	2.8
2. ICS score	0.17*	—		3.7	0.6
3. PROF score	0.16*	0.81*	—	3.7	0.6

ICS = interpersonal communication skills; PROF = professionalism; SVI = Standardized Video Interview.
*$p < 0.01$.

Based on the applicant ($n = 1,264$) gender, a series of independent-groups t-tests showed no statistically significant differences between female and male applicants on assessments of 1) PROF (male mean = 3.7 vs. female mean = 3.8, $t[1262] = 0.6$, $p = 0.57$); 2) ICS (male mean = 3.7 vs. female mean = 3.7, $t[1262] = 0.8$, $p = 0.41$); or 3) SVI score (male mean = 20.5 vs. female mean = 20.3, $t[1243] = 1.1$, $p = 0.25$).

Results for interviewer ($n = 151$) gender influences showed that male interviewers gave statistically significantly higher scores than female interviewers on 1) PROF (male mean = 3.9 vs. female mean = 3.8, $t[147] = 2.1$, $p = 0.04$); and 2) ICS (male mean = 3.9 vs. female mean = 3.7, $t[147] = 2.1$, $p = 0.03$) across all of the institutions, but these gender differences were small in magnitude. Male interviewers had significantly more years of experience than female interviewers (male mean = 9.3 years vs. female mean = 5.1 years, $t[148] = 3.1$, $p = 0.003$). However, interviewers' years of experience was not statistically significantly correlated with PROF scores ($r = 0.01$. $p = 0.85$) or ICS scores ($r = –0.06$. $p = 0.45$). No statistically significant effects of applicant–interviewer gender concordance on PROF or ICS scores were observed ($F < 1.0$, $p > 0.50$).

DISCUSSION

Our study presents data on comparisons between interview day assessments of residency applicants and the new SVI that demonstrates a small positive correlation between the SVI and interview day assessments of PROF and ICS. While SVI and interviewer assessments of PROF and ICS have some overlap, they are also potentially measuring separate domains. There are a number of potential reasons that could account for only finding a small degree of correlation between the two assessments. First, it is possible that residency

interviewers' assessments of PROF and ICS are fundamentally different than the manner in which the AAMC assesses these domains as measured by the SVI.[17] Since interview day assessments often rate applicants in multiple domains in addition to PROF and ICS, it is possible that interviewers are concurrently accounting for some of these SVI subcompetencies in other domains or possibly not at all. This explanation would also be concordant with observations by Schnapp et al.[24] who in a single-institution study showed no correlation between faculty global gestalt scores of PROF and ICS with the SVI. Second, the AAMC SVI utilizes behavioral and situational questions mapped to their clear definition of PROF and ICS. SVI examples include: "Imagine you are leading a multidisciplinary team composed of professionals with different areas of expertise. How should you make sure everyone works together effectively?" (ICS) and "One of your patients refuses treatment because it is incompatible with the patient's religious beliefs. What should you do in this situation?" (PROF).[17] Only two of our participating sites used structured or semistructured interviews, and all incorporated broad topics into questions beyond just PROF and ICS questions, which may have contributed to the small degree of correlation with the SVI. Our anchors for interview day assessments of ICS and PROF encompass a broader skill set than what is measured in the AAMC SVI. For example, our ICS rating anchors focus on the applicant having focused, articulate, effective, and poised communication, while our PROF ratings are less specific and include aspects of maturity and wisdom. Third, while AAMC SVI scoring relies on raters trained to emphasize content more than delivery, our interviewers are likely measuring components of both, including fluency of communication and nonverbal behaviors. Thus, it seems that these two assessments are measuring different, but related, aspects of ICS and PROF. However, further study may be required to truly understand what interviewers are measuring when asked to assess these domains.

It is also likely that factors such personality, similarities with the interviewer, and consideration of previous interactions, such as by e-mail or during clerkships, may influence interview day scores despite the presence of anchors for scoring. In addition, interviewers are not blinded to other elements of the residency application, which may create halo or horn effects on PROF and ICS assessments from the interview itself.[1] Thus, interviewer PROF and ICS scores may represent additional candidate factors that are not assessed in the SVI score. PROF can be a difficult domain to assess and measure, and our ability to do this effectively within a brief interview encounter may also be limited.

Of interest is the strong correlation ($r = 0.81$) between PROF and ICS assessments on our scoring instrument. This could suggest that raters, despite use of an anchored rating scale, have difficulty discriminating between behaviors that represent each domain and thus conceptualize them as a single entity. This observation is consistent with literature suggesting that ICS skills positively affect scoring in many other domains of the ACGME Core Competencies.[25] It is also possible that this correlation is an accurate representation of high covariance between the domains. Longitudinal research will be needed to clarify the discriminant validity of the PROF and ICS scores.

Interinstitutional correlations of scores from applicants completing interviews at more than one institution were significantly correlated and provide validity evidence for the interview day assessment tool. We chose not to place weight on the comparison of PROF and ICS scores of the same applicant who interviewed at more than one site in our study. We felt that the same applicant might perform very differently at different interview sites on different days due to a variety of factors, including interest in the program, physical factors such as illness or sleep adequacy, and external factors or stressors. As such, we viewed each interview as a discreet encounter that is potentially not comparable across sites.

The statistically significant association between AAMC SVI scores and ROL position is not surprising given that the SVI is intended to evaluate skills that are highly valued by EM program directors.[26] However, the scores for top-, middle-, and bottom-third applicants only revealed small absolute differences (0.4 and 0.7, respectively) and this small variation may not be meaningful in practice and is significant due to the large sample size. Interestingly, the SVI scores of the DNR applicants were identical to the lower-third group. We postulate that the DNR group is heterogeneous due to technical disqualifiers (e.g., lack of USMLE scores by rank list submission deadlines as required by the institutional selection policies, withdrawal by the applicant) as well as individuals with behavioral or academic concerns. In addition, ROL positioning, while it may emphasize academic traits, is subject to many influences.[27] Currently,

the SVI does not appear to be a useful discriminator for DNR positioning, and the practical significance of the SVI score differences by ROL position may be an avenue for future research.

Recent research has demonstrated gender discrepancies in ACGME EM Milestone proficiency levels assessments, as well as within letters of recommendation and the medical school performance evaluation.[19–21] We felt that it was important to understand whether gender bias could be a factor affecting interview day assessments, which, if present, could affect correlations with the SVI. We did not find any statistically significant differences in interview day assessment scoring related to applicant-interviewer gender concordance. We did find that male interviewers gave slightly higher ICS and PROF scores than female interviewers. However, this small difference in scores of 0.21 points and 0.12 points, respectively, while statistically significant, is likely not meaningful in a practical sense. It is possible that use of an anchored rating scale may provide one mechanism for residency programs to minimize gender bias within their applicant assessment processes. This process has been used successfully by the AAMC, which has found no evidence of gender bias in the AAMC SVI scores.[28]

In our results, program directors generally gave lower scores on both scales and higher scores were provided by nonphysicians, residents, and interviewers in the "other" category. This association is not entirely surprising as prior medical literature has shown differences in ratings between assessor groups and that assessors' interpretations are framed within their discipline, experience, and level of involvement with the learner.[29–32] Program directors may be more critical of applicants, knowing that they will have the responsibility of managing any remediation issues. They are also more likely to interview the largest number of candidates and thus may have a broader sample of applicants across which to calibrate their ratings. These differences in assessments, however, do not necessarily reflect bias or mean that one is more or less accurate than another. Literature suggests that differing assessments, as long as raters possess the skills and expertise to accurate judge the construct of interest, represent distinct but equally valid perspectives.[29,32–34] We did not find any statistically significant association between interviewer years of experience and PROF and ICS scores, suggesting that duration of experience is not the sole factor required to accurately assess these skills.

LIMITATIONS

There are many potential sources of bias inherent to the interview process for which we could not control.[1] Each site was free to conduct interviews per their normal process. The study protocol did not include scripted questions or formats; the only commonality was the PROF and ICS assessment tool. This design was utilized so that our study would represent "real-world" assessments rather than an artificial idealized interview state, with the goal of increasing the generalizability and external validity of the results as opposed to limiting them. This does mean that applicants were not all necessarily put in complex or stressful situations during all interviews. We also did not provide extensive rater training, as was done for the SVI. While additional training of interviewers could further standardize this process, we again elected not to do this in an attempt to represent real-world interview day assessments. Terms such as "minimum standards" and "expected standards" were not explicitly defined for the interviewers during the training process and were left up to individual interviewers' discretion when rating applicants although reference to Level 1 of the ACGME milestones was an implicit part of the concept. Further standardization would require a substantial investment of resources and could include standard setting exercises with a variety of in-person or recorded interview interactions.

Although the scoring instrument was created iteratively by the group in an effort to enhance validity, it did not undergo any formal piloting prior to implementation in actual interviews. In the experience of the authors, this is in line with standard practice of residency interview scoring; the tools used are generally based on content validity without further validity testing.

Our study population was preselected from review of ERAS application materials; this cohort may have different abilities than those not invited to interview. This may have impacted our overall data and its resultant correlations. For example, we observed a restricted score range of PROF and ICS assessments which likely attenuated our ability to identify true correlations that may be present in a study population that was not preselected.

The participating sites included five 4-year programs and two 3-year programs. While we found no differences between sites, it is possible that our skewed sample may make these findings more applicable to 4-year programs. Only two of the seven programs used semistructured interviews. It is unclear whether this

percentage is reflective of the overall EM community, and these results may be more applicable to programs using unstructured interviews.

CONCLUSIONS

In this multicenter study aiming to determine whether the Standardized Video Interview and usual interview day assessments of professionalism and interpersonal communication skills contribute similar or different data to the residency selection process, we found that interview day assessments using a novel tool have only a small, positive correlation with Association of American Medical Colleges Standardized Video Interview scores. It is therefore likely that both assessments provide meaningful, distinct information. For secondary objectives, there was strong interinstitutional correlation between interview day professionalism and interpersonal communication skills scores; a small but statistically significant correlation between Standardized Video Interview and rank order list position across institutions and no gender influences on interview day scores. However, the difference between a top-third candidate and a do-not-rank candidate was minimal, with only a 1.1 point Standardized Video Interview score difference. Similarly, the Standardized Video Interview could not distinguish between a bottom-third and a do-not-rank candidate. Further study is required to examine the predictive ability of both the Standardized Video Interview and well-designed interview day assessments on future clinical performance.

References

1. Stephenson-Famy A, Houmard BS, Oberoi S, Manyak A, Chiang S, Kim S. Use of the interview in resident candidate selection: a review of the literature. J Grad Med Educ 2015;7:539–48.
2. Kenny S, McInnes M, Singh V. Associations between residency selection strategies and doctor performance: a meta-analysis. Med Educ 2013;47:790–800.
3. Bhat R, Takenaka K, Levine B, et al. Predictors of a top performer during emergency medicine residency. J Emerg Med 2015;49:505–12.
4. ACGME Common Program Requirements. Available at: https://acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRs_2017-07-01.pdf. Accessed February 10, 2019.
5. Dunleavy D, Geiger T, Overton R, Prescott J. Results of the 2016 Program Directors Survey: Current Practices in Residency Selection. Available at: https://members.aamc.org/eweb/upload/Program%20Directors%20Survey%20Report.pdf Accessed October 27, 2018.
6. Dupras D, Edson R, Halvorsen A, Hopkins RH Jr, McDonald FS. "Problem residents": prevalence, problems and remediation in the era of core competencies. Am J Med 2012;125:421–5.
7. Zbieranowski I, Takahashi S, Verma S, Spadafora SM. Remediation of residents in difficulty: a retrospective 10-year review of the experience of a postgraduate board of examiners. Acad Med 2013;88:111–6.
8. Regan L, Hexom B, Nazario S, Chinai SA, Visconti A, Sullivan C. Remediation methods for milestones related to interpersonal and communication skills and professionalism. J Grad Med Educ 2016;8:18–23.
9. Silverberg M, Weizberg M, Murano T, Smith JL, Burkhardt JC, Santen SA. What is the prevalence and success of remediation of emergency medicine residents? West J Emerg Med 2015;16:839–44.
10. Papadakis M, Hodgson C, Teherani A. Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. Acad Med 2004;799:244–9.
11. Papadakis M, Teherani A, Banach M, et al. Disciplinary action by medical boards and prior behavior in medical school. N Engl J Med 2005;353:2673–82.
12. Papadakis M, Arnold G, Blank L, et al. Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. Ann Intern Med 2008;148:869–76.
13. Vukmir RB. Medical malpractice: managing the risk. Med Law 2004;23:495–513.
14. Joint Commission Center for Transforming Healthcare Releases Targeted Solutions Tool for Hand-Off Communications. Joint Commission Perspectives®, August 2012, Volume 32, Issue 8.
15. Panagioti M, Geraghty K, Johnson J, et al. Association between physician burnout and patient safety, professionalism, and patient satisfaction: a systematic review and meta-analysis. JAMA Intern Med 2018;178:1317–30.
16. Bird SD, Blomkalns A, Deiorio NM, Gallhue FE. Beyond test scores and medical knowledge: the Standardized Video Interview, an innovative and ethical approach for holistic assessment of applicants. Acad Med 2018;93:151.
17. The AAMC Standardized Video Interview: Essentials for the ERAS® 2019 Season. Available at: https://aamc-orange.global.ssl.fastly.net/production/media/filer_public/75/19/75198c9f-29d1-4884-b932-1e7975e4262a/standardized-video-interview-essentials-eras-2019-20180720.pdf. Accessed October 27, 2018.
18. Buckley RJ, Hoch VC, Huang RD. Lights, camera, empathy: a request to slow the emergency medicine Standardized Video Interview project study. AEM Educ Train 2018;2:57–60.

19. Dayal A, O'Connor DM, Qadri U, Arora VM. Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. JAMA Intern Med 2017;177:651–7.
20. Li S, Fant AL, McCarthy DM, Miller D, Craig J, Kontrick A. Gender differences in language of standardized letter of evaluation narratives for emergency medicine residency applicants. AEM Educ Train 2017;1:334–9.
21. Isaac C, Chertoff J, Lee B, Carnes M. Do students' and authors' genders affect evaluations? A linguistic analysis of medical student performance evaluations. Acad Med 2011;86:59–66.
22. ACGME Emergency Medicine Milestone Project. Available at: https://www.acgme.org/Portals/0/PDFs/Milestones/EmergencyMedicineMilestones.pdf. Accessed February 25, 2019.
23. Summary of AAMC® Standardized Video Interview™ Total Scores. Available at: https://aamc-orange.global.ssl.fastly.net/production/media/filer_public/5c/47/5c4790d6-e44f-43a0-8548-92e18243fb74/svi-scorereport-percentileranks-2017-2018.pdf. Accessed November 17, 2018
24. Schnapp BH, Ritter D, Kraut AS, Fallon S, Westergaard MC. Assessing residency applicants' communication and professionalism: Standardized Video Interview scores compared to faculty gestalt. West J Emerg Med 2019;20:132–7.
25. Haurani MJ, Rubinfeld I, Rao S, et al. Are the communication and professionalism competencies the new critical values in a resident's global evaluation process? J Surg Educ 2007;64:351–6.
26. National Resident Matching Program, Data Release and Research Committee: Results of the 2018 NRMP Program Director Survey. Available at: https://www.nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf. Accessed October 26, 2018.
27. Breyer MJ, Sadosty A, Biros M. Factors affecting candidate placement on an emergency medicine residency program's rank order list. West J Emerg Med 2012;13:458–62.
28. AAMC Standardized Video Interview Update. Association of Academic Chairs in Emergency Medicine. March 13, 2018. Available at: http://saem.org/docs/default-source/joint-retreat/2018/aacem-svi-update.pdf?sfvrsn=599935fd_2. Accessed November 17, 2018.
29. Bullock AD, Hassell A, Markham WA, Wall DW, Whitehouse AB. How ratings vary by staff group in multi-source feedback assessment of junior doctors. Med Educ 2009;43:516–20.
30. Risucci DA, Tortolania AJ, Ward RJ. Ratings of surgical residents by self, supervisors and peers. Surg Gynecol Obstet 1989;169:519–26.
31. Shay SB. The assessment of complex performance: a socially situated interpretative act. Harv Educ Rev 2004;74:307–29.
32. Govaerts MJ, Schuwirth LW, Van der Vleuten CP, Muijtjens AM. Workplace-based assessment: effects of rater expertise. Adv Health Sci Educ 2011;16:151–65.
33. Lance CE, Hoffman BJ, Gentry WA, Baranik LE. Rater source factors represent important subcomponents of the criterion construct space, not rater bias. Hum Resour Manag Rev 2008;18:223–32.
34. Crossley J, Jolly B. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. Med Educ 2012;46:28–37.

Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10346/full

Data Supplement S1. Pearson correlations between SVI, ICS, and PROF scores across sites.