

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

Objectives: The AAMC 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 (a) inter- and intra-institutional assessments of ICS and PROF; (b) correlation of SVI scores with Rank Order List (ROL) positions; and (c) 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 2017-2018 using a standardized instrument. Correlations between interview day PROF / ICS scores and the SVI were tested. A one-way ANOVA 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 ($r_s = .16$ and $.17$, respectively). ROL position was associated with SVI score ($p < .001$), 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.

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

INTRODUCTION

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/AET2.10346-19-002](https://doi.org/10.1002/AET2.10346-19-002)

29 Residency programs invest substantial resources into the high-stakes process of resident
30 selection. These efforts aim to identify individuals with a high likelihood of success during training.
31 Studies of residency applications show mixed results for predicting future performance including
32 medical school grades, United States Medical Licensing Examination (USMLE) performance, and letters
33 of recommendation.¹⁻³ Success in residency training and beyond likely requires a mixture of cognitive
34 and non-technical skills. As defined by the ACGME, Professionalism (PROF) requires a commitment to
35 carrying out professional responsibilities and an adherence to ethical principles; Interpersonal and
36 Communication Skills (ICS) requires the effective exchange of information and collaboration with
37 patients, their families, and health professionals.⁴ These elements are poorly represented in standard
38 application materials; however, these may be crucial determinants of professional success for future
39 trainees.⁵

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

51 In response to feedback from program directors about the lack of PROF and ICS data in
52 residency applications, the Association of American Medical Colleges (AAMC[®]) developed the
53 Standardized Video Interview[™](SVI). Using definitions identical to those of the ACGME, the SVI
54 attempts to provide information about PROF and ICS, through the use of behavioral and situational
55 interview questions, allowing programs to utilize this data when deciding who to invite for an
56 interview.¹⁷ The AAMC SVI became required for all applicants in EM for the 2017-18 residency
57 application season (2018 Match[®]), and expansion to other specialties is under consideration.¹⁷ The SVI
58 consists of 6 interactions scored by trained raters on a 5-point scale resulting in a numerical score
59 between 6 and 30. The AAMC has ongoing research on the predictive value of the SVI for residency
60 performance. There remain practical questions about how residency programs should utilize the SVI

61 score for applicant invitation and selection purposes as the predictive value of the SVI for future
62 performance is still an area of active research. Additionally, SVI field-testing has been met with some
63 controversy, so understanding its potential role in candidate selection is of crucial importance.¹⁸

64 Through a multi-site study with seven EM residency programs, we studied the correlation of the
65 SVI and interviewer assessments of PROF and ICS conducted during standard residency interviews. The
66 primary objective of this study was to assess the degree of correlation between the SVI and interviewer
67 assessments of PROF and ICS, in an effort to determine whether these assessments contribute similar or
68 different data to the residency selection process. A high correlation (generally defined as $r > 0.4$) would
69 indicate that the SVI and interview assessments are measuring similar candidate attributes, while a low
70 correlation (commonly defined as $r < 0.19$) would indicate they are measuring different attributes. We
71 hypothesized that we would observe a strong correlation. Secondary objectives included an
72 examination of (a) inter- and intra-institutional assessments of ICS and PROF; (b) correlation of SVI
73 scores with Rank Order List (ROL) positions; and (c) finally, due to recent research, we wanted to define
74 the potential influence of gender on interview day assessments for ICS and PROF.²⁵⁻²⁷

75

76 METHODS

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

82 All residency applicants who completed an interview at any one of the 7 programs during the
83 2018 Match[®] season and had an SVI score available in the Electronic Residency Application Service
84 (ERAS[®]) were eligible for inclusion. Applicants without a SVI score were excluded from the analysis.
85 Study approval and manuscript review were obtained from the AAMC[®] for use of the SVI data; however,
86 authors retained control over the data and final manuscript.

87 Two assessment tools were utilized in this study: (1) applicant SVI scores available through
88 ERAS[®] and (2) interviewer assessments of applicant PROF and ICS behaviors including interview
89 discussions and other interactions during on-site interviews. All 7 participating sites utilized the same 5-
90 point scoring rubric in order to standardize interviewer assessments (**Figure 1**). Use of a previously
91 validated tool was not possible, as there is no single validated assessment tool widely used. However,
92 many programs are attempting to measure these domains during interviews. Similarly, we did not

93 attempt to completely standardize the student interviews across all 7 sites, as we wanted to reflect the
94 usual process conducted by most residency programs. Our assessment scale was developed through an
95 iterative approach that included review of existing assessments from participating programs, the generic
96 SVI scoring algorithm, and group consensus from the study authors.¹⁷ Content validity was established
97 through review of this assessment tool by expert educators who have extensive residency leadership
98 and applicant interview experience as well as advanced medical education training. The assessment
99 items were determined to represent the construct being evaluated, with behavioral characteristics that
100 all program directors believed were critical to assess (**Figure 1**), and were considered similar constructs
101 of PROF and ICS that are intended to be assessed with the SVI.^{17,28} Response process validity was
102 addressed by a) ensuring that each interviewer completed their applicant ratings prior to discussion of
103 the applicants with other interviewers, b) attempting to blind interviewers to the applicants' SVI scores
104 during the immediate pre-interview file reviews by either providing the files as PDF documents, blocking
105 access to the SVI scores through ERAS, or requesting that program leadership (who may still have been
106 able to access SVI scores through ERAS) not look at the SVI scores proximate to interview day, and c)
107 training interviewers to utilize the anchors to inform their scoring through the use of brief site-specific
108 in-person as well as written methods prior to any interviews being conducted. Internal structure validity
109 evidence was provided by assessing inter-institutional correlations between ICS and PROF scores. All
110 sites assessed ICS and PROF during interview days prior to this study, and no other aspects of the
111 interview day were changed during this process (e.g. number of interviews, timing of interviews, etc).

112 Data were merged using the AAMC Identification Number to allow cross-referencing of
113 otherwise blinded data across sites. Apart from gender, all other individually identifying candidate
114 information was removed. Collected information about each interviewer included administrative title,
115 years of interview experience, and gender. We utilized data routinely collected as a part of residency
116 applications including the applicant's SVI score (possible range 6-30), applicant interview day PROF and
117 ICS scores (each scored on a separate 5-point scale), and the applicant's final position on the ROL (top $\frac{1}{3}$,
118 middle $\frac{1}{3}$, lower $\frac{1}{3}$, or do not rank (DNR)).

119 All data were maintained in Excel™ 2016. Statistical analysis was performed with IBM SPSS
120 Statistics for Windows, Version 25.0 (Armonk, NY: IBM Corp). We estimated zero-order Pearson's
121 correlations between interview day assessments of PROF and ICS, each using 5-point anchored scales
122 and the SVI score. Subject data for PROF and ICS were aggregated into a single rating by using the mean
123 of all ratings. We conducted a multiple regression analysis of PROF and ICS as predictors of SVI to gain
124 perspective on the degree of overlap between these assessments.

125 Analysis of the relationship between SVI score and ROL position used a one-way analysis of
126 variance (ANOVA), with ROL position as the independent variable. Follow-up pairwise comparisons were
127 performed with the Tukey HSD test when analyzing SVI scores for ROL position. Using the applicant and
128 the interviewer as the unit of analysis, applicant gender differences and interviewer assessment
129 differences in PROF, ICS, and SVI scores were assessed with a series of independent-groups *t*-tests. We
130 did not calculate inter-rater reliability among interviewers, as each interview is a distinct interaction and
131 applicants may perform differently in each interview. For testing our primary objective, the unit of
132 analysis was the applicant (N = 1,264). Tests for secondary objectives used the interviewer (N = 151) or
133 the interview (N = 4,854) as the unit of analysis where appropriate.

134

135 RESULTS

136

137 **Tables 1 and 2** summarize the descriptive statistics by sites. A total of 1,264 applicants were
138 interviewed by 151 interviewers across the 7 participating EM programs. This resulted in a total of 4,854
139 interviews, across 773 unique applicants. Missing data points were very infrequent. In descending order
140 of frequency, administrative positions included general faculty and fellows (49.7%), residents (16.6%),
141 assistant/associate program directors (15.9%), student clerkship leadership (5.3%), other leadership
142 including deans and chairs (5.3%), program directors (4.6%), and non-physicians (2.6%). Scores for
143 interview day assessments of ICS and PROF ranged from a low of 1 to a high of 5. However, very few
144 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
145 PROF. Interviewer role was statistically significantly associated with PROF scores, $F(6, 143) = 2.2, p =$
146 $.04$, and with ICS scores, $F(6, 143) = 2.5, p = .02$. Generally, lower PROF and ICS scores were provided by
147 program directors (mean = 3.6), and higher professionalism and communication scores were provided
148 by non-physicians, residents, and “others” (means = 3.9 – 4.0). There were strong inter-institutional
149 correlations between ICS and PROF scores all of which were statistically significant (r range = 0.38 to
150 0.68, all $p < 0.05$). Full data is provided in **Supplemental Table 1**. We calculated inter-institutional
151 correlations using data from students who completed interviews at more than one institution. In order
152 to achieve 80% power to detect a medium-to-large inter-institutional correlation of 0.40 at an alpha of
153 0.05, pairwise correlations with a sample size of 40 or higher were examined. Of the 42 possible inter-
154 institutional correlations, 14 met this inclusion criterion. As shown below, the 14 inter-institutional
155 correlations ranged from 0.38 to 0.68, and all were statistically significant at 0.05 (**Table 5**).

156 Results presented in **Table 3** indicate that, for our sample of applicants, SVI scores ranged from
157 12-28 and the mean AAMC SVI score was 20.4 (SD = 2.8), which was statistically significantly higher than
158 the mean score of 19.1 (SD = 3.1) reported for the AAMC SVI performance data in the 2018 Match[®].¹⁹
159 **Table 4** also shows a strong positive correlation between interview day assessments of PROF and ICS, r
160 = 0.81, $p < 0.001$ which was consistent across all interview sites (r s ranged from 0.71 to 0.89, median r =
161 0.81). Both PROF and ICS were positively and statistically significantly correlated with SVI score, but the
162 correlations were small in magnitude, r s = 0.16 (r s by site ranged from 0.11 to 0.26, median r = 0.18) and
163 0.17 (r s by site ranged from 0.05 to 0.29, median r = 0.19), respectively. These patterns of correlation
164 did not vary when examined by institution or by applicant gender (**Supplemental Table 1**). Only 2 of the
165 participating programs utilized interviews which included pre-defined behaviorally based interview
166 questions; in the other programs, questions asked were at the discretion of the individual interviewer.
167 When examined by interview method, the two programs utilizing structured interviews did not differ
168 from those with unstructured interviews. Interestingly, despite the small correlation between interview
169 day PROF and ICS scores and SVI scores, there was very little statistical overlap between PROF and ICS
170 scores and AAMC SVI scores when multiple regression analyses were conducted. Taken together, PROF
171 and ICS scores were associated with a statistically significant squared multiple correlation, adjusted R^2 =
172 0.028, $p < 0.001$, indicating that both variables together were associated with 2.8% of the variance in SVI
173 scores. Although statistically significant, these results indicate that over 97% of the variance in SVI scores
174 was not associated with PROF and ICS scores from the interview day assessments.

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

180 Based on the applicant (N = 1,264) gender, a series of independent-groups t -tests showed no
181 statistically significant differences between female and male applicants on assessments of a) PROF (male
182 $mean = 3.7$ vs. female $mean = 3.8$, $t[1262] = 0.6$, $p = 0.57$); b) ICS (male $mean = 3.7$ vs. female $mean =$
183 3.7 , $t[1262] = 0.8$, $p = 0.41$); or c) SVI score (male $mean = 20.5$ vs. female $mean = 20.3$, $t[1243] = 1.1$, $p =$
184 0.25).

185 Results for interviewer (N=151) gender influences showed that male interviewers gave
186 statistically significantly higher scores than female interviewers on a) PROF (male $mean = 3.9$ vs. female
187 $mean = 3.8$, $t[147] = 2.1$, $p = .04$); and b) ICS (male $mean = 3.9$ vs. female $mean = 3.7$, $t[147] = 2.1$, $p =$

188 0.03) across all of the institutions, but these gender differences were small in magnitude. Male
189 interviewers had significantly more years of experience than female interviewers (male *mean* = 9.3 years
190 vs. female *mean* = 5.1 years, $t [148] = 3.1, p = 0.003$). However, interviewers' years of experience was
191 not statistically significantly correlated with PROF scores ($r = 0.01, p = 0.85$) or ICS scores ($r = -0.06, p =$
192 0.45). No statistically significant effects of applicant-interviewer gender concordance on PROF or ICS
193 scores were observed ($F_s < 1.0, p_s > 0.50$).

194

195

196 Discussion

197

198 Our study presents data on comparisons between interview day assessments of residency
199 applicants and the new SVI which demonstrates a small positive correlation between the SVI and
200 interview day assessments of PROF and ICS. While SVI and interviewer assessments of PROF and ICS
201 have some overlap, they are also potentially measuring separate domains. There are a number of
202 potential reasons that could account for only finding a small degree of correlation between the two
203 assessments. First, it is possible that residency interviewers' assessments of PROF and ICS are
204 fundamentally different than the manner in which the AAMC assesses these domains as measured by
205 the SVI.¹⁷ Since interview day assessments often rate applicants in multiple domains in addition to PROF
206 and ICS, it is possible that interviewers are concurrently accounting for some of these SVI sub-
207 competencies in other domains, or possibly not at all. This explanation would also be concordant with
208 observations by Schnapp et al who in a single-institution study showed no correlation between faculty
209 global gestalt scores of PROF and ICS with the SVI.²⁰ Second, the AAMC SVI utilizes behavioral and
210 situational questions mapped to their clear definition of PROF and ICS. SVI examples include: "Imagine
211 you are leading a multidisciplinary team composed of professionals with different areas of expertise.
212 How should you make sure everyone works together effectively?" (ICS) and "One of your patients
213 refuses treatment because it is incompatible with the patient's religious beliefs. What should you do in
214 this situation?" (PROF).¹⁷ Only two of our participating sites used structured or semi-structured
215 interviews, and all incorporated broad topics into questions beyond just PROF and ICS questions which
216 may have contributed to the small degree of correlation with the SVI. Our anchors for interview day
217 assessments of ICS and PROF encompass a broader skill set than what is measured in the AAMC SVI. For
218 example, our ICS rating anchors focus on the applicant having focused, articulate, effective and poised
219 communication, while our PROF ratings are less specific and include aspects of maturity and wisdom.

220 Third, while AAMC SVI scoring relies on raters trained to emphasize content more than delivery, our
221 interviewers are likely measuring components of both, including fluency of communication and
222 nonverbal behaviors. Thus, it seems that these two assessments are measuring different, but related,
223 aspects of ICS and PROF. However, further study may be required to truly understand what interviewers
224 are measuring when asked to assess these domains.

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

232 Of interest is the strong correlation ($r = 0.81$) between PROF and ICS assessments on our scoring
233 instrument. This could suggest that raters, despite use of an anchored rating scale, have difficulty
234 discriminating between behaviors that represent each domain and thus conceptualize them as a single
235 entity. This observation is consistent with literature suggesting that ICS skills positively affect scoring in
236 many other domains of the ACGME Core Competencies.²¹ It is also possible that this correlation is an
237 accurate representation of high covariance between the domains. Longitudinal research will be needed
238 to clarify the discriminant validity of the PROF and ICS scores.

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

246 The statistically significant association between AAMC SVI scores and ROL position is not
247 surprising given that the SVI is intended to evaluate skills that are highly valued by EM program
248 directors.²² However, the scores for top-, middle-, and bottom- third applicants only revealed small
249 absolute differences (0.4 and 0.7 respectively) and this small variation may not be meaningful in practice
250 and is significant due to the large sample size. Interestingly, the SVI scores of the DNR applicants were
251 identical to the lower-third group. We postulate that the DNR group is heterogeneous due to technical

252 disqualifiers (e.g. lack of USMLE scores by rank list submission deadlines as required by the institutional
253 selection policies, withdrawal by the applicant) as well as individuals with behavioral or academic
254 concerns. In addition, ROL positioning, while it may emphasize academic traits, is subject to many
255 influences.²³ Currently, the SVI does not appear to be a useful discriminator for DNR positioning, and the
256 practical significance of the SVI-score differences by ROL position may be an avenue for future research.

257 Recent research has demonstrated gender discrepancies in ACGME EM Milestone proficiency
258 levels assessments, as well as within letters of recommendation and the medical school performance
259 evaluation (MSPE).²⁴⁻²⁷ We felt it was important to understand whether gender-bias could be a factor
260 affecting interview day assessments, which, if present, could affect correlations with the SVI. We did
261 not find any statistically significant differences in interview day assessment scoring related to applicant-
262 interviewer gender concordance. We did find that male interviewers gave slightly higher ICS and PROF
263 scores than female interviewers. However, this small difference in scores of 0.21 points and 0.12 points,
264 respectively, while statistically significant, is likely not meaningful in a practical sense. It is possible that
265 use of an anchored rating scale may provide one mechanism for residency programs to minimize
266 gender-bias within their applicant assessment processes. This process has been used successfully by the
267 AAMC, which has found no evidence of gender bias in the AAMC SVI scores.²⁷

268 In our results, program directors generally gave lower scores on both scales and higher scores
269 were provided by non-physicians, residents, and interviewers in the “other” category. This association
270 is not entirely surprising as prior medical literature has shown differences in ratings between assessor
271 groups, and that assessors’ interpretations are framed within their discipline, experience, and level of
272 involvement with the learner.²⁹⁻³² Program directors may be more critical of applicants, knowing that
273 they will have the responsibility of managing any remediation issues. They are also more likely to
274 interview the largest number of candidates, and thus may have a broader sample of applicants across
275 which to calibrate their ratings. These differences in assessments, however, do not necessarily reflect
276 bias or mean that one is more or less accurate than another. Literature suggests that differing
277 assessments, as long as raters possess the skills and expertise to accurately judge the construct of
278 interest, represent distinct but equally valid perspectives.³²⁻³⁵ We did not find any statistically significant
279 association between interviewer years of experience and PROF and ICS scores, suggesting that duration
280 of experience is not the sole factor required to accurately assess these skills.

281

282 **Limitations**

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

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

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

306 The participating sites included five 4-year programs and two 3-year programs. While we found
307 no differences between sites, it is possible that our skewed sample may make these findings more
308 applicable to 4-year programs. Only two of the seven programs used semi-structured interviews. It is
309 unclear whether this percentage is reflective of the overall EM community, and these results may be
310 more applicable to programs using unstructured interviews.

311

312 **Conclusions**

313 In this multicenter study aiming to determine whether the SVI and usual interview day
314 assessments of PROF and ICS contribute similar or different data to the residency selection process, we

315 found that interview day assessments using a novel tool have only a small, positive correlation with
316 AAMC SVI scores. It is therefore likely that both assessments provide meaningful, distinct information.
317 For secondary objectives, there was strong inter-institutional correlation between interview day PROF
318 and ICS scores; a small but statistically significant correlation between SVI and ROL position across
319 institutions; and no gender influences on interview day scores. However, the difference between a top
320 1/3 candidate and a do-not-rank candidate was minimal, with only a 1.1 point SVI score difference.
321 Similarly, the SVI could not distinguish between a bottom 1/3 and a do-not-rank candidate. Further
322 study is required to examine the predictive ability of both the SVI and well-designed interview day
323 assessments on future clinical performance.

324

325

326 **References**

327

328 1. Stephenson-Famy A, Houmard BS, Oberoi S, Manyak A, Chiang S, Kim S. Use of the interview in
329 resident candidate selection: a review of the literature. *J Grad Med Educ.* 2015; 7(4):539-548.

330

331 2. Kenny S, McInnes M, Singh V. Associations between residency selection strategies and doctor
332 performance: a meta-analysis. *Med Educ.* 2013; 47(8):790–800.

333

334 3. Bhat R, Takenaka K, Levine B, Goyal N, Garg M, Visconti A, et al. Predictors of a top performer
335 during emergency Medicine residency. *J Emerg Med.* 2015; 49(4):505-12.

336

337 4. ACGME Common Program Requirements

338 https://acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRs_2017-07-01.pdf Accessed
339 February 10, 2019.

340

341 5. Dunleavy D, Geiger T, Overton R, Prescott J. Results of the 2016 Program Directors survey:
342 current practices in residency selection.

343 <https://members.aamc.org/eweb/upload/Program%20Directors%20Survey%20Report.pdf>

344 Published September 2016. Accessed October 27, 2018.

345

- 346 6. Dupras D, Edson R, Halvorsen A, Hopkins RH Jr, McDonald FS. "Problem residents": prevalence,
347 problems and remediation in the era of core competencies. *Am J Med.* 2012;125(4):421–425.
348
- 349 7. Zbieranowski I, Takahashi S, Verma S, Spadafora SM. Remediation of residents in difficulty: a
350 retrospective 10-year review of the experience of a postgraduate board of examiners. *Acad*
351 *Med.* 2013;88(1):111–116.
352
- 353 8. Regan L, Hexom B, Nazario S, Chinai SA, Visconti A, Sullivan C. Remediation methods for
354 milestones related to interpersonal and communication skills and professionalism. *J Grad Med*
355 *Educ.* 2016;8(1):18-23.
356
- 357 9. Silverberg M, Weizberg M, Murano T, Smith JL, Burkhardt JC, Santen SA. What is the prevalence
358 and success of remediation of emergency medicine residents? *West J Emerg Med.* 2015
359 Nov;16(6):839-44.
360
- 361 10. Papadakis M, Hodgson C, Teherani A. Unprofessional behavior in medical school is associated
362 with subsequent disciplinary action by a state medical board. *Acad Med.* 2004; 799(3): 244-249.
363
- 364 11. Papadakis M, Teherani A, Banach M, et al. Disciplinary action by medical boards and prior
365 behavior in medical school. *N Engl J Med.* 2005;353(25):2673–2682.
366
- 367 12. Papadakis M, Arnold G, Blank L, et al. Performance during internal medicine residency training
368 and subsequent disciplinary action by state licensing boards. *Ann Intern Med.*
369 2008;148(11):869–876
370
- 371 13. Vukmir RB. Medical Malpractice: managing the risk. *Med Law.* 2004;23(3):495-513.
372
- 373 14. Joint Commission Center for Transforming Healthcare Releases Targeted Solutions Tool for
374 Hand-Off Communications Joint Commission Perspectives®, August 2012, Volume 32, Issue 8.
375
- 376 15. Panagioti M, Geraghty K, Johnson J, et al. Association between physician burnout and patient
377 safety, professionalism, and patient satisfaction: a systematic review and meta-analysis. *JAMA*

- 378 Intern Med. 2018;178(10):1317-1330.
379
- 380 16. Bird SD, Blomkalns A, Deiorio NM, Gallhue FE. Beyond test scores and medical knowledge: the
381 standardized video interview, an innovative and ethical approach for holistic assessment of
382 applicants. Acad Med. 2018;93(2):151.
383
- 384 17. The AAMC[®] Standardized Video Interview[™]: Essentials for the ERAS[®] 2019 Season.
385 [https://aamc-orange.global.ssl.fastly.net/production/media/filer_public/75/19/75198c9f-29d1-](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)
386 [4884-b932-1e7975e4262a/standardized-video-interview-essentials-eras-2019-20180720.pdf](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)
387 Accessed October 27, 2018.
388
- 389 18. Buckley RJ, Hoch VC, Huang RD. Lights, camera, empathy: a request to slow the emergency
390 medicine Standardized Video Interview project study. AEM Educ Train. 2018;2(1):57-60.
391
- 392 19. Summary of AAMC[®] Standardized Video Interview[™] Total Scores. [https://aamc-](https://aamc-orange.global.ssl.fastly.net/production/media/filer_public/5c/47/5c4790d6-e44f-43a0-8548-92e18243fb74/svi-scorereport-percentileranks-2017-2018.pdf)
393 [orange.global.ssl.fastly.net/production/media/filer_public/5c/47/5c4790d6-e44f-43a0-8548-](https://aamc-orange.global.ssl.fastly.net/production/media/filer_public/5c/47/5c4790d6-e44f-43a0-8548-92e18243fb74/svi-scorereport-percentileranks-2017-2018.pdf)
394 [92e18243fb74/svi-scorereport-percentileranks-2017-2018.pdf](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.
395
- 396 20. Schnapp, BH, Ritter D, Kraut AS, Fallon S, Westergaard, MC. Assessing residency applicants'
397 communication and professionalism: standardized video interview scores compared to faculty
398 gestalt. West J EM. 2018;X(X)X-X (accepted and epublished. Pending full citation).
399
- 400 21. Haurani MJ, Rubinfeld I, Rao S, et al. Are the communication and professionalism competencies
401 the new critical values in a resident's global evaluation process? J Surg Educ. 2007; 64(6):351-6.
402
- 403 22. National Resident Matching Program, Data Release and Research Committee: Results of the
404 2018 NRMP Program Director Survey. [https://www.nrmp.org/wp-](https://www.nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf)
405 [content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf](https://www.nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf) Accessed
406 October 26, 2018.
407
- 408 23. Breyer MJ, Sadosty A, Biros M. Factors affecting candidate placement on an emergency
409 medicine residency program's rank order list. West J Emerg Med 2012;13:458-62.

- 410
- 411 24. Dayal A, O'Connor DM, Qadri U, Arora VM. Comparison of male vs female resident milestone
412 evaluations by faculty during emergency medicine residency training. *JAMA Intern Med.*
413 2017;177(5):651-657.
- 414
- 415 25. Li S, Fant AL, McCarthy DM, Miller D, Craig J, Kontrick A. Gender differences in language of
416 standardized letter of evaluation narratives for emergency medicine residency applicants.
417 *AEM Educ Train.* 2017;1(4):334-339.
- 418
- 419 26. Isaac C, Chertoff J, Lee B, Carnes M. Do students' and authors' genders affect evaluations? A
420 linguistic analysis of medical student performance evaluations. *Acad Med.* 2011;86(1):59-66.
- 421
- 422 27. AAMC Standardized Video Interview Update. Association of Academic Chairs in Emergency
423 Medicine. March 13, 2018. [http://saem.org/docs/default-source/joint-retreat/2018/aacem-svi-
424 update.pdf?sfvrsn=599935fd_2](http://saem.org/docs/default-source/joint-retreat/2018/aacem-svi-update.pdf?sfvrsn=599935fd_2) Accessed November 17, 2018.
- 425
- 426 28. ACGME Emergency Medicine Milestones.
427 <https://www.acgme.org/Portals/0/PDFs/Milestones/EmergencyMedicineMilestones.pdf>
428 Accessed February 25, 2019.
- 429
- 430 29. Bullock AD, Hassell A, Markham WA, Wall DW, Whitehouse AB. How Ratings Vary by Staff
431 Group in Multi-Source Feedback Assessment of Junior Doctors. *Med Educ* 2009;43:516-520
- 432
- 433 30. Risucci DA, Tortolania AJ, Ward RJ. Ratings of Surgical Residents by Self, Supervisors and Peers.
434 *Surg, Gyn Ob* 1989;169:519-526
- 435
- 436 31. Shay SB. The Assessment of Complex Performance: A Socially Situated Interpretative Act. *Harv*
437 *Educ Rev.* 2004;74:307-329.
- 438
- 439 32. Govaerts MJB, Schuwirth LWT, Van der Vleuten CPM, Muijijens AMM. Workplace-based
440 assessment: Effects of Rater Expertise. *Adv Health Sci Educ* 2011;16:151-165
- 441

- 442 33. Lance CE, Hoffman BJ, Gentry WA, Baranik LE. Rater Source Factors Represent Important
443 Subcomponents of the Criterion Construct Space, not Rater Bias. *Hum Resour Manag Rev.*
444 2008;18:223-232.
- 445
- 446 34. Crossley J and Jolly B. Making Sense of Work-Based Assessment: Ask the Right Questions, in
447 the Right Way, About the Right Things, of the Right People. *Med Educ.* 2012; 46:28-37.
- 448
- 449 35. Bullock AD, Hassell A, Markham WA, Wall DW, Whitehouse AB. How ratings vary by staff group
450 in multi-source feedback assessment of junior doctors. *Med Educ.* 2009; 43(6): 516-20.

Author Manuscript

Table 1: Institutional characteristics by site detailing program format, applicant interviews, and SVI prevalence.

	Site A	Site B	Site C	Site D	Site E	Site F	Site G	Total
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	
# PGY-1 Positions	16	16	12	16	16	15	12	
# 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 M (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	
# missing SVI (%)	2 (1.0%)	2 (1.5%)	3 (1.5%)	2 (1.2%)	5 (2.8%)	1 (2.2%)	1 (0.6%)	
Legend: M= Mean; SD= Standard Deviation; SVI = Standardized Video Interview								

Author Manuscript

Table 2: Interviewer Demographics by Site

	Site A	Site B	Site C	Site D	Site E	Site F	Site G	Total
Interviewers (N)	32	36	25	10	15	11	22	151
Interviewer years experience M (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%
Legend: M= Mean; SD= Standard Deviation								

Table 3. Summary of SVI, ICS, and PROF scores by site and in aggregate

Site	Site A	Site B	Site C	Site D	Site E	Site F	Site G	Total
Applicants Interviewed (N)	203	132	196	171	177	229	156	1,264
SVI, M (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, M (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, M (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)
Legend: SVI= Standardized Video Interview; M= Mean; SD= Standard Deviation								

Author Manuscript

Table 4. Correlations between SVI, ICS, and PROF Scores ($N = 1,264$ Applicants)

Variable	1. SVI	2. ICS	3. Prof	M	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

* $p < .05$. ** $p < .01$.

Table 5. Inter-institutional correlations for ICS and PROF scores for applicants completing interviews at more than one participating institution where adequate sample size existed.

Institutions	<i>r</i> between ICS scores	<i>r</i> 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

*All inter-institutional correlations were statistically significant at $p < 0.05$.

Figure 1: ICS and Professionalism 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 professionalism ACGME milestones.²⁸

I. Verbal /Communication Skills

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

II. Professionalism/Maturity

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