The Standardized Video Interview: How Does It Affect the Likelihood to Invite for a Residency Interview?

Abbas Husain, MD,¹ Ida Li, MD,¹ Brahim Ardolic, MD,¹ Michael C. Bond, MD,² Jan Shoenberger, MD,³ Kaushal H. Shah, MD,⁴ Arlene S. Chung, MD,^{4,5} Jeffrey Van Dermark, MD,⁶ Jonathan M. Bronner, MD,⁷ Melissa White, MD,⁸ Todd Taylor, MD,⁸ Lukasz Cygan, DO,¹ William Caputo, MD,¹ Matthew Silver, MD,⁹ William C. Krauss, MD,⁹ Daniel J. Egan, MD,^{10,11}, and Moshe Weizberg, MD,¹

ABSTRACT

Background: The Association of American Medical Colleges instituted a standardized video interview (SVI) for all applicants to emergency medicine (EM). It is unclear how the SVI affects a faculty reviewer's decision on likelihood to invite an applicant (LTI) for an interview.

Objectives: The objective was to determine whether the SVI affects the LTI.

Methods: Nine Accreditation Council of Graduate Medical Education (ACGME)-accredited EM residency programs participated in this prospective, observational study. LTI was defined on a 5-point Likert scale as follows: 1 = definitely not invite, 2 = likely not invite, 3 = might invite, 4 = probably invite, 5 = definitely invite. LTI was recorded at three instances during each review: 1) after typical screening (blinded to the SVI), 2) after unblinding to the SVI score, and 3) after viewing the SVI video.

Results: Seventeen reviewers at nine ACGME-accredited residency programs participated. We reviewed 2,219 applications representing 1,424 unique applicants. After unblinding the SVI score, LTI did not change in 2,065 (93.1%), increased in 85 (3.8%) and decreased in 69 (3.1%; $p = 0.22$). In subgroup analyses, the effect of the SVI on LTI was unchanged by United States Medical Licensing Examination score. However, when examining subgroups of SVI scores, the percentage of applicants in whom the SVI score changed the LTI was significantly different in those that scored in the lower and upper subgroups ($p < 0.0001$). The SVI video was viewed in 816 (36.8%) applications. Watching the video did not change the LTI in 631 (77.3%); LTI increased in 106 (13.0%) and decreased in 79 (9.7%) applications ($p = 0.04$).

From the ¹Department of Emergency Medicine, Staten Island University Hospital–Northwell Health, Staten Island, NY; the ²Department of Emergency Medicine, University of Maryland School of Medicine, Baltimore, MD; the ³Keck School of Medicine of the University of Southern California, Los Angeles, CA; the ⁴Department of Emergency Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; the ⁵Department of Emergency Medicine, Maimonides Medical Center, Brooklyn, NY; the ⁶Department of Emergency Medicine, UT Southwestern Medical Center, Dallas, TX; the ⁷Department of Emergency Medicine, University of Kentucky, Lexington, KY; the ⁸Department of Emergency Medicine, Emory University School of Medicine, Atlanta, GA; the ⁹Department of Emergency Medicine, Southern California Permanente Medical Group, San Diego, CA; the ¹⁰Department of Emergency Medicine, Mount Sinai St. Luke's Roosevelt, New York, NY; and the ¹¹Columbia University Vagelos College of Physicians and Surgeons, New York, NY.

Received December 19, 2018; revision received February 7, 2019; accepted February 19, 2019.

This research was presented at the Council of Residency Directors in Emergency Medicine conference, San Antonio, TX, April 26, 2018; the Society for Academic Emergency Medicine Annual Meeting, Indianapolis, IN, May 15, 2018; and the New York Chapter of the American College of Emergency Scientific Assembly, Bolton Landing, NY, July 10, 2018.

The authors have no relevant financial information or potential conflicts to disclose.

Author contributions: Study concept and design—AH, BA, MCB, JS, DJE, and MW; acquisition of the data—AH, IL, MCB, JS, KHS, ASC, JVD, MW, TT, LC, WC, MS, WCK, DJE, and MW; analysis and interpretation of the data—AH, IL, BA, MCB, KHS, ASC, MW, DJE, and MW; drafting of the manuscript—AH, MCB, KHS, DJE, and MW; critical revision of the manuscript for important intellectual content—AH, MCB, JS, KHS, ASC, JVD, MW, TT, LC, WC, MS, WCK, DJE, and MW; statistical expertise—IL, KHS, and MW; administrative, technical, or material support—AH and MW; and study supervision—AH and MW. Supervising Editor: Daniel P. Runde, MD, MME, ?.

Address for correspondence and reprints: Abbas Husain, MD; e-mail: abbashu@gmail.com.

AEM EDUCATION AND TRAINING 2019;3:226–232

Conclusions: The SVI score changed the LTI in 7% of applications. In this group, the score was equally likely to increase or decrease the LTI. Lower SVI scores were more likely to decrease the LTI than higher scores were to increase the LTI. Watching the SVI video was more likely to increase the LTI than to decrease it.

In 2017, the Association of American Medical Colleges (AAMC) instituted the standardized video interview (SVI) as a pilot program for all applicants to emergency medicine (EM) residency programs. The SVI is an “online unidirectional video.”¹ Applicants are presented six questions via text prompts and have 30 seconds to review the question prior to responding. They then have up to 3 minutes to record an audio and video response to the question. The questions represent a mix of behavioral and situational questions to assess interpersonal and communications skills as well as knowledge of professionalism.¹ Applicants record six different 3-minute videos that are subsequently scored by two AAMC-trained expert reviewers. Each video is assigned a score of 1 representing rudimentary to 5 representing exemplary.² Thus, each applicant receives a total possible score of 6 to 30. Residency programs may view each applicant’s total score and also the entire video response of all six questions.

The SVI was created to provide faculty application reviewers an objective data point that assesses the applicants’ professionalism and interpersonal and communication skills. The intent was that this could widen the pool of applicants invited for an interview.¹ However, the decision on the likelihood to invite (LTI) an applicant for an interview for residency is multifactorial and varies among programs.^{2,3} Although the intent was to widen the applicant pool, the SVI theoretically has the ability to potentially limit an applicant as well. Therefore, in real-world use, it is unclear in which direction the SVI affects faculty application reviewers’ decisions on the LTI. We designed this multicenter study to investigate the impact of the SVI on a faculty reviewer’s decision on whether or not to offer an applicant an interview.

METHODS

Nine Accreditation Council of Graduate Medical Education (ACGME)-accredited EM residency programs participated in this prospective, observational study from September 15 through November 1, 2017. At each study site, faculty members who review residency applications reviewed each Electronic Residency

Application Service (ERAS) application. Application reviewers included program directors (PD), assistant PDs, and clerkship directors who are the decision makers involved in application review in their programs.

The LTI decision is multifactorial and varies among programs. Thus, to appropriately control for other factors affecting an institution’s LTI, such as United States Medical Licensing Examination (USMLE) scores, standard letters of evaluation (SLOEs), medical student performance evaluations (MSPEs), and medical school transcripts, we allowed reviewers to view all aspects of the ERAS application but blinded them to the SVI score. Blinding was accomplished by requiring the reviewers to assign an LTI without opening the SVI tab on the ERAS application. The faculty reviewers then assigned an LTI based on the traditional screening process. Faculty application reviewers were subsequently unblinded to the SVI score (without viewing the video) by opening the SVI tab in ERAS and were asked again to determine the LTI. We compared the post-SVI score LTI to the pre-SVI LTI. The LTI was determined on a 5-point Likert scale: 1 = definitely not invite, 2 = likely not invite, 3 = might invite, 4 = probably invite, 5 = definitely invite.⁴

In a subgroup of the study, faculty reviewers had the option of not only seeing the SVI score but also viewing the actual SVI video. Viewing the video was left to the discretion of the faculty reviewer. In the video group the faculty application reviewers assigned a pre-SVI LTI and a post-SVI video LTI. We then compared post-SVI video LTI to pre-SVI LTI. Ultimately this led to three groups for comparison: pre-SVI LTI, post-SVI score LTI, and post-SVI video LTI.

A predesigned standardized form (Data Supplement S1, Figure S1, available as supporting information in the online version of this paper, which is available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10331/full>) was designed by the principal investigators (face validity) to capture the decision-making process of a faculty application reviewer. It was piloted by four faculty application reviewers to ensure operational and construct validity. All faculty application reviewers were trained in the study protocol via a conference call and

in one-on-one sessions with an investigator to ensure consistency and accuracy.

There were 17 total faculty application reviewers throughout the study period from the nine institutions. Each application was screened by one faculty reviewer per site.

Inclusion criteria were applications received at any one of the participating EM residency programs that were selected to be screened for an interview. We viewed each application as a single data point. Thus, the same applicant could be reviewed at multiple sites and count as separate data points for the purpose of the study. We believe that this was appropriate because each site individually assesses each applicant independent of other sites. Exclusion criteria included applications which had no SVI score reported and applications in which the reviewer prematurely saw the SVI score prior to assigning an LTI.

Applicant demographic data were exported directly from ERAS applications. Reviewer demographic data were entered by the reviewers directly into a Google form. LTI was entered directly into a RedCAP (v 8.2.2) database.

Predetermined subgroup analyses were performed based on applicants' USMLE scores as follows: ≤ 200 , 201–220, 221–240, 241–260, and >260 ; and based on SVI scores as follows: 6–11, 12–17, 18–23, and 24–30. These SVI score ranges are described by the AAMC as representing different proficiency levels on the target competencies.²

Percentages of LTI were compared using chi-square. Mean SVI scores and LTIs were compared using t-test and ANOVA. Linear regression analysis was used to compare the change in LTI by SVI and USMLE score subgroups. All statistical tests were two-sided. A p-value of ≤ 0.05 was considered statistically significant. Statistics were calculated using GraphPad InStat.

A sample size analysis was performed prior to initiating the study. AAMC data state that 20% of EM applications result in an invitation to interview.⁵ Assuming an alpha of 0.05, to demonstrate that the SVI causes a 10% change in LTI would require a sample size of 1,537. The study was reviewed by the institutional review board at Northwell Health.

RESULTS

Seventeen faculty reviewers at nine ACGME-accredited residency programs participated in the study (Table 1). A total of 2,310 applications were reviewed. We

Table 1
Demographic of Residency Programs Completing the Screening Process

Residency programs	
Number of programs	9
Hospital type	
University	8 (89)
Community	1 (11)
Region	
Northeast	4 (40)
South	3 (30)
West	2 (30)
Training program	
3-year	7 (78)
4-year	2 (22)
Application reviewers	
Number of screeners	17
Position	
PD	6 (35)
Associate/assistant PD	8 (47)
Clerkship director	3 (18)
Years of experience screening ERAS applications, mean ($\pm$ SD); range	4.9 ($\pm$ 2.12); 1–14
Watched the SVI training videos provided by the AAMC	14 (77.8)

Data are reported as *n* (%) unless otherwise specified. AAMC = Association of American Medical Colleges; ERAS = Electronic Residency Application Service; PD = program director.

excluded 76 applications that had no SVI score and 15 in which the faculty reviewer was prematurely unblinded to the SVI score, leaving a study population of 2,219 applications (Figure 1).

There were 1,424 unique applicants and demographic data were available for 1,371 (Table 2). The remainder were unavailable as some institutions blocked ERAS demographics. These applications were not excluded from analysis as the demographics would not have affected the LTI.

The mean pre-SVI LTI was 3.0 ± 1.47 and the mean post-SVI score LTI was 3.0 ± 1.4 ($p = 0.8$). After the SVI score was revealed to faculty application reviewers, LTI was unchanged in 2,065 (93.1%, 95% confidence interval [CI] = 92.4%–94.5%), increased in 85 (3.8%, 95% CI = 3.08%–4.72%), and decreased in 69 (3.1%, 95% CI = 2.42%–3.925) applications ($p = 0.22$).

Subgroup analyses were performed by USMLE score and by SVI score. For all USMLE subgroups, SVI score did not lead to a change in LTI (Data Supplement S1, Table S1). However, in subgroup analysis by SVI score, the percentage of applications in which the SVI score changed LTI was significantly different

Figure 1. Flow of applications. SVI = standardized video interview.

among the subgroups ($p < 0.0001$). The largest affect was seen in the 6–11 score range (64.3% decreased; Table 3; Data Supplement S1, Figure S2).

There were 584 unique applicants that were reviewed at more than one site. The intraclass correlation coefficient (ICC) of the effect of SVI score on change in LTI was 0.129. This represents a poor correlation between sites.

The SVI video was viewed in 816 (36.8%) applications (814 viewed part of the video; two viewed the entire video). In the video group, we recorded a pre-SVI LTI while blinded to both the SVI score and the video and then unblinded the SVI score and then the video to compute a post-SVI score LTI and a post-SVI video LTI, respectively. There was no change in post-SVI video LTI in 631 (77.3%, 95% CI = 74.3%–80.2%), it increased in 106 (13.0%, 95% CI = 10.8%–15.5%) and decreased in 79 (9.7%, 95% CI = 7.73%–11.9%; $p = 0.04$). The percentage of applications in which the SVI video changed the LTI was significantly different for the various SVI subgroups ($p < 0.0001$; Table 4; Data Supplement S1, Figure S3).

DISCUSSION

Our study found that the SVI score changed the LTI in 7% of applications. However, there was no

Table 2
Demographic Characteristics of Applicants

Applicant Demographics	N = 1,371	Range	Mean $\pm$ SD
Age (years)		20–52	29.0 $\pm$ 3.3
Sex, n (%)			
Male	899 (66)		
Female	472 (34)		
Medical school region, n (%)			
Northeast	445 (32)		
Central	261 (19)		
South	419 (31)		
West	176 (14)		
International	70 (5)		
Medical school type, n (%)			
U.S. private	439 (32)		
U.S. public	698 (51)		
Osteopathic	164 (12)		
International	70 (5)		
Licensing examination score			
USMLE 1		186–271	230.3 $\pm$ 16.4
USMLE 2		187–282	243.0 $\pm$ 15.1
USMLE 2 CS			
Passed	99.1%		
Failed	0.9%		
COMLEX 1		405–790	576.1 $\pm$ 66.3
COMLEX 2		379–840	616.5 $\pm$ 89.5
COMLEX 2 PE			
Passed	99.2%		
Failed	0.8%		
SVI score		9–28	19.6 $\pm$ 3.0

COMLEX = Comprehensive Osteopathic Medical Licensing Examination; SVI = standardized video interview; USMLE = United States Medical Licensing Examination.

difference between the percentage of applications in which the LTI increased or decreased. When comparing SVI scores, lower scores (<11) were more likely to decrease the LTI. Higher SVI scores were not as impactful. This suggests that SVI scores less than 11 could potentially decrease the LTI.

Viewing the video was more likely to increase the LTI than to decrease it. Watching the applicant answer questions and seeing how they communicated and expressed professionalism was more likely to increase the LTI than was the numerical SVI score.

The SVI represents an AAMC initiative to provide objective data for residency application reviewers to evaluate when determining whether or not to interview an applicant. Much of an ERAS application is

Table 3
Effect of SVI Score on LTI Subdivided by SVI score

SVI SCORE	<i>n</i>	Pre-SVI LTI (Mean ± SD)	Post-SVI Score LTI (Mean ± SD)	p-value	% That Increased (95% CI)	% That Decreased (95% CI)	% Unchanged (95% CI)	p-value
6–11	14	2.43 ± 1.16	1.71 ± 0.99	0.09	0	64.3 (39.2 to 89.4)	35.7 (10.6 to 60.8)	<0.0001
12–17	500	2.75 ± 1.45	2.65 ± 1.43	0.25	0.4 (–0.15 to 0.95)	10.2 (7.68 to 13.2)	89.4 (86.4 to 92.0)	
18–23	1,483	3.05 ± 1.47	3.07 ± 1.47	0.71	2.9 (2.11 to 3.89)	0.61 (0.28 to 1.15)	96.5 (95.4 to 97.4)	
24–30	222	3.28 ± 1.43	3.47 ± 1.37	0.16	18 (13.2 to 23.7)	0	82.0 (76.3 to 86.8)	

Percentage that increased represents the percentage of applications in which the LTI increased after viewing the SVI score.

$p < 0.0001$ represents the test of differences in proportions.
LTI = likelihood to invite; SVI = standardized video interview;

Table 4
Effect of SVI Video on LTI Subdivided by SVI Score

SVI SCORE	<i>n</i>	Pre-SVI LTI (Mean ± SD)	Post-SVI LTI (Mean ± SD)	p-value	% That Increased (95% CI)	% That Decreased (95% CI)	% Unchanged (95% CI)	p-value
6–11	11	2.55 ± 1.21	1.73 ± 1.19	0.12	0	63.6 (35.2 to 92.0)	36.4 (7.97 to 64.8)	<0.0001
12–17	179	3.03 ± 1.39	2.80 ± 1.39	0.12	2.2 (0.05 to 4.35)	24.0 (18.0 to 31.0)	73.7 (66.6 to 80.1)	
18–23	521	3.10 ± 1.33	3.18 ± 1.34	0.33	12.9 (10.1 to 16.1)	5.4 (3.60 to 7.66)	81.8 (78.2 to 85.0)	
24–30	105	3.26 ± 1.23	3.63 ± 1.12	0.02	33.3 (24.4 to 43.2)	0.95 (–0.91 to 2.81)	65.7 (55.9 to 74.7)	

Percentage that increased represents the percentage of applications in which the LTI increased after viewing the SVI video.

$p < 0.0001$ represents the test of differences in proportions.
LTI = likelihood to invite; SVI = standardized video interview.

subjective (i.e., letters of recommendation, curriculum vitae) such that there is little objective information to help differentiate a higher-tier applicant from a lower one. Currently, the USMLE score is the most discerning part of the application as it correlates with board certification pass rates and success on in-training examinations.⁶

There are currently limited published data available on the SVI or its effects on LTI. The SVI has been shown to not correlate with USMLE scores thus providing an additional piece of data not previously available on ERAS applications.⁷ One single-site study found no correlation between the SVI score and faculty gestalt scores on communication and professionalism.⁸

The SVI was designed to quantify professionalism and interpersonal communication skills. As per the AAMC, the purpose of the SVI is to help residency programs widen the pool of applicants.¹ This study was designed to see how residency programs would use SVI data to decide whether an applicant warrants

an interview invitation. As a new construct, the SVI has not yet been proven to correlate with success or failure in residency training. Therefore, it is unclear how much weight PDs and application reviewers place on the SVI when determining whether to invite an applicant for an interview. Traditional factors outlined previously such as national board and EM testing scores, course clerkship scores, MSPEs, and SLOEs continue to be the major influencers on the LTI for PDs.⁹ It follows that the value that individual PDs place on the SVI should also be assessed.

Because of the growth of interview season and the numbers of programs applied to by each candidate, residency programs already divert numerous hours and resources away from their primary missions of education, research, and clinical care. A lengthy, time-intensive process of evaluating hundreds of students for each potential PGY-1 residency position has developed. One program in this study received over 1,500 applications which could mean over 450 hours of video. This additional time would add to the cost of

screening residency applications and is likely an insurmountable task for most residency programs. EM faculty spend an average of 250 hours per program on their current LTI decisions. In addition, EM residency programs and applicants appear to spend over US\$66 million per cycle on the interview process.¹⁰ To view every video for every applicant would represent an insurmountable cost to a residency program. However, a targeted approach to video reviews would be prudent and could be a future subject to investigate. It is important to remember that the AAMC recommends that the SVI not be used in isolation. Rather it should be interpreted in the context of the complete application.²

Future research can focus on whether the SVI correlates with real-world residency outcomes including performance on communication and professionalism milestones. Ultimately, it will take several years of data to answer this question.

LIMITATIONS

Each program has a different method of evaluating applicants and places varying values on different elements of the application. Thus, there were limitations in standardization of the application review process. In an attempt to overcome this, we allowed reviewers to view all elements of the ERAS application as they normally would before assigning the LTI. However, they remained blinded to the SVI score. The SVI score was then revealed and they reassessed LTI. Thus, the only variable that changed was viewing the SVI score. In addition, we enrolled several sites in varying geographic locations to capture the true impact of the SVI on an applicant's LTI.

Each site determined which applications would be screened for an interview based on their preexisting criteria. This was not standardized across the study sites. Thus, there may have been applicants who were not screened in our study due to other factors preventing them from receiving an interview (i.e., low USMLE scores). Also, not all applications at all sites were captured as not all faculty reviewers participated in the study.

The faculty application reviewers were not blinded to the purpose of the study. Thus, there was likely a Hawthorne effect on their decision on the LTI after seeing the SVI score. However, there is no way to effectively ask faculty reviewers to determine pre-SVI LTI and post-SVI LTI without them realizing that the only variable that changed was the SVI score. There is

also no way to blind them to the identity of the applicant because that would prevent them from having access to the personal statement, SLOEs, and MSPEs.

We viewed each application as a single data point. Thus, some applicants were reviewed at multiple sites and counted as separate data points for the purpose of the study. We believe this was appropriate because each site individually assesses each applicant independent of other sites.

We excluded applications in which the reviewer prematurely saw the SVI score prior to assigning an LTI. This was determined by self-reporting of the reviewer. It is possible that some reviews were prematurely unblinded to the SVI but were not excluded.

Our study was limited to nine programs. While we included many applications in the study, it is possible that other programs would value the SVI score differently and it may have a different impact on LTI in each institution. In fact, in our study, the ICC of the effect of the SVI score on change in LTI was poor. This suggests that various sites may use the SVI score in different ways.

Our study occurred during the pilot phase of the SVI. Thus, residency leaders had limited experience with the SVI and may have been unsure how to incorporate it into the application review process. In future years, as comfort with the SVI increases, the SVI may exert stronger influence on application review.

Watching the actual SVI video was optional in our study. We also did not standardize how long reviewers watched the videos. It is possible that by watching more of the video their LTI would have been different.

CONCLUSIONS

In a multicenter prospective observational study, the standardized video interview score changed the likelihood to invite in 7% of applications. In this group, the score was equally likely to increase or decrease the likelihood to invite. Lower standardized video interview scores were more likely to decrease the likelihood to invite than higher scores were to increase the likelihood to invite. The standardized video interview video was more likely to increase the likelihood to invite than to decrease it.

References

1. About the SVI. Association of American Medical Colleges. Available at: <https://students-residents.aamc.org/>

- applying-residency/article/about-svi/. Accessed March 13, 2018.
2. Using AAMC Standardized Video Interview Scores in Residency Selection: A Resource Guide. ERAS®2018 Season. Available at: https://higherlogicdownload.s3.amazonaws.com/AAMC/5ec089df-d33f-4763-a072-1c51350cf835/UploadedFiles/3Yzt9EJdRdWt6gNqBCu2_Using%20SVI%20Scores%20in%20Residency%20Selection_2017%2008%2008.pdf. Accessed March 31, 2018.
 3. Bhat R, Takenaka T, Levine B, et al. Predictors of a top performer in emergency medicine residency. *J Emerg Med* 2015;49:505–12.
 4. Artino AR Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach* 2014;36:6:463–74.
 5. Results and Data. 2017 Main Residency Match®. Available at: <http://www.nrmp.org/wp-content/uploads/2017/06/Main-Match-Results-and-Data-2017.pdf>. Accessed March 31, 2018.
 6. Harmouche E, Goyal N, Pinawin A, Nagarwala J, Bhat R. USMLE scores predict success in ABEM initial certification: a multicenter study. *West J Emerg Med* 2017;18:544–9.
 7. Egan D, Husain A, Bond M, et al. Standardized video interviews do not correlate to United States Medical Licensing Examination Step 1 and Step 2 scores. *West J Emerg Med* 2019;20:87–91.
 8. 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* 2018;20:132–7.
 9. Results of the 2018 NRMP Program Director Survey. National Resident Matching Program. Available at: <https://mk0nrmpcikgb8jxyd19h.kinstacdn.com/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf>. Accessed January 30, 2018.
 10. Van Dermark J, Wald DA, Corker JR, Reid DG. Financial implications of the emergency medicine interview process. *AEM Educ Train* 2017;1:60–9.

Supporting Information

The following supporting information is available in the online version of this paper available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10331/full>

Data Supplement S1. Supplemental material.