ORIGINAL ARTICLE

The Advanced Certifying Exam Simulation-Pro assessment instrument: evaluating surgical trainee examsmanship in virtual oral exams

Eileen R. Smith[1] · Julie M. Clanahan[1] · Annie Hess[1] · Britta Han[1] · Bradley S. Kushner[1] · T. K. Pandian[1] · Paul E. Wise[1] · Michael M. Awad[1] · Jessica B. Kramer[1]

Abstract

Purpose In response to the COVID-19 pandemic, many educational activities in general surgery residency have shifted to a virtual environment, including the American Board of Surgery (ABS) Certifying Exam. Virtual exams may become the new standard. In response, we developed an evaluation instrument, the ACES-Pro, to assess surgical trainee performance with a focus on examsmanship in virtual oral board examinations. The purpose of this study was two-fold: (1) to assess the utility and validity of the evaluation instrument, and (2) to characterize the unique components of strong examsmanship in the virtual setting, which has distinct challenges when compared to in-person examsmanship.

Methods We developed a 15-question evaluation instrument, the ACES-Pro, to assess oral board performance in the virtual environment. Nine attending surgeons viewed four pre-recorded oral board exam scenarios and scored examinees using this instrument. Evaluations were compared to assess for inter-rater reliability. Faculty were also surveyed about their experience using the instrument.

Results Pilot evaluators found the ACES-Pro instrument easy to use and felt it appropriately captured key professionalism metrics of oral board exam performance. We found acceptable inter-rater reliability in the domains of verbal communication, non-verbal communication, and effective use of technology (Guttmann's lambda-2 were 0.796, 0.916, and 0.739, respectively).

Conclusions The ACES-Pro instrument is an assessment with evidence for validity as understood by Kane's framework to evaluate multiple examsmanship domains in the virtual exam setting. Examinees must consider best practices for virtual examsmanship to perform well in this environment.

Keywords Mock oral exams · General surgery · Remote learning · Surgical education

Introduction

The general surgery board certification process includes a written Qualifying Exam and an oral Certifying Exam. Despite recent improvements in first-attempt passage rates, the oral examination component remains a challenge for trainees in achieving board certification. The American Board of Surgery (ABS) reports first-time failure rates ranging from 15 to 20% nationally [1]. General surgery programs have a vested interest in adequately preparing trainees for

these exams for two reasons. First, adequate preparation produces safe, competent board-certified surgeons. Second, the ACGME continues to consider board exam passage rates a key metric of program quality [2].

In recent years, simulated mock oral exams have been the primary method to prepare trainees for the high-stakes Certifying Exam [3]. Studies demonstrate that trainees find these experiences valuable, but more importantly, rigorous mock oral exam experiences have been shown to significantly improve first-time board passage rates [4–7]. While strong clinical knowledge is a critical element of successful oral board performance, trainee examsmanship including skills such as effective communication, organized thought process, and appropriate body language, is equally important for successful oral board completion [8]. At our institution, we previously developed and described the Advanced Certifying

✉ Eileen R. Smith
  e.r.smith@wustl.edu

1 Department of Surgery, Washington University School of Medicine, 660 South Euclid Street, Campus Box 8109, Saint Louis, MO 63110, USA

Exam Simulation (ACES) Program which utilized a public, in-person mock oral exam format with focused feedback on these examsmanship skills. This program led to a significant improvement in trainee passage rates when compared to traditional mock oral exams (100% versus 83.3%, $p = 0.049$) [8]. Though feedback in these sessions was delivered verbally to examinees immediately following the conclusion of the case scenarios, it was not standardized or organized with a formal assessment rubric. Other researchers have similarly found that identifying and addressing communication deficits remains a key strategy in optimizing trainee performance [9–12].

The COVID-19 pandemic has impacted how general surgery programs deliver educational content to trainees, necessitating creative remote learning solutions in many cases [13–15]. The ACES Program previously took place in person during regularly scheduled educational conference time [8]. However, in response to pandemic-driven limitations on in-person education, our program shifted our educational activities to a virtual setting using Zoom video conferencing (Zoom Video Communications Inc., 2019). This shift included both our ACES Program as well as the planned annual city-wide mock orals event which includes residents and faculty from another local general surgery program. The ABS has also administered its Certifying Exam online since 2020 to eliminate travel and face-to-face interactions and plans to continue to use virtual exams moving forward [16]. Given the global shift towards virtual examinations and mock orals, several studies have been published on initial experiences with this novel format. Two multi-institutional studies focused on vascular and cardiothoracic surgery trainees, for example, found the virtual mock oral platform to be both feasible and useful for participants [17, 18]. Similar pilot experiences have been published in the general surgery literature indicating that both faculty and trainee participants are satisfied with the experience and seek standardization of the process moving forward [19, 20]. In these pilot studies, trainees were primarily evaluated with pass–fail case rubrics and written feedback. To our knowledge, there have been no studies specifically evaluating the examsmanship component of the virtual examination environment.

Prior research has shown the importance of examsmanship skills in live certifying exam performance, and we hypothesized that these skills are also important for success in the virtual setting. In addition, new elements such as lighting, background, and internet connectivity are unique to the virtual environment and have yet to be examined for their impact on overall examsmanship performance. Therefore, the aim of this study is twofold. First, we aimed to characterize strong examsmanship performance in the setting of a virtual oral board exam. Second, we aimed to provide valid evidence for a new evaluation instrument—the ACES-Pro—to assess examsmanship performance and enable the delivery of standardized formative feedback on virtual exam performance.

Methods

Application of a validity framework

With the development of any new assessment tool, it is critical to collect and interpret evidence for the validity of its use. While there are many validity frameworks, herein we describe the application of Kane's framework to organize our evidence as well as identify evidence gaps. The elements of this framework include a clear statement of the proposed use of the assessment as described below in our assessment instrument development description and evaluation of four inferences: scoring, generalization, extrapolation, and implication [21].

Assessment instrument development

With input from both junior and senior staff surgeons at our quaternary care academic teaching hospital, we developed a new evaluation instrument to assess examsmanship in the virtual oral board exam setting. The proposed use of the assessment is to facilitate standardized feedback to trainees on their mock virtual oral board performance. Our instrument development group included two senior surgeons with prior experience as ABS examiners who served as subject matter experts and two junior surgeons who completed the oral board exam within the past five years. Those surgeons with direct experience as oral board examiners had the requisite three years elapse since their last involvement in this role. After a series of interviews with members of this faculty group, an initial assessment form was drafted to include both traditional professionalism metrics previously evaluated in the literature as well as new metrics focused on virtual examination. It was then subjected to repeated revisions until consensus on metrics was reached by the group (Fig. 1). The use of both subject matter experts and a review of relevant metrics in the literature for item construction helped provide evidence for content validity and support the scoring inferences of Kane's framework as well as provide a relevant test domain with the goal of supporting the extrapolation inference [21]. This assessment was designated as the ACES-Pro instrument in reference to our existing oral boards program, the Advanced Certifying Exam Simulation (ACES), and the new focus on professionalism metrics (Pro) in the virtual setting.

The final version of the ACES-Pro instrument (Online Appendix 1) consists of 15 standardized questions and a free text box for narrative feedback on performance and internet connectivity. The evaluation assesses three key
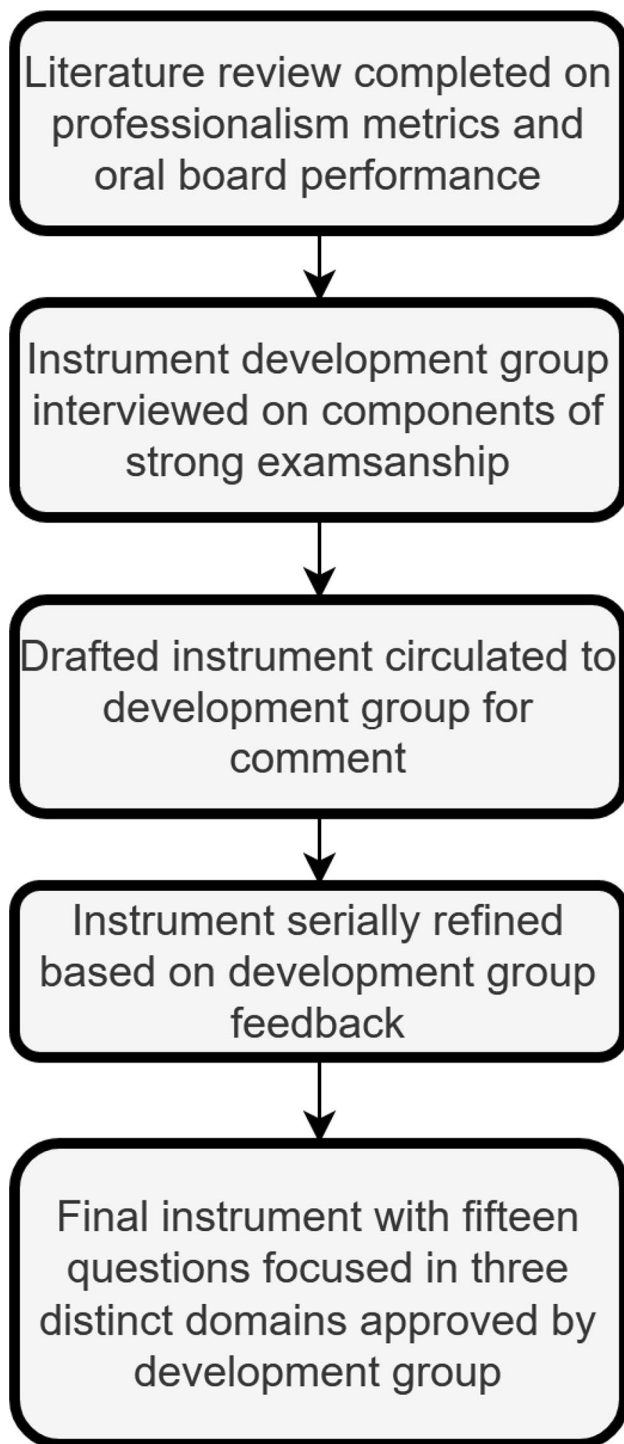
Literature review completed on professionalism metrics and oral board performance

↓

Instrument development group interviewed on components of strong examsanship

↓

Drafted instrument circulated to development group for comment

↓

Instrument serially refined based on development group feedback

↓

Final instrument with fifteen questions focused in three distinct domains approved by development group

Fig. 1 Instrument design process map

domains of virtual examsmanship: verbal communication, non-verbal communication, and effective use of technology. Specific metrics in these domains include both traditional elements such as clarity of speech, response organization, and use of filler words as well as those relevant and unique for the virtual setting, such as image framing, image background, and lighting.

Participating pilot evaluators

Following the development of the evaluation instrument, nine attending surgeons at our academic medical center participated in a pilot of the ACES-Pro tool to assess its usability and validity. This pilot study was approved by our institutional review board. We included seven junior surgeons (< 5 years in practice) and two senior surgeons (> 5 years in practice). Four of our nine pilot evaluators had assisted with the development of the ACES-Pro instrument. Areas of faculty specialization were diverse including pediatric surgery, minimally invasive surgery, colorectal surgery, endocrine surgery, and acute and critical care surgery. All participating surgeons were closely involved in general surgery resident education efforts and our existing oral boards curriculum. One of the senior faculty members had previously served as an oral board examiner for the ABS. This faculty member participated in both the development of the original ACES program and the ACES-Pro instrument. Both senior faculty surgeons who participated had also previously served as examiners in our formal city-wide mock oral exam program which simulates a full exam experience for both our program and another local surgical program. Of the junior attending surgeons who participated in our study, two of seven had previously participated as examiners in this same city-wide mock oral exam.

Assessment instrument pilot

To pilot the ACES-Pro instrument, our pilot evaluators reviewed four previously recorded virtual oral board exam scenarios. They did not receive any formalized training in the use of the assessment tool prior to completing the pilot. We chose to use previously recorded simulated oral exam sessions in our pilot to focus on characterizing the instrument's ease of use, reliability, and suitability for evaluating performance in the virtual exam format and to minimize other variables prior to real-time utilization with our trainees. The recordings used were generated with the examinee permission during prior remote learning sessions in the setting of a regional COVID-19 surge. Three of the examinees were post-graduate year (PGY) 5 residents, and the fourth was a PGY4 resident. The faculty examiners in these recordings were attending thoracic surgeons who had no role in developing the ACES-Pro instrument and were not among the pilot evaluators for this study. The case scenarios for this set of recorded exams focused on thoracic surgery topics and were chosen specifically to limit bias as thoracic surgery was outside the primary scope of practice for our pilot evaluators.

After viewing each recording, our pilot evaluators scored each examinee using the ACES-Pro instrument. Once all four recordings were viewed and scored, attending surgeons completed a brief survey about their overall experience using the instrument.

Statistical analysis

To address Kane's generalization inference, we evaluated evidence of reproducibility and reliability in assessment scores generated from the ACES-Pro tool. We chose not to use raw percentage agreement as this does not account for chance agreement. Instead, we utilized Gwet's Agreement Coefficient 2 (AC2) to assess rater agreement. Gwet's AC2 was chosen as it does not depend upon the assumption of independence between raters and can be used with ordinal data. Due to high levels of agreement between raters for individual items in our dataset—a result of ordinal scores in our instrument generated from a 2- or 3- point scale—traditional use of Cohen's Kappa was less reliable due to "Kappa paradox" [22]. Similar to traditional Kappa values, Gwet's AC2 values can be interpreted to indicate broad levels of agreement. In this study, we chose interpretations based on those used for traditional kappa values: $AC2 < 0.2 = Poor$; $0.2 < AC2 \leq 0.4 = Fair$; $0.4 < AC2 \leq 0.6 = Moderate$; $0.6 < AC2 \leq 0.8 = Good$; $0.8 < AC2 \leq 1 = Very good$ [23].

To assess reliability, the ACES-Pro instrument was divided into "factors" to evaluate each domain of assessment (e.g., verbal, non-verbal, and technology). Guttman's Reliability (Lambda-2) was then calculated to assess the overall reliability of each factor. Guttman's Lambda-2 was chosen as it does not rely on the assumptions of equal covariances with the "true" score and no error correlations. Lambda-2 is similar to the traditional statistic Cronbach's alpha, but rather than being an estimate of between-score correlation for randomly equivalent measures, Lambda-2 is an estimate of the between-score correlation for parallel measures. An alternative method to calculate reliability is Guttman's Reliability (Lambda), which is based on the split-half method. Guttman derived six different reliability measures and showed that Lambda-2 could be a more accurate reliability estimate than Cronbach's alpha [24]. Since our instrument's "factors" are essentially composite tasks, Lambda-2 was thought to be more robust than Cronbach's alpha.

Results

The ACES-Pro instrument examined three key domains of virtual oral board examsmanship: verbal communication, non-verbal communication, and effective use of technology.

Evidence for scoring

We examined pilot evaluator responses for inter-rater agreement (Table 1). With a single exception, all questions achieved at least moderate (AC2 > 0.45) inter-rater agreement. In the verbal communication domain, 3 out of 6 questions had very good or good (volume, clarity, pace) inter-rater agreement (AC2 > 0.75). The remaining 3 questions had moderate (AC2 > 0.45) inter-rater agreement. For the effective use of the technology domain, all questions had good (AC2 > 0.6) inter-rater agreement. Finally, inter-rater agreement was lowest for the non-verbal communication domain, with four questions that had moderate (AC2 > 0.45) inter-rater agreement and a single category (appearance) with only fair (AC > 0.2) inter-rater agreement.

With regard to scoring procedures, all pilot evaluators were able to complete the evaluation form in 5 min or less. A majority (66%) of our evaluators completed the form while viewing the oral board scenario, rather than completing the form at the conclusion of the scenario. Most pilot evaluators (60%) rated the instrument as extremely easy to use. Narrative feedback regarding the instrument itself stated that it was user-friendly and assessed the relevant metrics related to the verbal, non-verbal, and technology domains. Two faculty reported that ease of use could be increased by making the instrument a clickable electronic form rather than a digital word-processing document.

Evidence for generalization

To assess overall reliability, the ACES-Pro instrument was divided into "factors" to evaluate each domain of assessment

Table 1 Summary of rater agreement on ACES-Pro examsmanship assessment

Domain	Category	Gwet AC2	Agreement
Verbal	Volume	0.943	Very good
	Clarity	0.769	Good
	Pace	0.897	Very good
	Pauses	0.548	Moderate
	Filler Words	0.45	Moderate
	Appropriate medical language	0.479	Moderate
Technology	Framing	0.671	Good
	Background	0.656	Good
	Lighting	0.692	Good
Non-verbal	Eye contact	0.597	Moderate
	Appearance	0.204	Fair
	Posture	0.467	Moderate
	Organization	0.542	Moderate
	Movement	0.45	Moderate

(e.g., verbal, non-verbal, and technology). Across the three domains, our evaluators had acceptable overall reliability with Guttmann's Lambda-2 of 0.80 for verbal, 0.92 for technology, and 0.74 for non-verbal.

Evidence for implications

Several key themes emerged in the review of narrative qualitative feedback for examinees from pilot evaluators. In the domain of verbal communication, even when accounting for occasional connectivity issues, pilot evaluators indicated that they were able to assess speech volume, pace, and clarity as well as the use of medical language and filler words and the overall organization of response using the instrument. In regard to non-verbal communication, pilot evaluators delivered feedback on the body language of the examinees, assessing posture and use of movement or gestures as well as the dress of examinees. For example, all 4 examinees received narrative feedback noting casual attire for the mock exams. Further, pilot evaluators noted that some examinees appeared to be looking at their video cameras–giving the appearance of eye contact with the examiner–while others focused their gaze on their own computer screens. In the domain of effective use of technology, narrative feedback indicated that some evaluators thought the video angle resulting from holding a computer on a lap (versus on a desk) or utilizing a camera phone (versus a webcam connected to an actual computer) conveyed a more casual and less professional image.

Discussion

The COVID-19 pandemic has driven a shift in surgical education towards virtual formats that utilize video-conferencing, extending to the General Surgery Certifying Examination as administered by the ABS [13–15]. Professionalism metrics have been shown to play an important role in successful performance on this high-stakes exam [8–11]. Our work aimed to characterize successful examsmanship in the virtual oral exam setting as well as develop a standardized instrument for evaluating that performance.

Our results provide evidence for the use of the ACES-Pro assessment instrument to assess key examsmanship aspects of oral board performance in the virtual setting. The instrument was relatively easy to use and took only a few minutes to complete. With respect to scoring, Gwet's agreement coefficient and Guttman's Lambda-2 results suggest acceptable overall inter-rater reliability in the three domains of verbal communication, non-verbal communication, and effective use of technology. Acceptable inter-rater reliability provides evidence of internal validity for this novel assessment tool. Similarly, the three key domains of verbal communication,

non-verbal communication, and effective use of technology showed good reliability, supporting the stability and generalizability of these factors. These provide evidence that the selected test domains adequately reflect performance in a test setting, a key component of the generalization inference. Further, a thematic review of narrative feedback provided additional insights into both the usability of the ACES-Pro, as well as key challenges to evaluating virtual exam performance.

From this study, we propose a series of recommendations for best practices in the virtual exam environment with the aim of helping examinees utilize video conferencing technology effectively and convey an appropriate professional appearance in virtual exams (Table 2). Pilot evaluator feedback indicated that verbal communication skills were relatively easy to assess in the virtual oral exam format. In contrast, variations in faculty assessment of non-verbal communication and effective use of technology highlighted several key themes which guide our recommendations for optimizing performance in the virtual environment. For example, on the topic of eye contact, some faculty noted that examinees looked away from the camera and positioned their cameras in their screen periphery. To convey a sense of focused gaze, we recommend that trainees choose a specific camera or screen focus point in advance for the examination. In another example, pilot evaluators commented on poor image framing and lighting. We recommend that trainees assess their examination environment to optimize camera angle, lighting, and neutral background before participating in virtual oral exams. While we cannot say with certainty that an examinee's failure to optimize their performance in each of these three domains would negatively impact their exam score, it is possible that a lack of attention to these areas could negatively bias the examiner's scoring as professionalism metrics can play a role in the scoring of exam performance [8–11]. Future work should investigate the impact of these virtual professionalism metrics on examinee performance. A potential study could compare performance in a matched oral board scenario with standardized questions and examinee responses for case scenarios but differing virtual professionalism metrics as assessed by the ACES-Pro instrument. This evaluation could help characterize how virtual professionalism metrics impact overall performance in case scenarios.

As our examination efforts remain in a virtual setting, the video conferencing format has also allowed easy recording of exam sessions for later review by trainees or faculty evaluators. Prior work has indicated a gap between examinees' perception of their own performance before and after the video review [25]. With the virtual mock format, however, trainees can review their recorded performance in parallel with ACES-Pro assessment feedback and potentially improve their performance for future real-life examination.

Table 2 Summary of best examsmanship practices for oral exams utilizing video conferencing technology

Virtual mock oral exams: best practices for strong examsmanship	
Metric	Best practice
Exam setting	Select a quiet room, isolated from background noise or interruptions Prior to exam, conduct a test video call to make sure internet connectivity is appropriate for video calls
Background	Background should be neutral and minimally distracting—a blank wall is preferable
Image framing	Sit in a fixed chair; swivel office chairs may increase distracting movement Utilize a web camera on a computer, rather than a smartphone The computer with a camera should be fixed on a desk rather than held on a lap Sit far enough back to center head and shoulders in the frame Attention be paid to camera angle to avoid distorted perspective
Lighting	Avoid settings with strong backlighting and overhead lighting
Gaze	Select a point for fixed gaze, either looking directly at the web camera or at the image of the examiners
Volume	Complete a test of the microphone to ensure function prior to the exam
Verbal communication	Minimize the use of filler words and long pauses Avoid excessive use of clarifying questions Utilize appropriate medical language rather than lay terms Proceed through the response in organized and deliberate fashion
Nonverbal communication	Appear in professional dress Minimize movement or gestures Upright posture

While the ACES-Pro instrument was designed to assess performance in the virtual mock oral setting, it could easily be adapted to suit in-person examination performance as well, providing key feedback across both verbal and non-verbal skill domains that impact exam performance regardless of the test-taking environment.

Our study is not without limitations. This proof-of-concept study was aimed at instrument development and initial pilot use. Therefore, only a small number of pilot evaluators participated in the assessment of the ACES-Pro. In terms of Kane's validity framework, while we were able to generate evidence to support internal validity and potential generalizability of the assessment tool, evidence for extrapolation and implication inferences is currently limited. Future work with a larger number of faculty and examinees can provide a more robust assessment of scoring metrics.

Another aspect of Kane's extrapolation inference is the relationship between test performance and real-world performance. In this study, we did not generate evidence to evaluate this relationship. In the future, as we continue to use the ACES-Pro instrument in our virtual mock orals program, however, we can address this evidence gap by tracking multiple ACES-Pro scores for the same participants over time and examining the association between these scores and eventual pass/fail outcomes on the General Surgery Certifying Examination. Positive correlation would support the extrapolation of the tool. In this pilot study, we also generated little evidence to support the implications inference of Kane's Framework. However, recognition of this evidence gap again suggests a future direction for the work. Currently, in our mock oral program, general surgery residents receive

their ACES-Pro scores and feedback after their practice performance, but there is no further follow-up that occurs. Ideally, we could make note of domains in which trainees scored poorly, for example, and reassess these domains with the ACES-Pro at a subsequent assessment after the provision of feedback. Future work to formalize these feedback mechanisms and help make pass/fail decisions for individual examinees on the simulation overall would help us better understand the impact of ACES-Pro assessment scores on the learners.

An additional limitation is that faculty pilot evaluators in this study received no formal training on the appropriate use of the ACES-Pro instrument. Introduction of more formalized evaluator training could help to standardize the use of the assessment tool as well as mitigate the impact of implicit bias on performance evaluation. Such training would establish clear guidelines for instrument use and diminish examiner variability in interpreting the tool's domains. Existing video footage collected for this work could be utilized as an examiner training curriculum, allowing for dedicated practice using the ACES-Pro instrument prior to application in real-time oral board exam scenarios. Further, our study focused only on the faculty experience using the ACES-Pro assessment and did not collect data from trainees regarding the utility of this method of feedback. Understanding the utility of this assessment tool for trainees will be a key element to include in future studies, as actionable, formative feedback is crucial in the training environment.

Though we did not explicitly measure this, informal feedback from trainees who participated in prior in-person mock orals in our ACES program as well as our virtual mock orals

early in the pandemic indicated that even in the virtual environment, the presence of an audience was able to generate a similar "high-stakes" experience. In the future, we will also plan to closely simulate an actual oral exam experience by capturing multiple case scenarios for each examinee. In the video recordings used for the pilot, trainees only completed a single case scenario, which did not allow us to assess how they transitioned between scenarios. Anecdotally, it is well-known that transitions between case scenarios can be especially challenging if a trainee struggles with the preceding scenario. Further, the faculty examiners in the pre-recorded scenarios used for this pilot were well-known to our examinees, and this familiarity could have impacted trainee behavior during the exam scenarios. In addition, when our recordings were made, these sessions were informal without explicit expectation setting. This may have contributed to a more casual approach by examinees compared to a more high-stakes stimulation, which may have decreased trainee anxiety and stress levels during scenarios. The lack of clear guidelines may also have contributed to decreased inter-rater reliability, particularly in the area of trainee appearance. Moving forward, it will be important to set clear professional and technological expectations to better simulate the actual examination experience.

We recognize that this study does not quantify the impact of implicit bias on examsmanship scores, a topic that remains critical with the transition to a virtual format. Race, ethnicity, sex, and family status have been shown to impact the first pass rate of the ABS certifying exam [26]. The ABS has incorporated implicit bias training to help mitigate this impact, but the virtual format may complicate differences in examsmanship and professionalism components. To our knowledge, no studies have attempted to characterize professionalism in this virtual format. ACES-Pro was designed to provide specific and standardized feedback so as to limit the potential for implicit bias. However, additional work with a larger sample size is needed to determine if the tool can truly deliver on this goal. The ACES-Pro does incorporate elements such as internet speed, background, professional dress, etc., which could negatively impact trainees from lower socioeconomic backgrounds. To mitigate this impact, we recommend that trainees participate in virtual sessions at the hospital/medical campus with a reliable internet connection, accessible computers, and locations with professional backgrounds. Residency programs must also be willing to support and aid their trainees in this new era of technology dependence, and examiners must be trained to recognize their own biases.

Ultimately, we hope to incorporate the ACES-Pro into a robust virtual oral boards curriculum that would allow for multi-institutional board preparation. Both examinees and examiners would be able to review standardized instruction on both the professionalism and knowledge-based components of the exam. This curriculum would include video examples demonstrating both effective and ineffective performance. After a review of the video curriculum, both examinees and examiners could participate in paired oral board practice sessions across participating institutions. This would allow examinees to gain examination experience and receive feedback from examiners outside their home program.

Conclusion

Based on these pilot results for the ACES-Pro instrument, our team has continued to incorporate standardized examsmanship feedback into our existing mock orals program both within our own institution and locally at our multi-institution, city-wide mock oral events. The ACES-Pro can augment our existing mock oral exam approach to further equip our trainees for success in both a virtual and in-person exam environment. Future work will assess trainee satisfaction with the ACES-Pro instrument as well as trainee performance longitudinally. Our eventual goal is to utilize this work to launch a multi-institutional, virtual mock orals program to connect trainees and faculty nationwide and deliver a high-fidelity virtual oral board exam experience with actionable examsmanship feedback.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. General Surgery Exam Passage Rates. http://publicdata.absurgery.org/PassRates/. Accessed 26 Oct 2021.
2. ACGME. Common Program Requirements. 2020. https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRResidency2020.pdf. Accessed 13 Aug 2020.
3. Kimbrough MK, Thrush CR, Smeds MR, Cobos RJ, Harris TJ, Bentley FR. National landscape of general surgery mock oral examination practices: survey of residency program directors. J

Surg Educ. 2018;75(6):e54–60. https://doi.org/10.1016/j.jsurg.2018.07.012.

4. Aboulian A, Schwartz S, Kaji AH, de Virgilio C. The public mock oral: a useful tool for examinees and the audience in preparation for the American Board of Surgery Certifying Examination. J Surg Educ. 2010;67(1):33–6. https://doi.org/10.1016/j.jsurg.2009.10.007.

5. Fingeret AL, Arnell T, McNelis J, Statter M, Dresner L, Widmann W. Sequential participation in a multi-institutional mock oral examination is associated with improved American Board of Surgery Certifying Examination first-time pass rate. J Surg Educ. 2016;73(6):95–103. https://doi.org/10.1016/j.jsurg.2016.06.016.

6. Fischer LE, Snyder M, Sullivan SA, Foley EF, Greenberg JA. Evaluating the effectiveness of a mock oral educational program. J Surg Res. 2016;205(2):305–11. https://doi.org/10.1016/j.jss.2016.06.088.

7. Pennell C, McCulloch P. The effectiveness of public simulated oral examinations in preparation for the American Board of Surgery Certifying Examination: a systematic review. J Surg Educ. 2015;72(5):1026–31. https://doi.org/10.1016/j.jsurg.2015.03.018.

8. London DA, Awad MM. The impact of an advanced certifying examination simulation program on the American Board of Surgery Certifying Examination passage rates. J Am Coll Surg. 2014;219(2):280–4. https://doi.org/10.1016/j.jamcollsurg.2014.01.060.

9. Rowland-Morin PA, Burchard KW, Garb JL, Coe NP. Influence of effective communication by surgery students on their oral examination scores. Acad Med. 1991;66(3):169–71. https://doi.org/10.1097/00001888-199103000-00011.

10. Rowland-Morin PA, Coe NP, Greenburg AG, et al. The effect of improving communication competency on the certifying examination of the American Board of Surgery. Am J Surg. 2002;183(6):655–8. https://doi.org/10.1016/s0002-9610(02)00861-9.

11. Rowland PA, Trus TL, Lang NP, et al. The certifying examination of the american board of surgery: the effect of improving communication and professional competency: twenty-year results. J Surg Educ. 2012;69(1):118–25. https://doi.org/10.1016/j.jsurg.2011.09.012.

12. Houston JE, Myford CM. Judges' perception of candidates' organization and communication, in relation to oral certification examination ratings. Acad Med. 2009;84(11):1603–9. https://doi.org/10.1097/ACM.0b013e3181bb2227.

13. Chick RC, Clifton GT, Peace KM, et al. Using technology to maintain the education of residents during the COVID-19 pandemic. J Surg Educ. 2020;77(4):729–32. https://doi.org/10.1016/j.jsurg.2020.03.018.

14. Hintz GC, Duncan KC, Mackay EM, Scott TM, Karimuddin AA. Surgical training in the midst of a pandemic: a distributed general surgery residency program's response to COVID-19. Can J Surg. 2020;63(4):346–8. https://doi.org/10.1503/cjs.008420.

15. Coe TM, Jogerst KM, Sell NM, et al. Practical techniques to adapt surgical resident education to the COVID-19 era. Ann Surg. 2020;272(2):139–41. https://doi.org/10.1097/SLA.0000000000003993.

16. Exam Dates and Fees. http://www.absurgery.org/default.jsp?examdeadlines. Accessed 25 Oct 2021.

17. Zemela MS, Malgor RD, Smith BK, Smeds MR. Feasibility and acceptability of virtual mock oral examinations for senior vascular surgery trainees and implications for the certifying exam. Ann Vasc Surg. 2021. https://doi.org/10.1016/j.avsg.2021.03.005. (**published online ahead of print, 2021 Apr 7**).

18. Fiedler AG, Emerson D, Gillaspie EA, et al. Multi-institutional collaborative mock oral (mICMO) examination for cardiothoracic surgery trainees: results from the pilot experience. JTCVS Open. 2020;3:128–35. https://doi.org/10.1016/j.xjon.2020.07.007.

19. Huang IA, Lu Y, Wagner JP, et al. Multi-institutional virtual mock oral examinations for general surgery residents in the era of COVID-19. Am J Surg. 2021;221(2):429–30. https://doi.org/10.1016/j.amjsurg.2020.11.002.

20. Shebrain S, Nava K, Munene G, Shattuck C, Collins J, Sawyer R. Virtual surgery oral board examinations in the era of COVID-19 pandemic. How I do it! J Surg Educ. 2021;78(3):740–5. https://doi.org/10.1016/j.jsurg.2020.09.012.

21. Cook D, Brydges R, Ginsbur S, Hatal R. A contemporary approach to validity arguments; a practical guide to Kane's framework. Med Educ. 2015;49(6):560–75. https://doi.org/10.1111/medu.12678.

22. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543–9. https://doi.org/10.1016/0895-4356(90)90158-l.

23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.

24. Callender JC, Osburn HG. An empirical comparison of coefficient alpha, Guttman's lambda-2, and MSPLIT maximized split-half reliability estimates. J Educ Meas. 1979;16(2):89–99.

25. Kozol R, Giles M, Voytovich A. The value of videotape in mock oral board examinations. Curr Surg. 2004;61(5):511–4. https://doi.org/10.1016/j.cursur.2004.06.004.

26. Yeo HL, Dolan PT, Mao J, Sosa JA. Association of demographic and program factors with american board of surgery qualifying and certifying examinations pass rates. JAMA Surg. 2019;155(1):22–30. https://doi.org/10.1001/jamasurg.2019.4081.