Title: Feasibility Study of Self-Directed Mock Oral Examinations in the Assessment of General Surgery Resident Clinical Decision-Making

Article Type: Original Article

Author Names:

1. Terrance Peng
2. Ivy A. Huang
3. Megan Nguyen
4. Marilee Fisher
5. James X. Wu
6. Justin P. Wagner
7. Areti Tillou
8. Formosa Chen

Degrees and Affiliations:

1. MD, MPH. University of California Los Angeles David Geffen School of Medicine, Department of Surgery, Los Angeles, California, United States of America
2. MD. University of California Los Angeles David Geffen School of Medicine, Department of Surgery, Los Angeles, California, United States of America
3. Undergraduate Student. University of California Los Angeles, Los Angeles, California, United States of America
4. BSN, BA (Fourth-year medical student). University of California Los Angeles David Geffen School of Medicine, Department of Surgery, Los Angeles, California, United States of America
5. MD. University of California Los Angeles David Geffen School of Medicine, Department of Surgery, Los Angeles, California, United States of America
6. MD. University of California Los Angeles David Geffen School of Medicine, Department of Surgery, Los Angeles, California, United States of America
7. MD, MEd. University of California Los Angeles David Geffen School of Medicine, Department of Surgery, Los Angeles, California, United States of America
8. MD, MPH. University of California Los Angeles David Geffen School of Medicine, Department of Surgery, Los Angeles, California, United States of America

ORCID (Open Research and Contributor Identifier)

1. 0000-0003-1661-1730
2. 0000-0003-1697-075X

3. 0009-0009-8817-9829
4. 0000-0002-1078-8722
5. 0000-0001-5090-7573
6. 0000-0002-7383-9216
7. 0000-0003-4148-4999
8. 0000-0002-3303-8377

About the author: Marilee Fisher is currently a fourth-year medical student at the David Geffen School of Medicine at the University of California, Los Angeles (UCLA). She was the recipient of the 2025 UCLA Simulation Work-in-Progress Bridge Award and is beginning her general surgery residency training at Harbor-UCLA Medical Center in June 2026.

Corresponding author email: terrancepeng@mednet.ucla.edu

Acknowledgments: None

Financing: No funds, grants, or other support was received for conducting this study.

Conflict of Interest statement by authors: The authors have no relevant financial or non-financial interests to disclose.

Author Contribution Statement: Conceptualization: TP, IAH, FC. Data Curation: TP, IAH. Formal Analysis: MN. Investigation: TP, IAH, MF, JXW, JPW, AT, FC. Methodology: TP, IAH, MN. Project Administration: TP. Supervision: FC. Writing - Original Draft: TP, IAH, MF. Writing - Review Editing: TP, MN, MF, JXW, JPW, AT, FC.

Manuscript Word Count: 3242

Abstract Word Count: 240

Number of Figures and Tables: 7 total (3 figures + 4 tables)

Personal, Professional, and Institutional Social Network accounts:

- Twitter: @terrancepeng

Abstract

Background: This feasibility study describes the implementation of a self-directed mock oral examination (SD-MOE), which is administered in the absence of faculty examiners, and explores its potential as a supplemental tool in preparation for the American Board of Surgery Certifying Examination.

Methods: During a multi-institutional MOE, all examinees completed the conventional MOE (C-MOE) and were presented the option to additionally participate in the SD-MOE. Rates of technical issues were compared between formats. Exploratory analyses evaluated participant satisfaction, resource intensity, exam performance metrics, and interrater reliability.

Results: The C-MOE included 62 faculty examiners and 55 examinees, while the SD-MOE included 25 faculty examiners and 41 examinees. The frequency of technical issues did not differ between formats. Faculty examiner satisfaction was lower with the SD-MOE compared with the C-MOE (65% vs. 93%, $P<0.01$). Post-graduate year (PGY)-level correlated with mean case score (PGY-3: 2.45; PGY-4: 2.58; PGY-5: 2.70; $P=0.01$) and proportion of questions passed (PGY-3: 81.7%; PGY-4: 87.3%; PGY-5: 90.1%; $P<0.01$) in the C-MOE, but not the SD-MOE. PGY-level did not correlate with overall exam pass rates in either exam format. Interrater reliability did not differ significantly between the SD-MOE (weighted $\kappa=0.32$, 95% CI [0.19, 0.46]) and C-MOE (weighted $\kappa=0.25$, 95% CI [0.16, 0.35]).

Conclusion: The SD-MOE was successfully administered with minimal logistical challenges. While limited by optional SD-MOE participation and differences in case counts between formats, this feasibility study suggests the SD-MOE can be considered as a supplemental tool for resource-limited institutions.

Keywords: examinations, oral; educational measurement; clinical competence; internship and residency; education, medical, graduate

Abbreviations:

ABSCE: American Board of Surgery Certifying Exam

ANOVA: one-way analysis of variance

C-MOE: Conventional mock oral examination

MOE: Mock oral examination

PGY: Post-graduate year

SD-MOE: Self-directed mock oral examination

INTRODUCTION.

Mock oral examinations (MOEs) are administered by expert clinicians to prepare residents for the American Board of Surgery Certifying Examination (ABSCE), a summative assessment of safe clinical decision-making.[1] Resident participation in MOEs predicts first-time passage rates on the ABSCE[2–5] and informs program efforts to refine educational curricula.[6–8] In response to the COVID-19 pandemic, several programs successfully adapted MOEs to virtual platforms and demonstrated resident satisfaction and feasibility.[9–11] Nevertheless, MOE administration remains logistically challenging, limiting utility for resident education and assessment.

During conventional MOEs (C-MOE), two faculty examiners present case prompts to the examinee and score responses in real-time. Consequently, C-MOE administration requires numerous faculty examiners to be available simultaneously, which can be especially challenging if recruited from multiple institutions in an effort to better simulate ABSCE testing conditions. Thus, C-MOE implementation can be profoundly hindered by limited availability and time constraints among faculty and residents.[7,10,12,13] This study explores the implementation of a novel self-directed MOE (SD-MOE) format, in which examinees deliver video-recorded verbal responses to written prompts in the absence of a faculty examiner, as a potential supplemental tool given these obstacles. The SD-MOE offers several potential logistical advantages. Most notably, the use of video-recorded submissions would obviate the need for faculty examiners to personally administer the SD-MOE, instead allowing them to later view and score recordings at a time of their convenience. Although prior investigations of live versus video-based assessment have demonstrated agreement in scoring between methods, such studies have focused on assessment of technical skills rather than clinical decision-making.[14–16]

The primary objective of this study was to evaluate the feasibility of the SD-MOE format as a supplemental tool during a multi-institutional MOE. As an exploratory aim, this study compared performance metrics between formats, acknowledging that removal of the live probing that occurs naturally during examiner-examinee interactions could meaningfully alter the construct being assessed. The hypothesis of this study is that the SD-MOE can be administered with minimal logistical challenges while demonstrating favorable participant satisfaction ratings.

METHODS.

A multi-institutional virtual MOE was administered for surgical trainees from five regional general surgery residency programs. Eligible examinees included all general surgery residents of post-graduate year (PGY) levels 3, 4, and 5 from these institutions. All examinees participated in the C-MOE and were presented the option to additionally complete the SD-MOE. Surveys were completed by faculty examiners and resident examinees at the conclusion of the entire MOE program. All case topics for the C-MOE and SD-MOE were derived from the Surgical Council on Resident Education (SCORE) curriculum and selected through consensus by the MOE organizational committee faculty. Case scenarios, prompts, and

grading criteria for the C-MOE and SD-MOE were developed by expert surgeons on the organizational committee and reviewed for content accuracy and flow by authors FC, JW, and JPW. Approval was obtained from the University of California Los Angeles Institutional Review Board (IRB #21-000707). The procedures used in this study adhere to the tenets of the Declaration of Helsinki. This study is reported in accordance with the Guideline for Reporting Evidence-based Practice Educational interventions and Teaching (GREET) 2016 checklist.

Conventional Mock Oral Examination (C-MOE)

Modeled after the ABSCE, the C-MOE has garnered high satisfaction and ease-of-use ratings among participants.[9] Each examinee in the study completed 12 standardized C-MOE cases divided evenly across three categories: Trauma Surgery and Critical Care, Gastrointestinal and Acute Care Surgery, and Subspecialty Surgery and Surgical Oncology. One week prior to the examination, faculty were distributed a personalized link to a Qualtrics survey containing case information, prompts, and grading criteria (Qualtrics, Provo, UT).

The C-MOE was administered in Zoom breakout rooms corresponding to the three aforementioned clinical categories; in each breakout room, two faculty examiners read aloud case prompts to a single examinee who provided verbal responses (Zoom Communications, San Jose, CA). At the conclusion of each case, faculty examiners used a Qualtrics survey to independently grade responses according to the supplied rubric, assigning a case-specific score of 1 (failure), 2 (borderline performance), or 3 (pass). Examinees were allotted a maximum of 30 minutes to complete four total cases in each breakout room. At the conclusion of each 30-minute testing period, examinees proceeded to the next MOE session as dictated by their individual randomized schedule until completing all three assigned rooms.

Self-Directed Mock Oral Examination (SD-MOE)

Examinees were notified regarding the opportunity to participate in the optional SD-MOE first via email several weeks in advance and again verbally during orientation on the day of the MOE. During the SD-MOE, examinees were instructed to read case information shared on the screen of the testing computer and provide verbal responses to each prompt. Case development for the SD-MOE and C-MOE occurred via identical processes. No faculty examiners were present in the SD-MOE sessions, and examinees were not provided with any verbal cues or other interaction to guide case progression. Examinees were responsible for deciding when to advance to the next prompt after completing each verbal response. Participating examinees were allotted a maximum of 30 minutes to complete three cases in the SD-MOE, and the entirety of each SD-MOE was video-recorded. A trained, non-faculty proctor was present for the SD-MOE only to facilitate video-recording and provide technical support as needed. The scheduling sequence of the SD-MOE session relative to the C-MOE breakout rooms was randomized for each participant.

Faculty examiners for the SD-MOE were recruited initially from the pool of C-MOE examiners, with additional examiners being recruited as needed among faculty who had not participated in the C-MOE. Similar to the C-MOE, two faculty examiners were assigned to each SD-MOE session to review the associated video recording and score the examinee using a standardized rubric. SD-MOE faculty examiners were instructed to score each case while viewing the recording and to do so at their earliest convenience.

Scoring

Proportion of questions passed among questions answered and mean case scores were calculated for both MOE formats. Questions and cases that were not completed or recorded were excluded from the denominator when calculating these metrics; incomplete cases were not scored as failures, and this approach was applied consistently across both formats. The threshold for overall passing in both MOE formats was established as a mean case score ≥2.2. Pass/fail was considered a binary variable.

Resident and Faculty MOE Experience Surveys

Immediately following the MOE, anonymous non-piloted surveys (Online Resource 1) developed by the authors were emailed to examinees. Examinees who completed the SD-MOE reported their satisfaction with the exam format, ability to concentrate on the prompts, and perceived impact of video-recorded responses on exam efficiency and fairness. Technical issues were assessed via a survey item with three predefined response options: “Major issues, significant impact on examination performance”; “Minor issues, no significant impact on examination performance”; and “No issues.” *Examinees who participated only in the C-MOE completed a similar survey but were not presented with questions pertaining to the SD-MOE.* The examinee post-examination survey included 15 total items, with five Likert-scale items, six multiple choice items, and four open-ended response items. A separate anonymous non-piloted survey (Online Resource 2) developed by the authors was emailed to SD-MOE faculty examiners after scoring their assigned video-recorded submissions. SD-MOE faculty examiners were asked to report the average time required to score video-recorded submissions and rated their perception of the validity of scores obtained from the SD-MOE. The faculty post-scoring survey included eight items, with two Likert-scale items, five multiple choice items, and one open-ended response item. Missing survey data for examinees and examiners were managed by excluding non-respondents from item-specific denominator calculations.

Study Outcomes

The primary outcome of feasibility was defined as successful completion of the SD-MOE without major technical issues. Secondary exploratory outcomes included participant perceptions of the SD-MOE format, resource intensity (operationalized as the average time required by faculty examiners to score SD-MOE video recordings), interrater reliability, and performance metrics across formats. Examinee perceptions assessed in this study included satisfaction, ability to concentrate on prompts, perceived appropriateness of case content, and perceived impact of recording responses on examination efficiency and fairness. Faculty perceptions assessed in this study included overall satisfaction, perceived appropriateness of case content, perceived validity of SD-MOE scores, and ease of format.

Statistical Analysis

All analyses comparing participant perceptions, resource intensity, interrater reliability, and performance metrics across formats were exploratory in nature. Fisher's exact tests were used to compare resident survey responses across examination formats and report descriptive statistics for remaining responses. Pearson's Chi-square tests and one-way analysis of variance (ANOVA) were used to compare examinees who participated in both the SD-MOE and C-MOE with those who participated only in the C-MOE. The calculated Pearson's correlation coefficient (r) was used to report the correlation of continuous performance metrics across examination formats. Weighted Cohen's kappa coefficients (κ) with 95% confidence intervals (95% CI) were calculated to report pooled estimates of interrater reliability for each examination format. Pearson's Chi-square tests and one-way ANOVA were used to evaluate the association between PGY-level and performance metrics. Sensitivity analyses were performed to repeat pass/fail outcome analyses at mean case score thresholds of ≥2.0 (at or above borderline) and ≥2.5 (at or above the midpoint between borderline and full pass). P values <0.05 were considered statistically significant. All statistical analyses were conducted using R (RStudio, Boston, MA).

RESULTS.

A total of 55 examinees participated in the C-MOE, of which 41 (75%) also completed the SD-MOE. There was no significant difference in PGY-level between examinees who participated in the SD-MOE and those who participated only in the C-MOE (Table 1). A total of 62 faculty examiners participated in the C-MOE and 25 faculty examiners participated in the SD-MOE. Of the 25 faculty members participating in the SD-MOE, 19 (76%) also served as examiners in the C-MOE.

Of the 41 examinees who completed the post-examination survey (75% response rate), 32 (78%) participated in both the C-MOE and SD-MOE while nine (22%) participated only in the C-MOE (Table 2). There was no difference in the overall frequency of technical issues between the SD-MOE and C-MOE; six examinees reported minor technical issues with no significant impact in the SD-MOE, compared with one examinee who reported minor technical issues with no significant impact in the C-MOE. No examinees reported major technical issues in either examination format. There was no difference between formats in resident satisfaction, ability to concentrate on prompts, or perceived appropriateness

of case content. Of the 32 examinees who completed the SD-MOE, 18 believed scores being derived from video-recorded responses contributed to a more efficient (56%) and fair (56%) examination. Open-text feedback from examinees communicated a desire for real-time, post-case feedback and noted that SD-MOE case flow ambiguities were difficult to navigate without the ability to ask for clarification.

Among faculty examiners, 56 of 62 (90%) C-MOE participants completed the post-examination survey and 20 of 25 (80%) SD-MOE participants completed the post-scoring survey. On average, faculty examiners required 18.5 minutes to fully score each SD-MOE video recording (Table 3). A significantly lower proportion of faculty examiners reported overall satisfaction with the SD-MOE compared with the C-MOE (65% vs. 93%, $P<0.01$). There was no significant difference between formats in satisfaction with case content or ease of format among faculty examiners. The SD-MOE and C-MOE were perceived as equally valid by 50% of SD-MOE faculty examiners. Open-text feedback from examiners reflected an appreciation for logistical flexibility but noted technological difficulties and the inability to probe examinees as limitations.

There was no significant difference in proportion of questions passed, mean case score, or pass rates on the C-MOE between examinees who participated in both the C-MOE and SD-MOE compared with those who participated only in the C-MOE (Table 1). PGY-level was significantly associated with mean case score (PGY-3: 2.45; PGY-4: 2.58; PGY-5: 2.70; $P=0.01$) and proportion of questions passed (PGY-3: 81.7%; PGY-4: 87.3%; PGY-5: 90.1%; $P<0.01$) in the C-MOE, but not the SD-MOE (Table 4). PGY-level was not associated with overall exam pass rate in either format. Sensitivity analyses demonstrated that there remained no significant association between PGY-level and overall exam pass rate for either format at alternative mean case score thresholds of ≥2.0 and ≥2.5. There was no significant correlation between proportion of questions passed in each format ($r=0.19$, $P=0.24$) (Figure 1). A moderate positive correlation of mean case scores was identified between formats ($r=0.34$, $P=0.03$) (Figure 2). Interrater reliability was low in both formats, as reflected by fair agreement in the SD-MOE (weighted $\kappa=0.32$, 95% CI [0.19, 0.46]) and modest-to-fair agreement in the C-MOE (weighted $\kappa=0.25$, 95% CI [0.16, 0.35]). The wide overlap of these confidence intervals indicates no meaningful difference in interrater reliability between formats.

DISCUSSION.

In this feasibility study of a novel self-directed MOE format for use as a supplemental tool in preparation for the ABSCE, the SD-MOE was successfully implemented during a multi-institutional MOE program without the need for faculty examiners to be present for its administration (Figure 3). There were no significant differences in self-reported technical issues between the SD-MOE and C-MOE. Among faculty examiners, satisfaction ratings were lower for the SD-MOE compared with the C-MOE. PGY-level was a significant correlate of performance in the C-MOE, but not in the SD-MOE.

Among examinees, there was no significant difference between the SD-MOE and C-MOE with respect to overall satisfaction, ability to concentrate on prompts, and perceptions of case content appropriateness. In contrast, faculty examiner satisfaction ratings were significantly lower for the SD-MOE relative to the C-MOE. Open-text faculty survey responses indicated that some examiners were dissatisfied by the lack of bidirectional interaction, precluding their ability to probe for reasoning to better assess clinical judgment. Other examiners described difficulty navigating the video recording and scoring interface simultaneously, although some respondents appreciated the ability to replay responses. Additional sources of examiner dissatisfaction with the SD-MOE format were not clearly elucidated in this study, and potential explanations are consequently speculative. While there was no significant difference in reported ease of format between modalities, it is possible faculty examiners were dissatisfied by the asynchronous structure of the SD-MOE, which required setting aside a separate time to view and grade videos. Addressing potential sources of dissatisfaction would first require more detailed investigation into examiner perspectives, but strategies might entail both logistical and technical streamlining of the SD-MOE scoring process; for example, ensuring protected academic time to complete asynchronous grading and integrating SD-MOE video recordings and score input into a single webpage may help to improve the examiner experience.

In the C-MOE, PGY-level was correlated with both mean case score and percentage of questions passed but not with overall exam pass rates. In contrast, PGY-level was not significantly associated with any of these performance metrics in the SD-MOE. These divergent patterns are likely driven in part by the SD-MOE featuring fewer cases, thus contributing to lower statistical power. Additionally, they may also reflect that the SD-MOE and C-MOE potentially constitute different assessment constructs, not simply different delivery modalities. Navigating the unfamiliar assessment structure of the SD-MOE moreover may have introduced variability in performance and scoring among examinees and examiners, respectively. Additionally, examinees may have perceived the SD-MOE as having lower stakes and educational value in its profound departure from the standard oral examination structure, given that similarity to the ABSCE is a well-recognized benefit of C-MOEs.[17,18] It is possible that such perceptions may have attenuated examinees' level of engagement and effort on the SD-MOE.

The SD-MOE presents a theoretical opportunity to decrease bias by minimizing unintentional coaching from examiners through standardized prompt presentation; however, eliminating two-way communication also removes the ability to probe for understanding of key concepts and prompt examinees to demonstrate complex judgment, two factors that are key to oral assessments.[19,20] Given that interrater reliability has been described as poor even when experienced examiners grade in isolation, the lack of interaction between faculty examiners may have also negatively impacted interrater reliability in the SD-MOE.[21] Although beyond the scope of the present study, the inherent contrast between written SD-MOE prompts compared with verbal C-MOE prompts may have had varying impacts on performance across certain subgroups of examinees. For example, learners with dyslexia or attention deficit

hyperactivity disorder and those who are non-native English speakers may favor and perform better when case information is presented verbally, rather than in written form.[22–24]

There are multiple limitations to this feasibility study and the SD-MOE format. While participation in the SD-MOE was high at a rate of 75%, the opt-in nature of this study raises concern for potential selection bias. However, no significant differences were observed between examinees who participated in the SD-MOE compared with those who opted out with respect to PGY-level or C-MOE performance, as defined by mean case score, questions passed, and overall pass rate. Although the majority of SD-MOE faculty examiners participated in the C-MOE, non-identical groups of faculty examiners participating in the two MOE formats may have led to biases in scoring. In addition, a central limitation of this study is that there were fewer cases in the single SD-MOE session than each C-MOE breakout room, which likely led to compressed score ranges, attenuated correlations, and obscured PGY-level gradients. The small number of observations per SD-MOE examiner dyad also contributed to unstable estimates, which precluded reliable comparison of dyad-level kappa values between formats. Potential order effects on performance could not be investigated because the sequence in which individual examinees completed the SD-MOE relative to C-MOE cases was not recorded. While presenting identical cases in the SD-MOE and C-MOE format may have created opportunities for more robust direct comparison, this was not pursued as part of this feasibility study after considering the logistical limitations with respect to ensuring a standardized C-MOE experience for all examinees irrespective of their participation in the SD-MOE. Uneven survey response denominators between SD-MOE examinees and C-MOE-only examinees as well as between SD-MOE examiners and C-MOE examiners raise concern for potential non-response bias, limiting the reliability of between-format survey comparisons. Furthermore, survey items were not formally piloted before administration. These limitations thus create fragile survey comparisons and must be considered when interpreting results.

While the SD-MOE did not require faculty examiners for its administration, the examination process was not entirely autonomous, as non-faculty proctors were utilized in each SD-MOE session to facilitate video recording and provide logistical support as needed. Further optimizing SD-MOE administration such that support staff are not required would help to make this tool more accessible for prospective users at resource-limited institutions. The SD-MOE may also be further improved by including more dynamic functionality with standardized follow-ups to mimic examiner probing and adaptive branching logic based on examinee responses. With such improvements, residency programs could consider deploying the SD-MOE as a supplemental self-study resource in oral exam curricula and as a means of identifying residents who may benefit most from additional coaching while preparing for the C-MOE and ABSCE. Meaningful future research directions include a study with matched SD-MOE and C-MOE case counts and sets, integrating artificial intelligence-enhanced dynamic functionality, and evaluating the correlation of SD-MOE performance with ABSCE first-time pass rates.

This feasibility study demonstrates the successful implementation of the SD-MOE with minimal logistical challenges during a multi-institutional MOE. Administered in the absence of faculty examiners, the SD-MOE may offer valuable opportunities particularly in resource-limited environments for independent study, self-critique, and use of video recordings to educate other trainees.[3,25,26] Without criterion validity testing against performance on the ABSCE, however, the SD-MOE cannot be proposed as a substitute for the multi-institutional C-MOE and should be considered exclusively as a supplemental educational tool. Thus, the C-MOE should remain the preferred format for ABSCE preparation when possible.

SUMMARY – ACCELERATING TRANSLATION.

Title: Feasibility Study of Self-Directed Mock Oral Examinations in the Assessment of General Surgery Resident Clinical Decision-Making

Surgeons in the United States must pass a high-stakes oral examination, known as the American Board of Surgery Certifying Examination (ABSCE), to become fully board-certified. To prepare for the ABSCE, many general surgery residents participate in mock oral examinations (MOEs), which are intended to simulate the testing experience. However, MOEs are resource-intensive and require the simultaneous availability of numerous faculty to serve as live examiners. In this study, we sought to investigate whether a novel, self-directed MOE format (SD-MOE) could serve as a practical supplement to the conventional MOE (C-MOE). The primary aim was to determine whether the SD-MOE format was feasible and could be completed without significant technical problems. We additionally measured participant satisfaction, assessed interrater reliability, and compared performance metrics between examination formats.

This feasibility study took place during a multi-institutional MOE, which was administered virtually for trainees from five general surgery residency programs. During the C-MOE, two faculty examiners read standardized case prompts aloud and scored an individual resident examinee in real-time via a virtual video-conferencing platform. At the conclusion of each case, faculty examiners used a rubric to independently assign a case-specific score of pass, borderline, or fail. Examinees were allotted a maximum of 30 minutes to complete four total cases in each breakout room. At the conclusion of each 30-minute testing period, examinees proceeded to the next MOE session as dictated by their individual randomized schedule until completing all three assigned rooms.

Examinees were invited to participate in the optional SD-MOE first via email several weeks in advance and again verbally during orientation on the day of the MOE. In the SD-MOE, resident examinees instead read case prompts displayed on a computer screen and video-recorded their verbal answers without a faculty examiner present. Examinees were responsible for deciding when to advance to the next prompt after completing each verbal response. Participating examinees were allotted a maximum of 30 minutes to complete three cases in the SD-MOE, and the scheduling sequence of the SD-MOE session relative to the C-MOE breakout rooms was randomized for each participant. Two faculty examiners then later watched and scored the recordings using the same rubric at their earliest convenience. Faculty examiners for the SD-MOE were recruited initially from the pool of C-MOE examiners, with additional examiners being recruited as needed among faculty who had not participated in the C-MOE.

A total of 55 residents participated in the C-MOE, and 41 (75%) also volunteered to participate in the SD-MOE. A total of 62 faculty examiners participated in the C-MOE, and 25 faculty examiners participated in the SD-MOE. Of the 25 faculty members who participated in the SD-MOE, 19 (76%) also served as

examiners in the C-MOE. No major technical problems were reported. A small number of minor technical issues with no impact on the examination were reported in both the SD-MOE and C-MOE, with no significant difference between formats. Among resident examinees, there was no significant difference in satisfaction, ability to concentrate on prompts, and perceptions of case content appropriateness between the SD-MOE and C-MOE. In contrast, faculty examiners were less satisfied with the SD-MOE compared with the C-MOE (65% vs. 93%, $P<0.01$). On average, faculty spent approximately 18.5 minutes scoring each SD-MOE video recording. With respect to performance metrics, PGY-level was correlated with both mean case score and percentage of questions passed but not with overall exam pass rates in the C-MOE. In contrast, PGY-level was not significantly associated with any of these performance metrics in the SD-MOE. These divergent patterns are likely driven in part by the SD-MOE featuring fewer cases, thus contributing to lower statistical power. Additionally, they may also reflect that the SD-MOE and C-MOE potentially constitute different assessment constructs.

This study has multiple limitations. While participation in the SD-MOE was high at a rate of 75%, the opt-in nature raises concern for potential selection bias. Non-identical groups of faculty examiners participating in the two MOE formats may have led to biases in scoring. A central limitation is that there were fewer cases in the single SD-MOE session than each C-MOE breakout room, which likely led to compressed score ranges, attenuated correlations, and obscured PGY-level gradients. Potential order effects on performance could not be investigated because the sequence in which individual examinees completed the SD-MOE relative to C-MOE cases was not recorded. Uneven survey response denominators raise concern for potential non-response bias, limiting the reliability of between-format survey comparisons. Furthermore, survey items were not formally piloted before administration.

This feasibility study demonstrates the successful implementation of the SD-MOE with minimal logistical challenges during a multi-institutional MOE. Administered in the absence of faculty examiners, the SD-MOE may offer valuable opportunities particularly in resource-limited environments for independent study, self-critique, and use of video recordings to educate other trainees. The SD-MOE may also be improved by including more dynamic functionality with standardized follow-ups to mimic examiner probing and adaptive branching logic based on examinee responses. Further optimizing SD-MOE administration to obviate the need for support staff would also help to make this tool more accessible for prospective users at resource-limited institutions. Without criterion validity testing against performance on the ABSCE, however, the SD-MOE cannot be proposed as a substitute for the multi-institutional C-MOE and should be considered exclusively as a supplemental educational tool. With the elimination of two-way communication, the SD-MOE does not allow examiners to probe for understanding of key concepts and prompt examinees to demonstrate complex judgment, two factors that are key to oral assessments. The lack of interaction between faculty examiners may have also negatively impacted interrater reliability in the SD-MOE. Therefore, the C-MOE should remain the preferred format for ABSCE preparation when possible.

References

1. Higgins RM, Deal RA, Rinewalt D, Kuchta K, Luu MB, Statter MB, et al. The utility of mock oral examinations in preparation for the American Board of Surgery certifying examination. Am J Surg. 2016;211(2):416–20.

2. 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):e95–103.

3. Guzman E, Babakhani A, Maker VK. Improving outcomes on the ABS certifying examination: can monthly mock orals do it? J Surg Educ. 2008;65(6):441–4.

4. Nguyen J, Liu A, McKenney M, Elkbuli A. Predictive factors of first time pass rate on the American Board of Surgery certification in general surgery exams: a systematic review. J Surg Educ. 2021;78(4):1094–104.

5. Maker VK, Zahedi MM, Villines D, Maker AV. Can we predict which residents are going to pass/fail the oral boards? J Surg Educ. 2012;69(6):705–13.

6. Meyerson SL, Lipnick S, Hollinger E. The usage of mock oral examinations for program improvement. J Surg Educ. 2017;74(6):946–51.

7. 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.

8. Strickland C, Jensen A, McArthur T. Does the oral 'mock board' examination still have a role as a training tool? Acad Radiol. 2017;24(11):1463–7.

9. Huang IA, Lu Y, Wagner JP, Sandhu G, Chen DC, Jarman BT, 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.

10. Shebrain S, Nava K, Munene G, Shattuck C, Collins J, Sawyer R. Virtual surgery oral board examinations in the era of COVID-19 pandemic. J Surg Educ. 2021;78(2):365–8.

11. 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;76:14–9.

12. Smeds MR, Thrush C, Kimbrough MK, Moursi MM. Use of mock oral examinations in vascular surgery training programs: a nationwide survey. J Surg Res. 2018;232:94–8.

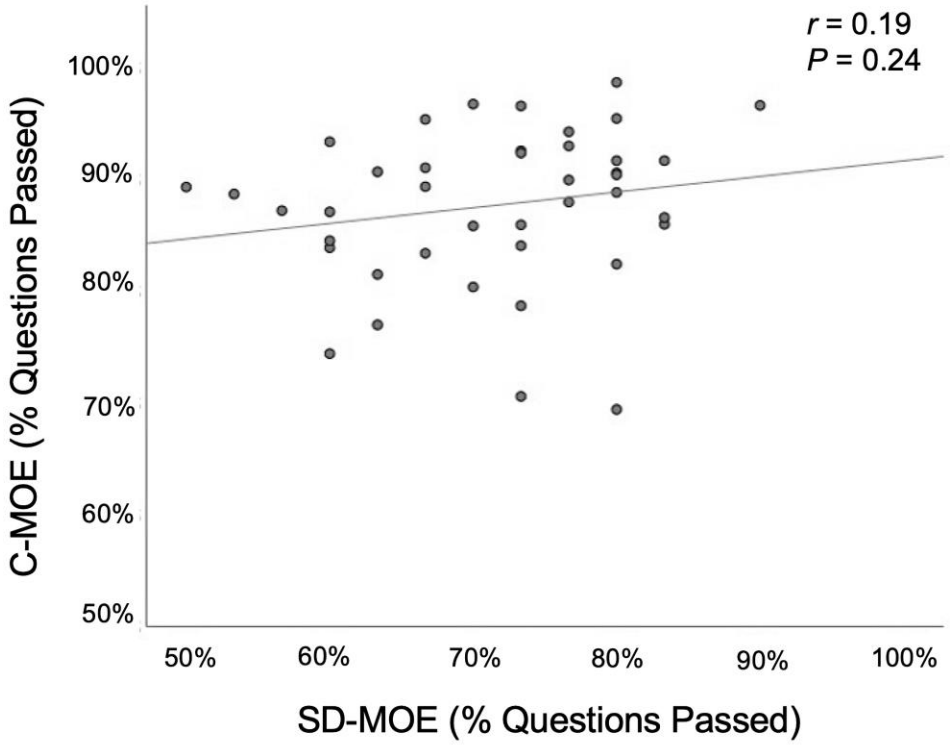
13. Peters GW, Decker RH, Park HS, Yu JB, Evans SB. Resident attitudes and benefits of mock oral board examinations in radiation oncology. BMC Med Educ. 2020;20(1):203.

14. House JB, Dooley-Hash S, Kowalenko T, Fales W, Jouriles N, Amini R, et al. Prospective comparison of live evaluation and video review in the evaluation of operator performance in a pediatric emergency airway simulation. J Grad Med Educ. 2012;4(3):312–6.

15. Patrawalla P, Eisen LA, Shiloh A, Ahmed A, Koenig SJ, Narasimhan M, et al. Development and validation of an assessment tool for competency in critical care ultrasound. J Grad Med Educ. 2015;7(4):567–73.
16. Scaffidi MA, Grover SC, Carnahan H, Khan R, Brar S, Jayaprakash A, et al. A prospective comparison of live and video-based assessments of colonoscopy performance. Gastrointest Endosc. 2018;87(3):766–75.
17. Lu Y, Miranda R, Quach C, Huang IA, Wagner JP, Chen DC, et al. Standardized multi-institutional mock oral examination: a feasible and valuable educational experience for general surgery residents. J Surg Educ. 2020;77(6):1568–76.
18. Subhas G, Yoo S, Chang YJ, Mittal VK, Healy M, Marcus C, et al. Benefits of mock oral examinations in a multi-institutional consortium for board certification in general surgery training. Am Surg. 2009;75(9):817–21.
19. Theobold AS. Oral exams: a more meaningful assessment of students' understanding. J Stat Data Sci Educ. 2021;29(2):156–9.
20. Huxham M, Campbell F, Westwood J. Oral versus written assessments: a test of student performance and attitudes. Assess Eval High Educ. 2012;37(1):125–36.
21. Jacobsohn E, Klock PA, Avidan M; Oral Examinations Group. Poor inter-rater reliability on mock anesthesia oral examinations. Can J Anaesth. 2006;53(7):659–68.
22. Waterfield J, West B. Inclusive Assessment in Higher Education: A Resource for Change. Plymouth: University of Plymouth; 2006.
23. Ramella D. Oral exams: a deeply neglected tool for formative assessment in chemistry. In: Blaser M, Clark T, Lamont L, Stewart JJ, editors. ACS Symposium Series. Vol 1340. Washington, DC: American Chemical Society; 2019. p. 79–89.
24. Weston TJ. The validity of oral accommodation in testing [dissertation]. Boulder (CO): University of Colorado; 2002.
25. Kozol R, Giles M, Voytovich A. The value of videotape in mock oral board examinations. Curr Surg. 2004;61(5):511–4.
26. Ruiz TL, Sellers B, Devarakonda A, Wehrle CJ, Arora TK. A novel mock oral curriculum for senior surgery residents: results of a pilot study. J Surg Res. 2022;277:92–9.

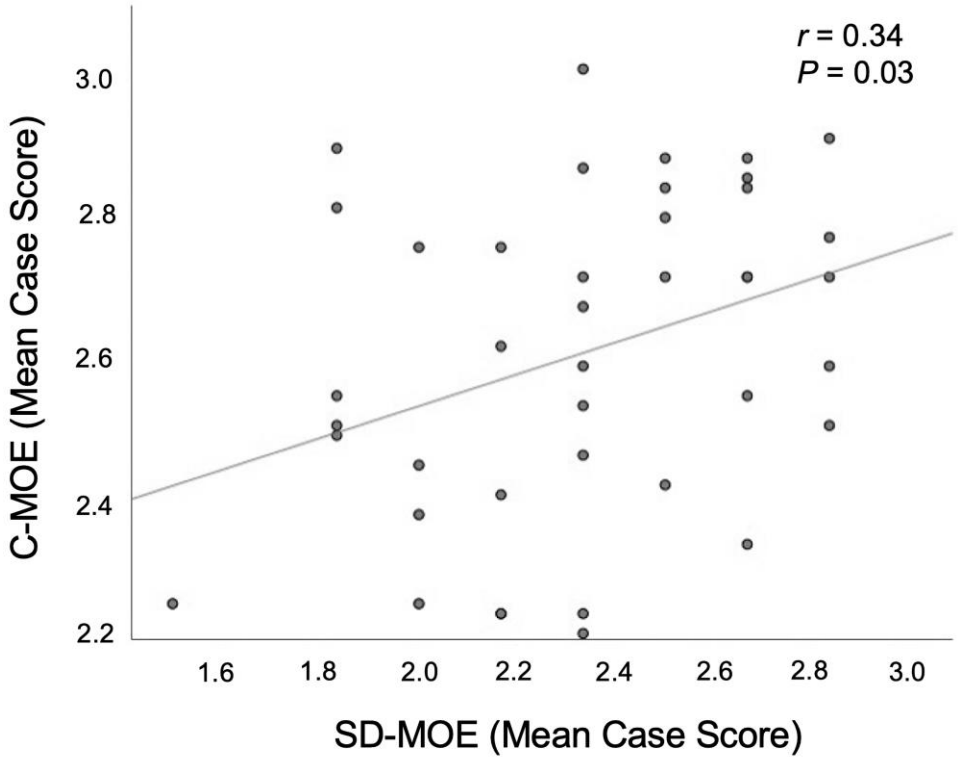
FIGURES AND TABLES.

Figure 1. Correlation of Proportion of Questions Passed Across Examination Formats.



This figure depicts the correlation between the proportion of questions passed on the C-MOE and SD-MOE. Each data point represents one examinee's performance on each examination.
Abbreviations: C-MOE: conventional mock oral examination; SD-MOE: self-directed mock oral examination

Figure 2. Correlation of Mean Case Score Achieved Across Examination Formats.



This figure depicts the correlation between the mean case score on the C-MOE and SD-MOE. Each data point represents one examinee's performance on each examination.
Abbreviations: C-MOE: conventional mock oral examination; SD-MOE: self-directed mock oral examination

Figure 3. Utility and Feasibility of the Self-Directed Mock Oral Examination

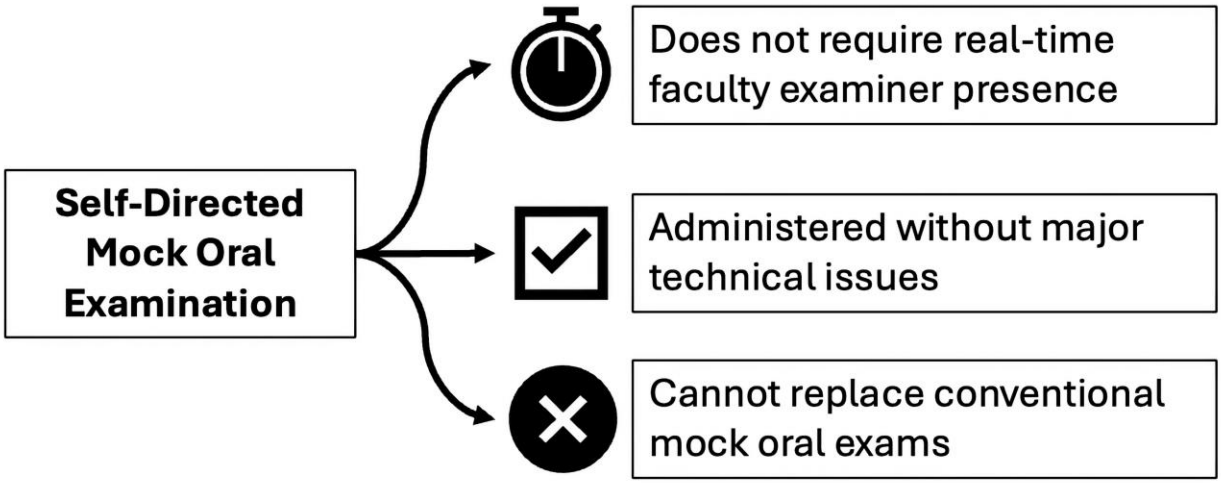


Table 1. Comparison of SD-MOE Participating Residents with Non-Participating Residents

	C-MOE and SD-MOE (N=41)	C-MOE Only (N=14)	*P* value
Post-Graduate Year, (No., %)			0.07
PGY-3	8 (19.5%)	4 (28.6%)	
PGY-4	10 (24.4%)	7 (50.0%)	
PGY-5	23 (56.1%)	3 (21.4%)	
Mean Case Score on C-MOE[a]	2.60	2.60	0.37
Questions Passed on C-MOE (%)	87.7%	85.0%	0.36
Pass Rate on C-MOE (%)	95.0%	92.9%	0.09

Abbreviations: C-MOE: conventional mock oral examination; SD-MOE: self-directed mock oral examination; PGY: post-graduate year

[a]Maximum case score was 3 points

Table 2. Resident Agreement with Survey Statements Regarding the C-MOE and SD-MOE

	C-MOE (N=9)[a]	SD-MOE (N=32)[b]	*P* value
Satisfied with the examination (No., %)	8/9 (89%)	22/31 (71%)	0.40
Able to concentrate on the prompts	8/9 (89%)	27/32 (84%)	0.99
Appropriate case content	8/9 (89%)	27/32 (84%)	0.99
Technical Issues Experienced			0.99
Major, had an impact on the examination	0/9 (0%)	0/32 (0%)	
Minor, had no impact on the examination	1/9 (11%)	6/32 (19%)	
No technical issues	8/9 (88%)	26/32 (81%)	
Recording responses makes the examination more efficient	--	18/32 (56%)	--
Recording responses makes the examination more fair	--	18/32 (56%)	--

Abbreviations: C-MOE: conventional mock oral examination; SD-MOE: self-directed mock oral examination

[a]A total of 14 residents completed only the C-MOE, and 9 of those completed the post-examination survey.

[b]A total of 41 residents completed the both the C-MOE and SD-MOE, and 32 of those completed the post-examination survey. One resident in this group did not respond to the satisfaction item (n=31).

Table 3. Faculty Agreement with Survey Statements Regarding the C-MOE and SD-MOE

	C-MOE (N=56)[a]	SD-MOE (N=20)[b]	*P* value
Satisfied with the examination (No., %)	52/56 (93%)	13/20 (65%)	<0.01
Appropriate case content	52/56 (93%)	16/20 (80%)	0.99
Ease of format	47/56 (84%)	16/20 (80%)	0.73
Validity of scores derived from scoring recordings compared with scoring in real-time			--
Less valid	--	4/20 (20%)	
Equally valid	--	10/20 (50%)	
More valid	--	3/20 (15%)	
No response	--	3/20 (15%)	
Average time to grade a recording (minutes)	--	18.5	--

Abbreviations: C-MOE: conventional mock oral examination; SD-MOE: self-directed mock oral examination

[a]56 of 62 faculty examiners for the C-MOE completed the post-examination survey.
[b]20 of 25 faculty examiners for the SD-MOE completed the post-scoring survey.

Table 4. Correlation of Examination Outcomes with PGY for the C-MOE and SD-MOE

	C-MOE (N=55)	*P* value	SD-MOE (N=41)	*P* value
Mean Case Score		0.01		0.93
PGY-3	2.45		2.35	
PGY-4	2.58		2.30	
PGY-5	2.70		2.35	
Questions Passed (%)		<0.01		0.07
PGY-3	81.7%		65.0%	
PGY-4	87.3%		71.0%	
PGY-5	90.1%		73.8%	
Pass Rate (%)		0.22		0.76
PGY-3	87.5%		66.7%	
PGY-4	91.7%		75.0%	
PGY-5	100.0%		79.2%	

Abbreviations: C-MOE: conventional mock oral examination; SD-MOE: self-directed mock oral examination; PGY: post-graduate year

Online Resource 1. Resident Post-Examination Survey on the Mock Oral Examination

This resource was the resident survey on the mock oral examination, which was completed online via Qualtrics (Qualtrics, Provo, UT). Residents who responded on Question 5 that they participated in SD-MOE were presented with Questions 6 and 8.

MOE Resident Survey

Start of Block: MOE Resident Survey

Q0.
Mock Oral Examination Resident Survey
Thank you for your participation in the mock oral examination. Please answer the survey questions below. This survey will take approximately 5 minutes.

Q1. What is your PGY (in clinical years)?

- ○ PGY-3
- ○ PGY-4
- ○ PGY-5

Q2. If not yet committed, please indicate your top choice specialty.

_____

Page Break

Q3. How many standardized mock orals examinations have you participated in, ***prior to this current examination***?

- ○ None, today was my first
- ○ 1
- ○ 2
- ○ 3+

Display This Question:

If How many standardized mock orals examinations have you participated in, prior to this current examination? != None, today was my first

Q4. Did you participate in our virtual mock orals examination last year?

- ○ Yes
- ○ No

Q5. Did you participate in the self-directed mock oral examination room this year?

- ○ Yes
- ○ No

Page Break

Display This Question:

If Did you participate in the asynchronous room this year? = Yes

Q6. Regarding the **self-directed MOE only...**

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was satisfied with the self-directed examination	○	○	○	○	○
The content of cases was appropriate	○	○	○	○	○
I was able to concentrate on the clinical prompts	○	○	○	○	○
I felt comfortable in the examination room	○	○	○	○	○
Recording responses makes the examination more efficient	○	○	○	○	○
Recording responses makes the examination more fair	○	○	○	○	○
I am confident I will pass the ABSCE on my first attempt	○	○	○	○	○
This examination helped me identify areas of my knowledge that need improvement	○	○	○	○	○

Display This Question:

If Did you participate in the self-directed room this year? = No

Q7. Regarding the **mock oral examination...**

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I was satisfied with the examination overall	○	○	○	○	○
The content of cases was appropriate	○	○	○	○	○
I was able to concentrate on the clinical prompts	○	○	○	○	○
I felt comfortable in the examination rooms	○	○	○	○	○
I am confident I will pass the ABSCE on my first attempt	○	○	○	○	○
This examination helped me identify areas of my knowledge that need improvement	○	○	○	○	○

Display This Question:

If Did you participate in the self-directed room this year? = Yes

Q8. Did you experience any technical issues with the audio or video connection in the **self-directed MOE room**? If so, please describe the issues you experienced.

- ○ Major issues, significant impact on examination performance_____
- ○ Minor issues, no significant impact on examination performance_____
- ○ No issues

Display This Question:

If Did you participate in the self-directed room this year? = No

Q9. Did you experience any technical issues with the audio or video connection during the examination? If so, please describe the issues you experienced.

- ○ Major issues, significant impact on examination performance_____
- ○ Minor issues, no significant impact on examination performance_____
- ○ No issues

Page Break

Q10. In your opinion, how was your performance in each case category?

	Poor	Somewhat poor	Neither good nor poor	Somewhat good	Very good
Trauma Critical Care	○	○	○	○	○
Acute Care and GI Surgery	○	○	○	○	○
Subspecialty and Oncologic Surgery	○	○	○	○	○

Q11. In your opinion, how likely are you to have passed the **mock oral examination** today?

- ○ Very unlikely
- ○ Somewhat unlikely
- ○ Neither likely nor unlikely
- ○ Somewhat likely
- ○ Very likely

Q12. In your opinion, how likely are you to pass the **ABS Certifying Examination** if you took it today?

- ○ Very unlikely
- ○ Somewhat unlikely
- ○ Neither likely nor unlikely
- ○ Somewhat likely
- ○ Very likely

Q13. Which topics would you like to have covered in future exams?

_____

_____

_____

_____

_____

Q14. Please enter in any feedback for your examiners (anything they did that was very helpful or not helpful for the examination)

_____

Q15.
Please provide any additional comments (e.g. case flow, case content, etc.).

_____

_____

_____

_____

_____

Online Resource 2. Faculty Post-Examination Survey on the SD-MOE

This resource was the faculty survey on the SD-MOE, which was completed online via Qualtrics (Qualtrics, Provo, UT). Only faculty examiners who participated in the SD-MOE were asked to completed this survey.

MOE Faculty Survey (SD-MOE)

Q0. Please complete the following survey only **ONE TIME** **after scoring all recordings assigned to you.**

Q1. Prior to this year, how many times had you been an examiner for **in-person** standardized mock oral examinations?

- ○ 0, I have never been an examiner for in-person MOE
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ >5 (fill in approximate # times if >5)

_____

Q2. Prior to this year, how many times had you been an examiner for **virtual/online** standardized mock oral examinations?

- ○ 0, this is my first year as an examiner for virtual/online MOE
- ○ 1
- ○ 2
- ○ >3 (fill in approximate # times if >3)

_____

Display This Question:

If Prior to this year, how many times had you been an examiner for in-person standardized mock oral... != 0, I have never been an examiner for in-person MOE

Q3. How *valid* are scores derived from scoring recorded responses of clinical decision-making, when compared to scoring in real-time?

- ○ Not valid
- ○ Less valid
- ○ Equally valid
- ○ More valid
- ○ Much more valid

Q4. On average, how much time (minutes) did it take to score a resident's recording?

0 6 12 18 24 30 36 42 48 54 60

Time (minutes) ()



Q5. Did you listen to recordings at normal playback speed or increase the video playback speed (e.g. 1.5x speed)?

- ○ Normal playback speed
- ○ Increased playback speed to: _____

Q6. Please rate the following

	Extremely dissatisfied	Somewhat dissatisfied	Neither satisfied nor dissatisfied	Somewhat satisfied	Extremely satisfied
Overall Satisfaction With Self-Directed Room	○	○	○	○	○
Variety and content of cases	○	○	○	○	○
Ease of format	○	○	○	○	○

Q7. Would you rate yourself as a hawk (strict grader) or a dove (lenient grader) when assessing resident performance in general?

☐ **Hawk** (strict grader)

☐ **Dove** (lenient grader)

Q8. Please provide any additional comments you wish to share.

_____