






INNOVATIONS REPORT

The OSSE: Development and validation of an “Objective Structured Supervision Evaluation”

Nicholas Pokrajac MD¹  | Kelly N. Roszczynialski MD, MS¹ | Ashley Rider MD, MEHP¹ | Cori Mc Clure Poffenberger MD¹  | Sarah Williams MD, MHPE¹ | Mc Kenzie Eakin MD² | Richard W. Sapp MD, MS²  | Eric Jenkins MEd² | Michael A. Gisondi MD^{1,3}  | Kimberly Schertzer MD, MS¹ 

¹Department of Emergency Medicine, Stanford University School of Medicine, Palo Alto, California, USA

²Stanford University School of Medicine, Palo Alto, California, USA

³The Precision Education and Assessment Research Lab, Department of Emergency Medicine, Stanford University, Palo Alto, California, USA

Correspondence

Nicholas Pokrajac, MD, Department of Emergency Medicine, Stanford University School of Medicine, Palo Alto, CA 94395, USA.

Email: pokrajac@stanford.edu

Abstract

Background: Trainee supervision and teaching are distinct skills that both require faculty physician competence to ensure patient safety. No standard approach exists to teach physician supervisory competence, resulting in variable trainee oversight and safety threats. The Objective Structured Teaching Evaluation (OSTE) does not adequately incorporate the specific skills required for effective supervision. To address this continuing medical education gap, the authors aimed to develop and identify validity evidence for an “Objective Structured Supervision Evaluation” (OSSE) for attending physicians, conceptually modeled on the historic OSTE.

Methods: An expert panel used an iterative process to create an OSSE instrument, which was a checklist of key supervision items to be evaluated during a simulated endotracheal intubation scenario. Three trained “standardized residents” scored faculty participants' performance using the instrument. Validity testing modeled a contemporary approach using Kane's framework. Participants underwent simulation-based mastery learning (SBML) with deliberate practice until meeting a minimum passing standard (MPS).

Results: The final instrument contained 19 items, including three global rating measures. Testing domains included *supervision climate*, *participant control of patient care*, *trainee evaluation*, *instructional skills*, *case-specific measures*, and *overall supervisor rating*. Reliability of the assessment tool was excellent (ICC range 0.84–0.89). The assessment tool had good internal consistency (Cronbach's $\alpha = 0.813$). Out of 24 faculty participants, 17 (70.8%) met the MPS on initial assessment. All met the MPS after SBML and average score increased by 19.5% (95% CI of the difference 10.3%–28.8%, $p = 0.002$).

Supervising Editor: Dr. Lalena Yarris.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *AEM Education and Training* published by Wiley Periodicals LLC on behalf of Society for Academic Emergency Medicine.

Conclusions: The OSSE is a promising tool to assess faculty supervision performance. Further study should evaluate the effect of OSSE training on supervision in the clinical environment, which may have measurable changes in patient-centered outcomes and resident teaching.

NEED FOR INNOVATION

Patient safety mandates that emergency physicians competently supervise junior trainees, particularly during high-risk procedures, yet teaching faculty must balance the need to deliver safe, high-quality care with entrustment decisions that promote trainee autonomy and learning. While conceptual frameworks of clinical supervision offer insights for faculty development,¹⁻⁶ there is no standardized, objective approach to teaching faculty to supervise procedures competently.

BACKGROUND

Moore's expanded outcomes framework for planning and assessing continuing medical education emphasizes that competence develops when participants demonstrate learning and receive expert feedback.⁷ The Objective Structured Teaching Evaluation (OSTE) is an example of one such training model tailored to faculty learners. The OSTE uses a standardized student in a scripted scenario to assess faculty members, give direct feedback, and provide opportunity to practice teaching skills.^{8,9} However, the OSTE focuses on teaching and is not designed to address the unique cognitive tasks required for effective trainee supervision. Teaching and supervision are distinct skill sets that are often conflated because outstanding clinician educators may teach and supervise somewhat seamlessly. However, in the clinical learning environment, effective teachers may be subpar supervisors and vice versa. Therefore, the OSTE is an insufficient tool for faculty members to be competent supervisors. There remains a need for training opportunities focused on the role of the supervising physician.

OBJECTIVE OF INNOVATION

Using the OSTE as a conceptual model, we aimed to design and evaluate an "Objective Structured Supervision Evaluation" (OSSE) to assess and teach attending supervisory skills during a procedure.

DEVELOPMENT PROCESS

As proof of concept, we chose endotracheal intubation for the design of our OSSE given the risk level of the procedure and evidence of improved patient outcomes with its performance under direct

supervision.¹⁰ Instrument development consisted of: (1) literature review, (2) adaptation of a published OSTE exercise¹¹ fit to our purposes, (3) selection of 10 experts in medical education and airway management for content review, (4) a Delphi process¹² conducted by the expert panel to reach consensus on instrument items, and (5) testing for evidence of validity incorporating Kane's framework (Figure 1).¹³

THE IMPLEMENTATION PHASE

Eligible participants included volunteer emergency medicine faculty members in a residency program at an academic medical center. All participants provided informed consent to have their assessment data analyzed. Each participated in a 20-minute scripted OSSE designed to simulate a realistic clinical scenario (Appendix S1). The simulations occurred in winter 2021. Our institutional review board (IRB) deemed this research protocol exempt (IRB 54403).

Participants supervised intubation of an elderly woman in acute respiratory failure by a standardized resident in a simulated environment. Each session included a faculty participant, a standardized resident, and a trained faculty proctor. One of three volunteer senior medical students from our institution assumed the role of the standardized resident. One of four trained peer faculty proctored all exercises. The OSSE included several scripted patient safety threats in which the standardized resident attempted to perform unsafe actions; the expectation was that the supervisor would correct them. Participant performance was not shared with department leaders and had no bearing on employment status.

After assessment, participants were given direct feedback, time for deliberate practice, and repeat testing until a minimum passing standard (MPS) was met using mastery learning principles.¹⁴ The MPS for the assessment tool was set by our expert panel using the Mastery Angoff approach, with additional determination of items critical for patient safety.¹⁵

We trained volunteer medical students and faculty proctors prior to participant testing. Standardized resident training consisted of two separate 2-h sessions that included: (1) review of the OSSE simulation case and assessment instrument, (2) practice of endotracheal intubation using relevant equipment, and (3) calibration of the instrument via mock sessions. Training of faculty proctors included: (1) a review of the OSSE and simulation case, (2) guidance on scripted prompts to use when participants directed the standardized resident to perform tasks outside the scope of the learning objectives (for example, instructing the standardized resident to use unavailable equipment), and (3) practice with a mock participant with one of the

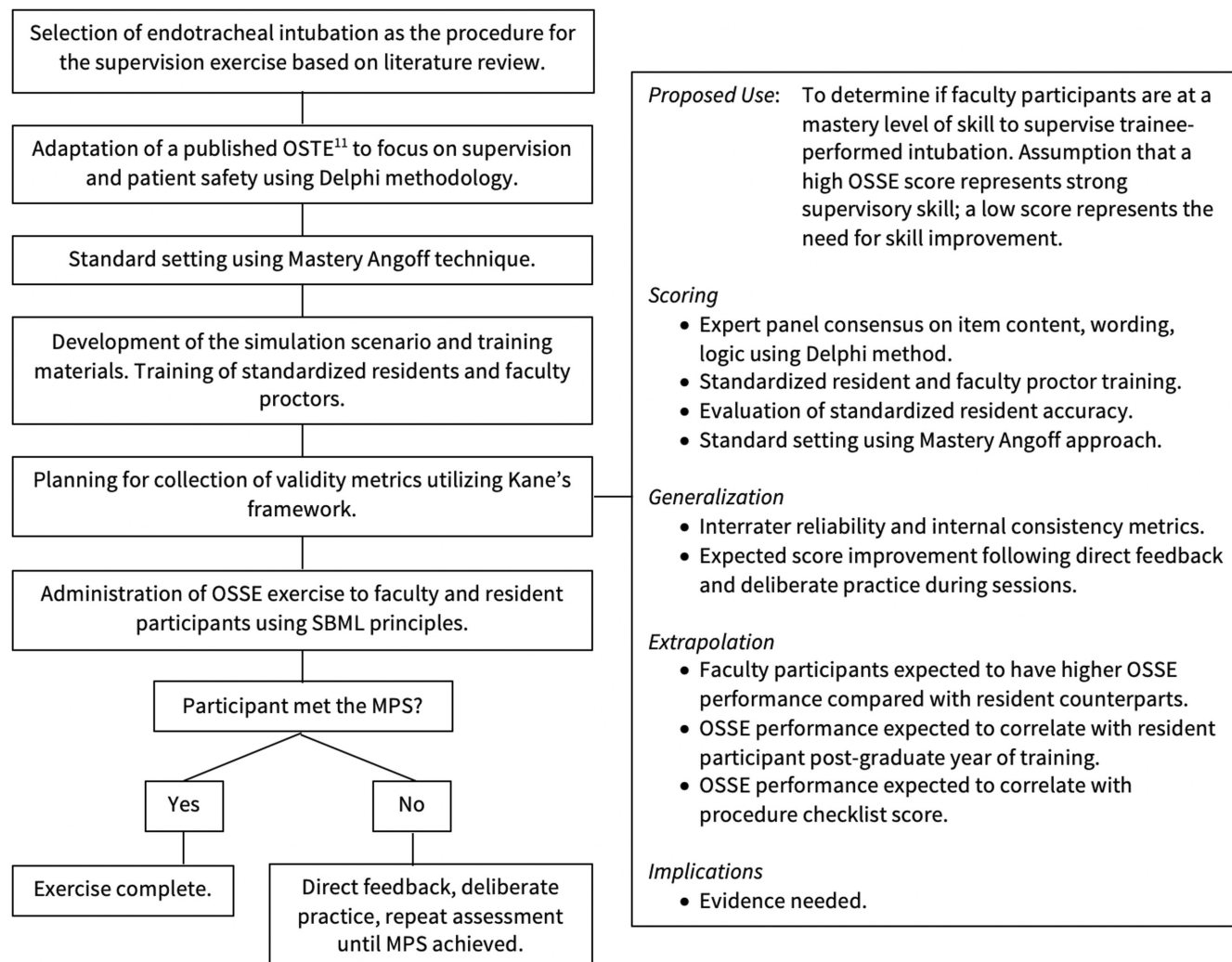


FIGURE 1 Flowchart of OSSE development and validity evidence. MPS, minimum passing standard; OSSE, Objective Structured Supervision Evaluation; OSTE, Objective Structured Teaching Evaluation; SBML, simulation-based mastery learning.

trained standardized residents. Standardized residents and proctors referenced scripts as needed during the exercise ([Appendix S1](#)). Proctors used a script to introduce the scenario and provide a description of available resources to each participant. OSSEs used an in situ HAL® patient simulator (Gaumard Scientific).

As a measure of validity, we analyzed the accuracy and reproducibility of standardized resident actions. A separate assessor evaluated one-third of participant sessions via video review, using a checklist of scripted actions to determine standardized resident accuracy during each recorded OSSE session ([Appendix S1](#)). We also evaluated internal structure of the OSSE assessment instrument by analyzing interrater reliability and internal consistency. One trained reviewer evaluated one-third of recorded sessions to independently score faculty participants using the instrument to determine interrater reliability for each standardized resident. Internal consistency analysis utilized assessment scores from all faculty participants.

No assessment tool currently exists for objective evaluation of attending supervision that could be used for comparison to our OSSE assessment instrument. Therefore, two assumptions were made to

further assess validity. First, we expected that supervision performance would positively correlate with an evaluation of endotracheal intubation competency. During the OSSE session, assessors graded faculty-standardized resident pairs with a previously derived procedure checklist used at our institution for resident training ([Appendix S1](#)). We then correlated OSSE scores with those from our intubation checklist. Second, we expected that trainees would not perform as well as attending physicians in an assessment of supervision. To measure this, we evaluated four resident volunteers (one from each year of post-graduate training) to act as supervisors of a standardized resident using our assessment tool. We then compared scores between resident and faculty volunteer participants. Residents consented to participation and their assessments were not shared with program directors.

Data analysis

We used SPSS Statistics for Macintosh, Version 27 (IBM Corp). Standardized resident accuracy measurements included: (1) overall

adherence to scripted actions across all sessions, (2) percentage agreement between each standardized resident, and (3) percentage agreement between recorded sessions of individual standardized residents. Analysis of inter-rater reliability utilized intraclass correlation coefficients between each standardized resident and an independent reviewer. Internal consistency of our instrument was measured by Cronbach's alpha. Correlations utilized Pearson's coefficient. We used the Mann-Whitney U-test to compare assessment scores between trainee and faculty member participants. Baseline assessment scores were compared with those following deliberate practice using a paired t-test.

OUTCOMES

The final OSSE assessment instrument domains included *supervision climate*, *control of patient care*, *trainee evaluation*, *instructional skills*, *case-specific measures*, and *overall supervisor rating*. The assessment tool included three global rating scales (GRS) among 19 items (Appendix S1). Our standard setting yielded a separate MPS for checklist items and GRS items. The MPS was 78.2% (21/26) for checklist items and 76.1% (12/15) for GRS items. Judges deemed three items critical: "immediately corrected erroneous actions," "prioritized patient safety," and "supervisor assumed control at appropriate time."

We administered OSSE sessions for 24 faculty participants. Overall, 17/24 (70.8%) participants met the MPS on initial assessment. Average initial assessment score was 76.2%. After feedback and deliberate practice, absolute participant scores increased 19.5% (95% CI of the difference 10.3%–28.8%, $p = 0.002$), and all participants met the MPS.

Overall standardized resident adherence to scripted actions during reviewed sessions was 97.1% (range 93.0%–98.2%). Intraclass agreement of each standardized resident ranged from 93.0% to 96.5%. Reliability of the OSSE assessment instrument was excellent. Intraclass correlation coefficients between each standardized resident and an independent reviewer ranged from 0.84 to 0.89. Internal consistency of the assessment instrument was good (Cronbach's $\alpha = 0.813$).

There was a moderate, positive correlation between participant OSSE and intubation procedure checklist scores (Pearson's $r = 0.435$; 95% CI 0.049–0.709; $p = 0.03$). Year of postgraduate training had a strong, positive correlation with resident OSSE score (Pearson's $r = 0.976$; 95% CI 0.080–0.999; $p = 0.02$). As expected, faculty members had a significantly higher median OSSE score than resident volunteers (85.4% vs. 65.9%; estimated 95% CI of the difference 7.3%–36.6%; $p = 0.01$).

REFLECTIVE DISCUSSION

Our proof-of-concept model demonstrates validity evidence for the OSSE to assess and teach faculty physician supervision of a high-risk procedure. The OSSE provides opportunity for participants to

demonstrate resident supervision in an educational setting, a Level 4 outcome of competence in Moore's framework.⁸ The OSSE as an evolution of the historic OSTE format, specifically incorporating distinct supervisory skills to ensure patient safety.

Although our results provide validity evidence for the OSSE model, further testing is required to demonstrate improvement in patient-centered outcomes.¹⁶ Such evaluation is necessary before implementation as a routine form of continuing professional development for faculty members. Our preliminary data also suggest that OSSE methodology may be appropriate for training senior resident physicians to be competent supervisors of junior trainees, although this cohort requires further study. Our small number of resident study participants limits interpretation of graded performance and comparison to attending physician subjects.

It is unclear whether supervision performance differs in the clinical environment among faculty members who participated in our OSSE compared to their non-OSSE-trained peers. This requires development of appropriate real-time evaluations of supervision or potentially observation of supervisors by peer coaches using the OSSE tool on shift. In the future, the OSSE may be modified for oversight of advanced residents in alignment with graded responsibility or as a training tool for residents serving in a supervisory capacity.

AUTHOR CONTRIBUTIONS

Nicholas Pokrajac and Kimberly Schertzer conceived and designed the study, with Michael A. Gisondi providing subject expertise. All authors except Michael A. Gisondi and Kimberly Schertzer contributed significant time to assess study participants and collect data. Nicholas Pokrajac recruited participants and oversaw data collection. Nicholas Pokrajac primarily performed data analysis. Nicholas Pokrajac, Kimberly Schertzer, and Michael A. Gisondi drafted the manuscript, and all authors contributed substantially to its revision. Nicholas Pokrajac takes responsibility for the manuscript as a whole.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

ORCID

Nicholas Pokrajac  <https://orcid.org/0000-0002-3033-3995>

Cori Mc Clure Poffenberger  <https://orcid.org/0000-0003-0461-1071>

Richard W. Sapp  <https://orcid.org/0000-0002-2381-0300>

Michael A. Gisondi  <https://orcid.org/0000-0002-6800-3932>

Kimberly Schertzer  <https://orcid.org/0000-0003-4375-8180>

REFERENCES

1. Kilminster S, Cottrell D, Grant J, Jolly B. AMEE guide no. 27: effective educational and clinical supervision. *Med Teach*. 2009;29(1):2-19.
2. Sterkenburg A, Barach P, Kalkman C, Gielen M, ten Cate O. When do supervising physicians decide to entrust residents with unsupervised tasks? *Acad med*. 2010;85:1408-1417.
3. Kennedy TJT, Lingard L, Baker GR, Kitchen L, Regehr G. Clinical oversight: conceptualizing the relationship between supervision and safety. *J Gen Intern med*. 2007;22(8):1080-1085.

4. Sebok-Syer SS, Chahine S, Watling CJ, Goldszmidt M, Cristancho S, Lingard L. Considering the interdependence of clinical performance: implications for assessment and entrustment. *Med Educ*. 2018;52(9):970-980.
5. Goldszmidt M, Faden L, Dornan T, van Merriënboer J, Bordage G, Lingard L. Attending physician variability: a model of four supervisory styles. *Acad med*. 2015;90:1541-1546.
6. Farnan JM, Johnson JK, Meltzer DO, et al. Strategies for effective on-call supervision for internal medicine residents: the superb/safety model. *J Grad med Educ*. 2010;2(1):46-52.
7. Moore DE, Green JS, Gallis HA. Achieving desired results and improved outcomes: integrating planning and assessment throughout learning activities. *J Contin Educ Health Prof*. 2009;29(1):1-15.
8. Julian K, Appelle N, O'Sullivan P, Morrison EH, Wamsley M. The impact of an objective structured teaching evaluation on faculty teaching skills. *Teach Learn med*. 2012;24(1):3-7.
9. McSparron JI, Ricotta DN, Moskowitz A, et al. The ProSTE: identifying key components of effective procedural teaching. *Ann Am Thorac Soc*. 2015;12(2):230-234.
10. Farnan JM, Petty LA, Georgitis E, et al. A systematic review: the effect of clinical supervision on patient and residency education outcomes. *Acad med*. 2012;87:428-442.
11. Osman C, Dembitzer A, Zabar S, Tewksbury L. Using Objective Structured Teaching Exercises for faculty development. *MedEdPORTAL*; 2015 Oct 28. [10.15766/mep_2374-8265.10258](https://doi.org/10.15766/mep_2374-8265.10258)
12. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995;311(7001):376-380.
13. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49(6):560-575.
14. McGaghie WC, Barsuk J, Wayne DB. *Comprehensive Healthcare Simulation: Mastery Learning in Health Professions Education*. Springer; 2020.
15. Yudkowsky R, Park YS, Lineberry M, Knox A, Ritter EM. Setting mastery learning standards. *Acad med*. 2015;90:1495-1500.
16. McGaghie WC, Issenberg SB, Barsuk JH, Wayne DB. A critical review of simulation-based mastery learning with translational outcomes. *Med Educ*. 2014;48(4):375-385.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pokrajac N, Roszczynialski KN, Rider A, et al. The OSSE: Development and validation of an "Objective Structured Supervision Evaluation". *AEM Educ Train*. 2022;00:e10784. doi: [10.1002/aet2.10784](https://doi.org/10.1002/aet2.10784)