

ENTRUST: A Serious Game-Based Virtual Patient Platform to Assess Entrustable Professional Activities in Graduate Medical Education

Dana T. Lin¹, MD, FACS
Edward F. Melcer², PhD
Oleksandra Keehl³, MS
Hyrum Eddington⁴, BS
Amber W. Trickey⁵, PhD, MS, CPH

Jason Tsai, MS
Fatyma Camacho, MS
Sylvia Bereknyei Merrell⁶, DrPH, MS
James Korndorffer Jr⁷, MD, MHPE, FACS
Cara A. Liebert⁸, MD, FACS

ABSTRACT

Background As entrustable professional activities (EPAs) are implemented in graduate medical education, there is a great need for tools to efficiently and objectively evaluate clinical competence. Readiness for entrustment in surgery requires not only assessment of technical ability, but also the critical skill of clinical decision-making.

Objective We report the development of ENTRUST, a serious game-based, virtual patient case creation and simulation platform to assess trainees' decision-making competence. A case scenario and corresponding scoring algorithm for the Inguinal Hernia EPA was iteratively developed and aligned with the description and essential functions outlined by the American Board of Surgery. In this study we report preliminary feasibility data and validity evidence.

Methods In January 2021, the case scenario was deployed and piloted on ENTRUST with 19 participants of varying surgical expertise levels to demonstrate proof of concept and initial validity evidence. Total score, preoperative sub-score, and intraoperative sub-score were analyzed by training level and years of medical experience using Spearman rank correlations. Participants completed a Likert scale user acceptance survey (1=strongly agree to 7=strongly disagree).

Results Median total score and intraoperative mode sub-score were higher with each progressive level of training ($\rho=0.79$, $P<.001$ and $\rho=0.69$, $P=.001$, respectively). There were significant correlations between performance and years of medical experience for total score ($\rho=0.82$, $P<.001$) and intraoperative sub-scores ($\rho=0.70$, $P<.001$). Participants reported high levels of platform engagement (mean 2.06) and ease of use (mean 1.88).

Conclusions Our study demonstrates feasibility and early validity evidence for ENTRUST as an assessment platform for clinical decision-making.

Introduction

Graduate medical education (GME) is moving toward a competency-based paradigm and utilizing entrustable professional activities (EPAs) as a means of granting increasing levels of autonomy to trainees as they acquire proficiency in key clinical tasks and domains.^{1,2} Multiple core GME specialties have defined and plan to integrate EPAs into residency and fellowship programs to ensure that trainees can safely and independently perform the essential work activities that embody a particular specialty.³⁻⁶ In 2018, the American Board of Surgery (ABS) identified 5 initial core general surgery EPAs for implementation in surgical residency.⁷ Recently, after conducting a 2-year nationwide pilot, the ABS announced the launch of 18 EPAs for all training programs by 2023.^{8,9} Several ABS task

forces and working groups are currently working to formalize requirements and develop best practices for implementation.¹⁰

The determination of entrustment is predicated on direct observation and assessment by faculty of behaviors in the clinical setting. While frequent, real-time micro-assessments are ideal, there are significant barriers to implementation.^{11,12} This approach places a sizeable burden on faculty amidst competing clinical demands. To assuage this faculty load and efficiently capture high volumes of performance data, mobile applications have been developed to facilitate short, frequent, workplace-based assessments. However, assessment tools deployed via mobile apps have limitations of being too broad or too specific, and the feedback provided too brief or unactionable.¹¹

Given these challenges, many ABS pilot institutions have operationalized EPAs by focusing on operative autonomy, as readily available tools exist to measure this construct, such as SIMPL and the Ottawa Surgical Competency Operating Room

DOI: <http://dx.doi.org/10.4300/JGME-D-22-00518.1>

Editor's Note: The online version of this article contains the survey used in the study.

Evaluation.^{13,14} While such apps have demonstrated validity evidence for evaluating technical skills performance,^{13,15-17} they do not assess clinical decision-making. However, competency in each ABS EPA domain requires a trainee to make informed, safe, complex, and highly nuanced decisions across a spectrum of clinical presentations, not only intraoperatively, but also pre- and postoperatively. Therefore, there is a critical need for evidence-based EPA-aligned tools that specifically measure entrustment for clinical decision-making as a fitting complement to technical skills evaluations.

Clinical decision-making has traditionally been tested by methods such as oral boards questioning, a time- and resource-intensive process that is subject to bias.¹⁸ Virtual patient simulations have the benefit of testing learner competency through standardized scenarios, enabling trainees to demonstrate their diagnostic and clinical acumen in an observable, objective, and measurable way. Implicit bias and subjectivity are minimized, protecting against “halo effect.”¹⁹ This approach also decompresses the assessment burden off faculty raters and away from fast-paced clinics. Game-based assessment also provides a more authentic and immersive context for assessing competency, which is crucial for acquiring a more accurate gauge of skill.²⁰ Furthermore, the intrinsic engaging and motivating nature of game-based assessment appears to increase learners’ flow experience and reduce test anxiety, leading to improved examination performance.²¹ Virtual reality programs and serious game-based assessments are gaining increasing recognition,²¹⁻²⁴ specifically in medicine²⁵⁻²⁷ and postgraduate medical education in light of the EPA framework.²⁸

To address this need to develop objective and rigorous yet efficient and scalable tools to assess surgical decision-making, our group developed ENTRUST, an innovative serious game-based platform to author and deploy virtual patient simulation case scenarios to assess decision-making competence for EPAs. It allows medical experts without background in computer programming to create case scenarios customized to a clinical area of interest. Once activated on ENTRUST, diagnostic workup and treatment actions may be directly observed and captured. This data may inform entrustment decisions and level of autonomy granted to a trainee. In this study, we created and iteratively developed via expert consensus a single case scenario for the ABS EPA domain Inguinal Hernia. The case was deployed on ENTRUST and piloted by participants to demonstrate proof of concept and collect preliminary validity evidence using Messick’s framework.^{29,30}

Objectives

We report the development of and preliminary validity evidence for ENTRUST, an innovative virtual patient simulation platform designed to assess clinical decision-making competence for entrustable professional activities.

Findings

Participants found ENTRUST engaging and easy to use, and performance on ENTRUST correlated positively with level of medical expertise.

Limitations

Limitations include a small sample size comprising participants from a single institution.

Bottom Line

ENTRUST has potential to be utilized by training programs to assess clinical acuity and inform entrustment decisions for trainees.

Methods

Platform Development

ENTRUST was developed in collaboration with a software engineering team at the Baskin School of Engineering, University of California, Santa Cruz (UCSC). The concept, features, and functionality of the platform were envisioned and guided by 2 board-certified general surgeons with expertise in surgical education (D.T.L., C.A.L.). The software development team was assembled and led by a game developer at UCSC (E.F.M.). The team consisted of a project manager, full-stack programmer, user interface designer, and graphic artist. Platform development was supported through seed grant funding from the Division of General Surgery and Department of Surgery at Stanford University School of Medicine and by the Mark Freidell Research Grant from the Association of Program Directors in Surgery, an amount totaling \$95,500.

Platform Description

Assessment Platform: The online ENTRUST assessment platform features 2 primary phases of game play:

1. Preoperative Simulation Mode (FIGURE 1): Patient scenarios begin in the preoperative setting, where the examinee is presented with a brief clinical vignette and initiates workup of the patient. The examinee can elicit physical examination findings, order diagnostic tests, administer fluids and medications, and perform bedside procedures. The patient’s vital signs change dynamically based on the patient’s clinical status and interventions performed. Points are earned for ordering relevant labs and key interventions; conversely, points are lost

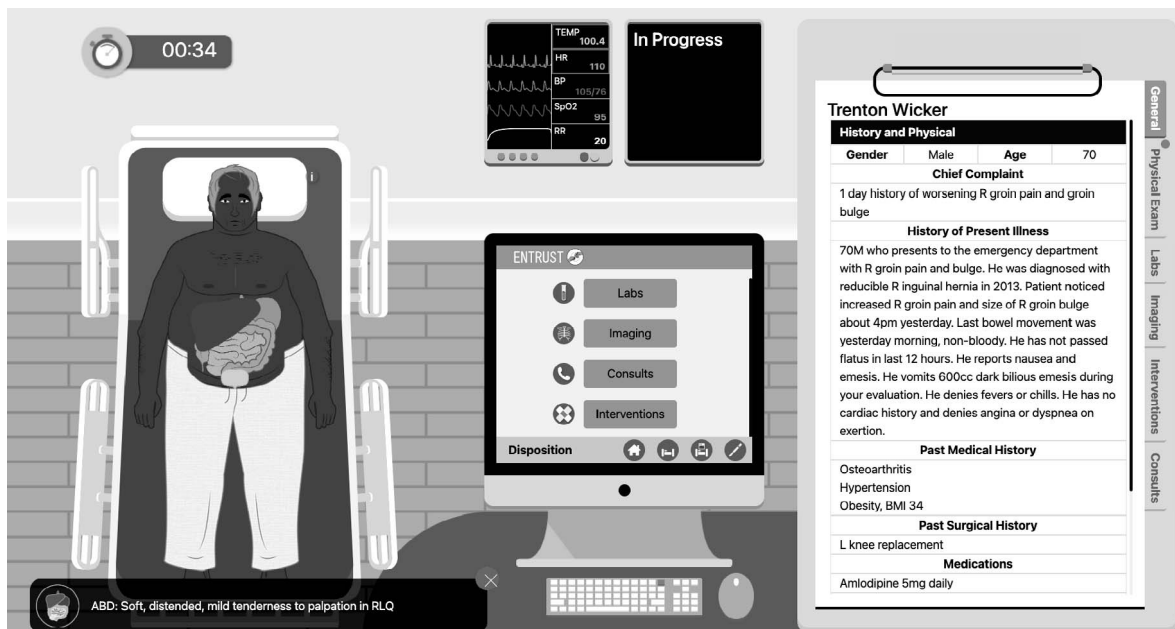


FIGURE 1
ENTRUST Assessment Platform: Serious Game-Based Preoperative Simulation Mode

for performing inappropriate, unnecessary, or harmful actions.

2. Intraoperative/Postoperative Question Mode (FIGURE 2): When the examinee proceeds to the operating room, the case scenario transitions to the Question Mode where the examinee is tested on intraoperative decision-making and management of postoperative complications via a series

of single best answer multiple-choice questions (MCQs). Points are awarded for answering correctly and deducted for answering incorrectly.

Authoring and Administration Portal: ENTRUST features an online authoring portal for clinicians and content experts without a background in

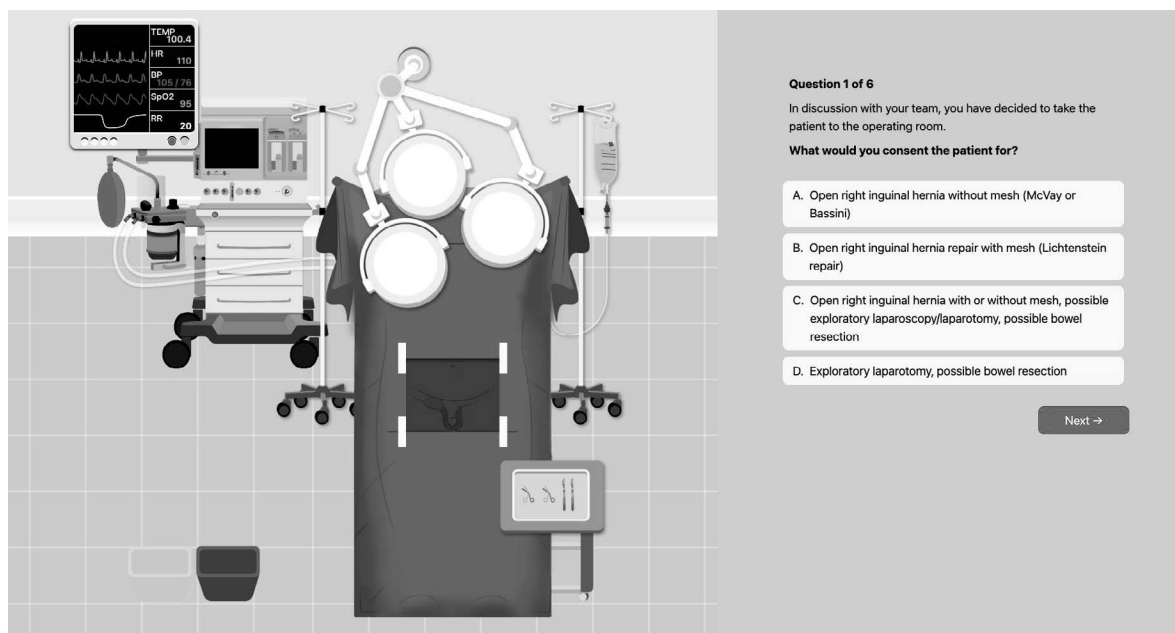


FIGURE 2
ENTRUST Assessment Platform: Intraoperative Multiple-Choice Question Mode

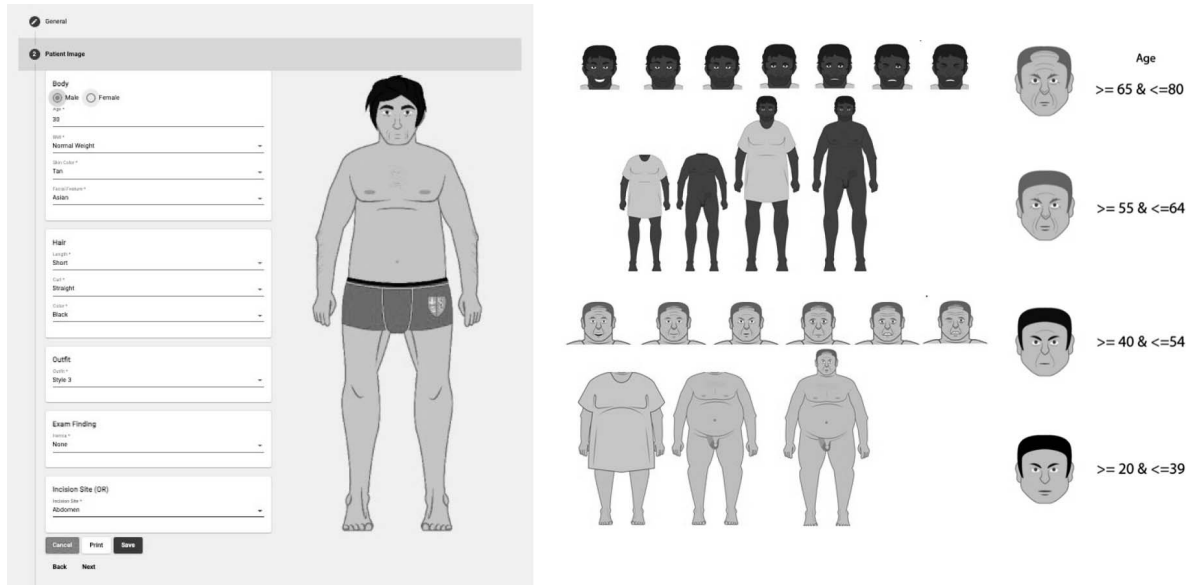


FIGURE 3

ENTRUST Authoring Platform: Case Authoring Tool for Patient Character Generation and Entry of Clinical Data and Scoring Algorithm for Virtual Patient

computer programming to create and deploy new case scenarios (FIGURE 3). Numerous customization options allow for nearly unlimited cases to be crafted. Patient avatars are designed via a patient character generating tool that can depict a diversity of ages, body habitus, skin tones, facial expressions, hair color/styles, and apparel. Specialized labs and orders can be added to basic default choices in the diagnostic and intervention menus. Clinical vignettes and MCQs are entered and edited via a structured template. Clinical photographs and radiology images can be uploaded to be interpreted by the examinee. Preprogrammed vital sign algorithms may be selected to reflect the clinical trajectory of the patient, including *stable patient*, *respiratory decompensation*, and *septic shock*, among others. Authors designate effects of interventions on vital signs and assign point values to actions based on a tiered scoring system. Individual cases can be linked together to create examinations. For secure examination deployment, administrators can create and distribute unique login credentials to examinees and schedule specific times during which the user credentials and the examinations themselves are activated.

Back-end Database: A secure back-end database logs detailed trainee performance data, including a time stamp of all actions, points awarded or deducted for the action or intervention, and responses to all MCQs. The database may be queried to extract data

in either individual or aggregate format for program-specific or research purposes.

Technology Specifications: ENTRUST works on most modern browsers (Chrome, Firefox, and Edge) and is distributable to participants through a web link. The platform can be run on almost any modern computer with internet connectivity. It is currently optimized for desktop or laptop computers with a mouse or touchpad; future iterations may enable use on touchscreen tablet and smartphone devices.

Case Creation and Scoring Algorithm

As an ABS EPA pilot⁸ institution, our program was assigned the Inguinal Hernia EPA as a subject area. Therefore, a case scenario featuring a patient with a strangulated inguinal hernia was authored and aligned with the ABS EPA definition and essential functions for inguinal hernia.⁷ The case content was then reviewed and discussed in an in-person meeting by an expert panel of 5 board-certified general surgeons representing academic and community practice settings. Key/critical actions were identified through unanimous agreement by the panel; differences in practice patterns were catalogued. The MCQs were vetted to ensure unanimous agreement on content and answer choice. The case was iteratively revised based on this feedback, with the final case scenario reviewed and approved by authors who are also board-certified general surgeons (C.A.L., D.T.L.). A scoring algorithm for ENTRUST was developed by C.A.L. and D.T.L. to

reflect appropriateness of actions. The scoring framework was designed to be intuitive to the case author ascribing the point values, starting with neutral/optional actions at 0 with graded intervals in either direction commensurate to the clinical relevancy and effect on the patient's health status. The virtual patient case scenario, MCQs, and scoring attributions were uploaded to the ENTRUST assessment platform via the authoring portal. Case writing, vetting, revision, and scoring required approximately 4 to 5 hours. Uploading of case content onto the ENTRUST authoring portal was completed in 1 hour. The virtual patient simulation case was functional and deployable immediately thereafter. Technical beta-testing was conducted by the development team to ensure that the platform was in a stable, feature-complete state.

Initial Validity Study

Participants were recruited to complete the ENTRUST inguinal hernia case scenario in January 2021. Medical students (MS) or physician assistants (PA) and general surgery residents rotating on core general surgery services were invited to participate, as were practicing surgeons at our institution. Participants first completed a survey querying demographic information, education and training, and prior video game experience. After viewing a standardized video tutorial, participants completed a non-scored practice case which enabled them to familiarize themselves with the platform interface and functionality. Participants then completed the ENTRUST strangulated inguinal hernia case scenario in a proctored setting. Following the simulation, an online user-acceptance survey was administered to obtain feedback on usability and user experience through a 7-point Likert scale (1=strongly agree to 7=strongly disagree).

Descriptive statistics for total and sub-scores, including range, median, and interquartile range, were calculated for each surgical training group (students [S], residents [R], and attending surgeons [A]). Spearman rank calculations were conducted to assess the relationship between total score and years of medical training, which was visualized using locally estimated scatterplot smoothing (LOESS). To assess the potential of ENTRUST to discriminate between different levels of surgical training, we performed known-groups analysis comparing the total case score and sub-scores between groups. Posttest user survey data were analyzed to assess ease of use, platform engagement, and participant response processes.

In view of Messick's contemporary framework,²⁹ preliminary validity evidence for ENTRUST was collected from 3 sources: content, relationship to other variables, and response process. Content evidence was established by alignment of case content with published ABS EPA definitions⁷ and consensus-driven expert review of case content and scoring algorithm. Relationship to other variables was investigated by comparing ENTRUST scores to demographic factors and level of surgical expertise. Response process evidence was collected via the posttest user survey.

Results

A total of 19 participants completed this study: 6 students (31.6%; 3 MS3, 2 MS4, and 1 PA student), 7 residents (36.8%; 2 postgraduate year (PGY)-1, 2 PGY-2, 2 PGY-3, 1 PGY-4), and 6 attending surgeons (31.6%). The mean age was 33.3 years; 9 of the participants were female (47.4%); 1 identified as Latino (5.3%), 1 Black or African American (5.3%), 4 Asian (21.1%), and 13 White (68.4%). The self-reported prior video game experience of the participants ranged from 0 to 6 hours per week with mean 1.17 (SD 2.12). There was no correlation between total score and prior video game experience ($\rho=0.47$, $P=.23$). Total scores ranged from -800 to +2250, with preoperative sub-scores ranging from -1200 to +1050 and intraoperative sub-scores ranging from -800 to +1200. Years of medical training was positively correlated with ENTRUST total score ($\rho=0.82$, $P<.001$, FIGURE 4) and intraoperative question mode sub-score ($\rho=0.70$, $P<.001$). Progressive increases in total score were observed at each level of medical training (median: S=175, R=400, A=1500; $\rho=0.79$, $P<.001$; FIGURE 5a) and intraoperative sub-score (median: S=0, R=400, A=1200; $\rho=0.69$, $P=.001$; FIGURE 5c). There was no statistically significant difference in preoperative simulation sub-score by level of training (median: S=-175, R=0, A=300; $\rho=0.39$, $P=.10$; FIGURE 5b).

A posttest online survey queried participants' ability to navigate the ENTRUST simulation platform by rating statements on a 7-point Likert rating scale (1=strongly agree to 7=strongly disagree). The survey was completed by 16 of the 19 (84%) participants in the study. Overall, participants found it easy to interpret the patient's clinical state from the vital signs and patient medical record (75.0% agree or strongly agree; mean=1.88, SD=0.81), perform a physical examination (81.3% agree or strongly agree; mean=1.81, SD=1.33), and order labs, medications, and other interventions (75.0% agree or strongly

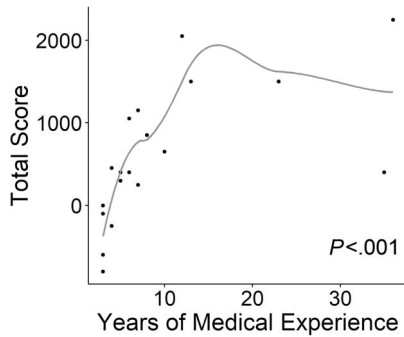


FIGURE 4
ENTRUST Total Score by Cumulative Years of Medical Training and Experience

agree; mean=1.88, SD=1.15). Participants reported a high level of engagement while using ENTRUST (81.3% agree or strongly agree; mean=2.06, SD=1.29).

Discussion

We developed ENTRUST, an innovative virtual patient authoring and assessment platform to deploy rigorous, case-based patient simulations for evaluation of EPAs. This study verifies the usability and functionality of both the ENTRUST authoring portal and assessment platform. Our data indicate that ENTRUST possesses the ability to differentiate between levels of surgical training, offering initial evidence of its relationship to other established variables in surgical education. A significant increase in total score and intraoperative question mode sub-score was noted with successively higher levels of surgical expertise. Statistical significance was not achieved for the preoperative sub-score, which is most likely attributed to small sample size. A more pronounced difference was seen between groups for the intraoperative question mode sub-score, likely

reflecting that the intraoperative question mode necessitates more nuanced and sophisticated surgical decision-making ability as it queries how the examinee would proceed in response to the varying clinical findings in the operating room. In addition, the high usability and acceptability ratings indicate positive response process evidence.

Several publications have leveraged simulation to formatively assess EPAs in undergraduate medical education and other clinical specialties and demonstrated feasibility of the assessments and acceptability among key stakeholders,³¹⁻³³ with one reporting early validity evidence.³⁴ Time and resource-intensive, in-person simulation scenarios were employed for most, limiting the ability to scale across training programs. One utilized a web-based virtual simulation case platform to evaluate incoming interns' ability to diagnose and manage common chief complaints in emergency medicine. Their work highlighted the benefit of psychological safety in a virtual environment as well as the ability to generate individualized feedback to identify deficiencies and tailor learning. The virtual platform interface possessed acceptable fidelity for users.³³ The ENTRUST platform is unique among these simulation-based assessments in its potential for scale and broad applicability to diverse specialties and learner populations. The authoring and administration capabilities empower faculty educators to custom create and deploy case scenarios relevant to their specialties and program leadership to track and integrate objective performance data with other work-based assessments to guide entrustment decisions.

The limitations of this study include a small sample size comprising participants from a single institution. While development of ENTRUST necessitated a high upfront cost, the inherent design of the platform allows for widespread utilization and access without a

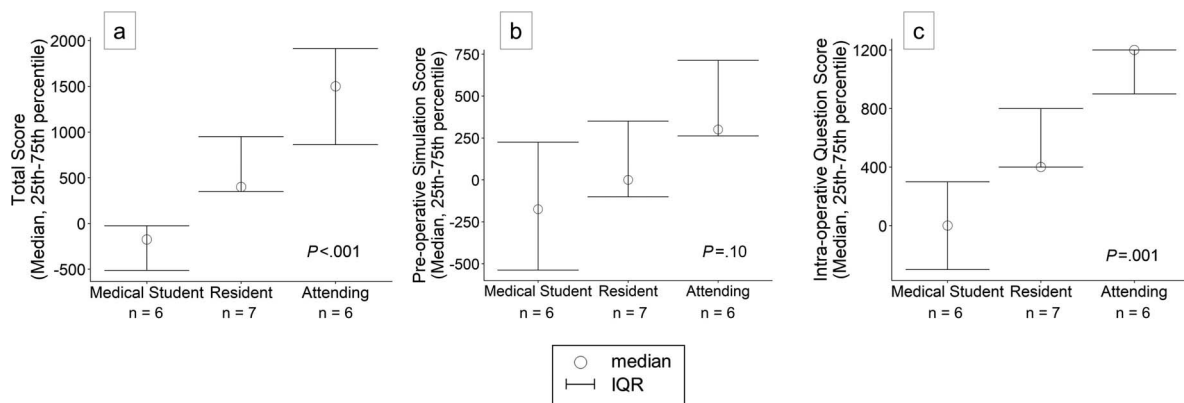


FIGURE 5
ENTRUST Total Score by Level of Training

Note: (a) Total Score; (b) Preoperative Simulation Mode Sub-Score; (c) Intraoperative Question Mode Sub-Score.

significant need for additional resources. The greatest barrier to implementation of ENTRUST in training programs may be the personnel to create robust, evidence-based case scenarios to assess the full suite of EPA domains in a given specialty area.

This study will inform future larger scale studies featuring a repertoire of cases to encompass the 18 core general surgery EPAs. We plan to further collect validity evidence, conduct standard setting, and map game play patterns and actions to EPA levels. We envision ENTRUST as a complement to work-based micro-assessments to provide trainees and program leadership an objective measure of clinical decision-making competence. While micro-assessments afford frequent, formative feedback to trainees, ENTRUST may serve as a higher-stakes summative assessment providing more detailed, granular feedback on trainees' readiness for entrustment for each EPA domain.

We also plan to expand the capability of ENTRUST by developing additional environments, assets, and functionality to accommodate higher acuity clinical scenarios situated in the trauma bay and intensive care unit settings. Heightened security infrastructure, score, and feedback reports for trainees, as well as EPA dashboards for program leadership are also underway.

Conclusions

This initial study of ENTRUST demonstrates the feasibility of delivering immersive, evidence-based clinical decision-making assessments in an efficient and scalable manner, as well as promising preliminary validity evidence for ENTRUST as an objective measure of clinical competence.

References

- ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med*. 2007;82(6):542-547. doi:10.1097/ACM.0b013e31805559c7
- ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. Curriculum development for the workplace using entrustable professional activities (EPAs): AMEE Guide No. 99. *Med Teach*. 2015;37(11):983-1002. doi:10.3109/0142159X.2015.1060308
- Hart D, Franzen D, Beeson M, et al. Integration of entrustable professional activities with the milestones for emergency medicine residents. *West J Emerg Med*. 2019;20(1):35-42. doi:10.5811/westjem.2018.11.38912
- American Board of Pediatrics. Entrustable professional activities for general pediatrics. Accessed January 2, 2023. <https://www.abp.org/content/entrustable-professional-activities-general-pediatrics>
- Deitte LA, Gordon LL, Zimmerman RD, et al. Entrustable professional activities: ten things radiologists do. *Acad Radiol*. 2016;23(3):374-381. doi:10.1016/j.acra.2015.11.010
- Fessler HE, Addrizzo-Harris D, Beck JM, et al. Entrustable professional activities and curricular milestones for fellowship training in pulmonary and critical care medicine: report of a multisociety working group. *Chest*. 2014;146(3):813-834. doi:10.1378/chest.14-0710
- Brasel KJ, Klingensmith ME, Englander R, et al. Entrustable professional activities in general surgery: development and implementation. *J Surg Educ*. 2019;76(5):1174-1186. doi:10.1016/j.jsurg.2019.04.003
- American Board of Surgery. Entrustable professional activities pilot to begin this July. Published January 13, 2021. Accessed January 12, 2021. <http://www.absurgery.org/quicklink/absnews/absupdate0518.html#epa>
- New model of surgical resident autonomy coming in 2023. *ACS Clinical Congress News*. Published October 23, 2021. Accessed January 21, 2022. <https://www.acscconews.org/new-model-of-surgical-resident-autonomy-coming-in-2023/>
- American Board of Surgery. Entrustable professional activities. Accessed January 3, 2023. <https://www.absurgery.org/default.jsp?epahome>
- Young JQ, Sugarman R, Schwartz J, McClure M, O'Sullivan PS. A mobile app to capture EPA assessment data: utilizing the consolidated framework for implementation research to identify enablers and barriers to engagement. *Perspect Med Educ*. 2020;9(4):210-219. doi:10.1007/s40037-020-00587-z
- Malone K, Supri S. A critical time for medical education: the perils of competence-based reform of the curriculum. *Adv Health Sci Educ*. 2012;17(2):241-246. doi:10.1007/s10459-010-9247-2
- Bohnen JD, George BC, Williams RG, et al. The feasibility of real-time intraoperative performance assessment with SIMPL (System for Improving and Measuring Procedural Learning): early experience from a multi-institutional trial. *J Surg Educ*. 2016;73(6):e118-e130. doi:10.1016/j.jsurg.2016.08.010
- Saliken D, Dudek N, Wood TJ, MacEwan M, Gofton WT. Comparison of the Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) to a single-item performance score. *Teach Learn Med*.

- 2019;31(2):146-153. doi:10.1080/10401334.2018.1503961
15. George BC, Bohnen JD, Schuller MC, Fryer JP. Using smartphones for trainee performance assessment: a SIMPL case study. *Surgery*. 2020;167(6):903-906. doi:10.1016/j.surg.2019.09.011
 16. Eaton M, Scully R, Schuller M, et al. Value and barriers to use of the SIMPL tool for resident feedback. *J Surg Educ*. 2019;76(3):620-627. doi:10.1016/j.jsurg.2019.01.012
 17. Spencer M, Sherbino J, Hatala R. Examining the validity argument for the Ottawa Surgical Competency Operating Room Evaluation (OSCORE): a systematic review and narrative synthesis. *Adv Health Sci Educ Theory Pract*. 2022;27(3):659-689. doi:10.1007/s10459-022-10114-w
 18. 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*. 2020;155(1):22-30. doi:10.1001/jamasurg.2019.4081
 19. DiBrito SR, Hicks CW. The American Board of Surgery Certifying Exam: embrace technology, eradicate bias. *Surgery*. 2021;169(4):994-995. doi:10.1016/j.surg.2020.07.026
 20. Seelow D. The art of assessment: using game based assessments to disrupt, innovate, reform and transform testing. *J App Test Technol*. 2019;20(suppl 1):1-16.
 21. Mavridis A, Tsiatsos T. Game-based assessment: investigating the impact on test anxiety and exam performance. *J Comput Assist Learn*. 2017;33(2):137-150. doi:10.1111/jcal.12170
 22. Nguyen THD, Melcer E, Canossa A, Isbister K, El-Nasr MS. Seagull: a bird's-eye view of the evolution of technical games research. *Entertain Comput*. 2018;26:88-104.
 23. Melcer E, Nguyen THD, Chen Z, Canossa A, El-Nasr MS, Isbister K. Games research today: analyzing the academic landscape 2000-2014. *Network*. 2015;17:20.
 24. DiCerbo KE. Game-based assessment of persistence. *J Educ Technol Soc*. 2014;17(1):17-28.
 25. Charlier N. Game-based assessment of first aid and resuscitation skills. *Resuscitation*. 2011;82(4):442-446. doi:10.1016/j.resuscitation.2010.12.003
 26. Vallejo V, Wyss P, Rampa L, et al. Evaluation of a novel serious game based assessment tool for patients with Alzheimer's disease. *PloS One*. 2017;12(5):e0175999. doi:10.1371/journal.pone.0175999
 27. Lin DT, Park J, Liebert CA, Lau JN. Validity evidence for surgical improvement of clinical knowledge ops: a novel gaming platform to assess surgical decision making. *Am J Surg*. 2015;209(1):79-85. doi:10.1016/j.amjsurg.2014.08.033
 28. Graafland M, ten Cate O, van Seventer JP, Schraag JMC, Schijven MP. Mapping the demand for serious games in postgraduate medical education using the entrustable professional activities framework. *Games Health J*. 2015;4(5):381-386. doi:10.1089/g4h.2014.0129
 29. Messick S. Standards of validity and the validity of standards in performance assessment. *Educ Meas Issues Pract*. 1995;14(4):5-8. doi:10.1111/j.1745-3992.1995.tb00881.x
 30. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ Theory Pract*. 2014;19(2):233-250. doi:10.1007/s10459-013-9458-4
 31. Herrigel DJ, Donovan C, Goodman E, et al. Simulation as a platform for development of entrustable professional activities: a modular, longitudinal approach. *Cureus*. 2020;12(10):e11098. doi:10.7759/cureus.11098
 32. Kman NE, Thompson L, Hess J, et al. Entrustable professional activity 10: case simulation and assessment-STEMI with cardiac arrest. *MedEdPORTAL*. 2016;12:10517. doi:10.15766/mep_2374-8265.10517
 33. Peng CR, Schertzer KA, Caretta-Weyer HA, et al. Assessment of entrustable professional activities using a web-based simulation platform during transition to emergency medicine residency: mixed methods pilot study. *JMIR Med Educ*. 2021;7(4):e32356. doi:10.2196/32356
 34. Dwyer T, Wadey V, Archibald D, et al. Cognitive and psychomotor entrustable professional activities: can simulators help assess competency in trainees? *Clin Orthop*. 2016;474(4):926-934. doi:10.1007/s11999-015-4553-x



Dana T. Lin, MD, FACS, is Clinical Associate Professor, Department of Surgery, Stanford University School of Medicine; **Edward F. Melcer, PhD**, is Assistant Professor and Director of the Alternative Learning Technologies and Games (ALT Games) Lab, Baskin School of Engineering, University of California, Santa Cruz; **Oleksandra Keehl, MS**, is a PhD Student, Baskin School of Engineering, University of California, Santa Cruz; **Hyrum Eddington, BS**, is Data Manager and Data Analyst, Stanford-Surgery Policy Improvement Research and Education Center; **Amber W. Trickey, PhD, MS, CPH**, is Senior Biostatistician, Stanford-Surgery Policy Improvement Research and Education Center; **Jason Tsai, MS**, is a Master's Student, Baskin School of Engineering, University of California, Santa Cruz; **Fatyma Camacho, MS**, is a Master's Student, Baskin School of Engineering, University of California, Santa Cruz; **Sylvia Berekyei Merrell, DrPH, MS**, is a Research Scholar, Department of Pediatrics, Stanford University School of Medicine; **James**

Korndorffer Jr, MD, MHPE, FACS, is Vice Chair of Education and Associate Professor of Surgery, Department of Surgery, Stanford University School of Medicine; and **Cara A. Liebert, MD, FACS**, is Clinical Assistant Professor, Department of Surgery, Stanford University School of Medicine.

Funding: This study was supported by the Mark Freidell Research Grant from the Association of Program Directors in Surgery and through seed grant funding from the Division of General Surgery and Department of Surgery at Stanford University School of Medicine.

Conflict of interest: The authors declare they have no competing interests.

This work was presented virtually as an oral abstract presentation at the American College of Surgeons Clinical Congress, October 23-27, 2021.

The authors would like to thank Yulin Cai, Ruonan Chen, Melissa Lee, and Samuel Shields for their contributions to the development of the ENTRUST platform.

Corresponding author: Dana T. Lin, MD, FACS, Stanford School of Medicine, danalin@stanford.edu, Twitter @danalinmd

Received July 6, 2022; revisions received November 22, 2022, and January 12, 2023; accepted January 19, 2023.