Published in final edited form as:

Simul Healthc. 2015 December ; 10(6): 345–351. doi:10.1097/SIH.000000000000108.

AN INTERNAL MEDICINE SIMULATED PRACTICAL EXAM FOR ASSESSMENT OF CLINICAL COMPETENCY IN 3RD YEAR MEDICAL STUDENTS:

AN INTERNAL MEDICINE SIMULATED PRACTICAL EXAM

Cheryl Bodamer, PhD MPH RN [Assistant Professor],

Center for Human Simulation and Patient Safety, School of Medicine, Virginia Commonwealth University

Moshe Feldman, PhD MS [Assistant Professor],

Office of Assessment and Evaluation Studies, Assistant Director of Research and Evaluation, Center for Human Simulation and Patient Safety, School of Medicine, Virginia Commonwealth University

Jeffrey Kushinka, MD [Assistant Professor of Medicine],

Department of Internal Medicine, Virginia Commonwealth University

Ellen Brock, MD, MPH [Professor of Obstetrics and Gynecology],

Director, Center for Human Simulation and Patient Safety, School of Medicine, Virginia Commonwealth University

Alan Dow, MD, MSHA [Associate Professor of Medicine],

Assistant Vice president of Health Sciences for Interprofessional and Collaborative Care, Department of Internal Medicine, Virginia Commonwealth University

Jessica A. Evans, BS [Research assistant], and

School of Medicine, Virginia Commonwealth University

Gonzalo Bearman, MD, MPH, FACP [Professor of Medicine]

Department of Internal Medicine, Virginia Commonwealth University

Abstract

Introduction—Achieving standardized assessment of medical student competency in patient care is a challenge. Simulation may provide unique contributions to overall assessment. We developed an Internal Medicine Standardized Simulation Based Examination (SSBE) for the 3rd year clerkship to assess students' medical knowledge, diagnostic skills, and clinical management skills. We assessed convergent and test-criterion validity by comparing the relationship of SSBE

Corresponding Author: Cheryl Bodamer PhD MPH RN, Assistant Professor, Center for Human Simulation and Patient Safety, VCU School of Medicine, P.O. Box 980038, Richmond, VA 23298, Phone: 804-628-4511, Fax: 804-827-1367, ccunnaal@vcu.edu.

Institution:

Virginia Commonwealth University

Supplemental Digital Files

Supplemental Digital Content 1, adobe acrobat, photoquiz p.6

Supplemental Digital Content 2, word, exam case questions p.7

Supplemental Digital Content 3, ppt, exam case questions p. 8

scores with USMLE Step 2 clinical knowledge, shelf exam, eQuiz, OSCE, ward evaluation scores, and overall clerkship grades. We hypothesize that use of the SSBE will allow for a more reliable assessment of these competencies and add value to existing assessments.

Methods—A prospective study design was used. The SSBE consisted of a computer based photo quiz and cases on high fidelity simulators. Performance on the SSBE was compared to standardized examinations, clinical evaluations, and overall clerkship grades. Students completed an evaluation of the experience.

Results—Two hundred seven students completed the SSBE, with a mean score of 76.69 (SD 7.78). SSBE performance was positively related to other assessments of medical knowledge (eQuiz scores ($r(203) = .33, p < .01$), shelf exam scores ($r(158) = .53, p < .01$), and clinical performance (ward scores) ($r(163) = .31, p < .01$) but not to OSCE scores. There was a positive relationship to final class grades ($r(163) = .45, p < .01$), shelf exam ($r(158) = .52, p < .01$) and Step 2 clinical knowledge scores ($r(76) = .54, p < .01$). The majority (93%) of students agreed that it was a fair exam.

Conclusion—Our results provide validity evidence for the SSBE as an additional assessment tool that uses a novel approach for evaluating competency in patient care at the clerkship level.

Background

With the widespread adoption of the American Council on Graduate Medical Education (ACGME) core competencies, medical educators are increasingly striving to evaluate students using competency-based assessment.^{1,2,3,4} For some core competencies, such as Medical Knowledge, standardized exams from the National Board of Medical Examiners (NBME) and the United States Medical Licensing Examination (USMLE) provide a reliable and valid standard for student ranking. Other core competencies, like Professionalism and Interpersonal and Communication Skills, are evaluated in a standardized fashion by the Objective Structured Clinical Examination (OSCE)^{5,6} culminating in the USMLE Step 2 Clinical Skills (CS) exam.^{2,3} A third group of core competencies that include Patient Care, Systems-Based Practice, and Practice-Based Learning and Improvement are often assessed through clinical evaluations of faculty and supervising house staff. Assessments for this last group of core competencies are prone to unreliability. The processes are not standardized, thus comparison of learners as well as ensuring the attainment of competency is difficult.^{7,8,9,10} Measurement tools, most notably the mini-clinical examination (mini-CEX),¹¹ have been developed to improve assessment but many important areas of competency remain poorly evaluated. In addition, Systems-Based Practice and Practice-Based Learning and Improvement are newer concepts to medical education and less integrated in clerkship curricula.¹² The modern physician needs to be able to work within complex systems (Systems-Based Practice) and learn from the process and outcomes of care delivery (Practice-Based Learning and Improvement) to provide better care to patients (Patient Care).

To reach these educational goals, a more reliable and standardized evaluation of how care is delivered would be of interest to medical schools and to residency programs and would also align with the goals of the American Board of Internal Medicine (ABIM) Milestones project, which seeks to advance residents on the basis of achieved competencies rather than

time spent at a given level of training.^{13,14} To address the deficiency in high quality standardized assessments of the ability to deliver competent patient care, we sought to develop the Standardized Simulation Based Examination (SSBE), a novel examination consisting of patient cases using high-fidelity simulators. These cases require participants to demonstrate the application of medical knowledge to patient care within a complex inpatient environment. This form of assessment has the possibility to enhance medical education by providing new discriminative information about learner competency at the bedside.

Our goal was to assess validity evidence of the SSBE by comparing the relationship of SSBE scores to other commonly used medical student assessments and testing the unique contribution of SSBE scores in explaining student outcomes. Specifically, we expected that SSBE scores would have moderate correlations or convergent validity with other medical student assessments of medical knowledge (USMLE/shelf exam/eQuiz) and clinical reasoning (ward scores), and lower correlations with assessments targeting other competencies such as communication and interpersonal skills. Finally, we expected to find test-criterion validity by investigating whether SSBE scores provide unique information about medical student success on their final grade and USMLE Step 2 clinical knowledge scores.

Methods

The Standardized Simulation Based Examination (SSBE) was conducted using third year medical students enrolled in the Internal Medicine Clerkship. The SSBE is a three-part examination consisting of a photo quiz and patient cases on high fidelity adult simulators, conducted over 150 minutes. Each student rotated through the three components of the exam in small groups of six randomly assigned to their groups ahead of time and instructed to work independently with no discussion allowed. Students were evaluated individually.

Component one consisted of a 52-item multiple choice photo quiz tested over 50 minutes (See adobe acrobat file, Supplemental Digital Content 1, photoquiz). Students logged into a secure site at their own computer station to complete the quiz. The quiz assessed medical knowledge and clinical skills not routinely evaluated in a standardized method during the clerkship, including interpretation of ten chest radiographs, thirteen electrocardiograms (ECGs), and the recognition and diagnosis of twelve common dermatologic and seven ophthalmologic conditions.

The second component comprised of 6 cases over 30 minutes using the Harvey® Cardiopulmonary Simulator to assess clinical competency in the diagnosis of hypertension, aortic stenosis, cardiomyopathy, acute pericarditis, a normal heart, and aortic regurgitation (See document file, Supplemental Digital Content 2, exam case questions). In this session, the instructor entered the appropriate diagnosis code on the keypad, read the case questions from the large LCD screen, and positioned the stethoscope on Harvey. The students listened to the heart sounds with the stethophones then recorded their answers to the exam questions.

The following is an example case question:

A 50 year-old man who hasn't seen a physician in 15 years presents to your office for establishment of regular care and to update his health maintenance. He denies known medical problems and feels well overall. His vital signs reveal a blood pressure of 153/95 mm Hg, a heart rate of 77/min, a respiratory rate of 16/min, and a temperature of 37.2 Celsius. You auscultate his heart. What is the abnormal finding and it's most likely cause?

- A. Physiologic S3, no underlying disease
- B. S4, due to essential hypertension
- C. S3, due to chronic systolic heart failure
- D. Fixed splitting of S2, due to atrial septal defect
- E. S4, due to hypertrophic obstructive cardiomyopathy (HOCM)

The third component of the exam was comprised of three simulated patient cases with a 16-item test on common acute medical conditions; administered on the iSTAN® adult simulator over 45 minutes. The iSTAN® component provided a standardized, simulated, and real-time assessment of diagnostic and management skills of acute conditions such as asthma, myocardial infarction and sepsis. The experience was designed to require participants to interpret clinical findings, determine diagnoses, and determine the appropriate evaluation and management plan.

Scenarios were developed by internal medicine faculty and simulation experts and piloted with multiple clinical content experts before they were used for the exam. The scenarios were programmed using Muse/TouchPro® software (Table 1).

Two mannequins and their patient monitors were set up with their laptops in a large room and run by the simulation technologists. Multiple choice exam case questions were displayed on a large LCD screen along with the lab results and diagnostic images (See PowerPoint file, Supplemental Digital Content 3, exam case questions). The instructor read the case and exam questions displayed on the screen. Students examined the iStan® mannequin, auscultated heart and lung sounds, looked at vital signs on the monitor, and recorded their answers to the questions. Students were provided with a clipboard and a blank answer key. At the end of exam session, students entered their answers electronically. Additionally, they completed an evaluation survey consisting of a 17 item Likert scale questionnaire evaluating students' perceptions and attitudes about simulation technologies for assessment.

The SSBE was developed in coordination with the faculty from the Center for Human Simulation and Patient Safety in the Virginia Commonwealth School of Medicine. The study was approved by the VCU Institutional Review Board.

Statistical Analyses

We used the framework described by Downing¹⁵ to evaluate the validity of the SSBE. This framework specifies five domains of validity related to content, response process, internal structure, relationship with other variables, and consequences. SSBE scores, evaluation surveys, and medical school performance outcomes were entered into a secure Red Cap

database¹⁶ and analyzed using IBM SPSS version 21. Descriptive statistics were calculated for evaluation survey items to evaluate face validity of the SSBE. Student comments from the evaluation survey were assigned content-specific codes and grouped into overarching categories following a conventional qualitative content analysis approach to better understand students' impressions of the SSBE.¹⁷

Correlations among study variables and regression analyses were used to investigate the relationship of SSBE performance with other student performance metrics used by medical educators for assessing medical knowledge, clinical skills, and patient care (ward evaluations, shelf examination scores, and OSCE scores). In addition, multivariable modeling with logistic regression analysis was used to evaluate the added benefit of SSBE scores in predicting student outcomes such as clerkship grades.

Results

Two hundred seven third year medical students participated in the SSBE from March 2011-June 2013 during their internal medicine clerkship rotation. Fourteen students participated during the 2010–2011 academic year, 65 during the 2011–2012 academic year and 128 during the 2012–2013 academic years. A one-way analysis of variance (ANOVA) on mean SSBE scores did not reveal any significant difference between cohorts; hence all cohort data were analyzed as one group.

Performance on the SSBE

SSBE scores are summarized in Table 2. The mean SSBE score was 77% (SD = 7.78) ranging from 53% to 93%. Scores on component I (medical knowledge) ranged from 42% to 90% with the top quartile of students scoring 81% or above and the bottom quartile of student scoring below a 69%. Scores on component II (cardiovascular knowledge and clinical reasoning) ranged from 33% to 100% with the top quartile of students scoring 81% or above and the bottom quartile of student scoring below a 50%. Scores on component III (clinical knowledge and decision making) ranged from 33% to 100% with the top quartile of students scoring 88% or above and the bottom quartile of student scoring below a 69%. Mean scores were significantly higher for component III (M=79.5, SD=11.47) compared to component II (M=75.33, SD=18.31) ($t(205)=3.06, p<.01$) and component I (M=75.15, SD=8.43) ($t(203)=5.17, p<.01$). Our results showed that students scored significantly higher on component III (clinical knowledge and decision making) than other sections of the SSBE. Faculty can use significant performance differences between components to identify learning needs and inform clerkship curriculum planning.

Relationship with medical student assessments

SSBE performance was strongly related to other assessments of medical knowledge including eQuiz scores ($r(203) = .33, p<.01$) and shelf exam scores ($r(158)=.52, p<.01$). SSBE scores were also positively related to clinical performance as measured by ward evaluation scores ($r(163) = .31, p<.01$) demonstrating some convergent validity with other assessments used to measure similar knowledge and skills targeted by the SSBE.

The medical knowledge component I (photo quiz) was most strongly associated with the shelf exam ($r(157) = .47$, $P < .01$) and USMLE Step 2 clinical knowledge ($r(163) = .43$, $p < .01$), and showed more moderate correlation with ward scores ($r(162) = .31$, $p < .01$). There were no significant correlations between SSBE scores and performance on the OSCE ($r(163) = .05$, $p = .26$).

All SSBE component scores were significantly positively related to eQuiz scores, shelf exam scores, and ward evaluation scores with the photo quiz and iStan® components showing stronger correlations than the Harvey® component. Table 3 presents descriptive statistics and correlations among study variables.

A linear regression analysis was then conducted to test whether SSBE scores predicted unique variance in ward evaluation scores. The eQuiz and SSBE were entered as predictor variables because they both correlated with ward evaluation scores. SSBE scores were found to predict unique variance in ward evaluations ($\beta = .27$, $t(2,162) = 3.48$, $p < .01$) after accounting for eQuiz scores thus supporting the hypothesis that the SSBE assessed a unique clinical domain other than medical knowledge measured by the eQuiz assessment.

Relationship with medical student outcomes

Table 4 shows the mean SSBE scores and the mean final grades in the IM clerkship. Students are assigned one of 5 grade levels at the end of their clerkship including honors, high pass, pass, marginal, or incomplete. SSBE scores were strongly positively related to final class grades ($r(163) = .45$, $p < .01$), shelf exam scores ($r(158) = .52$, $p < .01$) and USMLE Step 2 clinical knowledge scores ($r(76) = .54$, $p < .01$). A one-way analysis of variance (ANOVA) was conducted to test whether students with higher class grades scored higher on the SSBE. Results of the ANOVA showed that students receiving 'Honors' or 'High Pass' final grades scored significantly higher on the SSBE than students receiving 'Pass' or 'Marginal' final grades ($F(4,160) = 10.33$, $p < .01$).

Multivariate Logistical Regression on final grades—A logistical regression was conducted to test whether SSBE scores could be used to predict overall student performance (Table 5). Students receiving 'High Pass' or 'Honors' final grades were categorized into the referent group and compared to students receiving 'Pass', 'Marginal' or 'Incomplete' final grades. SSBE was a significant predictor of students receiving a 'High Pass' or 'Honors' final grade after accounting for OSCE, ward evaluation, and eQuiz scores indicating that SSBE scores add unique value for predicting final grades. Specifically, for every one-unit increase in percent correct on SSBE score students were 1.08 times (or 8%) more likely to receive a 'High Pass' or higher final grade.

Linear regression on shelf exam scores—A stepwise linear regression analysis was conducted to test whether SSBE scores predicted performance on the shelf exam separately from other predictors (Table 6). Ward evaluation scores, OSCE, eQuiz, and SSBE were entered as predictor variables in the model because they were positively correlated with shelf exam scores. All variables in the model were significant accounting for an adjusted $R^2 = .45$ or approximately 45% of the variance in shelf exam scores. SSBE was a significant

predictor after accounting for other variables and showed the strongest relationship with shelf exam performance ($\beta = .35$, $t(4,155)=5.37$, $p < .01$).

Linear regression on USMLE Step 2 Clinical Knowledge Scores—A stepwise linear regression was conducted to test whether the SSBE predicted unique variance in USMLE Step 2 clinical knowledge scores. There were 78 students who had already taken the Step 2 clinical knowledge exam included in the analysis (Table 7). eQuiz and OSCE scores were entered into the model because they were both positively related to Step 2 clinical knowledge scores, but only eQuiz scores were retained. A second model was then tested to measure whether eQuiz scores and SSBE scores predicted Step 2 clinical knowledge scores. Results showed that eQuiz scores ($\beta = .32$, $t(2,75)=3.26$, $p < .01$) and SSBE scores ($\beta = .42$, $t(2,75)=4.34$, $p < .01$) both significantly predicted unique variance in USMLE Step 2 clinical knowledge scores (*adjusted* $R^2 = .37$, $p < .01$). These results suggest that the SSBE can measure knowledge and skills related to USMLE Step 2 performance that is not captured using other assessment methods.

SSBE Student Evaluations

One hundred eighty two evaluations were completed for a total response rate of 88%. Overall, students were satisfied with the content of the exam and rated the SSBE better for testing targeted knowledge and skills compared to paper based measures (Table 8). The Harvey component was rated the most favorably with the photo quiz component rated least favorably. When compared to paper-based exam questions, 89% agreed that photo quiz questions are better for testing image recognition, and 79% reported that they more adequately test physical exam findings.

A qualitative conventional content analysis approach¹⁵ was used to further evaluate students' reactions to the SSBE. Two trained evaluators reviewed 371 comments and identified four overarching categories: 1) Assessment technology, 2) curriculum design, 3) organization of the SSBE, and 4) overall testing experience. Comments were then assigned to each category through a consensus process.

Most of the comments, 35% ($n=135$), were related to assessment technology. Students reported difficulty discerning mechanical artifacts from cardiac and respiratory sounds on the iStan® simulator and some reported difficulty with the quality of pictures in the photo quiz. Nineteen percent favored more integration of simulation within the internal medicine clerkship. Eighteen percent desired modification of the testing conditions including more time, space, and immediate feedback. Lastly, 16% conveyed a high level of satisfaction with the testing experience as a whole; "I think this exam is a more comprehensive method to test skills; and I thought the simulation sessions were great and tested various components of clinical development."

Discussion

Traditional standardized written exams have been used for decades to assess medical knowledge, but whether they accurately assess overall clinical competency is much less certain. In a healthcare environment that is increasingly complex, the traditional shelf

examination most likely does not accurately assess all knowledge and skills important to the development of clinical competency. “Competency based medical education, CBME, necessitates a robust and multifaceted assessment system that is designed with key features and components considered as best practice.” (Holmboe et al, 2010)⁴

The “utility index” is a formula used to measure the quality of the assessment; “utility = validity × reliability × educational impact × acceptability × cost effectiveness”^{18,4} and is one indicator of best practice. Based on the pilot, the SSBE has attributes that would contribute to a high utility index.

Our results provide some validity evidence for the SSBE as an additional assessment tool for evaluating competency in patient care. Current assessment of clinical competence in the IM Clerkship relies heavily on the ward evaluation score and the OSCE. Our OSCE assessed interpersonal skills, bedside manner, and limited components of the physical examination; we did not attempt to simulate most pathologic physical exam states (e.g. abnormal cardiac sounds) with standardized patients.

SSBE scores were correlated with ward evaluation scores. In contrast, SSBE scores were uncorrelated with assessments related to interpersonal and communication skills as measured by the OSCE, providing discriminant validity evidence. Unlike the clinical ward evaluation, the standardized nature of the SSBE minimizes subjectivity in learner assessment. SSBE scores were moderately correlated with other measures of medical knowledge including shelf exam, equip, and USMLE Step 2 clinical knowledge scores, providing convergent validity evidence. Most notably, while SSBE scores were correlated with ward evaluation scores and Step 2 clinical knowledge scores, ward evaluation and Step 2 clinical knowledge scores were not correlated with each other, suggesting that the SSBE measures a construct that independently overlaps the constructs measured by ward evaluations and Step 2 scores. The lack of correlation between Step 2 clinical knowledge scores and ward evaluation scores is particularly remarkable because of the significant correlation between most other assessments. Under the above framework, the SSBE may provide a novel, standardized, and additional metric with which to globally measure competence, both in patient care and medical knowledge.

Students perceived the simulated portion of the exam favorably for assessing skills related to competency in patient care. Comments reflected some technological limitations related to interpreting cardiopulmonary findings on the high fidelity iStan mannequin, the size of student groups, and the time allotted per question. These deficiencies were remediated by shrinking the size of groups, making changes to the simulators, and updating the examination software. Importantly, the SSBE allowed students to self-identify their areas of weakness in which to focus their studies. Faculty were provided with insights about the need for additional learning experiences in simulation prior to the exam with high fidelity mannequins, as well as more instruction in the areas of cardiology, ophthalmology, and dermatology. This feedback is crucial because the SSBE is designed to allow case content and correct responses to be titrated to the level of learner to provide a developmentally appropriate, criterion-based assessment.

Implementing the SSBE required technical as well as economic considerations. Simulation space, equipment, and technologists as well as faculty time are necessary to develop the three components and administer the exam. To accommodate 200 students, four daylong sessions were conducted throughout the academic year. Because the resource requirements are significant, utilizing this type of assessment requires the commitment of educational planners and institutional leaders. In addition, the SSBE remains a pass/fail assessment at our institution as we establish the psychometric characteristics of this form of assessment.

Limitations of this study are several. Content validity for the assessments relied on the expertise of internal medicine faculty. Because no gold standard for patient care competency exists, we relied on ward evaluation scores. Correlating SSBE scores to observed clinical care and patient outcomes is a direction for further study. In addition, because students were only assessed once, we have no data on reliability of this assessment over time. This study was limited to one institution and one clerkship so findings of the SSBE may not be generalized to other clerkships or other levels of learners. Finally, a detailed cost estimate is not available for the SSBE.

Our study adds to the body of literature on the validity of simulation technologies for assessment in medical education.^{19,20,21} In a recent systematic review of technology-enhanced simulation to assess health professionals, Cook et al. found that validity evidence for these types of assessments is limited; most of the evidence is focused on specific tools for procedural skills, and lies within certain medical specialties.²² Furthermore, the use of simulated technologies for student assessment appears to be even more limited; out of 417 studies, only 115 involved medical students (N=3,197).²² In addition, the research methodology and reporting outcomes for the majority of the studies reviewed needed improvement.²² No additional studies were found related to the use of simulated technologies for assessment at the clerkship level.

In conclusion, we propose that simulated technologies, when bundled as an SSBE, may offer a novel, standardized method of assessing competency in patient care. The examination format was well received by the students and is feasible with current simulation technologies. There remains a need for more collaborative research on assessment tools that rise to meet the challenge of determining expertise in medicine.⁴ We are seeking to expand, refine, and validate the SSBE for competency-based assessment of patient care, an area currently lacking reliable, valid and feasible assessment tools.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Declaration of Interest section:

Alan Dow has been supported by a faculty fellowship grant from the Josiah H. Macy, Jr Foundation and by grants from the Donald W. Reynolds Foundation and Health Resources and Services Administration (UD7HP26044A0). The contents are solely the responsibility of the authors and do not necessarily represent official views of the Josiah H. Macy Jr. Foundation, the Donald W. Reynolds Foundation, or the Health Resources and Services Administration.

Moshe Feldman received financial support for the project from a Virginia Commonwealth University (VCU) Clinical and Translational Research Award **Number UL1TR000058 from the NCRR**.

References

1. Langan TS, Rigby IJ, Walker IW, Howes D, Donnon T, Lord JA. Simulation-based training in critical resuscitation procedures improves residents' competence. *CJEM*. 2009; 11(6):535–539. [PubMed: 19922713]
2. Tervo RC, Dimitrievich E, Trujillo AJ, Whittle K, Redinius P, Wellman L. The Objective Structured Clinical Examination (OSCE) in the clinical clerkship: an overview. *S D J Med*. 1997 May; 50(5): 153–156. [PubMed: 9155233]
3. Green ML, Aagaard EM, Caverzagie KJ, Chick DA, Holmboe E, Kane G, Smith CD, Lobst W. Charting the road to competence: developmental milestones for internal medicine residency training. *JGME*. 2009 Sep; 1(1):5–20.
4. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach*. 2010; 32(8):676–682. [PubMed: 20662580]
5. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ*. 1979 Jan; 13(1):41–54. PMID:763183. [PubMed: 763183]
6. Patricio MF, Juliao M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Medical Teacher*. 35:503–514. [PubMed: 23521582]
7. Herbers JE Jr, Noel GL, Cooper G, Harvey J, Pangaro LN, Weaver MJ. How accurate are faculty evaluations of clinical competence? *Journal of General Internal Medicine*. 1989; 4(3):202. [PubMed: 2723833]
8. Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. *Acad Med*. 2009; 84(3):301–309. [PubMed: 19240434]
9. Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. Effect of rater training on reliability and accuracy of mini-CEX scores: A randomized, controlled trial. *Journal of Internal Medicine*. 2008; 24(1):874–879.
10. Feldman M, Lazzara EH, Vanderbilt AA, Diaz Granados D. Rater training to support high-stakes simulation-based assessments. *Journal of Continuing Education in the Health Professions*. 2012; 32(4):279–286. [PubMed: 23280532]
11. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA*. 2009 Sep 23; 302(12):1316–1326. [PubMed: 19773567]
12. Berwick DM, Finkelstein JA. Preparing medical students for the continual improvement of health and health care: Abraham Flexner and the new "public interest". *Acad Med*. 2010 Sep; 85(9 Suppl):S56–S65.
13. Miller GE. The Assessment of Clinical Skills/Competence. *Performance Acad Med*. 1990; 65(9): 63–67.
14. Pangaro L, Cate OT. Frameworks for learner assessment in medicine: AMEE Guide No.78. *Medical Teacher*. 2013; 35:e1197–e1210. Tables. [PubMed: 23676179]
15. Downing SM. Validity on meaningful interpretation of assessment data. *Medical Education*. 2003; 37:830–837. [PubMed: 14506816]
16. Harris PA, Taylor R, Thielke JP, Gonzalez N, Conde JG. Research electronic capture (REDCap) – A metadata driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009 Apr; 42(2):377–381. [PubMed: 18929686]
17. Hsieh HF, Shannon SE. Three Approaches to Qualitative Content Analysis. *Qual Health Res*. 2005; 15:1277. [PubMed: 16204405]
18. van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ*. 1996; 1(1):41–67.
19. Andreatta PB, Gruppen LD. Conceptualizing and classifying validity evidence for simulation. *Medical Education*. 2009; 43(11):1028–1035. [PubMed: 19874494]

20. Scalese RJ, Obeso VT, Issenberg SB. Simulation Technology for Skills Training and Competency Assessment in Medical Education. *JGIM*. 2007; 23(1):46–49. [PubMed: 18095044]
21. McGaghie W. Use of Simulation to assess competence and improve healthcare. *Journal of the International Association of Medical Sciences Educators*. Downloaded from http://www.iamse.org/artman/publish/printer_622.shtml.
22. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-Enhanced Simulation to Assess Health Professionals: A systematic Review of Validity Evidence, Research methods, and Reporting Quality. *Academic Medicine*. 2013; 88(6):872–883. [PubMed: 23619073]

Table 1

Simulator Operation: IStan

Case 1: Asthmatic	
State 1. Asthma Baseline; HR 100, BP110/75, SpO2 95%, RR 20, Temp 98.6 (37C), Lung sounds-wheezes bilaterally, EKG rhythm- normal sinus, vocal sounds "I can't breathe"	Trigger=Go to state 2 for question 4
State 2. Asthma Q4: HR 130, BP100/60, SpO2 89%, RR35	End after question 6
Case 2: Acute Coronary Syndrome	
State 1. ACS Baseline: HR 95, BP 150/100, SpO2 98%, RR 17, Temp 98.7 (37.1C), lung sounds- crackles (all fields), EKG rhythm- normal sinus, vocal sounds- "my chest hurts".	Trigger =go to state 2 for question 2
State 2: ACS Q2: HR 90, BP 145/85, SpO2 93%, RR26, EKG rhythm- PVC's, heart sounds- pan systolic murmur, vocal sounds -none.	Trigger=go to state 3 for question 4
State 3: ACS Q4: HR: 70, BP 105/50, SpO2=89%, RR 40, EKG rhythm-PVC's transition to VFib after 10 sec	End after question 5
Case 3: Pneumonia with Septic Shock	
State 1. Baseline pneumonia: HR 147, BP 90/50, SpO2 90%, RR 33, Temp 103.6 (39.8C), lung sounds- crackles on the right, decreased sounds on the left, EKG rhythm-normal sinus	Trigger=go to state 2 for question 4
State 2. Pneumonia Q4: HR 137, BP 80/40, SpO2 90%, RR31, other –cyanosis of fingers on both hands	End after question 5

TABLE 2

Summary of SSBE Scores

	Mean	Median	Mode	SD
SSBE Competency Score	76.7%	76.4%	76.1%	7.8
Component I: Medical knowledge	75.1%	76.0%	75.0%	8.4
Component II: Cardiovascular knowledge and diagnostic reasoning (Harvey®)	75.3%	83.3%	83.3%	18.3
Component III: clinical knowledge and decision making (iStan®)	79.5%*	81.3%	81.3%	11.5

TABLE 3

Descriptive statistics and correlations among study variables

	N	Mean	Median	SSBE	Photo quiz	Harvey	iStan	eQuiz	OSCE	Ward Score	Shelf	Final Grade	Step 2
SSBE	207	76.69	76.41	-									
Photoquiz	204	75.15	76.00	.76**	-								
Harvey@	206	75.32	83.33	.57**	.26**	-							
iStan@	207	79.50	81.25	.75**	.27**	.19**	-						
eQuiz	205	73.40	74.00	.33**	.26**	.16*	.24**	-					
OSCE	165	85.06	85.67	.05	-.01	.12	.00	.29**	-				
Ward score	165	8.34	8.44	.31**	.25**	.16*	.21**	.20**	.05	-			
Shelf Examination Score	160	79.09	80.00	.52**	.47**	.26**	.32**	.43**	.27**	.47**	-		
Final Grade	160	2.47-	3.00	-.45**	-.43**	-.32**	-.21**	-.36**	-.26**	-.47**	-.71**	-	
Step 2 USMLE Score	78	241.37	242.00	.54**	.43**	.37**	.30**	.48**	.37*	.18	.67**	-.55**	-

Note.

* = $p < .05$.** = $p < .01$.

TABLE 4

Mean SSBE Score and Final Grade in the IM Clerkship

Final Grade	N	SSBE (Mean)
Honors	31	81.31
High Pass	44	78.74
Pass	82	74.70
Marginal	3	69.95
Incomplete	5	64

TABLE 5
Logistic Regression Analysis of SSBE Components as predictors of Final Grade in the IM Clerkship

Variable	B	se(b)	Wald	OR estimate	p
OSCE	.15	.04	12.63	1.163	.00
Ward score	2.9	.55	27.47	18.01	.00
eQuiz (%)	.04	.03	2.46	1.04	.12
SSBE(%)	.07	.03	6.23	1.08	.01

TABLE 6

Linear Regression Analysis of SSBE Components as predictors of Shelf Examination Score in the IM Clerkship

Variable	<i>B</i>	SE	<i>t</i>	<i>p</i>
Ward score	.29	1.00	4.64	.00
OSCE	.17	.10	2.83	.01
eQuiz (%)	.21	.07	3.29	.00
SSBE (%)	.35	.07	5.37	.00

Note. N= 160. Adjusted R^2 = .45

TABLE 7

Linear Regression Analysis of SSBE Components as predictors of Step 2 USMLE Scores

Variable	<i>B</i>	SE	<i>t</i>	<i>p</i>
eQuiz (%)	.32	.22	3.26	.00
SSBE (%)	.42	.24	4. 4.34	.00

Note. N = 78. Adjusted R^2 = .37.

Table 8

Percentage agreement with evaluation survey items.

Evaluation Survey Item	% Agreement
The exam content was appropriate.	91%
The use of simulation was a better test of my knowledge compared to using no simulation for this exam.	89%
The amount of content covered was appropriate.	96%
I felt well prepared for this exam.	89%
I felt comfortable working with the human patient mannequins/part task trainers.	87%
The Harvey simulator is a good tool for testing my knowledge of cardiac murmurs.	95%
The iStan simulator is a good tool for testing my knowledge of disease states.	81%
Photo quiz questions are better for testing image recognition compared to paper based quiz questions.	89%
Photo quiz questions are better for testing physical exam findings compared to paper based quiz questions.	79%
I feel the IM Clerkship prepared me for the Harvey component of this exam.	98%
I feel the IM Clerkship prepared me for the iStan component of this exam.	80%
I feel the IM Clerkship prepared me for the photo quiz component of this exam.	38%
This exam contributed to my professional development.	87%
Overall, I thought this was a fair exam.	93%

N = 182.