Jefferson.

Thomas Jefferson University
Jefferson Digital Commons

Department of Surgery Faculty Papers | Department of Surgery

6-1-2011

Evaluation of the use of patient-focused simulation for student assessment in a surgery clerkship.

Gerald A Isenberg
Thomas Jefferson University, gerald.isenberg@jefferson.edu

Katherine W Berg
Thomas Jefferson University, katherine.berg@jefferson.edu

Judith A Veloski
Thomas Jefferson University, judith.veloski@jefferson.edu

Dale D Berg
Thomas Jefferson University, dale.berg@jefferson.edu

J J. Veloski
Thomas Jefferson University, jon.veloski@jefferson.edu

See next page for additional authors

Let us know how access to this document benefits you

Follow this and additional works at: http://jdc.jefferson.edu/surgeryfp

Part of the Surgery Commons

Recommended Citation

Isenberg, Gerald A; Berg, Katherine W; Veloski, Judith A; Berg, Dale D; Veloski, J J.; and Yeo, Charles, "Evaluation of the use of patient-focused simulation for student assessment in a surgery clerkship." (2011). *Department of Surgery Faculty Papers.* Paper 41.
http://jdc.jefferson.edu/surgeryfp/41

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's Center for Teaching and Learning (CTL). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Department of Surgery Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: JeffersonDigitalCommons@jefferson.edu.

Authors
Gerald A Isenberg, Katherine W Berg, Judith A Veloski, Dale D Berg, J J. Veloski, and Charles Yeo

This article is available at Jefferson Digital Commons: http://jdc.jefferson.edu/surgeryfp/41

Accepted for publication: *The American Journal of Surgery* February 16, 2010

Published as:

Evaluation of the Use of Patient-focused Simulation for Student Assessment in a Surgery Clerkship

Volume 201, Issue 6, June 2011, Pages 835-840

DOI: 10.1016/j.amjsurg.2010.01.034

Gerald A. Isenberg, M.D., Katherine W. Berg, M.D., Judith A. Veloski, R.N., B.S.,

Dale D. Berg, M.D., J. Jon Veloski, M.S., Charles J. Yeo, M.D.

Jefferson Medical College of Thomas Jefferson University
1025 Walnut Street
Philadelphia, Pennsylvania 19107

Address Reprint Requests to:

Gerald A. Isenberg, M.D.
Associate Professor

Department of Surgery
Jefferson Medical College
Thomas Jefferson University
1025 Walnut Street
Philadelphia, PA 19107

Email: Gerald.Isenberg@jefferson.edu
Business Telephone: 215-955-6879
Fax: 215-955-2404

Corresponding Author

Jon Veloski
Jon.Veloski@jefferson.edu
215-955-7901

Acknowledgment

As Director of the Surgery Clerkship until the fall of 2007, Philip J. Wolfson, M.D. led the planning, development and implementation of the surgical clinical skills assessment described in this paper. Were it not for his untimely accidental death in the fall of 2007, Dr. Wolfson would have been the lead author of this study, which he conceived and nurtured.

The methods and preliminary results of the first year of the study were presented at the annual meeting of the Association of Standardized Patient Educators, July 2007, Toronto, Canada.

Key words

Clerkship; Simulation; Clinical Skills; Medical Students

Summary

Simulated patients and simulation devices were used to test the clinical skills of 670 students at the end of their core surgery clerkship over three years. Scores for procedures and communications skills were reliable and valid in relation to other measures of performance. It is feasible to combine simulated patients with simulation devices to strengthen the validity of assessments of students' clinical skills.

Abstract

Purpose: The purpose of this study was to evaluate the use of simulated patients in conjunction with anatomical and tissue task-training models to assess skills.

Methods: Faculty reviewed the objectives of the clerkship to identify skills to be acquired. Three cases were developed related to rectal examination, suturing and inserting intravenous lines and nasogastric tubes. Student scores were based on their ability to gather data from simulated patients and perform procedures on simulation models.

Results: A total of 670 students were assessed between 2006 and 2009. Alpha reliability coefficients were 0.97 for Communication/ Interpersonal Skills, 0.71 for Procedures and 0.58 for Data Gathering. Students receiving low ratings from faculty in the clerkship had significantly($p<0.001$) lower simulation scores. There were significant($p<0.001$) relationships between scores and grades in other clerkships.

Conclusions: The combination of simulated patients and simulation models yielded reliable scores for procedural and interpersonal skills, and evidence of validity related to clinical ratings.

Introduction

National surveys indicate that formal clinical skills assessments, such as objective structured clinical examinations (OSCE), are being used for student assessment and evaluation of the clinical curriculum in the majority of medical schools.[1, 2] Although many schools are developing comprehensive assessments to evaluate the clinical curriculum and assure that students are well-prepared for the clinical skills component of the United States Medical Licensing Examination (USMLE Step 2 CS), many others have used this type of performance-based testing within specific clerkships such as family medicine[3, 4] and obstetrics/gynecology.[5]

A 2007 survey reporting the most important characteristics of surgery clerkships in North America revealed that 38% of US schools and 69% of Canadian schools used OSCEs for student assessment.[6] Although there have been numerous published reports of the use of simulated patients to assess surgical residents,[7-10] and medical students in surgical clerkships,[11-13] there has been less attention directed toward the assessment of trainees' surgical skills, and the related development of OSCE stations that combine both simulated patients and simulation models. [14-17] Naylor et al recently reported on the use of simulation models to assess skills proficiency,[18] and LeBlanc et al reported on a study of 16 medical students and residents in a validation of an integrated assessment of third-year students' technical and communication skills.[19]

The goal of this study was to develop an OSCE by using a combination of simulated patients in conjunction with anatomical and tissue task-training simulation models to assess medical students' clinical skills at the end of a required surgery clerkship. In particular, we wanted to focus on the measurement of students' proficiency in performing common procedures that were to be learned during the clerkship. We planned to examine the reliability of these measurements, and to examine their construct validity in relation to students' clinical performance in surgery

and in other clerkships, and in relation to National Board of Medical Examiners' (NBME) subject examination scores.

Methods

Participants

Participants included 670 third-year students at a large private medical school in the U.S. The Institutional Review Board of the university determined that this retrospective use of routine evaluation data collected as part of the educational program was exempt from human subjects review.

Study design

The data were collected as part of the required surgery clerkship rotation completed by third-year medical students during the 2006-07, 2007-08 and 2008-09 academic years. Before 2006-2007 the routine assessments in the surgery clerkship had already included global clinical performance ratings by faculty, and use of the NBME Subject Examination in Surgery. However, there was a consensus that the overall validity of the assessment of students' competence could be further strengthened by adding a more formal clinical skills assessment at the end of the clerkship. Since the objectives of the clerkship included the development of specific technical skills such as inserting intravenous lines, a decision was made to use simulated patients in conjunction with anatomical and tissue simulation models. It has been reported that this combination of patients and simulation models, referred to as patient-focused simulation, can trigger authentic responses from trainees on a level that computers or models alone are unable to achieve, and thus strengthen the validity of these tests.[17]

Approximately six months before the beginning of the study in 2006, members of the surgery faculty reviewed the educational objectives and required clinical experiences during the surgery clerkship to identify the skills that would be most appropriate for assessment in a formal clinical skills examination.[20, 21] The objectives of the clerkship required that students develop proficiency in a number of specific technical procedures. Three cases were developed related to the diagnosis of gastrointestinal bleeding, the acute management of a forearm laceration, and the evaluation and acute management of abdominal pain. Each case required that students gather clinical data from the simulated patient and then perform procedures including rectal examination, IV insertion, nasogastric tube insertion, urinary catheter insertion and suturing using anatomical and tissue simulation models.

During the orientation sessions to the clerkship over the three years the students were given formal instruction on the performance of technical procedures, and their performance was monitored throughout the clerkship using a computer-based tracking system in which students logged their experience with each procedure and transferred data into a central monitoring system.[22] The cases required that students perform rectal examinations using a Rectal Examination Model (manufactured by Life/form), insert IVs using a Multi-venous IV Training Arm (manufactured by Laerdal), insert nasogastric tubes using an NG Tube and Tracheostomy Care Simulator (manufactured by Laerdal), insert a urinary catheter using the Advanced Catheterization Trainer (manufactured by Limbs and Things) and perform suturing using a Suture Pad (manufactured by Limbs and Things).

A total of 15 simulated patients were trained to set up the simulation models, present the cases, and complete the checklists and rating scales to assess the students' data gathering skills, proficiency in performing the selected procedures and ability to communicate effectively. The

simulated patients were also trained to be able to perform the procedures themselves, so that they would be able to understand how to evaluate the students' performance. One of the authors (KWB) evaluated each simulated patient's proficiency in performing each of the selected procedures to confirm that they would be able to accurately assess the students' proficiency.

At the end of each teaching block across the academic year the three cases were administered to approximately 25 students in two, one-half day testing sessions. The encounters were recorded for quality assurance and feedback to students using two video cameras and a microphone in each room. To assure that the students were familiar with the models they were allowed to practice performing procedures on the models shortly before the OSCE.

Performance measures

The following three scores were calculated for each student: Data Gathering, Procedural Skills, and Communication/Interpersonal Skills. A Data Gathering score was calculated from the checklists as the percentage of history items obtained from the patient and the percentage of physical examination maneuvers performed correctly in each case. The Procedural Skills scores were calculated based on checklists of the critical steps required to perform each procedure successfully.[23] For example, the suturing case included specific items such as "Inspected site of the laceration," "Cleaned wound with betadine," "Put on sterile gloves," "Injected local anesthetic," and "Held and used needle driver correctly." Table 1 lists the checklist items that were used to assess the students' ability to insert a nasogastric tube. Communication/ Interpersonal Skills Scores were calculated using Likert scale weights yielding scores on a 0 to 100 scale developed for the medical school's annual comprehensive clinical skills assessment.[24] Other available measures of students' performance included final clinical grades in the surgery

clerkship, and five other core clerkships in family medicine, internal medicine, obstetrics/gynecology, pediatrics and psychiatry, based on global ratings assigned by faculty members at the end of each clerkship. The four-point grading scale in descending order includes Honors, Excellent, Good and Fail.

Statistical analyses

Item analysis procedures were used to evaluate the psychometric performance of individual checklist items. Reliability was computed using Cronbach's Coefficient Alpha. Product-moment correlations were calculated to describe relationships between scores. Construct validity was determined based on analysis of variance and t-tests at alpha=0.01 for differences in mean scores on the three components of the OSCE among levels of the clinical performance ratings that faculty assign to students during clerkship rotations. Calculations were performed using Stata 10.0 (College Station, TX).

Results

A total of 670 students completed the surgery OSCE during the three academic years. The means and standard deviations of their scores for Data Gathering, Procedural Skills and Communication/ Interpersonal Skills were 78.1 (11.7), 87.9 (8.6), and 78.5 (8.6), respectively. There was no significant difference in the students' mean scores on the three scales across the three academic years and across the blocks of the academic years.

Reliability

Cronbach's Coefficient Alpha is an index of the precision and reproducibility of test scores. Values range from 0 to 1.0, with high values indicating that students would be expected to obtain

similar scores on repeated test administrations. Values close to 0 indicate that the scores are not reproducible, and for practical purposes random. The values of Alpha for the surgery OSCE ranged from 0.58 for Data Gathering, to 0.71 for Procedures and 0.97 for Communication/ Interpersonal Skills.

Construct validity

The construct validity of an instrument is judged by examining the empirical relationships between its scores and other external variables that are logically related to the construct the instrument purports to measure. For example, if the surgery OSCE is designed to measure students' clinical competence one would expect to find positive relationships between scores on the OSCE and faculty ratings of the students' clinical performance during the Surgery clerkship.

Table 2 shows the students' mean scores on the surgery OSCE for groups of students based on their overall clinical grade in the surgery clerkship, which is derived from faculty ratings of their clinical performance during the rotation. The value of Wilks' Lambda ($0.96, p<0.001$) from multivariate analysis of variance indicated significant overall differences for the mean OSCE scores across the three levels of clinical grades in the surgery clerkship. The results of t-tests of the differences between the mean OSCE scores for the group with the highest grade of Honors and those with more average grades of Excellent were not significant. However, the t-tests indicate that the differences between the group with the low clinical performance rating of Good and the group with the rating of Excellent were significant.

In order to judge construct validity, one would also expect find positive relationships between performance in the surgery OSCE and clinical performance in other clerkships. Table 3 shows the relationship between the number of low clinical grades received by students in the five other

core clerkships (Family Medicine, Internal Medicine, Obstetrics/Gynecology, Pediatrics and Psychiatry) and their scores on the surgery OSCE. Previous research has shown that the number of low clinical performance ratings a student receives across clerkships is an important predictor of subsequent weak clinical performance in residency, whereas the number of Excellent or Honors ratings across clerkships is not.[25] The results indicate that the scores on Data Gathering and Communication/Interpersonal Skills in the surgery OSCE were associated with performance ratings in the other clerkships. However, as one might expect there was no relationship between low ratings in the other clerkships and Procedural Skills scores on the surgery OSCE.

Finally, in order to judge construct validity, one would also examine relationships between the surgery OSCE scores and scores on the comprehensive OSCE administered to all students at the end of the third year curriculum and scores on the NBME Subject Examination in Surgery. Table 4 shows that the highest correlation of 0.28 is between Surgery Communication/ Interpersonal Skills and the comparable score on the annual comprehensive third-year OSCE. There is also a high correlation of 0.21 between the Data Gathering scores on the two assessments. At the other extreme, the lowest correlations between 0.09 and 0.11 are found between the Procedures score on the surgery OSCE and the other measures. There are also low correlations ranging from 0.10 to 0.12 between the three surgery OSCE scores and the Documentation score on the comprehensive OSCE.

Comments

The purpose of this study was to evaluate the routine use of simulated patients in conjunction with anatomical and tissue task-training simulation models to assess medical students' clinical skills at the end of a required surgery clerkship. While there have been multiple published

reports of the use of OSCEs in surgery clerkships, there has been limited published evidence of the effectiveness of OSCEs using a combination of simulated patients and anatomical and tissue simulation models.

Previous studies of OSCEs for surgical residents have underscored the importance of the fundamental criterion of reliability for these types of examinations to confirm that the scores are precise and reproducible.[7] We found extremely high reliability for Communication/ Interpersonal Skills scores based on a five-point rating scale. We found acceptable reliability for Procedural Skills scores based on only 3 cases. The lower reliability coefficient for Data Gathering scores is consistent with the uncertainty surrounding the quality of the checklists developed for measuring history and physical examination skills in these types of tests, and the difficulty in applying these checklists objectively and reproducibly.[24, 26-30]

We found small positive and statistically-significant relationships between performance in the surgery OSCE and faculty ratings of performance in the surgery clerkship. These findings are consistent with an early study based on five years of data that underscored the complementary nature of clerkship ratings and SP clinical skill assessments.[31] Clerkship ratings are based on multiple observations of students' knowledge, skills and personal characteristics over a period of six weeks, but limited by a lack of objective metrics and interpersonal relationships. The OSCE, while designed to be a reliable measure of defined set of skills in a controlled and potentially objective setting, is limited to just a few hours on a single day at the end of the six-week clerkship.

We also found modest, but statistically-significant correlations between the surgery OSCE scores for Data Gathering and Communication/Interpersonal Skills and comparable scores on the

comprehensive OSCE administered at the end of the third year curriculum. This is not surprising since the content of the comprehensive third-year OSCE spans multiple clerkships across the third year. However, it appears that scores for Communication/Interpersonal Skills may have the greatest predictive validity in relation to performance in other clerkships.

Our goal of administering the OSCE every six weeks imposed several practical constraints and required efficient use of time and resources. First, we used a small number of cases in order to be able to test about 30 students per block in one or two days during a six week cycle. The decision to use only three cases was supported in part by the findings of Nackman et al,[16] who studied the effect of the use of a human clinical simulator in a surgery clerkship and reported significant effects using an OSCE with only three cases. However, if this type of clinical skills assessment is to be used for grading individual students, then a larger of number of cases will be needed in order to cover the broad content of surgery and to achieve acceptable reliability.

A second decision related to resources was the decision to repeat the same three cases each block in order to minimize training costs and simplify the equating of students' scores across blocks. Two concerns are that students may divulge the content of the exam to their classmates or that they may acquire these skills in other clerkships. Consistent with the findings of Niehaus who studied this issue over a decade ago, we found no significant difference in scores across the year when the same three cases were repeated every block suggesting that neither concern is an issue.[32]

A third consideration was the decision to train the simulated patients rather than faculty observers to fill out the checklists used to assess the students' skills in performing the procedures on the anatomical and tissue simulators. While some may argue that faculty members are better

able to evaluate students' procedural skills, questions have been raised about the risk of bias when faculty members are asked to objectively rate the performance of students they have encountered in the clinical setting.[33] It is important to emphasize that although the simulated patients had been trained to perform each of the clinical tasks they were observing the students perform, they were not rating the quality of the students' performance. They were instructed to indicate whether or not each of the discrete steps of the procedures had been performed properly, i.e., they answered yes/no questions and did not rate proficiency. One recent study in emergency medicine demonstrated good inter-rater reliability between expert and first and second year college students trained to rate physicians' competence in procedural skills.[34]

When used in the literature the term simulator can refer to the entire spectrum of simulation devices ranging from anatomical and tissue simulators, through electromechanical simulators to high fidelity computer-based mannequins.[35, 36] Although the present study used simple anatomical and tissue models, we believe that the methods developed in this study and our findings in conjunction with simulated patients may be transferable to more elaborate simulations that have been, and continue to be evaluated experimentally in surgery clerkships.[16]

The results of this three-year study support the feasibility of combining simulated patients with anatomical and tissue simulation models to produce a clinical skills assessment with greater validity than assessment using simulated patients alone. This combination is particularly relevant in surgery where the objectives of the clinical clerkship include procedural skills that cannot be assessed using simulated patients alone. Relationships between the OSCE scores and faculty ratings of students' clerkship performance in the surgery clerkship as well as other clerkships in the third year provided evidence of the OSCE's construct validity. Three cases yielded acceptable reliability for the measurement of Procedural Skills and very high reliability

for Communication/Interpersonal Skills. The reliability of the Data Gathering scores was too low to be used for decisions on individual examinees. A larger number of cases is needed to strengthen the reliability of Procedural Skills and to produce reliable Data Gathering scores that can be used for grading individual students.

References

1. Hauer KE, Hodgson, CS, Kerr, KM, Teherani, A, and Irby, DM. A national study of medical student clinical skills assessment. Acad Med. 2005;80:S25-S29.

2. Hauer KE, Teherani, A, Kerr, KM, O'Sullivan, PS, and Irby, DM. Impact of the United States Medical Licensing Examination Step 2 Clinical Skills exam on medical school clinical assessment. Acad Med. 2007;81:s13-s16.

3. Prislin MD, Fitzpatrick, CF, Lie, D, Giglio, M, Radecki, S, and Lewis, E. Use of an objective structured clinical examination in evaluating student performance. Fam Med. 1998;30:338-44.

4. Ferrell B and Thompson, BL. A clinical performance assessment for a third-year required clerkship in family medicine. Fam Med. 1993;25:262-3.

5. Bienstock JL, Martin, S, Tzou, W, and Fox, HE. Medical students' gender is a predictor of success in the obstetrics and gynecology basic clerkship. Teaching and Learning in Medicine. 2002;14:240-3.

6. Lind DS, Deladisma, AM, Cue, JI, Thomax, AM, MacFadyen, BV, and Nesbit, RR. Survey of student education in surgery. J Am Coll Surgeons. 2007;204:969-76.

7. Cohen R, Reznick, RK, Taylor, BR, Provan, J, and Rothman, A. Reliability and validity of the objective structured clinical examination in assessing surgical residents. Am J Surg. 1990;160:302-5.

8. Sloan DA, Donnelly, MB, Schwartz, RW, and Strodel, WE. The Objective Structured Clinical Examination: The new gold standard for evaluating postgraduate clinical performance. Ann Surg. 1995;222:735-42.

9. Martin JA, Regehr, G, Reznick, RK, MacRae, HM, Murnaghan, J, Hutchinson, C, and Brown, M. Objective structured assessment of technical skills (OSATS) for surgical residents. British Journal of Surgery. 1997;84:273-8.

10. MacRae HM, Regehr, G, Leadbetter, W, and Reznick, RK. A comprehensive examination for senior surgical residents. Am J Surg. 2000;179:190-3.

11. Merrick HW, Nowacek, G, Boyer, J, and Robertson, J. Comparison of the objective structured clinical examination with the performance of third-year medical students in surgery. Am J Surg. 2000;179:286-8.

12. Merrick HW, Nowacek, GA, Boyer, J, Padgett, B, Francis, P, Gohara, SF, and Staren., ED. Ability of the Objective Structured Clinical Examination to differentiate surgical residents, medical students, and physician assistant students. J Surg Res. 2002;106:319-22.

13. Sachdeva AK, Wolfson, PJ, Blair, PG, Gillum, DR, Gracely, EJ, and Friedman, M. Impact of a standardized patient intervention to teach breast and abdominal examination skills to third-year medical students at two institutions. Am J Surg. 1997;173:320-5.

14. Brunt LM, Halpin, VJ, Klingensmith, ME, Tiemann, D, Matthews, BD, Spitler, JA, and Pierce, RA. Accelerated skills preparation and assessment for senior medical students entering surgical internship. J Am Coll Surgeons. 2008;206:897-907.

15. Engum SA, Jeffries, P, and Fisher, L. Intravenous catheter training system: computer-based education versus traditional learning methods. Am J Surg. 2003;186:67-74.

16. Nackman GB, Bermann, M, and Hammond, J. Effective use of human simulators in surgical education. J Surg Res 2003;115:214-8.

17. Kneebone R, Nestel, D, Wetzel, C, Black, S, Jacklin, R, Aggarwal, J, Yadollahi, F, Wolfe, J, Vincent, C, and Darzi, A. The human face of simulation: patient-focused simulation training. Acad Med. 2006;81:919-24.

18. Naylor RA, Hollett, LA, Valentine, RJ, Mitchell, IC, Bowling, MW, Ma, AM, Dinneen, SP, Bruns, BR, and Scott, DJ. Can medical students achieve skills proficiency through simulation training? Am J Surg 2009;2009.

19. LeBlanc V, Tabak, D, Kneebone, R, Nestel, D, MacRae, HM, and Moulton, C. Psychometric properties of an integrated assessment of technical and communications skills. Am J Surg 2008;197:101.

20. Committe on Testing and Evaluation, Association for Surgical Education. The Objective Structured Clinical Examination. 2001.

21. Cerilli GJ, Merrick, HW, and Staren, ED. Surgical educator preferences regarding key Objective Structured Clinical Examination topics. J Surg Res 2001;101:124-9.

22. Rattner SL, Louis, DZ, Rabinowitz, C, Gottlieb, J, Nasca, TJ, Markham, FW, Gottlieb, RP, Caruso, JW, Lane, JL, Veloski, JJ, Hojat, M, and Gonnella, JS. A new system for documenting and comparing medical students' clinical experiences. JAMA. 2001;286:1035-40.

23. Marbas, L. and Case, E. Blueprints: Clinical Procedures. New York: Blackwell Publishing, 2004.

24. Worzala K, Rattner, SL, Boulet, JR, Majdan, JF, Berg, D, Robeson, MR, and Veloski, JJ. Evaluation of the congruence between students' post-encounter notes and standardized patients' checklists in a clinical skills examination. Teaching and Learning in Medicine. 2008;20:31-6.

25. Herman MW, Veloski, JJ, and Hojat, M. Validity and importance of low ratings given to medical school graduates in noncognitive areas. J Med Educ. 1983;58:837-43.

26. Hodges B, Regehr, G, McNaughton, N, Tiberius, R, and Hanson, M. OSCE checklists do not capture increasing levels of expertise. Acad Med. 1999;74:1129-34.

27. Gorter S, Rethans, J, Scherpbier, A, van der Heijde, D, van der Vleuten, C, and van der Linden, S. Developing case-specific checklists for standardized-patient-based assessments in internal medicine: A review of the literature. Acad Med. 2000;75:1130-7.

28. Clauser BE, De Champlain, AF, Swartz, MH, and Colliver, JA. Further discussion of SP checklists and videotaped performances. Acad Med. 2000;75:315-6.

29. De Champlain AF. Further discussion of SP checklists and videotaped performances. Acad Med. 2000;75:316-7.

30. Swartz MH and Colliver, JA. Further discussion of SP checklists and videotape performances. Acad Med. 2000;75:317-8.

31. Vu NV. Consistency of pass-fail decisions made with clinical clerkship ratings and standardized patient examination scores. Acad Med. 1994;69:S40.

32. Niehaus AH. Is test security a concern when OSCE stations are repeated across clerkship rotations? Acad Med. 1996;71:287.

33. Martin JA, Reznick, RK, Rothman, A, Tamblyn, R, and Regehr, G. Who should rate candidates in an objective structured clinical examination? Acad Med. 1996;71:170-5.

34. Bullock G, Kovacs, G, Macdonald, K, and Storey, BA. Evaluating procedural skills competence: inter-rater reliability of expert and non-expert observers. Acad Med. 1999;74:76-8.

35. Issenberg SB, McGaghie, WC, Hart, IR, Mayer, JW, Felner, JM, Petrusa, ER, Waugh, RA, Brown, DD, Safford, RR, Gessner, IH, Gordon, DL, and Ewy, GA. Simulation technology for health care professional skills training and assessment. JAMA. 1999;282:861-6.

36. Issenberg SB, McGaghie, WC, Petrusa, ER, Gordon, DL, and Scalese, RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Medical Teacher. 2005;27:10-28.

Table 1.- Examples of Checklist Items Used in Clinical Skills Assessment - Insertion of Nasogastric Tube

1. Washed hands before beginning procedure.

2. Estimated the length of NG tube that would be going into the patient.

3. Applied lubricant to the tube.

4. Inserted the tube via the nares, and asked the patient to swallow as the tube was being advanced.

5. Checked final placement of the tube by injecting air with a syringe and listening over the stomach with a stethoscope.

6. Secured the tube to the nose with tape.

Table 2.- Mean Scores on Surgery Clinical Skills Assessment within Groups Based on Students' Overall Clinical Grade in Surgery Clerkship

Overall Clinical Grade in Surgery Clerkship	n	Surgery Clinical Skills Assessment: Data Gathering	Procedures	Communication/ Interpersonal Skills
Honors	197	79.1	89.1	79.6
Excellent	393	78.5	87.8	78.8
Good	80	74.6	86.1	74.3
Fail	0			

Clinical grades are based on global ratings of Honors, Excellent, Good, or Fail assigned to each student by faculty in the surgery clerkship. Other grades based on test scores on the National Board of Medical Examiners' Subject Examination in Surgery are reported separately from each student's clinical grade.

Wilks' Lambda=0.96, F=4.73, $p<0.001$

The results of t-tests for differences between the mean for the Honors group versus the mean for the Excellent group for each score were not significant.

The results of t-tests for differences between the mean for the Excellent group versus the mean for the Good group were significant for Data Gathering ($p<0.01$) and for Communication/Interpersonal Skills ($p<0.001$), but not for Procedures.

Table 3.- Mean Scores on Surgery Clinical Skills Assessment within Groups Based on the Number of Low Clinical Grades Students Received in Five Other Core Clerkships

Number of Low Clinical Grades	n	Surgery Clinical Skills Assessment: Data Gathering	Procedures	Communication/ Interpersonal Skills
0	377	79.6	88.3	79.9
1	164	77.7	88.2	79.0
2	57	77.0	87.7	76.6
3 - 5	72	73.6	86.1	71.4
p-value		0.001	0.23	0.001

Clinical grades in each core clerkship are based on faculty global ratings of Honors, Excellent, Good, or Fail for each student based on their clinical performance.

The Number of Low Clinical Grades is the number of Good or Fail ratings accumulated by a student in the Family Medicine, Internal Medicine, Obstetrics/Gynecology, Pediatrics and Psychiatry clerkships.

The p-values are based on analysis of variance of scores among the four levels of Number of Low Ratings.

Table 4.- Product-moment Correlations for 670 Students Scores on Surgery Clinical Skills Assessment, NBME Subject Examination and Annual Third-year Clinical Skills Assessment

Surgery Clinical Skills Assessment	NBME Subject Exam	Annual Third-year Clinical Skills Assessment		
		Data Gathering	Communication/ Interpersonal Skills	Documentation
Data Gathering	0.16	0.21	0.13	0.12
Procedures	0.11	0.09	0.10	0.10
Communication/ Interpersonal Skills	0.15	0.18	0.28	0.12