The American Journal of Surgery (2015) 210, 603-609

The American Journal of Surgery®

Association of Women Surgeons

Development and evaluation of a simulation-based continuing medical education course: beyond lectures and credit hours

Carla M. Pugh, M.D., Ph.D.[a,*], Fahd O. Arafat, M.D.[b], Calvin Kwan, B.S.[a], Elaine R. Cohen, M.Ed.[a], Yo Kurashima, M.D.[c,d], Melina C. Vassiliou, M.D., M.Ed.[c,d], Gerald M. Fried, M.D.[c,d]

[a]Department of Surgery, University of Wisconsin School of Medicine and Public Health, 600 Highland Avenue, Madison, WI 53792, USA; [b]Department of Surgery, Northwestern University, Chicago, IL, USA; [c]The Department of Surgery, Steinberg-Bernstein Centre for Minimally Invasive Surgery & Innovation, McGill University, Montréal, Quebec, Canada; [d]The Arnold and Blema Steinberg Medical Simulation Centre, McGill University, Montréal, Quebec, Canada

KEYWORDS:
Simulation;
Validity;
Training;
CME;
Evaluation;
Laparoscopy

Abstract

BACKGROUND: The aim of our study was to modify our previously developed laparoscopic ventral hernia (LVH) simulator to increase difficulty and then reassess validity and feasibility for using the simulator in a newly developed simulation-based continuing medical education course.

METHODS: Participants (N = 30) were practicing surgeons who signed up for a hands-on postgraduate laparoscopic hernia course. An LVH simulator, with prior validity evidence, was modified for the course to increase difficulty. Participants completed 1 of the 3 variations in hernia anatomy: incarcerated omentum, incarcerated bowel, and diffuse adhesions. During the procedure, course faculty and peer observers rated surgeon performance using Global Operative Assessment of Laparoscopic Skills–Incisional Hernia and Global Operative Assessment of Laparoscopic Skills rating scales with prior validity evidence. Rating scale reliability was reassessed for internal consistency. Peer and faculty raters' scores were compared. In addition, quality and completeness of the hernia repairs were rated.

RESULTS: Internal consistency on the general skills performance (peer $\alpha = .96$, faculty $\alpha = .94$) and procedure-specific performance (peer $\alpha = .91$, faculty $\alpha = .88$) scores were high. Peers were more lenient than faculty raters on all LVH items in both the procedure-specific skills and general skills ratings. Overall, participants scored poorly on the quality and completeness of their hernia repairs (mean = 3.90/16, standard deviation = 2.72), suggesting a mismatch between course attendees and hernia difficulty and identifying a learning need.

CONCLUSIONS: Simulation-based continuing medical education courses provide hands-on experiences that can positively affect clinical practice. Although our data appear to show a significant

The study was funded by the Department of Defense USAMRMC grant W81XWH0710190. The Department of Defense had no involvement in interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

* Corresponding author. Tel.: +1-608-265-9574; fax: +1-608-252-0936.

E-mail address: pugh@surgery.wisc.edu

Manuscript received February 18, 2015; revised manuscript May 15, 2015

http://dx.doi.org/10.1016/j.amjsurg.2015.05.034

mismatch between clinical skill and simulator difficulty, these findings also underscore significant learning needs in the surgical community.

A commitment to continuous learning and practice improvement is essential in everyday clinical practice. However, the traditional implementation of continuing medical education (CME) allows course attendance and course credits to overshadow measured achievement and internal motivation for excellence. Continuing medical education is part of a rapidly changing system of professional development that includes assessment, remediation, and reassessment. Although jurisdictional requirements may vary from state to state, most simply require a number of hours of CME to be completed annually to satisfy licensure needs.[1] Currently, the larger focus for CME is the hours needed to maintain licensure and certification. However, the motivation behind the original development of CME programs was to inspire lifelong learning. With the increasing emphasis on quality in health care, CME has great potential to move from a basic focus on maintenance of licensure to improving quality in clinical practice and ensuring ongoing physician competence.

Although simulation has been widely accepted as a training and assessment modality in graduate medical education, use in CME is not as common. In addition, the external drivers for objective assessment of clinical skill during residency training are on the rise. Many of the residency review programs are requiring documentation of annual evaluations of skill in a hands-on setting, away from direct patient care.[2] These evaluations include assessments of knowledge, skills, and attitudes. However, after completion of training, these types of assessments are sparse. The exception is the newly implemented maintenance of certification program in the United States.[3] These programs, as defined by specialty board organizations, are bringing focus to the need for ongoing assessment after residency training.

In the medical literature, there is a paucity of research on the use of simulation-based technology in CME courses.[4,5] One study used a proficiency-based curriculum as part of a CME course. This group demonstrated the feasibility of a half-day CME course to improve performance in laparoscopic suturing.[6] The American Board of Anesthesia has also made some headway in developing simulation-based assessments for maintenance of certification. Overall, research in this field is largely based on self-report satisfaction data and lacks any standardized performance evaluation.[7]

The American College of Surgeons supports the use of simulation-based surgical education to enhance patient safety, meet the requirements for maintenance of certification and address the core competencies that all surgeons and trainees are required to achieve.[8] Despite the American College of Surgeons' support, development, implementation, and evaluation of simulation-based CME courses are lacking. To achieve these goals, valid and reliable measures of performance are necessary.

Our prior work using the laparoscopic ventral hernia simulator revealed the importance of intraoperative decision making for this procedure. This work underscored the need for decision-based metrics in addition to those used to assess technical skills.[9] The aim of our present study was to modify our previous laparoscopic ventral hernia simulator to increase difficulty and then reassess validity and feasibility for use in a CME course. Specifically, we sought to assess the following: (1) the validity and reliability of a previously developed procedure-specific (Global Operative Assessment of Laparoscopic Skills–Incisional Hernia [GOALS-IH]) rating scale and global (Generic GOALS) rating scale largely used for graduate medical education and (2) to document validity support for simulated laparoscopic ventral hernia scenarios of increasing difficulty.

Methods

Setting and participants

This study was performed at the 97th Clinical Congress in San Francisco, California in 2011. The Clinical Congress is designed to provide individuals with a wide range of learning opportunities, activities, and experiences that will match their educational and professional development needs. Practicing surgeons attending the conference, who signed up for the hernia course, served as participants. Minimally invasive surgery–trained surgeons served as faculty raters. Participant data were collected over a 1 day period. The Northwestern University Institutional Review Board approved the study, and all participants provided informed consent.

Protocol

This was an evaluation study to assess the feasibility of creating a simulation-based CME course for practicing surgeons performing laparoscopic ventral hernia repairs. Course objectives included demonstrating (1) proper port placement strategies; (2) efficient and strategic adhesiolysis; and (3) effective mesh management.

The course was limited to 30 participants. Before beginning, participants were randomly placed into 10 groups of 3. Five groups worked simultaneously during the 1st half of the course. The 2nd cohort completed the task during the 2nd half of the course. Each group was assigned to 1 of the 5 faculty raters. While 1 participant performed a laparoscopic ventral hernia repair on 1 of the 3 variations in hernia anatomy (incarcerated omentum, incarcerated bowel, and diffuse adhesions), another participant navigated the camera, and the other participant served as a peer-rater. Participants rotated until they played all 3 roles, and the 3 variations of the hernia repairs were performed (Fig. 1). Individual hernia skins were retained for postuse evaluation.

Each participant was allotted 30 minutes to complete the repair. Throughout the simulated procedure, the faculty rater and participant serving as the peer-rater completed the GOALS–IH module and GOALS rating scale assessments to evaluate performance. After the conference, a blinded minimally invasive surgery–trained faculty member rated the quality and level of completeness of the hernia repairs based on the skin used by each participant.

Materials

The laparoscopic ventral hernia simulator consists of a plastic box trainer base covered by a fabric skin that contains a ventral hernia defect, Fig. 2. The box trainer is intended to mimic the abdominal cavity extending from the diaphragm to the upper plane of the pelvic cavity. The abdominal cavity is layered with various simulated organs to mimic the gross anatomy of the abdominal cavity, including the peritoneum, mesentery, and bowel. In our 1st study, participants rated the model as having similar fidelity and utility to an animal model for hernia repair.[9]

Although the base configuration of the laparoscopic ventral hernia simulator was used and showed evidence of validity in another study,[9] for this project, we modified the pre-existing simulator to include 2 additional variations in hernia adhesions (Fig. 3). The incarcerated omentum configuration had been previously used in a course evaluating chief residents.[8] This hernia type was fabricated using loosely woven cotton with simulated blood vessels. The omentum was attached to the hernia sac using fabric glue. The 2 additional variations included diffuse adhesions and incarcerated bowel. The diffuse adhesion simulator was fabricated using self-adherent translucent material to simulate the adhesions. This material was placed over the anterior abdominal peritoneum and omentum within the hernia sac, thus creating a combination of filmy and dense adhesions. Finally, the incarcerated bowel hernia type was fabricated using polar fleece fabric for the bowel and mesentery. The bowel was attached to the hernia sac using Krazy Glue (Westerville, OH), which resulted in a dense, firm adhesion to the hernia sac. Although the hernia adhesion types varied, the size and location of the hernia defect was standard across all users. In addition, once the materials and fabrication procedures were finalized, the models for this study were mass produced by a commercial vendor, Busy Bee Enterprises, (Yorba Linda, CA).

Based on our prior experience with the omentum adhesion model, we became aware that the difficulty of this type of adhesion could be increased by modifying the type of glue or adhesive used to attach the omentum inside the hernia defect. In essence, double-sided tape allowed for an easy adhesiolysis, whereas Krazy Glue made for a more a more difficult adhesiolysis. Despite the use of Krazy Glue in this model, we hypothesized that the omentum variation would be less difficult than the incarcerated bowel model. Likewise, we hypothesized that the incarcerated bowel model would be less difficult than the diffuse adhesion model. The evidence that we sought in this project to assess our hypothesis was the amount of time spent during the adhesiolysis and the performance ratings on the GOALS–IH module and GOALS rating scale assessments.

Each simulator was prepared as if the patient was prepped, draped, and preinsufflated. Participants were given a standard laparoscopic instrument set that included a 30° rigid laparoscope (KARL STORZ, El Segundo, CA), VersaStep trocars (Covidien, Norwalk, CT), open abdomen needle drivers and hemostats, Maryland dissectors (KARL STORZ), Endo Shears (Covidien), ProTack (Covidien), Endo Graspers (Covidien), Carter-Thomason suture passer (Inlet Medical, Eden Prairie, MN), and Parietex mesh (Covidien). Operative tasks included port placement, dissection of the adhesions, mesh preparation, and fixation for completion of the hernia repair.

Outcome measures

Faculty and peer-raters assessed the hernia repairs using 2 rating scales with prior validity evidence, The GOALS–IH is a 7-item global rating scale developed by experts to evaluate the steps of LIHR (placement of trocars, adhesiolysis, estimation of mesh size and shape,

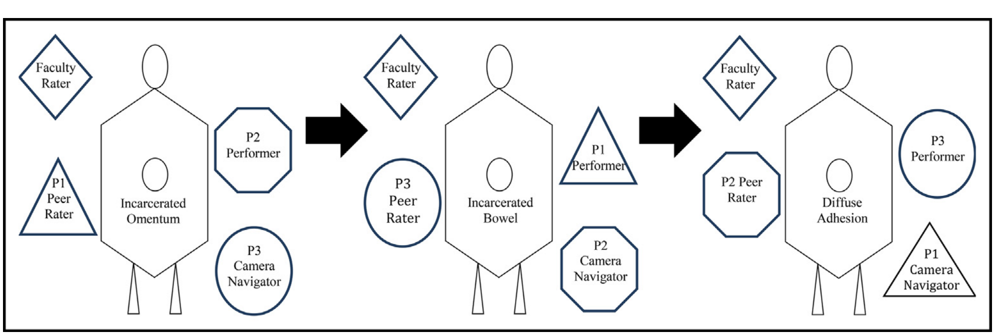

Figure 1 Station role rotation of participants.

Figure 2 Participant repairing hernia defect using LVH box trainer.

mesh orientation and positioning, mesh fixation, knowledge and autonomy in use of instruments, overall competence), each rated on a 5-point Likert scale.[9] The generic GOALS checklist was also used.[10] This checklist evaluates general laparoscopic performance in 5 domains (depth perception, bimanual dexterity, efficiency, tissue handling, and autonomy).These items are also rated on a 5-point Likert scale: 1 (poor), 3 (average), 5 (excellent), with descriptors to anchor ratings 1, 3, and 5.

Participants' laparoscopic ventral hernia simulator skins were analyzed and individually graded in 4 major areas; suturing, tacking, ports, and mesh placement. For suturing, the score was based on the number of sutures placed and tied and on the location of suture placement with respect to the edge of the mesh (maximum 6 points possible). For tacking, the score was based on the number of quadrants completed and the location of tack placement with respect to the edge of the mesh (maximum 5 points possible). For port placement, participants were scored on the absolute number of ports (<3 ports $= 0$, $\geq 3 = 1$) and whether there was triangulation (yes $= 1$, no $= 0$; maximum 2 points). For mesh placement, the score was based on whether the mesh was skewed or flat with respect to the hernia defect (2 points). Finally, participants were graded on whether they cut the peritoneum during the lysis of adhesions (1 point). Participants could score up to 16 points depending on the quality and completeness of their hernia repair.

Data analysis

GOALS–IH module and GOALS rating scale reliabilities were assessed for internal consistency using Cronbach's alpha. Peer and faculty raters' item scores for each scale were assessed separately with their own reliability estimate. Peer and faculty raters' scores were compared using a paired samples *t* test. Correlations between peer and faculty scores were also assessed. Hernia repair scores were averaged and later stratified based on adhesion type. Stratification results were used to understand the effects of simulator difficulty on hernia repair scores. Data analysis was performed using SPSS, version 20 (IBM Corp, Armonk, NY).

Results

Internal consistency on the general skills performance and procedure-specific performance for GOALS–IH (peer $\alpha = .91$, faculty $\alpha = .88$) and GOALS (peer $\alpha = .96$, faculty $\alpha = .94$) was very high. Peers were more lenient than faculty raters on all items in both the procedure

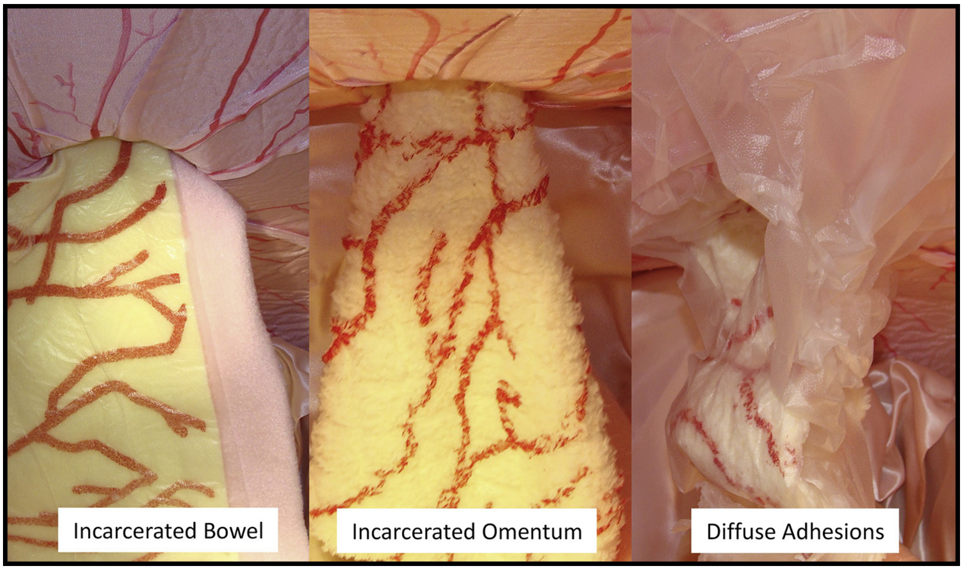

Figure 3 Three variations of hernia repairs: incarcerated omentum, incarcerated bowel, and diffuse adhesions.

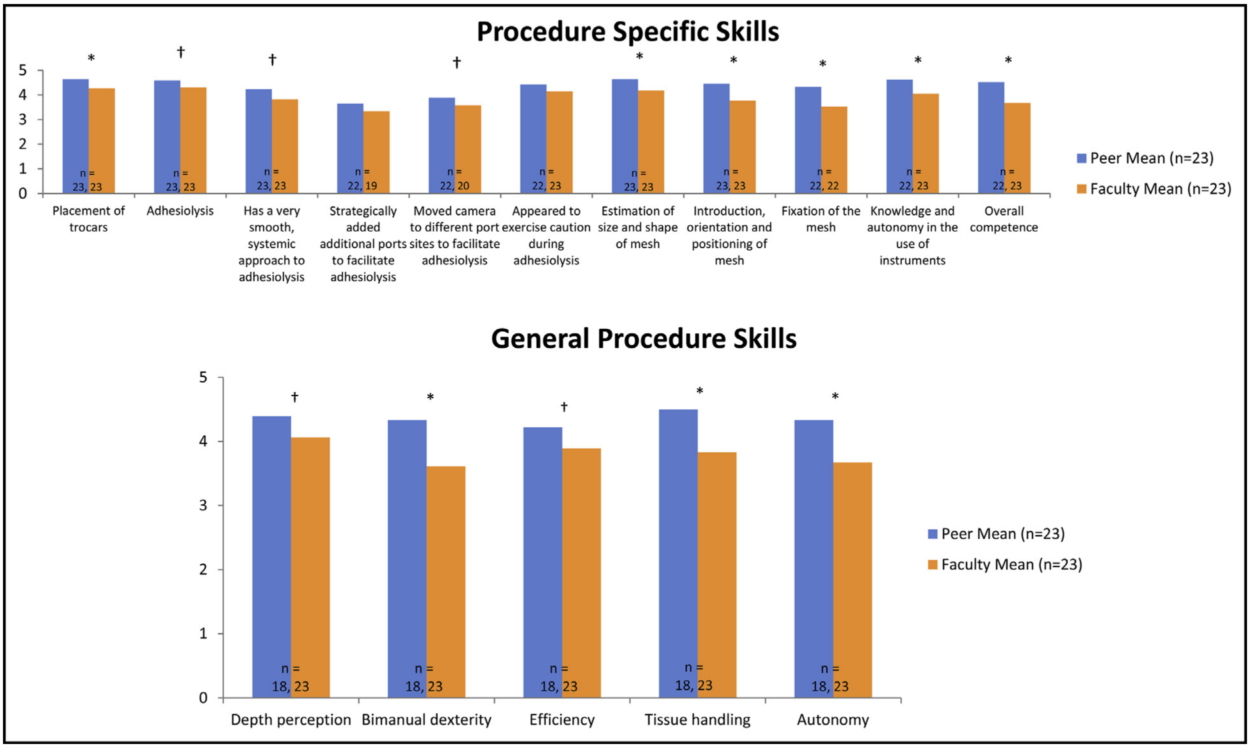

Figure 4 Comparison of peer and faculty ratings. $^*p < .05$; $^\dagger p < .01$.

specific skills and general procedure skills ratings (Fig. 4). In addition, correlations were generally low between the peer and faculty scores with the exception of a few moderate correlations.

Overall, out of a possible 16 points, participants scored poorly on the quality and completeness of their hernia repairs (mean = 3.90, standard deviation [SD] =2.72). Key performance areas where most participants lost points are as follows: (1) 70% of pzarticipants failed to triangulate their ports; (2) 39% of participants spent most of the time dissecting adhesions and never progressed to the mesh preparation step in the allotted time. For GOALS grading and participant evaluation purposes, these persons were allowed to execute the last three steps of the procedure at the end of the learning experience.

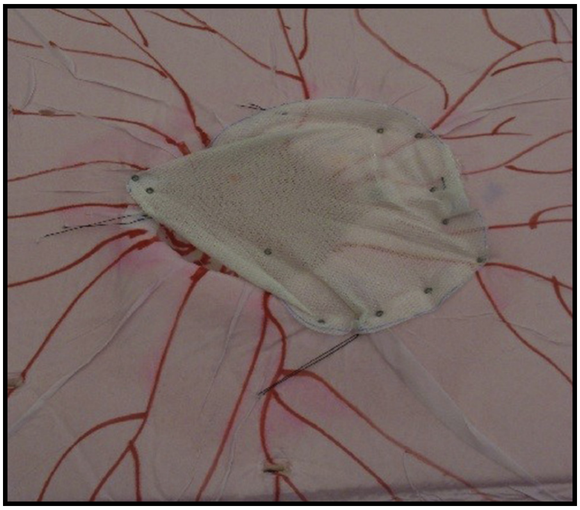

Figure 5 Skewed mesh with poor suture placement.

Of those who progressed to the mesh preparation step in the allotted time, 43% failed to conduct proper planning for the mesh landing zone (Fig. 5). Participants repairing the diffuse adhesions had the most difficult time completing the task (56% failed to reach the mesh preparation step). When comparing scores between hernia repair types, participants scored 4.21 (SD = 3.18) on the incarcerated omentum, 3.81 (SD = 2.77) on the incarcerated bowel, and 3.66 (SD = 2.38) on the diffuse adhesions. Although there is a trend toward lower scores with increased adhesion complexity, this was not statistically significant.

Comments

Simulation-based CME courses have great potential to change the traditional lecture-based implementation of continuing education for practicing physicians. The purpose of this study was to modify our laparoscopic ventral hernia simulator to increase difficulty and reassess validity and feasibility of a newly developed CME course for practicing surgeons. In this study, we also used 2 rating scales (GOALS–IH module and GOALS), both of which have shown validity evidence in prior work. As part of our study, we reassessed the validity and reliability of the GOALS–IH module and GOALS rating scales for a CME course as prior use has largely focused on residents.[10,11] Our results show that these rating scales can serve as valid and reliable measures for CME courses. Faculty and peer raters were able to use the surveys for CME purposes. This provides validity support for use of the GOALS ratings scales in a different setting and with a variety of simulation scenarios. Additional validity support was evidenced by the more stringent ratings by faculty raters on the rating scale assessments. Finally, despite the differences in context and raters, the GOALS rating scales maintained reliability levels as previously reported.[10,11] These findings may help facilitate the integration of simulation into clinical curricula that include certification and recertification[4] without the need for development of new rating scales. The differences in peer and faculty evaluation warrant further investigation whether these reflect undue rater stringency or lack of insight among the participants about what constitutes good quality performance. This has implications for self-assessment and for feedback.

Most of our faculty raters were minimally invasive surgery–trained surgeons with extensive experience in advanced laparoscopic procedures and in training residents and fellows in laparoscopic hernia surgery. This may have created high expectations for performance of practicing surgeons and some bias toward lower ratings. For the peer evaluators, there is the possibility that they were not sufficiently knowledgeable about what constitutes a high-quality repair for these more difficult simulated scenarios. Laparoscopic adhesiolysis requires good planning and a set of psychomotor skills and visual perception that can add significant difficulty to a simple hernia repair. This makes a procedure with multiple steps and a variety of surgical instruments and materials sufficiently complex. Future work is needed to understand whether future courses should incorporate more gradual increases in the level of difficulty and to better understand whether raters' expectations were excessively high for the learners involved in such courses.

Our original laparoscopic ventral hernia simulator was shown to have validity evidence in a prior study that used the simulator for 4th and 5th year surgical residents.[9] For the current project, we wanted to increase the difficulty of the scenarios with the assumption that the CME course attendees would have more laparoscopic experience. In our evaluation of simulation difficulty, we noted that the surgeons who had the diffuse type of adhesions spent more time and were at greater risk of not finishing the procedure. Although the simulation was more difficult, it is duly noted that the surgeons spent extra time carefully dissecting the adhesions. As such, future use of this model should allow for more time to complete the entire procedure or have course objectives that are less dependent on task completion rates. In addition, future choice of the adhesion variations should be based on better assessment of the experience level of course attendees. Despite the mismatch between simulator difficulty and course participant skill, most course participants reported a high level of satisfaction with this course. Moreover, although our data appear to show a significant mismatch between clinical skill and simulator difficulty, these findings also underscore significant learning needs in the surgical community.

The implications of our study relate to the potential for simulation to be used for basic and advanced CME courses. The first step in course development is highly dependent on learner characteristics. Once this is determined, the learning objectives can be more clearly aligned. Use of off the shelf and locally fabricated simulations can greatly facilitate achievement of a wide variety of learning objectives. Although we are in the beginning phases of exploring all the possibilities, there is clear support for use of simulation as a teaching and assessment modality for practicing surgeons. Our evaluation of this simulation-based CME course provided useful feedback regarding use of rating scales, fabrication of operative scenarios, and the range of skill levels in a small cohort of practicing surgeons. This study added validity evidence to the current simulation-based assessment by showing that simulator difficulty can be modified with sufficient realism for different skill levels. In addition, the GOALS–IH and GOALS rating scales were adopted for use in a CME course. Finally, the rating scale results confirmed simulator difficulty.

CME has resurfaced in its level of importance, as maintenance of certification has become a new requirement for most clinical specialties. Maintenance of certification involves 4 specific requirement areas designed to assess physician competencies. These include maintenance of an unrestricted license, hospital privileges, and satisfactory references; continuing education and periodic self-assessment; satisfactory performance on a secure, standardized examination; and participation in outcome registries and quality improvement programs.[12] If designed correctly, simulation-based scenarios can be implemented as a form of continuing education, self-assessment, and standardized examination.

This study has several limitations. First, it was a small group of surgeons at a single course. However, the multiple sources of evaluation helped to facilitate an objective course evaluation and performance assessment. Simulator design had an effect on completion times. The design of adhesions involved the exploration and investigation of a number of different adhesives. The one chosen as our final product turned out to be quite strong and more difficult to dissect. We did not assess the level of difficulty of the newly developed adhesions before the session. As some adhesions were quite difficult, additional time may have allowed more users to complete the hernia repair. In addition, more time might have yielded a more complete assessment of each user's ability to perform a complete hernia repair. As case difficulty affects performance, specific measures (beyond previous number of cases and years in practice) should be included in course planning and possibly judged based on a structured, detailed preassessment of the learner and learner needs. In addition, 39% of participants spent most of the time dissecting adhesions and failed to progress to the mesh preparation step in the allotted time. For GOALS grading and participant evaluation purposes, these persons were allowed to execute the last 3 steps of the procedure at the end of the learning experience. As a result, the last 3 procedure specific items include peer and faculty ratings under "second chance" circumstances and fall out of the original testing and evaluation time. Despite this nuance, the trend in faculty and peer rating remained the same.

Overall the course appeared to be a success. Course instructors and participants were engaged in the tasks, and there were high-level discussions regarding operative approaches and choices. Our findings show promise for the use of previously developed graduate medical education assessments in CME courses. In addition, we were able to easily modify a previously developed simulator to achieve significant variations in scenario difficulty. Although the most effective delivery of advanced CME courses for surgeons is still unknown, it is clear that there is a need to promote continued development and evaluation of a variety of learning and assessment modalities and approaches. As the use of simulation continues to increase, we predict this will change the focus and delivery of CME courses in the future. Achieving this goal will help to move CME beyond a time and credit-based requirement to an experience that motivates ongoing physician competence and improves quality in clinical practice.

References

1. Miller SH, Thompson JN, Mazmanian PE, et al. Continuing medical education, professional development, and requirements for medical licensure: a white paper of the conjoint committee on continuing medical education. J Contin Educ Health Prof 2008; 28:95–8.
2. Accreditation Council for Graduate Medical Education (ACGRADUATE MEDICAL EDUCATION) Program Requirements for Surgery. https://www.acGraduateMedicalEducation.org/acGraduateMedicalEducationweb/tabid/150/ProgramandInstitutionalAccreditation/SurgicalSpecialties/Surgery.aspx. Accessed April 9, 2015.
3. Spivey BE. Continuing medical education in the United States: why it needs reform and how we propose to accomplish it. J Contin Educ Health Prof 2005;25:134–43.
4. Scott DJ. Patient safety, competency, and the future of surgical simulation. Simul Healthc 2006;1:164–70.
5. McGaghie WC, Siddall VJ, Mazmanian PE, et al, American College of Chest Physicians Health and Science Policy Committee. Lessons for continuing medical education from simulation research in undergraduate and graduate medical education: effectiveness of continuing medical education: American College of Chest Physicians Evidence-Based Educational Guidelines. Chest 2009;135(3 Suppl):62S–8S.
6. Stefanidis D, Sierra R, Korndorffer Jr JR, et al. Intensive continuing medical education course training on simulators results in proficiency for laparoscopic suturing. Am J Surg 2006;191:23–7.
7. McIvor W, Burden A, Weinger MB, et al. Simulation for maintenance of certification in anesthesiology: the first two years. J Contin Educ Health Prof 2012;32:236–42.
8. Sachdeva AK, Pellegrini CA, Johnson KA. Support for simulation-based surgical education through American College of Surgeons–accredited education institutes. World J Surg 2008; 32:196–207.
9. Pugh C, Plachta S, Auyang E, et al. Outcome measures for surgical simulators: is the focus on technical skills the best approach? Surgery 2010;147:646–54.
10. Ghaderi I, Vaillancourt M, Sroka G, et al. Evaluation of surgical performance during laparoscopic incisional hernia repair: a multicenter study. Surg Endosc 2011;25:2555–63.
11. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005;190:107–13.
12. The American Board of Surgery. Maintenance of Certification (MOC) Overview. http://www.absurgery.org/default.jsp?exam-moc. Accessed March 15, 2013.