The American Journal of Surgery (2012) 203, 54–62

The American Journal of Surgery®

The Association for Surgical Education

Construct and face validity of the American College of Surgeons/Association of Program Directors in Surgery laparoscopic troubleshooting team training exercise

Nabeel A. Arain, M.D., M.B.A.[a], Deborah C. Hogg, B.S.[a], Rajiv B. Gala, M.D.[b], Ravi Bhoja, M.D.[c], Seifu T. Tesfay, R.N., M.S.[a], Erin M. Webb, B.S.[a], Daniel J. Scott, M.D.[a,*]

[a]Department of Surgery, [b]Department of Gynecology, and [c]Department of Anesthesiology, Southwestern Center for Minimally Invasive Surgery, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390-9156, USA

KEYWORDS:
Team training;
Laparoscopic troubleshooting;
Communication;
Patient safety;
Virtual operating room (VOR);
Simulation;
Crew resource management (CRM)

Abstract

BACKGROUND: Our aim was to develop an objective scoring system and evaluate construct and face validity for a laparoscopic troubleshooting team training exercise.

METHODS: Surgery and gynecology novices (n = 14) and experts (n = 10) participated. Assessments included the following: time-out, scenario decision making (SDM) score (based on essential treatments rendered and completion time), operating room communication assessment (investigator developed), line operations safety audits (teamwork), and National Aeronautics and Space Administration–Task Load Index (workload).

RESULTS: Significant differences were detected for SDM scores for scenarios 1 (192 vs 278; P = .01) and 3 (129 vs 225; P = .004), operating room communication assessment (67 vs 91; P = .002), and line operations safety audits (58 vs 87; P = .001), but not for time-out (46 vs 51) or scenario 2 SDM score (301 vs 322). Workload was similar for both groups and face validity (8.8 on a 10-point scale) was strongly supported.

CONCLUSIONS: Objective decision-making scoring for 2 of 3 scenarios and communication and teamwork ratings showed construct validity. Face validity and participant feedback were excellent.

The Accreditation Council for Graduate Medical Education has established 6 core competencies, including patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice.[1] Traditional didactic teaching methods address many of these areas and simulation increasingly has been used to teach technical skills. Indeed, best practices for task-based technical skills training using simulators have evolved over the past 2 decades.[2] However, by comparison there has been relatively little work in the surgical simulation arena designed to teach or assess interpersonal and communication skills and professionalism.

The American College of Surgeons (ACS) along with the Association of Program Directors in Surgery (APDS) have established one of the most robust simulation libraries currently available.[3] Phase III of the ACS/APDS Surgical Skills

Presented at the Association for Surgical Education (ASE) Meeting, 2011 Surgical Education Week, March 22–26, Boston, MA.
* Corresponding author. Tel.: +1-214-648-2677; fax: +1-214-648-2301.
E-mail address: Daniel.Scott@UTSouthwestern.edu
Manuscript received April 18, 2011; revised manuscript August 6, 2011

0002-9610/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.amjsurg.2011.08.010

Figure 1 VOR showing the surgeon trainee along with OR personnel during performance of a simulated laparoscopic cholecystectomy.

Curriculum for residents includes team training curricula that specifically aim to address such nontechnical skills. National initiatives also have been promoted through grant support from the Agency for Healthcare Research and Quality (AHRQ), which is part of the US government's Department of Health and Human Services. In recent years, the AHRQ sponsored investigations in hopes of enhancing patient safety through team training simulations. As part of an AHRQ grant–funded project, our research team developed a team training exercise that was designed to teach decision making and communication skills related to laparoscopic physiology and equipment troubleshooting. This module was included as part of the ACS/APDS national curriculum.[4]

Despite our intensive earlier work, our ability to assess performance in the team training environment was limited. We therefore embarked on developing and refining such methods. We had several aims for this study. First, we wished to develop an objective scoring system for our exercise, whereby decision making regarding best practices and patient safety could be measured. Second, we aimed to determine construct validity of this objective scoring system, as well as other new and established rating tools for communication, teamwork, and workload. Third, we wanted to evaluate face validity and participant impressions regarding this module. For these purposes, construct validity was defined as the ability of a test to measure the trait that it purports to measure, as commonly established by detecting statistically significant performance differences between groups of varying experience.[5,6] Face validity was defined as the extent to which this simulation exercise was a good method for learning and assessing technical, communication, and teamwork skills, as well as patient safety practices.

Methods

This study was approved by the University of Texas Southwestern Medical Center's Institutional Review Board and was a collaborative project between investigators from the departments of surgery, gynecology, and anesthesiology conducted in a virtual operating room (VOR) at the Southwestern Center for Minimally Invasive Surgery. The VOR used in this study was an animal OR that was modified accordingly to accommodate our team training exercises (Fig. 1). Equipment included an anesthesia mannequin (SimMan; Laerdal Medical AS, Stavanger, Norway), a box-trainer with porcine explant specimens, and actual laparoscopic instrumentation.

Team training protocol

We enrolled participants from general surgery and gynecology consisting of novices (postgraduate year [PGY] 1–2, n = 14) and experts (faculty, n = 10). A surgeon-only model was selected for this study, whereby each team consisted of one surgeon participant along with 3 role-playing facilitators (confederates) including an anesthesiologist, a scrub technician who served as the camera driver and first assistant, and a circulating nurse. Before the study, the confederates were trained to facilitate the scenarios in a standardized fashion according to predetermined choreography. For general surgery, the surgical model was cholecystectomy (liver and gallbladder explants) and for gynecology the model was tubal ligation (uterus/uterine horn explants).

Each team participated in a single uninterrupted iteration of the laparoscopic troubleshooting module, as described in detail elsewhere.[4] Participants were oriented to the VOR set-up and provided standardized instructions regarding the training exercise, as well as a scripted introduction regarding the patient's history and medical data. The exercise (Video, available online) included a time-out (Table 1) and 3 scenarios (Figs. 2 and 3).

The time-out has been debated extensively within the surgical community and presents an area of much controversy in that a unanimous consensus is lacking regarding

Table 1 Time-out scoring

	Criteria met	
	Yes	No
1. Held before starting the surgery		
2. Patient name		
3. Type of surgery		
4. Antibiotics administered before incision		
5. Appropriate antibiotic selection confirmed		
6. SCDs in place		
7. SCDs activated before induction		
8. Patient's medical history and anticipated problems discussed		
9. Specific surgical plan, duration, equipment discussed		

Score = (number of criteria met/9) × 100.
SCD = sequential compression device.

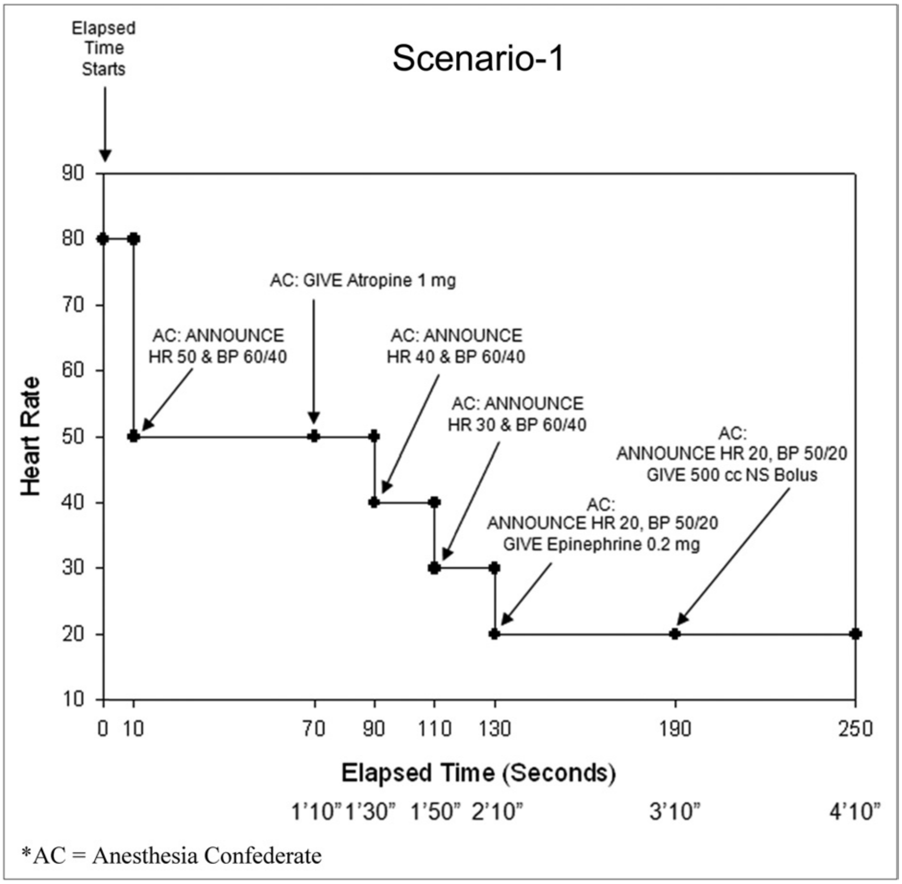


Figure 2 Simulation scenario 1 (bradycardia/hypotension) showing a strategic timeline of simulated physiologic events, anesthesia confederate (AC)-administered treatments, and AC announcements at certain intervals.

what elements should be included in the time-out.[7–10] Based on a review of the literature and consensus among our multidisciplinary research team, we considered 9 elements as being appropriate for inclusion (Table 1). Because our patient simulator did not include legs, we opted to incorporate sequential compression devices being in place and activated before induction as part of our patient stem; thus, these elements were not evaluated.

In scenario 1 (Fig. 2) the patient becomes bradycardic and hypotensive, and in scenario 3 (Fig. 3) the patient becomes hypercarbic with increased peak inspiratory pressures. For both of these physiologic simulations, the participant is expected to render 4 mandatory treatments (Table 2) in a timely fashion; at predetermined intervals, the anesthesia confederate administers 3 of these treatments if the participant has not done so by a predetermined point. However, only the surgeon participant is allowed to desufflate the abdomen. As soon as all 4 treatments are rendered, the scenario is completed and the next scenario is started. If the 4 treatments are not rendered, then the next scenario is started after the cut-off time (4 min and 10 s) is reached.

In scenario 2, the image on the monitor suddenly becomes quite dark, such that the surgery may not continue. The participant is expected to systematically troubleshoot the laparoscopic equipment and render 4 mandatory actions (Table 2). Unlike scenarios 1 and 3, the confederates are not allowed to provide solutions if the participant is unable to do so; if the participant does not render the 4 mandatory actions within the cut-off time (4 min and 10 s), then the next scenario is started.

Although in previous iterations of this module we had used a lengthy debriefing component, in this study we used a truncated debriefing after each simulation, given our primary objective of investigating the time-out and scenario components; accordingly, the debriefing was not evaluated as part of this study.

Each exercise lasted approximately 30 minutes. In addition to the confederates, a single designated simulation director (D.J.S. for surgery and R.B.G. for gynecology) rated each participant's performance by direct observation. A faculty anesthesiologist performed real-time physiologic manipulations using the SimMan simulator. Often, an additional member of the research team was available to assist with conducting the exercise. These personnel were located in the control room during the exercise.

Performance evaluation

For the time-out component, we developed an objective scoring formula by giving credit for the percentage of criteria met (Table 1). For the 3 scenarios, we developed a slightly more complex objective scoring formula based on

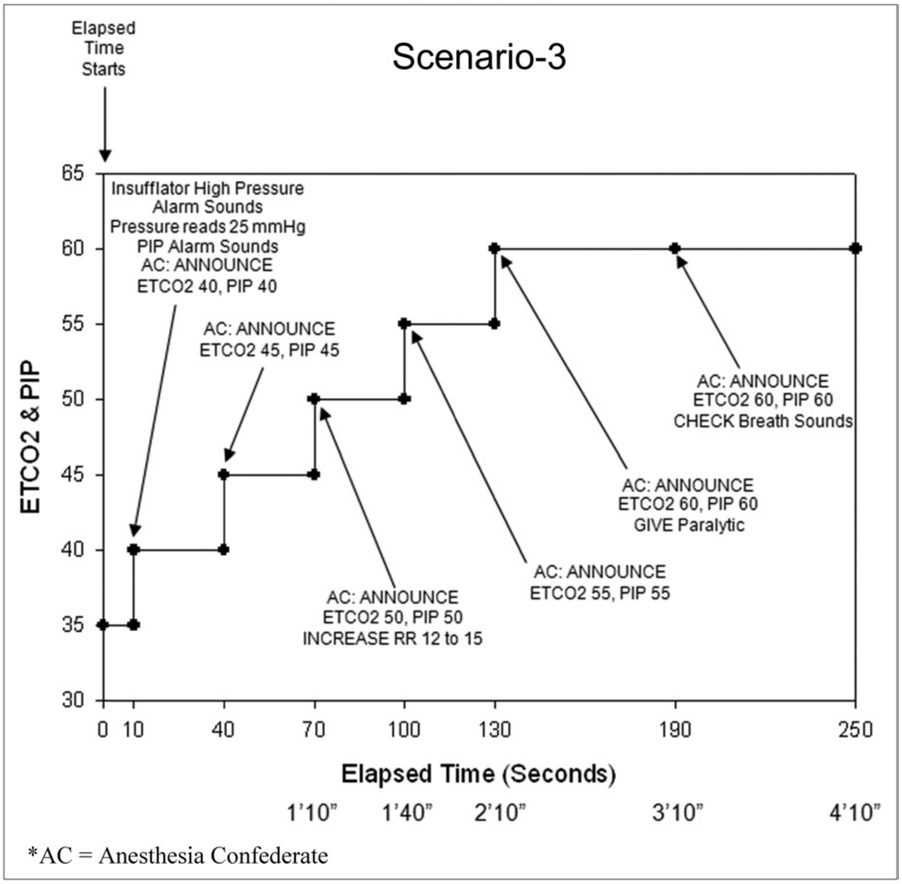


Figure 3 Simulation scenario 3 (hypercarbia) showing a strategic timeline of simulated physiologic events, anesthesia confederate (AC)-administered treatments, and AC announcements at certain intervals.

the number of mandatory treatments administered and the speed with which they were given. After considering several possibilities, we defined the scenario decision making (SDM) score as 250: completion time + (62.5 × [number of treatments rendered]). Because all scenarios included 4 mandatory treatments, the multiplier of 62.5 allowed the treatments to be weighted equally with the completion time.

A 16-item operating room communication assessment (ORCA) instrument was developed according to literature reviews and multidisciplinary consensus by our research team; the tool was used and refined in an iterative fashion over a 15-month period as part of our prior study.[4] This instrument was designed to assess communication, interpersonal, professionalism, and leadership skills in the OR environment with an emphasis on crew resource management (CRM) principles including closed-loop communication, anticipation, situational awareness, and vigilance (Appendix A). Each of the 16 domains was rated on a scale of 1 to 5 (with a higher score indicating superior performance) with anchored (explicit descriptors) end points and midpoints, similar in structure to the Objective Structured Assessment Technical Skills instrument as described by investigators at the University of Toronto.[5] Prior use of this instrument was viewed favorably by our team but no validation data were available before this study. An ORCA score was developed by transforming the scale of 1 to 5 to a scale of 0 to 4 (thus, allowing a zero score), then these transformed values were summed across the 16 domains, divided by the highest possible rating of 64, and multiplied by 100 to create a scaled percentile score.

To assess teamwork skills, we elected to use the line operations safety audits (LOSA) behavioral marker rating scale (Appendix B), which has been validated previously to evaluate CRM skills in the field of aviation.[11] Although this

Table 2 Scenario-specific mandatory treatments

Scenario 1 (bradycardia/ hypotension)	Atropine
	Epinephrine
	Fluid bolus
	Desufflate abdomen*
Scenario 2 (loss of visualization)	Check camera box and cord*
	Check light-source box and cord*
	Check (clean/inspect) or replace laparoscope (use spare)*
	Exchange and use spare camera*
Scenario 3 (hypercarbia)	Increase minute ventilation
	Paralytic
	Check breath sounds

*This action may be performed only by the surgeon and not by the confederates.

rating tool originally was developed to assess teamwork within a flight crew cockpit, we modified it to be applicable for the OR environment. We retained all 13 markers from the original LOSA instrument but changed wording as needed; for instance, "cockpit" was changed to "OR," "pilot" was changed to "surgeon," and so forth. Each marker was associated with defined observations and rated on a scale of 1 to 4. Similar to the ORCA scoring method, the LOSA ratings were transformed to a zero scale, divided by the highest possible rating of 39, and then multiplied by 100 to create a scaled percentile score.

To measure workload, each participant completed the previously validated National Aeronautics and Space Administration–Task Load Index (NASA-TLX) rating scale[12] (Appendix C). This instrument also was developed for aviation CRM and is designed to be used before debriefing, as was performed in this study. Subjects rated the mental, physical, and temporal demands of the task, as well as their performance, effort, and frustration level. Each of the 6 domains has a descriptor and was rated on a scale of 1 to 10 (with a higher score indicating a higher workload). Similar to the earlier-described methods, the NASA-TLX ratings were transformed to a zero scale, divided by the highest possible rating of 54, and then multiplied by 100 to create a scaled percentile score.

Participants also completed a postprocedure survey (investigator developed, Appendix D). The survey consisted of 11 statements with responses rated on a scale of 1 to 10, with 1 indicating "strongly disagree" and 10 indicating "strongly agree." Four survey items evaluated face validity in terms of technical, communication, teamwork, and patient safety skills. Three items evaluated realism in terms of the VOR environment, suspended disbelief, and the laparoscopic model. Three items evaluated educational benefit in terms of patient care, benefit, and unique learning experience. One item evaluated quality of the educational activity. Data from individual items were averaged to provide ratings for each category.

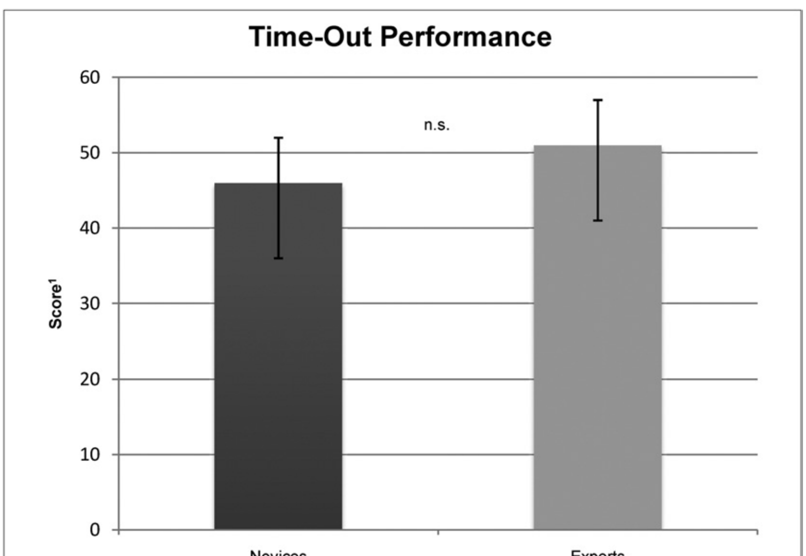
Assessments were completed in the following order: time-out was performed first, then SDM for each of the 3 scenarios in order; these ratings were performed real-time as scores were assigned for these components by the designated proctor under direct/live observation. After completion of the exercise, participants completed the NASA-TLX and the postprocedure survey and the proctor completed the ORCA and LOSA rating tools. Efforts were made to alternate the order in which the postexercise instruments were completed to avoid a halo bias; however, a formal randomization structure for this purpose was not followed.

Construct validity was evaluated by comparing differences in performance between novices and experts. Mann–Whitney U tests were performed using SigmaStat (Systa Software, Inc, Chicago, IL); a P value of less than .05 was considered significant and data are reported as mean $\pm$ SD.

Results

Surgery participants included 8 PGY-1 residents and 6 faculty; gynecology participants included 5 PGY-1 and 1 PGY-2 resident and 4 faculty. All participants completed the entire exercise and complete data were available.

Time-out performance scores (Fig. 4) for novices (46 $\pm$ 6) and experts (51 $\pm$ 10) were not significantly different (n.s.). Interestingly, all participants met the first 3 criteria but only 21% of novices and 50% of experts met criteria 4, no novices and only 10% of experts met criteria 5, and no novices or experts met criteria 8 or 9.



[1]Score = (# criteria met/9) x100; A higher score indicates superior performance

Figure 4 Time-out performance scores for novices and experts.

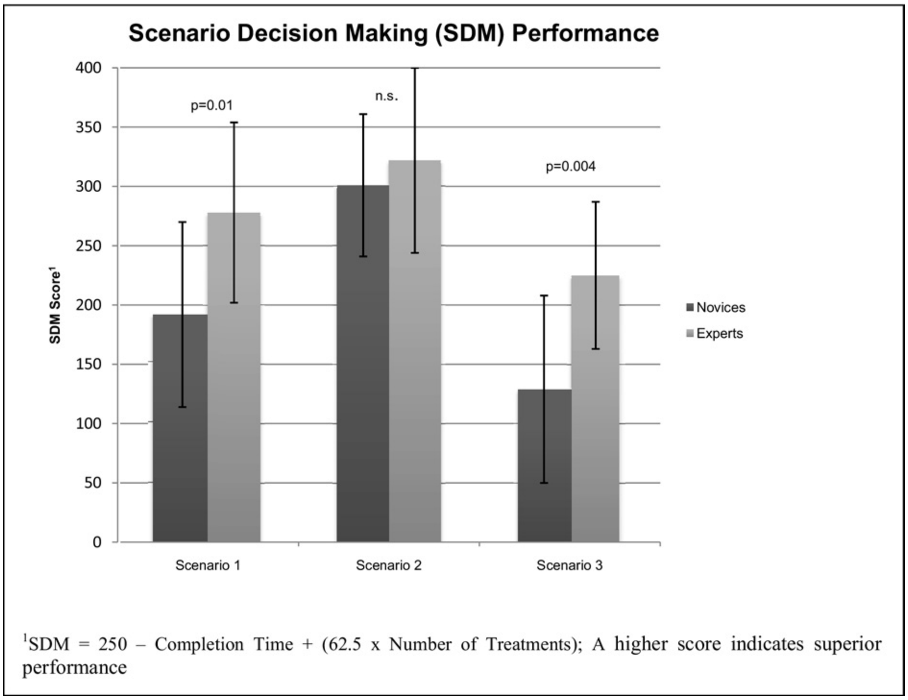


Figure 5 SDM performance scores for novices and experts for scenarios 1 through 3.

According to SDM scores, significant performance differences were detected for scenarios 1 (192 ± 78 vs 278 ± 76; $P = .01$) and 3 (129 ± 79 vs 225 ± 62; $P = .004$) between novices and experts, respectively (Fig. 5). These differences were attributable to the novices completing fewer mandatory treatments (2.1 vs 2.8 for scenario 1 and 1.4 vs 2.2 for scenario 3) and requiring longer completion times (192 vs 147 s for scenario 1 and 210 vs 162 s for scenario 3) compared with experts. However, no statistically significant differences were detected between novices and experts for scenario 2 (301 ± 60 vs 322 ± 78, n.s.).

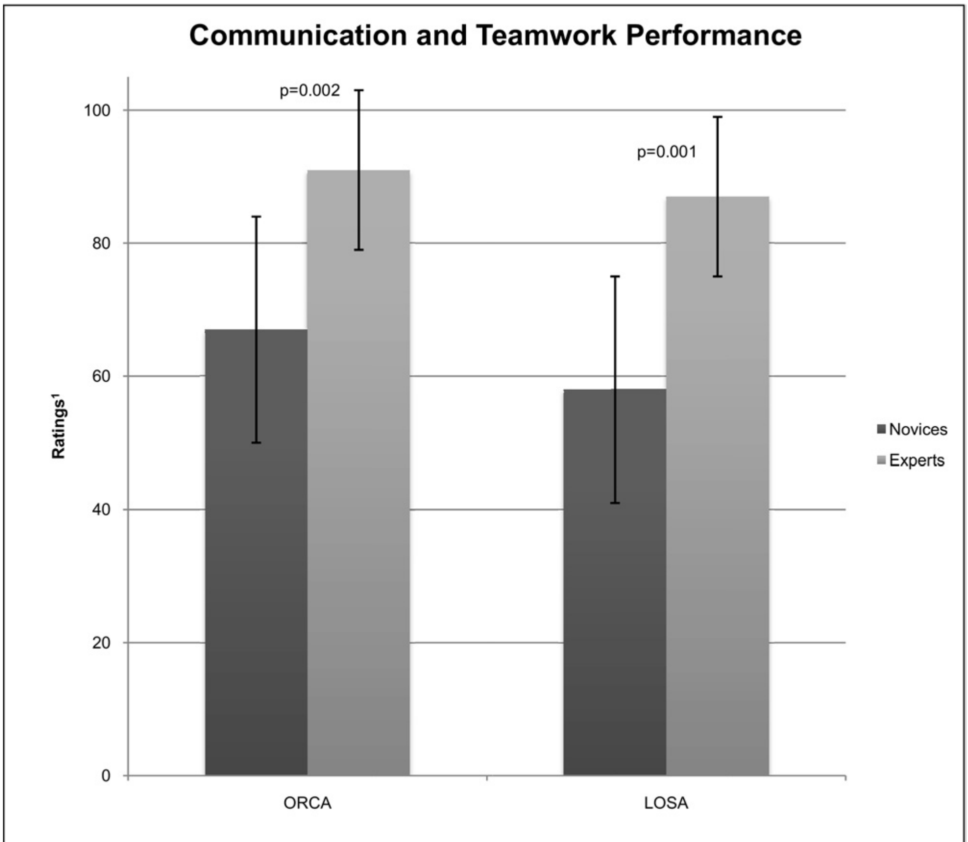
According to ORCA scores, significant performance differences were detected between novices and experts (67 ± 17 vs 91 ± 12; $P = .002$) (Fig. 6). The largest difference for any single ORCA item was seen for anticipation (which related to preventing delays or adverse events via anticipation) with novices achieving a rating of 2.7 versus 4.5 for experts. Similarly, LOSA scores (Fig. 6) showed significant differences between novices and experts (58 ± 17 vs 87 ± 12; $P = .001$). The largest difference for any single LOSA item was seen for automation management (which related to balancing automation with situational requirements) with novices achieving a rating of 2.5 versus 3.8 for experts.

The NASA-TLX workload ratings were not significantly different (62 ± 10 vs 58 ± 13) for novices and experts, respectively. The postprocedure survey data (Fig. 7) strongly supported face validity (8.8 ± 1.3), realism (7.1 ± 1.8), educational benefit (8.5 ± 1.6), and quality (8.8 ± 1.2). In addition, participant feedback and impressions were excellent.

Comments

As advances are made in surgical education, our understanding of competency and the importance of decision making, communication, and leadership skills continues to evolve. Indeed, deficits in these areas have been implicated as reasons for poor clinical outcomes.[13–15] At the same time, surgical simulation has become a mainstream part of both training and assessment. For example, the Fundamentals of Laparoscopic Surgery (FLS) program has proven highly effective in fostering skill acquisition and robust performance benchmarks have been established, such that high-stakes certification is available.[16,17] Moreover, given the extensive validation of this program, documentation of competency via certification is required by the American Board of Surgery.[18] However, simulation-based surgical team training is a much less mature field. The best methods for implementing exercises and performing assessments in these environments have not yet been established and much research is ongoing.

We undertook this study to build on our prior work regarding a laparoscopic troubleshooting module that was designed to be relevant to events that may occur in the clinical environment with some frequency.[4] Over the course of implementing this exercise for numerous teams over a 2-year period, we thoroughly refined the module into its current configuration. Namely, we constructed the scenarios such that they were highly structured and limited by cut-off times. Specifically, we wished to mimic our work in task simulation whereby object performance scoring could be applied. Arguably, assessments have been the most difficult aspect of team training to develop. This study documented the suitability of our strategy in fostering the development of objective scoring based on time and



[1]A higher score indicates superior performance

Figure 6 Communication (ORCA) and teamwork (LOSA) overall performance scores for novices and experts.

errors, in essence using a similar strategy as the tremendously successful FLS program. Even with relatively small cohorts of 14 novices and 10 experts, we detected significant performance differences in decision making for scenarios 1 and 3, which both simulated physiologic abnormalities; thus, construct validity was supported for the SDM scoring system. Although we did not detect significant differences in performance for scenario 2, this likely was related to the relatively simplistic nature of this component and the ability of even novice residents to troubleshoot the equipment satisfactorily. By using completion time as a performance measure, some studies[19] have detected construct validity for surgical team training, whereas others have not.[20] However, no studies to date have integrated completion time with errors into a single decision-making score. Such a strategy allows real-time scoring of the exercise without the need for subsequent video review or sophisticated equipment. Thus, we anticipate that learning curves and proficiency benchmarks may be generated using this strategy, facilitating both training and assessment in this environment.

Numerous studies have used a variety of rating tools to evaluate communication, teamwork, and CRM principles. Some studies have simply surveyed participants after carefully simulated training exercises.[21] Other studies have used instruments borrowed from the aviation community, with variable results. Two instruments have been used mostly widely, including NOn-TECHnical Skills evaluation (NOTECHS)[19,22–24] and LOSA.[11] The NOTECHS instrument has shown construct validity in some studies,[20,25] but not in others.[19,22] Much less information in surgery is available for LOSA, but favorable results have been reported in anesthesia and other fields.[26] Our study supports construct validity for LOSA in the context of surgical team training.

However, in our opinion, neither NOTECHS or LOSA provide thorough evaluations of communication; data collected using these instruments are quite limited and essentially cover 3 to 4 domains, most of which directly relate to teamwork. Other reports describe favorable results for communication rating tools that evaluate an entire surgical team.[13,27,28] However, validated rating tools to assess an individual's communication skills are lacking. Through an iterative process and according to multidisciplinary consensus, our team developed such a tool—the ORCA instrument—which carefully defined descriptors for 16 distinct communication elements. Our study documented construct validity for this instrument, which was quite interesting.

Our NASA-TLX data indicated that workload was similar for novices and experts, indicating that enrollment of a wide variety of learner levels into this simulation exercise may be reasonable. The postprocedure survey data provided favorable ratings (7.1–8.8) for face validity, realism, educational benefit, and quality of the exercise, and novice and expert ratings were similar (Fig. 7). These data suggest that this simulation exercise is a good method for learning and

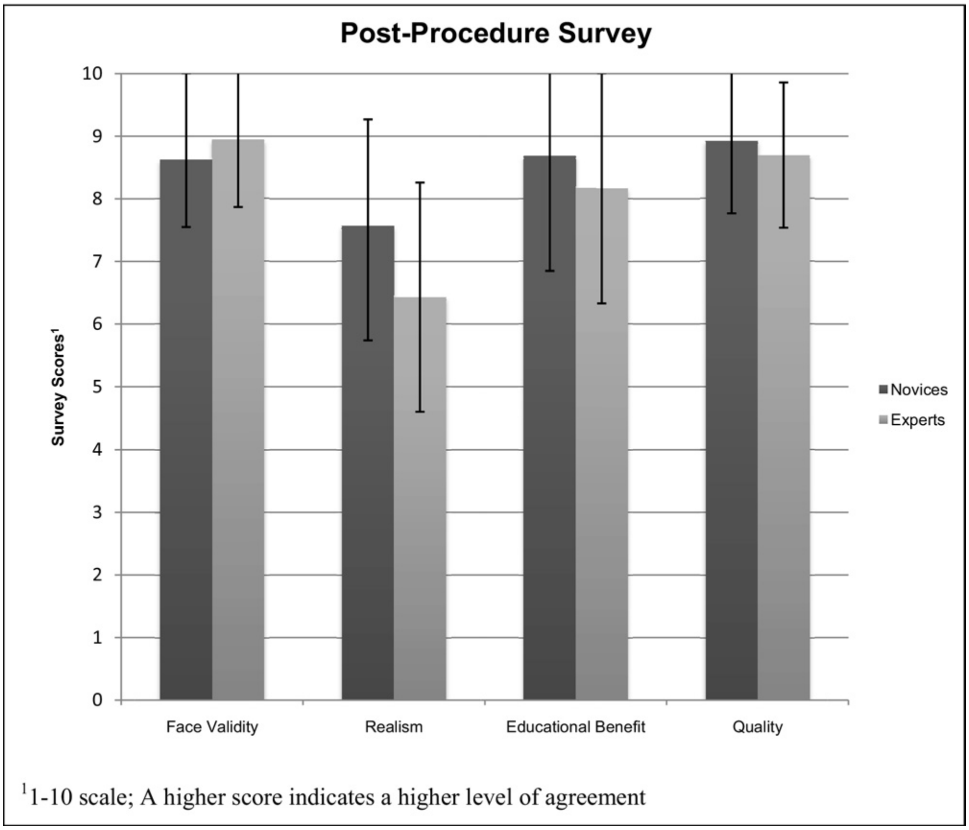


Figure 7 Postprocedure survey scores for novices and experts for face validity, realism, educational benefit, and quality.

assessing technical, communication, and teamwork skills, as well as patient safety.

Interestingly, we detected no significant performance differences between groups according to time-out scores (Fig. 4). As mentioned, defining the ideal or even essential elements of a time-out may be controversial.[7–10] From prior work, our group was quite comfortable with the 9 elements we defined as being appropriate for inclusion (Table 1). However, both novices and experts met only about half of the defined criteria for our time-out. Thus, a gap may exist either between our expectations and the performance of both novice and expert clinicians. Alternatively, the performance of both novice and expert clinicians simply may be suboptimal and in need of improvement. For instance, none of our participants met criteria 8 regarding a team discussion about pertinent medical problems of the patient; this is despite our case stem stipulating that the patient undergoing surgery had a known cardiomyopathy. Certainly, it seems reasonable to expect the surgeon to initiate at least a brief conversation with the anesthesiologist regarding this pre-existing condition. In any case, additional studies in this area and further consensus building for essential and important time-out components are needed.

Several limitations of this study are worth mentioning. Various aspects of validity were evaluated but no reliability investigations were conducted; such studies are necessary and planned. We overcame this constraint by using a single rater for all performance measures within the specified disciplines (surgery and gynecology). Although the time-out and SDM ratings were objective, the ORCA and LOSA ratings were subjective and the raters were not blinded to the experience level (novice or expert) of the participants; potentially introducing bias or confounding our results, thus, additional validation studies will be needed. Another limit was that we did not incorporate evaluation of technical skills into this study. Although we aimed to perform such assessments when we first began our work in team training,[4] we rapidly realized that given our focus on decision making and communication and the frequent interruptions choreographed during our scenarios, that evaluation of technical skill was not facilitated by the structure of our module. Thus, in contrast to other groups,[19,20,22] we did not assess this performance domain. Indeed, we received the lowest face validity ratings for technical skill (8.3) compared with communication (9.0), teamwork (8.9), and patient safety practices (8.8), corroborating that technical skill was a secondary priority for our exercise.

Finally, a word regarding logistics, both for conducting this type of training, as well as this type of research. Typically for task training, a proctor-to-trainee ratio of 1:3 to 1:5 may be used for programs such as FLS.[29] In contrast, team training is much more personnel-intensive, requiring up to 3 confederates and 2 to 3 additional proctors for a single participant. The development of methods whereby fewer personnel are required may be quite helpful. Certainly, OR personnel such as scrub technicians and nurses, or other health care providers such as medical, physician assistant, or nursing students may be used instead of confederates. However, our experience in using whole teams for these simulations has been quite troublesome from a scheduling standpoint. In addition, using whole teams instead of a confederate model introduces significant variability and obviates the educator's ability to reproduce the simulation exercise in a structured fashion; in the context of using such simulations for the purpose of assessment, limiting the variability as was performed in our study may be quite important.

In conclusion, this study yielded positive validation results regarding our previously reported laparoscopic troubleshooting module. Construct validity was supported for the objective scoring we developed for decision making for the 2 physiologic scenarios and for the ORCA and LOSA instruments. Although construct validity was not supported for the time-out or scenario 2 SDM scores, further refinements for these components may prove useful. Regarding face validity and participant feedback, data were quite positive. Further adoption and investigations regarding this curriculum are encouraged.

Acknowledgments

This study was funded by a research grant from the Agency for Healthcare Research and Quality.

References

1. Accreditation Council for Graduate Medical Education. Core competencies. Available from: http://www.acgme.org/acWebsite/RRC_280/280_coreComp.asp. Accessed: April 15, 2011.
2. Scott DJ. Proficiency-based training for surgical skills. Semin Colon Rect Surg 2008;19:72–80.
3. Scott DJ, Dunnington GL. The new ACS/APDS skills curriculum: moving the learning curve out of the operating room. J Gastrointest Surg 2008;12:213–21.
4. Scott DJ. Laparoscopic troubleshooting module. In: Dunnington G, ed. American College of Surgeons (ACS)/Association of Program Directors in Surgery (APDS) Surgical Skills Curriculum for Residents Phase III, 2008. Available from: http://www.facs.org/education/surgicalskills.html. Accessed: April 18, 2011.
5. Wanzel KR, Ward M, Reznick RK. Teaching the surgical craft: from selection to certification. Curr Probl Surg 2002;39:573–659.
6. Scott DJ, Jones DB. Virtual reality training and teaching tools. In: Soper NJ, Swanstrom LL, Eubanks WS, eds. Mastery of Endoscopic and Laparoscopic Surgery. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2005:146–69.
7. Sehgal NL, Fox M, Sharpe BA, et al. Critical conversations: a call for a nonprocedural "time out". J Hosp Med 2011;6:157–62.
8. Faltz LL, Morley JN, Flink E, et al. The New York model: root cause analysis driving patient safety initiative to ensure correct surgical and invasive procedures. In: Henriksen K, Battles JB, Keyes MA, et al, eds. Advances in Patient Safety: New Directions and Alternative Approaches. Vol. 1. Rockville, MD: Agency for Healthcare Research and Quality; 2008.
9. Cronen P, Qadan M, Hicks NZ, et al. Standards of surgery beyond metropolitan centers: a fresh look at perioperative quality measures in small-town America. Am J Surg 2010;200:97–104.
10. Watkins JM, Qadan M, Battista C, et al. A closer look at surgical quality measures across different surgical specialties. Am J Surg 2010;200:90–6.
11. Helmreich RL, Merritt AC, Wilhelm JA. The evolution of crew resource management training in commercial aviation. Int J Aviat Psychol 1999;9:19–32.
12. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, eds. Human Mental Workload. Amsterdam: Elsevier; 1987.
13. Halverson AL, Casey JT, Andersson J, et al. Communication failure in the operating room. Surgery 2011;149:305–10.
14. Disease-Specific Care Certification—National Patient Safety Goals. Oak Brook Terrace, IL: The Joint Commission; 2008. Available from: http://www.jointcommission.org. Accessed: March 22, 2011.
15. Sexton JB, Makary MA, Tersigni AR, et al. Teamwork in the operating room: frontline perspectives among hospitals and operating room personnel. Anesthesiology 2006;105:877–84.
16. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. Ann Surg 2004;240:518–25.
17. Mashaud LB, Castellvi AO, Hollett LA, et al. Two-year skill retention and certification exam performance after fundamentals of laparoscopic skills training and proficiency maintenance. Surgery 2010;148:194–201.
18. Fundamentals of Laparoscopic Surgery. Available from: http://www.flsprogram.org/. Accessed: April 15, 2011.
19. Moorthy K, Munz Y, Forrest D, et al. Surgical crisis management skills training and assessment: a stimulation-based approach to enhancing operating room performance. Ann Surg 2006;244:139–47.
20. Powers KA, Rehrig ST, Irias N, et al. Simulated laparoscopic operating room crisis: an approach to enhance the surgical team performance. Surg Endosc 2008;22:885–900.
21. Paige J, Kozmenko V, Morgan B, et al. From the flight deck to the operating room: an initial pilot study of the feasibility and potential impact of true interdisciplinary team training using high-fidelity simulation. J Surg Educ 2007;64:369–77.
22. Moorthy K, Munz Y, Adams S, et al. A human factors analysis of technical and team skills among surgical trainees during procedural simulations in a simulated operating theatre. Ann Surg 2005;242:631–9.
23. Avermate JAG, Kruijsen EAC. NOTECHS: the evaluation of non-technical skills of multi-pilot air crew in relation to the JAR-FCL requirements. Research Report for the European Commission. National Aerospace Laboratory, Amsterdam: DGVII NLR-CR-98443; 1998.
24. Sevdalis N, Davis R, Koutantji M, et al. Reliability of a revised NOTECHS scale for use in surgical teams. Am J Surg 2008;196:184–90.
25. Black SA, Nestel DF, Kneebone RL, et al. Assessment of surgical competence at carotid endarterectomy under local anaesthesia in a simulated operating theatre. Br J Surg 2010;97:511–6.
26. Gaba DM, Howard SK, Flanagan B, et al. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. Anesthesiology 1998;89:8–18.
27. Hull L, Arora S, Kassab E, et al. Observational teamwork assessment for surgery: content validation and tool refinement. J Am Coll Surg 2011;212:234–43.
28. Sevdalis N, Lyons M, Healey AN, et al. Observational teamwork assessment for surgery: construct validation with expert versus novice raters. Ann Surg 2009;249:1047–51.
29. Scott DJ, Jones DB, Fried GM. Basic laparoscopic skills module. In: Dunnington G, ed. American College of Surgeons (ACS)/Association of Program Directors in Surgery (APDS) Surgical Skills Curriculum for Residents Phase 1, 2007. Available from: http://www.facs.org/education/surgicalskills.html. Accessed: April 18, 2011.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.amjsurg.2011.08.010.