

Identifying the Need and Content of an Advanced Laparoscopic Skills Curriculum: Results of a National Survey

Dmitry Nepomnayshy MD, Lahey Hospital, Adnan Alseidi MD, Virginia Mason, Shima Fitzgibbons MD, Georgetown, Dimitrios Stefanidis MD, Carolinas Medical Center

INTRODUCTION: A recent survey of fellowship directors suggested significant deficits in the laparoscopic technical skills of graduated general surgery residents. Our aim was to define the need for and possible content of a simulation based curriculum in advanced laparoscopic surgical skills (ALS).

METHODS: An anonymous on-line survey was developed and distributed to all Fellowship Council program directors (PD), current fellows (CF) and recent fellowship graduates (FG). The survey was designed to assess the perceived need, possible content, and implementation challenges of an ALS curriculum. Images and descriptions of recently developed simulation based advanced laparoscopic tasks included off-angle camera work and restricted space suturing. These tasks were then evaluated by respondents with additional suggestions solicited via free response.

RESULTS: Of 186 respondents (RR=64%), 40% were CF, 22% were FG and 37% were PD. Respondents primarily self-identified as minimally invasive (MIS) and/or bariatric surgeons (78%) and as hepatobiliary surgeons (12%). 73% of those surveyed identified a need for an ALS curriculum. All three respondent groups cited laparoscopic needle positioning and suturing (78%), and bimanual coordination during dissection and retraction (72%), as the skills with the most need for improvement. In addition, the majority of responding PDs identified "lack of familiarity with anatomy and procedure" (74% of PDs) and "lack of proficiency at laparoscopic bowel anastomosis" (59% of PDs) as problem areas. Respondents felt that successful implementation of an ALS curriculum depended on both overall feasibility and the ability for repeated practice, and should not be dependent on cost. Thematic analysis of free responses revealed the following priorities for skills and tasks to be included in an ALS curriculum: 1) difficult dissections and difficult exposure 2) forehand/backhand and under tension suturing 3) non-dominant hand drills 4) working with an off-set camera and 5) suturing and handling fragile tissue with properties similar to peritoneum or bowel.

CONCLUSION: This needs assessment survey identified several specific skill sets such as advanced suturing, bimanual coordination and managing difficult anatomy that graduating surgery residents need to improve. As a result, the development of an advanced laparoscopic skills curriculum is currently underway to address these deficits.

Systematic Development of a Proctor Certification Exam for the Fundamentals of Laparoscopic Surgery Testing Program.

E. Matthew Ritter MD FACS¹, Inga C. Brissman BS², ¹Department of Surgery, Walter Reed National Military Medical Center / Uniformed Services University, Bethesda MD, ²Fundamentals of Laparoscopic Surgery Program, Society of American Gastrointestinal and Endoscopic Surgeons, Los Angeles CA

Objective: The Fundamentals of Laparoscopic Surgery (FLS) exam is administered by a network of proctors at institutions throughout the world and is required by the American Board of Surgery. Incorrect administration of the exam represents a significant threat to FLS test validity as the program continues to grow. The objective of this project was to develop a test blueprint and accompanying computer based test to assess FLS proctors' knowledge and skills relating to FLS test administration.

Methods: The Twelve Steps for Effective Test Development outlined by Downing were used as the conceptual framework. Eleven subject matter experts (SME) were involved in creation of a job/task analysis blueprint of the knowledge and skills required to be an effective FLS proctor. Each job/task was rated by SME's with respect to frequency of occurrence and importance (3 point scale for each). Using the weighted blueprint, a 45-question exam was created using online test delivery software and pilot tested with 62 current FLS proctors. Pass/Fail standard setting was done using a modified 3-way Angoff technique involving 5 SME's.

Results: The blueprint generated 44 items mapped to 4 domains. SME weighting ranged from 1-9. Topics with a weighing of < 3 were eliminated from the test while those with a weighting factor >8.5 were represented by two questions on the test. Items included on the test had an average content validity index of 0.91. The mean score during beta testing was 32/45 (71%, range 19 -41). Reliability of the exam as measured by Cronbach's alpha was excellent at 0.72. Mean item difficulty was 0.71 and mean item discrimination as measured by point biserial correlation was 0.28. The pass/fail cut score was set at 28/45 (62%) which resulted in a pass rate of 83% during beta testing.

Conclusion: We have developed a test representing the skill set of the FLS proctor that demonstrates content, internal structure, and consequence validity evidence. Incorporation of this test into the FLS proctor training program should reduce the threats to response process validity for the high stakes FLS exam by improving the accuracy and reducing the variability of proctor performance.

Crowd-Sourced Assessment of Technical Skills (C-SATS): Solving the assessment dilemma in surgical skills evaluation.

Shanley B. Deal MD¹, Justin Warren², Mohamad I. Haque MD, LTC³, Timothy Brand MD³, Bryan Comstock, MS², Thomas S. Lendvay MD², Adnan Alseidi MD, Med¹, ¹Virginia Mason Medical Center, Seattle, WA, ²University of Washington, ³Madigan Army Medical Center

Background: Objective and unbiased assessment of surgical skills remains an ongoing challenge in surgical education. One of the greatest barriers is the time required of faculty experts to observe and assess learners. We sought to evaluate the feasibility and accuracy of crowd sourcing for assessment of technical skills (CSATS).

Methods: Seven volunteer general surgery interns were instructed and given ample time for proctored training on peg transfer, precision cutting, and intra-corporeal knot-tying. Interns then performed these tasks on the electronic data generator for evaluation (EDGE) laparoscopic trainer, individually, in a test setting. Edited and de-identified videos were uploaded to a private, secure website in a standardized format. Five surgical faculty content experts (FE) randomly evaluated 21 video clips. Amazon.com Mechanical Turk crowd workers (CW), generated a minimum of 30 evaluations on each of the same videos. CW were included based on rater performance history as well as passing a calibration and attention test. Both groups used the global objective assessment of laparoscopic skills (GOALS) validated rating instrument. Krippendorff's alpha coefficient and the intra-class correlation coefficient evaluated inter-rater reliability. Linear mixed effects models derived an average CW rating and FE rating for each video clip and Pearson correlation coefficients if inter-rater reliability was found to be greater than $\alpha > 0.7$.

Results: Within 19 hours and 15 minutes we received 662 ratings from CW and received 105 ratings from FE over the course of 10 days. FE ratings were found to be of borderline internal-consistency across videos (Krippendorff's $\alpha = 0.55$). Faculty expert ratings were highly correlated with crowd worker ratings (Pearson's correlation coefficient = 0.78, $p < 0.001$).

Conclusions: Crowd worker assessment of technical performance ratings may be a reliable basic surgical skills assessment tool. Crowd sourcing solves the inherent problem of inter-rater and inter-institution reliability error. We propose the use of crowd-sourced assessment of technical skills (CSATS) as a strong candidate to standardize the evaluation of technical skills in General Surgery residency training.

Towards an Advanced Laparoscopic Skills Curriculum: Which tasks measure expert skill?

Richard Birkett, Theodore Delmonico, James Whittleage, Ruth Ruthazer, Neal Seymour, Dmitry Nepomnayshy, Lahey Hospital, Burlington, MA

BACKGROUND: In order to develop an advanced laparoscopic skills (ALS) curriculum, we previously described an initial validity trial of a novel fundamentals of surgery (FLS) based task set that was able to discriminate performance of general/minimally invasive (MIS) surgeons at higher levels of expertise whereas FLS could not. However, the entire task set took 6-7 times longer to complete than the FLS tasks. The purpose of this study was to isolate specific ALS tasks that could discriminate experience level in order to consolidate the entire skill set and focus ALS skills simulation training.

METHODS: General surgery residents, fellows and MIS general surgeons were recruited at 2 national meeting and 2 institutions. Subjects were timed on a series of 10 tasks; 3 FLS tasks and 7 ALS tasks using a previously described FLS based scoring system of time and errors. Groups were defined as expert (PGY 5+), intermediate (PGY 3-4) and novice (PGY 1-2). Scores between groups were compared using a non-parametric Kruskal-Wallis test.

RESULTS: 51 General/MIS surgeons completed the trial. 15 attendings (29.4%), 5 fellows (9.8%), and 31 residents (60.8%) stratified into novice (n=14), intermediate (n=13), and expert (n=24) groups. Pairwise comparisons showed no differences on FLS between groups. However, when pairwise comparisons of groups were carried out for ALS tasks, the experts out-performed both the novices (P=0.0003) and the intermediate group. (P=0.0256). When analyzing tasks individually, only suturing tasks revealed a difference between groups; experts had significantly better scores than novices on offset forehand suture (P=0.0001), offset backhand suture (P=0.0163) and confined space suture (P=0.0099) tasks. In addition, experts had better scores than the intermediates on offset suture (P=0.0162) and backhand suture (P=0.0465) tasks.

CONCLUSION: Out of a 7 task ALS set, only 3 tasks discriminate performance differences between levels of expertise. These tasks focus on advanced suturing skills with an offset camera, backhand suture and working in a restricted space. Our data supports the incorporation of these tasks when building an ALS curriculum.

Psychometric Properties of the Global Operative Assessment of Laparoscopic Skills using Item Response Theory

Yusuke Watanabe¹, E Bilgic¹, Amin Madani¹, Katherine C. Mckendy¹, Lawrence Lee¹, Yoichi Ito², Melina C. Vassiliou¹, ¹The Steinberg-Bernstein Centre for Minimally Invasive Surgery and Innovation, McGill University Health Centre, Montreal, Canada, ²Department of Translational Research Cooperation Promotion, Hokkaido University Graduate School, Sapporo, Hokkaido, Japan

Background: The Global Operative Assessment of Laparoscopic Skills (GOALS) assessment tool makes use of 5 equally weighted items to evaluate laparoscopic skills based on a 5-point Likert scale. However, the actual contribution of each item may vary based on the difficulty of the skill being measured and its discriminative power. The purpose of this study was to determine the level of difficulty and discriminative power of each of the 5 GOALS items using item response theory (IRT).

Methods: total of 322 GOALS assessments used to evaluate intra-operative performance were included for a range of different laparoscopic procedures. Each of the 5 items (depth perception, bimanual dexterity, efficiency, tissue handling, and autonomy) on the tool was analyzed using IRT. Threshold parameters of item difficulty (corresponding to each of the 5-point Likert scale levels) and discrimination (slope parameter) were estimated for each item. Lower values in the threshold parameters denote easier skills being measured.

Results: Table 1 shows the threshold parameters and the slope parameter for each of the 5 GOALS items. Depth perception, tissue handling, and autonomy displayed similar item difficulty distributions and discriminative ability. The higher slope parameters seen with bimanual dexterity and efficiency are indicative of greater discriminative ability than the other items.

Conclusions: RT psychometric analysis of GOALS reported that its five items did not demonstrate uniform difficulty and discriminative power, suggesting that these five items should not be scored equally. Bimanual dexterity and efficiency seem to have stronger discrimination to estimate performance. Weighted scores based on these findings could improve the accuracy of estimating individual laparoscopic skills.

Table 1. Threshold parameters and slope estimates for each of the 5 GOALS items using IRT (graded response model)

Items	Threshold Parameter				Slope Parameter
	1/2	2/3	3/4	4/5	
Depth perception	-9.2	-6.1	-2.8	0.7	2.9
Bimanual dexterity	-12.9	-6.9	-1.5	4.6	4.7
Efficiency	-13.0	-8.0	0.0	5.8	5.6
Tissue handling	-8.4	-5.3	-1.9	1.8	2.7
Autonomy	-6.3	-3.6	0.0	3.0	2.4

Optimizing Training Cost-Effectiveness of Simulation-Based Laparoscopic Inguinal Hernia Repairs

Roberto Hernandez Irizarry, MD Benjamin Zendejas, MD Shahzad Ali, MD, David Farley, MD, Department of Surgery College of Medicine, Mayo Clinic, Rochester, MN

Objective: We sought to compare whole- versus part-task simulation-based training (SBT) of laparoscopic inguinal hernia repairs (IHR), with the aim of improving training cost-effectiveness.

Summary Background Data: Little is known about what makes SBT more or less cost-effective. Motor learning theory suggests that highly-complex tasks are probably best trained under conditions of part-task (PT), as opposed to whole-task (WT) training. Within PT, random practice of tasks, instead of in sequence, has been shown to lead to improved skill retention and transfer.

Methods: General surgery (GS) residents were equally randomized to PT versus WT, mastery learning type, simulation based training of laparoscopic IHR. Training time and resources used to reach mastery (skill acquisition), performance at one month testing (skill retention), and intraoperative time and performance scores (skill transfer) were compared between groups.

Results: Forty-four GS trainees were randomized. No significant demographic, previous clinical experience or skill performance differences existed at baseline. All residents achieved mastery benchmarks. Trainees in the PT group achieved mastery on average 17 minutes faster (60.2 ± 23.8 vs 77.1 ± 24.8 minutes, $p=0.02$, saving 6.2 instructor hours), employed fewer material resources (curricular cost savings of \$2,380 or \$121 per learner, US Dollars), and were more likely to retain mastery level performance at one month retention testing (59% vs 22.7% $p=0.03$). No significant differences in intraoperative performance were encountered.

Conclusions: For highly complex surgical tasks, such as laparoscopic IHR, random part-task SBT seems to be more cost-effective, when compared to whole-task training. Such findings have important implications in the design of future SBT curricula.

Construct Validity of a Novel, Objective Evaluation Tool for the Basics of Laparotomy Training (BOLT) using a Simulated Model

Lawrence Greenawald, MD¹, Mohammad Shaikh, MD¹, Jorge Uribe, MD¹, Faiz Shariff, MD¹, Barry Mann, MD², Christopher Pezzi, MD³, Andres Castellanos, MD¹, D. Scott Lind, MD¹, ¹Department of Surgery, Drexel University College of Medicine. Philadelphia, PA 19102, ²Department of Surgery, Lankenau Medical Center, Wynnewood, PA 19096, ³Department of Surgery, Abington Memorial Hospital, Abington, PA 19001

Objective: While the majority of abdominal operations are still performed using an “open” technique, there remains no validated, objective measure of the fundamental surgical skills necessary for laparotomy. Therefore, we describe our initial success in designing and implementing an objective evaluation of opening and closing a simulated abdomen.

Design: The surgical literature was reviewed to compile a comprehensive list of the steps involved in performing a laparotomy. Then, using the modified Delphi method, a survey was first sent to local surgeons and then to the members of the Association for Surgical Education (ASE) and the Association of Program Directors in Surgery (APDS) asking them to rank the essential steps of laparotomy using a 5 point Likert-type scale (1=least essential, 5=most essential). Based on our survey results, items with a score of 4 or more were included in the final evaluation tool. Following this, 10 novices (PGY-1 residents) and 10 experts (PGY 5 residents and faculty) performed a videotaped, simulated laparotomy on a viscoelastic model (*Figure 1*). Videos were then reviewed by 3-blinded expert raters using a 16-item binary checklist. Results for both groups were compared using chi-square tests for binary outcomes and unpaired t-tests for continuous values.

Results: Experts performed better than novices opening (*0.94 vs. 0.51; p<0.0001*), closing (*0.85 vs. 0.16; p<0.0001*) and in overall performance (*0.88 vs. 0.27; p=0.06*) on the simulated abdomen. In addition, novices caused injury to simulated intra-abdominal contents more frequently (*6 vs. 1; p=0.01*) and took longer to open the abdomen (*6:06 vs. 3:43; p=0.01*) than experts (*Table 1*).

Conclusion: We demonstrate initial construct validity of an objective evaluation tool for opening and closing a simulated abdomen. Ongoing efforts include enhancing the fidelity of the viscoelastic simulated laparotomy model and integrating this evaluation tool into competency-based surgical curricula.

This work was funded by a grant from the Association for Surgical Education (ASE) and the Association of Program Directors in Surgery (APDS).

Validation of a novel intraoperative assessment tool: The Surgical Procedure Feedback Rubric

Ayca Toprak MD¹, Ulemu Luhanga¹, Sarah Jones MD², Andrea Winthrop MD¹, Laura McEwen MD¹, ¹Queen's University, Kingston, ON, CN, ²University of Western Ontario, London, ON, CN

Background: Competency-based training models in surgical education challenge programs to adequately document and assess trainees' clinical decision-making, problem solving, and procedural skills in the operating room. The Surgical Procedure Feedback Rubric (SPR) is designed to document resident performance during a single, directly observed operative encounter and provide targeted feedback to support learning. It differs from other assessment tools because it defines performance criteria by increasing complexity through the use of behavioural anchors, thus embedding standards of performance in the tool. This study reports results of a one year multi-method observational study collecting validity evidence for the SPR.

Methods: This study was guided by the validity frameworks developed by Kane (2013) and Stobart (2009). The SPR was introduced to three post-graduate surgical training programs at our institution: General Surgery, Orthopaedics, and Obstetrics and Gynecology. The internal structure of the SPR was examined using exploratory factor analysis. The effects of post-graduate training year (PGY), program, role in surgery and procedural complexity on SPR scoring were examined using a four-way ANOVA. To explore SPR use, perceived effectiveness, quality of feedback, and the impact on faculty-resident interactions document analysis and semi-structured interviews with staff and residents were conducted.

Results: Over 500 SPRs were completed for 20 General Surgery, 19 Orthopaedic, and 6 Obstetrics and Gynecology residents. All PGY levels were represented. The number of SPRs completed per resident ranged from 2-30. Results to date, pending the completion of classification of procedural complexity, demonstrated the utility of the SPR in distinguishing the intraoperative performance of different PGY levels ($F(5,359)=4.07$, $p<0.001$). Specifically, post-hoc analyses indicated that differences in SPR scores were observed between junior and senior residents. Document analysis and interviews with faculty and staff revealed that the introduction of the SPR increased the frequency and timeliness of feedback on intraoperative performance to residents.

Conclusion: The SPR is a novel direct observation assessment tool to document resident performance in the operating room and to provide targeted feedback to support learning in a competency-based model of surgical education.

Reliable Assessment of Operative Performance

Elif Bilgic, Yusuke Watanabe, Katherine M. McKendy, Amani Munshi, Gerald M. Fried, Liane S. Feldman, Melina C. Vassiliou Steinberg-Bernstein Centre for Minimally Invasive Surgery and Innovation, McGill University Health Centre, Montreal, QC, Canada

Introduction: Surgical education is moving increasingly to a proficiency-based approach. This requires evidence of satisfactory performance for progression through the curriculum. There is a lack of consensus regarding the number of intraoperative assessments required to reliably measure a trainee's performance. Different factors in the clinical environment may impact performance, making it challenging to reliably assess skill. We evaluated the impact of case factors on performance assessment using Generalizability Theory, and determined the number of cases needed to establish reliable assessments in surgical training.

Methods: Surgical trainees were assessed by attendings during direct observation of laparoscopic surgery over 2 months using the Global Operative Assessment of Laparoscopic Skills (GOALS). Data were collected prospectively, assessing each trainee multiple times. Each variance component and the impact of each factor on assessment scores were calculated using Analysis of Variance. The reliability coefficient was calculated according to Generalizability Theory (GT), with participants, cases, and attendings as factors, along with their interaction terms. The number of cases needed to achieve a reliability of 0.8 was determined using a decision study.

Results: Eight trainees (3 PGY2, 1 PGY3, 3 PGY5 and 1 Fellow) received a median of 3 GOALS assessments (range 2-15) each (total of 48 assessments). The reliability of one assessment per trainee was 0.73. "Case factors" accounted for 16.8% of total variance, while "attending surgeon factors" and "interaction between trainee and attending surgeon" accounted for 8.7% and 1.2% respectively. Increasing the number of cases per trainee assessed by a single attending increased reliability of the GOALS assessment incrementally: 0.73 (1 case), 0.80 (2 cases), 0.83 (3 cases), and 0.84 (4 cases).

Conclusion: These preliminary data suggest that case factors accounted for most of the variance in assessment scores. More than two assessments per trainee should be required to obtain reliable performance scores using GOALS. This methodology may be used to determine the number of assessments needed per procedure to provide reliable assessment of technical proficiency in laparoscopic surgery.

A Sensitivity Analysis and Opportunity Cost Evaluation of the Surgical Council on Resident Education Curriculum

John L. Falcone, M.D., M.S.^{1,2}, Richard S. Feinn, Ph.D.³ ¹Owensboro Health, Owensboro, KY, ²University of Louisville, Louisville, KY, ³ Quinnipiac University, Frank H. Netter, M.D. School of Medicine, North Haven, CT

Introduction: The Surgical Council on Resident Education (SCORE) is responsible for a unified national general surgery residency training educational curriculum. There is evidence that examinees subscribed to the SCORE curriculum have a 1.6% higher pass rate on the American Board of Surgery Qualifying Examination (ABS-QE). These examinees have a significantly higher adjusted raw score. The purposes of this study are to perform statistical sensitivity analyses, and an opportunity cost analysis of the SCORE curriculum with regard to the ABS-QE outcomes.

Methods: In this study, statistical sensitivity chi-square analyses were performed to evaluate the minimum ABS-QE pass rate differences that would have resulted in a significant outcome favoring SCORE subscription. Sample size analysis was also performed, using an overall weighted pass rate of 86.1%. Statistics were performed using an $\alpha = 0.05$. Opportunity cost analysis was performed with regard to SCORE subscription costs by subscribing residency programs.

Results: The pass rate on the ABS-QE for the 893 residents with SCORE subscriptions was 86.3%, and the pass rate for the 139 residents without SCORE subscriptions was 84.7% ($p = 0.43$). Sensitivity analysis showed that a minimum ABS-QE difference of 6.4% would have favored the SCORE curriculum ($p = 0.04$). Sample size analysis showed that a total of 15K-16K resident outcomes would have been necessary for a 1.6% ABS-QE difference to be significant ($p < 0.05$). This represented a 1400% increase (~15 years). The opportunity cost for all SCORE-subscribing residency programs was \$747,425 dollars, or \$467,141/percentage passage increase on the ABS-QE. The opportunity cost over 15 years would be an expected \$11.2 million.

Conclusions: Although resident examinees with SCORE subscriptions have higher raw scores on the ABS-QE, there are no significant pass rate differences between resident cohorts. Sensitivity analyses show that although the power of the analyses is somewhat limited, there would have to be a significant increase in sample size for a SCORE subscription to be associated with significant increases in ABS-QE outcomes. The opportunity cost of SCORE subscription overall is substantial. Individual residency programs with more limited resources should determine if the cost of subscription is financially sound.

Impact of a novel on-screen overlay frame of reference system for orientation during intraoperative laparoscopic surgery

Jen Hoogenes, Soojin Kim, Rami Elias, Kevin Kim, Bobby Shayegan, Kevin Piercey, Anil Kapoor, Ranil Sonnadara, Jacqueline Wilcox, Edward D. Matsumoto (ALL authors are affiliated with McMaster University, Department of Surgery, Hamilton, Ontario, Canada)

Introduction: Surgical educators are faced with unique challenges when teaching laparoscopic skills. Because the surgical field is projected onto the endoscopic screen in the operating room (OR) and is used as the focal point for performing the procedure, laparoscopic teaching involves various types of communication to direct the learner. Directional terminology can get confusing when the spatial orientation of the on-screen target differs from each individuals' vantage point. We explored how teaching occurs in the laparoscopic setting both with and without a directional tool designed to improve communication in the OR.

Methods: We previously showed that using a combination of standardized verbal commands and a transparent monitor overlay (a clock and an alphanumeric grid) significantly improved teaching with laparoscopic simulators. We have videotaped 19 laparoscopic teaching cases in the OR to serve as a baseline comparator group. Participants were blinded to whether the video camera was recording. We then introduced the alphanumeric overlay into the OR. After minor adjustments to the tool to avert interference, we have since videotaped eight teaching cases (with four different surgeons). Videotapes were transcribed verbatim and qualitatively analyzed. Follow-up interviews were conducted with participants after each overlay case.

Results: Analysis of the baseline cases revealed that teaching involved the instructor pointing to the screen, the patient, and the instrument(s); holding the instrument to guide learners' movements; and displaying the type of hand or arm movement required. Directional terms such as "right angle", "up", "down", "medial", etc. were also frequently used. Learners were often told to visualize the movement before it was completed. The alphanumeric grid is utilized more frequently during the teaching of junior residents and all participants have found it to be useful for practice. The preciseness of the grid format allows the learner to move directly the focal point, essentially eliminating the need for pointing and gesturing.

Conclusions: We will continue to evaluate the alphanumeric grid in practice and eventually add and evaluate overlays of a clock with x:y triangulation and one with a dartboard design. Results of these additional observations will be presented, along with photographs and videotapes of the system in use in the OR.

Resident Selection: Associations among psychological factors and candidate rank to a general surgery residency program

Kimberley R. Monden, PhD, Thomas R. Cox, PsyD, James W. Fleshman, MD - Baylor University Medical Center

Objective: The purpose of this study was to explore the relationship among psychological variables associated with well-being and a candidates ranking to a general surgery residency (GSR) training program. A secondary purpose was to explore potential differences in personality traits and psychological well-being (PWB) between high and low ranking candidates.

Methods: This was a prospective study of 92 applicants to a GSR program at an urban community hospital in the Southwestern United States. Applicants were scored across standard criteria (e.g. USMLE scores, letters, interview) and ranked by interviewers. Applicants also completed an assessment of personality and psychological well-being (PWB). To assess perceptions of personality characteristics believed to be most vital in a general surgery resident, 22 faculty members ranked, in order from most to least important, the five personality traits they preferred residents to possess.

Results: Comparison of standard ranking criteria and psychological variables indicates that conscientiousness, extraversion, and PWB (specifically environmental mastery) are correlated with faculty rank of GSR candidates. However, these psychological variables do not provide any predictive value about a candidate's rank above and beyond academics and interview. There is a difference in personality traits and PWB between the top, middle, and bottom third ranking candidates. In descending order of importance, faculty ranked conscientiousness, openness, extraversion, and agreeableness as the most desirable traits in a GSR.

Conclusions: Results indicate that conscientiousness, extraversion, and PWB (specifically environmental mastery) are correlated with faculty assigned rank of GSR candidates. However, in this sample, these psychological variables do not provide any predictive value above and beyond academics and interview. Faculty also rated conscientiousness, openness, and extraversion as the top three traits they prefer in general surgery residents and appear to do a sufficient job of selecting candidates with these traits, as 77% of the top third ranked applicants were high in conscientiousness and 70% were high in openness. Application of these findings to resident selection in GSR programs will be discussed.

Top of their Game: Lessons Learned from Surgeons who were Music Virtuosos, Olympians, and Navy Seals

Siobhan Deshauer¹ Dorotea Mutabzic, MD²; Carol-Anne Moulton, MD, MEd, PhD², ¹McMaster University, Hamilton, Canada, ²University of Toronto, Toronto, Canada

Background/Purpose: A priority on mental skills among top performers in other fields has prompted surgical educators to explore their use in surgery. Contextual differences, however, might limit the degree to which these skills can and should be applied. We identified surgeons with a previous record of 'elite' performance in other fields to specifically explore: 1) which skills were learned in their previous training, 2) how were these skills developed, 3) how, if at all, have these skills been adapted to surgery, and 4) what recommendations would they have for surgical educators.

Methods: A constructivist grounded theory study was conducted. Semi-structured interviews were conducted with 17 surgeons (5 athletes, 6 musicians, and 6 military personnel) from across Canada and the United States, with follow-up interviews performed to clarify and further probe emerging constructs. Data were coded and analyzed iteratively until emerging themes were saturated. A reflexive approach was adopted throughout.

Results: A wide range of mental skills adapted from non-surgical training was described (e.g., self-talk, mental rehearsal, and reframing). Short term benefits, seemed to fit one of two goals: 1) optimization of surgical performance, and 2) increased resilience. Longer-term benefits were described as continued personal engagement with work (often through achieving the mental state of 'flow'; "using a pre-performance routine minimizes the chaos, it allows me to get into that zone"), stress management strategies (such as "tactical breathing" or "putting the mind in neutral and the body in drive"), recovering after complications in surgery (often reframing as a "learning experience") and consequently burnout prevention. The application of specific mental skills in surgical practice was variable (e.g. routine vs. non-routine cases); "even if it's routine, I still think things can be optimized and that's why I enjoy the routine things".

Conclusion: Mental skills learned in other domains appear to have some applicability to surgery, potentially improving surgical performance, resilience and surgeon wellness. These findings might be a valuable addition to current conversations on burnout prevention among surgeons. Further research is needed to explore how mental skills training is best delivered within the field of surgery.

Adaptive versus volume-based training in simulation: a randomized controlled trial

Yinin Hu, MD; Kendall D. Brooks; Helen Kim; Adela Mahmutovic; Joanna Choi; Ivy A. Le; Sara K. Rasmussen, MD, PhD, University of Virginia School of Medicine, Charlottesville, VA

Objectives: Current operative training is primarily based on length of experience or numbers of cases performed. This paradigm fails to account for learning rate variability. Cumulative sum (Cusum) is a quality-control tool that tracks proficiency in real-time and is suited for adaptive, individualized curricula. The purpose of this study is to determine whether Cusum can streamline procedural training compared to a traditional, volume-based paradigm.

Methods: First- and second-year medical students were randomly assigned to Cusum or Control arms. Participants repeatedly executed three simulated, invasive techniques with one-on-one instruction: suturing, intubation, and central venous catheterization (CVC). Feedback using weighted checklists was provided after every practice attempt for both groups. Control participants were each required to practice 8-9 hours. Cusum participants practiced until proficiency—dictated by Cusum—was attained in all tasks. An experienced surgeon blinded to study arm assignment conducted post-test evaluations. Evaluations included a composite overall score and task-specific sub-scores. The study size was designed to detect a 15% practice time difference with 80% power. Group comparisons were performed using Wilcoxon rank-sum.

Results: Forty-eight participants completed the study (24 per group). Average post-test overall score was 93.1% for Cusum and 92.1% for Control ($p = 0.710$). Cusum group practiced 20.9% fewer hours on average than Control group (6.83 vs 8.63 hours, $p < 0.001$). Cusum participants required between 4.5 to 9.3 hours (SD 1.36) to attain proficiency in all tasks, indicating substantial learning rate variability. Task-specific sub-scores demonstrated non-significant trends favoring Cusum (Figure 1).

Conclusions: This is the first randomized, controlled trial to compare Cusum-guided and volume-based practice. Traditional protocols based on volume or time do not account for variations in learning rate. By implementing an adaptive paradigm to account for this variability, Cusum promoted more efficient time utilization while maintaining excellent results

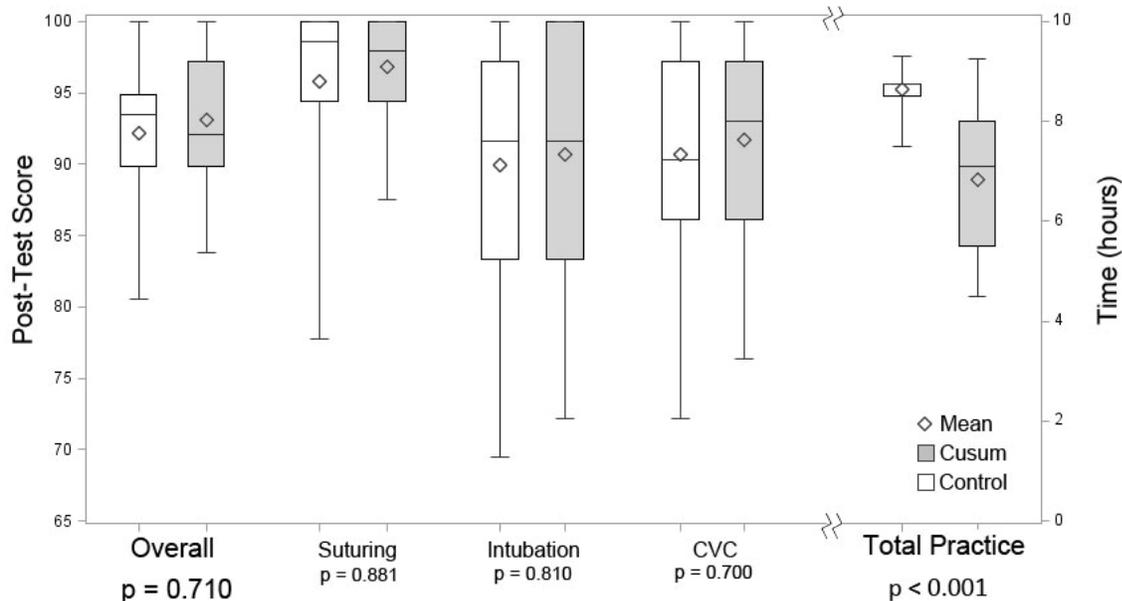


Figure 1: Post-test performance and total practice time for Control (white) and Cusum (gray) groups. Box-and-whisker plots denote median, first and third quartiles, and range. Markers denote means.

The Complex Phenomenon of Stress in the Operating Room: Cool Under Pressure?

Natashia Seemann MD¹, Vicki R LeBlanc PhD¹, Tamara Gimon², Carol-anne Moulton MD, PhD MEd¹, ¹Department of Surgery, University of Toronto, The Wilson Centre, ²Faculty of Medicine, University of Toronto, Toronto, CN

Background/Purpose: Recent interest in surgeon wellness has motivated conversations regarding stress in the operating room (OR). While most research has focused on the physiologic manifestations of stress, recent studies have highlighted the complexity of stress, including the cognitive, emotional and sociocultural influences. The purpose of this study was to refine a methodology to capture the complexity of stress in the OR and to explore the experience of stress by surgeons.

Methods: This exploratory mixed-methods study collected real-time intra-operative physiologic data (heart rate variability, respiratory rate, sweat gland activity, thoracic impedance and salivary cortisol) from six surgeons during multiple operations (2-4 cases per surgeon, elective and emergency). The short-form State-Trait Anxiety Inventory (STAI) was used to assess surgeons' perceived stress. Pre-and post-operative interviews explored surgeons' experience of stress. A grounded theory approach guided qualitative data collection and analysis. Stressful events were identified by predetermined parameters of physiologic change, observer (intra-operative notes) or surgeon (discrete foot pedal, interview).

Results: Twenty-two stressful events were identified. For each stressful event, triangulation of datasets (physiologic, observer, interview) provided means for exploring different facets of stress. For example, physiologic data identified stress best through heart rate variability analysis. Observer data identified stress as an emotional phenomenon, as demonstrated in an observer fieldnote, "surgeon very angry after trainee scrub nurse trips on the cord for his headlamp" [P3, OR1]. Interview data, revealed the complex sociocultural phenomenon of stress, as stated in a pre-operative interview "I see I have a good team today, which reduces my anxiety right off the bat.." [speaking about fellow, resident, scrub nurses and anaesthetist, P5, OR2]. Collecting intra-operative STAI's and salivary cortisol samples proved non-feasible in most cases, resulting in these data not being related to specific events, and thus irrelevant. There was great inter-surgeon variability in not only the physiologic manifestation of stress, but also the perceived experience of stress.

Conclusion: This study demonstrated the ability to identify stress in surgeons by combining objective physiologic and subjective perceived stress data, leading to a refined and integrated mixed methodology to study stress among surgeons and surgical teams in the future.

Making Average Performance Excellent: Implementation Results of a Mental Skills Curriculum during Simulator Training

Stefanidis D¹, Anton N¹, Howley L¹, Pimentel M¹, Davis , Sevdalis N², Arora S², Brown C¹, ¹Carolinas HealthCare System, Charlotte, NC, ²Imperial College of London, UK

Introduction: Stressful events in the operating room may negatively impact surgical performance and lead to errors. Mental skills training refers to the implementation of performance-enhancing and stress-coping strategies to enable individuals achieve optimal performance under any (including stressful) condition. Our objective was to develop a surgery-specific mental skills curriculum (MSC) and obtain initial evidence of effectiveness.

Methods: Through collaborative work among a performance psychologist, a curriculum expert, and a surgical educator a comprehensive surgery-specific MSC was developed. This curriculum consisted of 8 performance-enhancing modules, including introduction to mental skills, goal setting, activation management, attention management, imagery, refocusing strategies, and preoperative mental routines. The MSC effectiveness was assessed during proficiency-based acquisition of laparoscopic skill on simulators by novices; it was delivered via video modules, a learner workbook, and individualized practice sessions with a mental skills trainer. Participant performance before and after training was assessed using the FLS suturing task, the modified test of performance strategies (TOPS), d2 test of attention, and mental imagery questionnaire (MIQ) and stress using the state trait anxiety inventory (STAI). Skill transfer was assessed on a porcine lap Nissen model. Feedback on the MSC was sought from participants. A paired t-test was used to analyze the results.

Results: Nine surgical novices (age 23 ±7 years, 55% women, 78% right handed) completed the curriculum. Compared with baseline participants improved significantly their laparoscopic performance and mental skills at training completion (see table). During the porcine test participant suturing score was 244±143 and STAI scores changed insignificantly (15±3vs16±2;p=n.s.); 89% participants were observed to use taught mental skills. Feedback on the curriculum was very positive

Table

	Baseline	End of Training	p-value
FLS Intracorporeal Suturing Score	13±39	328±202	P<0.01
TOPS	219±18	263±23	P<0.01
d2	538±64	595±60	P<0.001
MIQ	3.8±2.3	5.8±0.7	P<0.001

Higher scores indicate better performance

Conclusions: A surgery-specific simulator-based mental skills curriculum was developed and its effectiveness in improving mental skills and surgical performance demonstrated during a transfer test. Additional high quality evidence is needed to document the benefits of this curriculum for surgeons and is currently underway at our institution.

Creating an institutional-specific program for teaching professionalism: Learning from the professionals

Afif N. Kulaylat MD, Danni U. Qin MS, Susie Sun MD, Christopher S. Hollenbeak PhD, Jane R. Schubart PhD, Antone J. Aboud PhD, Donald J. Flemming MD, Peter W. Dillon MD MSc, Edward R. Bollard MD, David C. Han MD
Penn State Hershey Medical Center, Hershey, PA

Introduction: Organizational ethics is a term used by industry to express organizational values irrespective of governmental and/or regulatory law. Vignettes are commonly used in onboarding processes to communicate this professional culture. In medical training, mistreatment of learners remains a significant barrier to achieving a culture of respect. Our aim was to develop and assess institutionally relevant clinical vignettes as part of the orientation of incoming trainees.

Methods: Experts from the University's School of Labor and Employment Relations spent eight weeks embedded in the care teams of the institution. Based on their observations, six thematic areas were identified as sources of potential or significant mistreatment: verbal abuse, specialty choice discrimination, non-educational tasks, withholding/denying learning opportunities, neglect, and gender/racial discrimination. Vignettes were created and presented to incoming trainees during orientation, followed by facilitated discussions with selected faculty and residents from different departments across the institution. Perceptions of the appropriateness of the interactions depicted in the vignettes were measured on a 5-point Likert scale before and after the discussions, and compared using t-tests. Baseline demographic and history of mistreatment data were also collected. Post-intervention feedback was reviewed to identify possible themes.

Results: There were 144 residents and fellows that participated, with 77% reporting either having been mistreated or having observed mistreatment at least once. Post-discussion scores demonstrated decreased variability in perceptions of mistreatment across vignettes. Two vignettes demonstrated significant differences ($p < 0.05$) in pre- and post-discussion scores (non-educational tasks and gender/racial discrimination). Primary themes identified from feedback were the value of exchanging diverse perspectives among peers and faculty and a desire for more ambiguous scenarios.

Conclusions: Clinical vignettes provide a useful mechanism to discuss professionalism. Their value appears to lie in a robust facilitated discussion, and generating both self- and environmental awareness of contextually relevant issues such as mistreatment. Decreased variability in perceptions of behavior suggests a more aligned definition of professional behavior among trainees, a critical step in establishing a defined organizational ethic. Assessing differences in perception can also identify areas for focused education. Frequently viewed as culprits of mistreatment or unprofessional behavior, surgeons have the opportunity to provide leadership in this domain.

Working volume: An indirect measure of situational awareness and operative safety

Anne-Lise D. D'Angelo, MD¹, Drew N. Rutherford, MS^{1,2}, Rebecca D Ray, PhD^{1,3}, Andrea Mason, PhD², Carla M. Pugh, MD, PhD¹ ¹University of Wisconsin, School of Medicine and Public Health ²University of Wisconsin, Department of Kinesiology ³University of Wisconsin, Department of Statistics

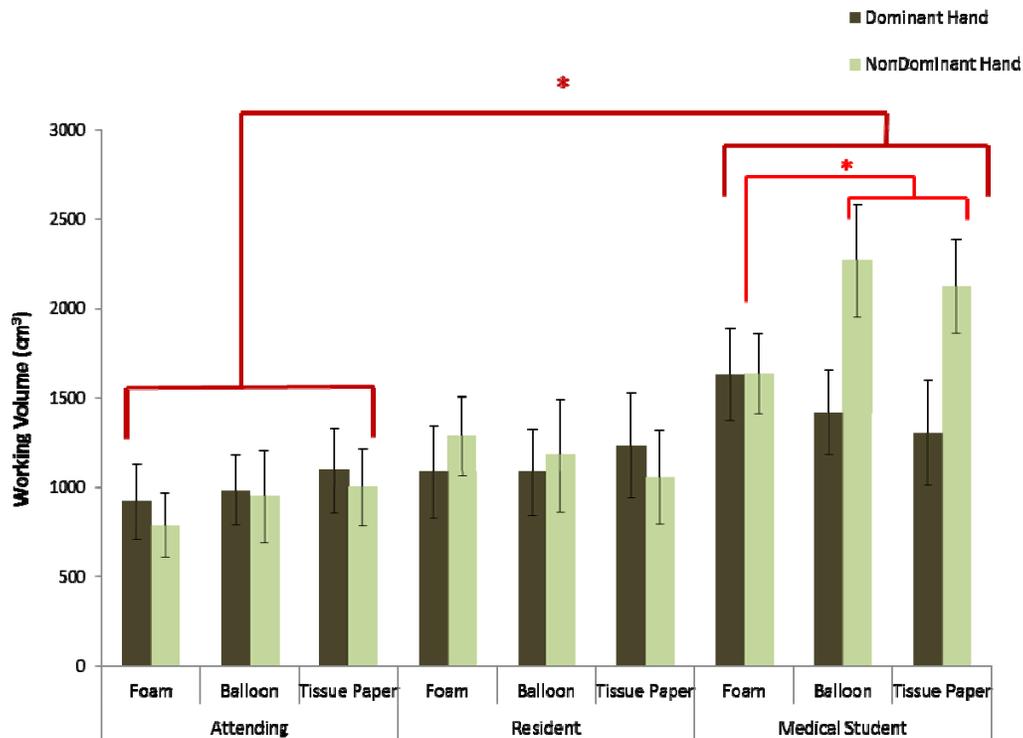
Introduction: Motion tracking technology has been used to assess psychomotor skills performance. However, more work needs to be done to better understand its potential as an assessment tool. The study aim was to evaluate working volume as a potential assessment metric. Our hypothesis is that more experienced surgeons will have smaller working volumes than less experienced surgeons.

Methods: Surgical attendings (N=6), residents (N=4) and medical students (N=5) performed a suturing task on simulated connective tissue (foam), artery (rubber balloons), and friable tissue (tissue paper). Using a motion tracking system, effective working volume was calculated based on average distance from the center position of the three-dimensional path of each hand. Analysis of variance was used to assess differences in working volume by experience level, dominant/non-dominant hand and tissue type. Videos were reviewed to provide context.

Results: Analysis revealed a linear relationship between experience and working volume. Attendings had the smallest working volume (M=954.42cm³, SD=382.84) and students had the largest (M=1728.37cm³, SD=428.03, t(1)= -3.13, p=.01). The three-way interaction of experience, hand and tissue type shows that the attending surgeons and surgical residents maintained a similar working volume between both of their hands with all three tissue types. In contrast, medical students' non-dominant hand covered significantly larger working volumes for the balloon and foam materials (F(2,11)=9.71, p=0.004; p<.05 for pairwise comparisons) (Figure 1). Qualitative video review revealed awkward postures and wide needle movements by medical students.

Conclusions: This study provides known groups validity evidence for the use of working volume as a psychomotor metric for an open surgical skill. Working volume may provide a means for assessing surgical efficiency, ability to operate in confined anatomical spaces and operating room safety regarding needle stick injuries.

Figure 1. Working volume by experience level, tissue type and dominant/non-dominant hand



* p<.05

SURGICAL EDUCATION RESEARCH FELLOWSHIP: PROGRAM OUTCOMES AND PARTICIPENT REPORTED BARRIERS TO FELLOWSHIP COMPLETION

Alfred J. Croteau, MD¹, Kimberly D. Schenarts², PhD John L. Falcone, MD, MS³, ¹University of Vermont, Burlington, VT, ²University of Nebraska Medical Center, Omaha, NE, ³University of Louisville, Owensboro, KY

Introduction: For the past 20 years the Association for Surgical Education (ASE) has offered the Surgical Education Research Fellowship (SERF)--a one year, home-site fellowship designed to equip investigators with the skills and knowledge needed to plan, implement and report research studies in surgical education. Successful completion of this fellowship requires participation in didactic sessions, working with an assigned mentor, and completion of an academic product that culminates in a peer reviewed manuscript or presentation. The purpose of this study is to evaluate the rate of SERF completion and describe participant characteristics, demographics of success, and barriers to program completion.

Methods: We electronically surveyed participants from 1994-2011 the SERF rosters and collected retrospective outcomes data from the ASE database. Survey items were developed using a modified Delphi technique. Outcome variables included completion, academic rank, effect on career development, and self-identified primary and secondary barriers to fellowship completion. Chi-square tests were used to compare completion rates, reported barriers, and differences in career development amongst participant groups.

Results: During the study period there were 151 SERF fellows. Sixty-one percent of fellows completed the program. Program completion rates did not vary by participant group (faculty, clinical fellow, resident) ($p=0.84$). Completion rate of our electronic survey was 42% with 79% of respondents being those who completed the fellowship. Eighty-three percent of fellows completing the fellowship, presented work at a peer reviewed regional or national meetings; 71% additionally published this work in a peer-reviewed journal. Key barriers noted by participants included insufficient support from advisors 28% and insufficient support from home institutions 19%. Reporting frequency for these barriers was similar amongst all participant groups ($p=0.17$). Additionally eighty-five percent of participants reported a positive impact on career development.

Conclusions: Overall the Association for Surgical Education SERF program has been success and has contributed significantly to both development of participants' careers and surgical education literature. Equal proportions of faculty, surgical fellows and residents successfully completed the SERF program. Insufficient institutional and advisor support was reported as the primary barrier to success. Careful initial advisor/advisee pairing and rigorous assessment of interval progress may boost successful outcomes.

The impact of goal setting and goal orientation on performance during a clerkship surgical skills training program

Aimee K. Gardner, PhD, Diana Diesen, MD, Deborah Hogg, BS, Sergio Huerta, MD, UT Southwestern Medical Center, Dallas TX

Introduction: Acquiring basic surgical skills before beginning a surgical clerkship rotation ensures that medical students can participate in important patient care activities and feel like a valued member of the surgical team. The purpose of this study was to integrate relevant goal setting theory and to identify if trainees' goal orientations have an impact on the assigned goals → performance relationship.

Methods: Third-year medical students (MS3s) participated in surgical skills training prior to beginning their clerkship rotation. All MS3s underwent baseline knot tying (KT) and camera navigation (CN) skill assessment and completed a demographics/goal orientation questionnaire. MS3s then participated in one of three one-hour training programs according to their rotation schedule. All began with the same video tutorial, but were provided with no specific goals and told to "do your best" and to make the most of the training time (DYB group); provided with performance goals regarding time to completion and metrics to achieve (performance group); or provided with learning goals in which they were encouraged to develop personal goals for mastering the processes required to perform well (mastery). Skills tests were again completed after training. Pre/post tests were video-recorded and de-identified for evaluation by a single blinded instructor under an IRB-approved protocol.

Results: 127 MS3s (Age: 25 ±2.6; 54% women) participated in the training program. Pre- to post-training performance changes were significant for all groups on both tasks ($p < .01$), but the increase was significantly greater ($p < .01$) for the Mastery group (DYB $KT\Delta = 2.14$, $CN\Delta = 1.69$; Performance $KT\Delta = 2.49$, $CN\Delta = 2.24$; Mastery $KT\Delta = 3.04$, $CN\Delta = 2.76$, on a 1-5 Likert scale). Hierarchical regression analyses were used to examine the interaction of goal orientation and goal condition. Results indicate that individuals endorsing a learning goal orientation perform significantly better in the mastery goal condition ($p < .01$), worse in the performance condition ($p < .05$), and exhibited no significant difference in the DYB condition. Performance-oriented individuals performed worse in the DYB condition ($p < .05$), but exhibited no significant difference in the performance or mastery conditions.

Conclusions: These data indicate that ample consideration of goal type and trainee goal orientation must be considered during curriculum development in order to maximize educational value.

Impact of Structured Video Debriefing on Trauma Teams' Ability to Self-assess Teamwork

Gene Kurosawa MS¹, Alexander Wei MS¹, Nina Ho BS¹, Eunjung Lim PhD¹, Ben Berg MD¹, Kara Terada RN², Susan Steinemann, MD^{1,2}, ¹University of Hawaii, ²The Queen's Medical Center, Honolulu, HI

INTRODUCTION: Optimal performance of multidisciplinary trauma teams requires accurate assessment and formative feedback on non-technical skills. We investigated the ability of teams to self-assess the non-technical skills of trauma resuscitation before and after video debriefing. We hypothesized that immediate post-resuscitation self-assessment would differ from that of expert observers, but might approach experts' assessment after video debriefing.

METHODS: Experienced (Level II Center) trauma nurses were taught teamwork principles and non-technical skills assessment via a web-based didactic and lecture, using five behavioral domains of T-NOTECHS (Am J Surg, 2012;203:69-75). Subsequently, groups of 3-4 nurses joined multidisciplinary teams including surgeon and emergency physician confederates, and completed four, ten-minute, videotaped human patient simulator-based blunt trauma resuscitation scenarios. Immediately after each scenario, teamwork was independently assessed by nurses (self-assessment) as well as two physician trauma teamwork experts ("experts") using T-NOTECHS. Nurses then participated in a 30-minute structured video debriefing of the scenario, and then repeated their self-assessment using T-NOTECHS. Nurses' average T-NOTECHS ratings were compared to experts' (gold standard) ratings via t-test. The study was approved by our Institutional Review Committee.

RESULTS: 34 nurses consented to participate. Immediately after each scenario (prior to video debriefing), nurses' overall T-NOTECHS ratings were significantly higher than experts ($p < .001$). Post-video debriefing, nurses' rerating of overall teamwork was not significantly different from experts. When evaluating differences by behavioral domain, nurses consistently rated all domains other than "Leadership" higher than experts. Significant differences in rating the areas of "Communication" and "Resource Management" persisted even after video debriefing ($p < .01$). Analysis of a subgroup of six nurses (18%) who had undergone simulation-based T-NOTECHS training with video debriefing five years ago, revealed mean T-NOTECHS ratings closer to the experts, with a significant difference only in the domain of "Communication."

CONCLUSIONS: Trauma team self-assessment of non-technical skills may be inaccurate. Nursing team members tend to be more generous in their estimation of team communication than expert raters. Structured video debriefing appears to improve overall teamwork assessment skills following simulated trauma resuscitation. Our findings may provide a rationale for incorporating resuscitation video review in the trauma quality improvement process.

The Surgery Core Clerkship Flipped Classroom: A Prospective Cohort Trial

Cara A. Liebert, MD; Dana Lin, MD; Laura Mazer, MD, MS; James N. Lau, MD; Department of Surgery, Stanford University School of Medicine, Stanford, CA

Background: The flipped classroom, a blended learning paradigm employing pre-session online videos reinforced with interactive small group sessions, has been proposed as an alternative to traditional didactic lectures. Although this model has been applied to pre-clinical curricula, it has not been previously applied to a surgery core clerkship.

Methods: A prospective cohort of medical students (n=38) enrolled in the surgery core clerkship from March to August 2014 was taught utilizing a flipped classroom approach, with outcomes compared to a matched retrospective cohort (n=42) enrolled in the traditional didactic curriculum from March to August 2013. The flipped classroom consisted of eight modules covering core general surgery topics, each comprised of a pre-test, pre-session online videos, in-class clinical reasoning and skills session, and post-test. Pre- and post-test performance as well as NBME shelf exam scores were analyzed to assess knowledge acquisition and application. End-of-clerkship survey data was analyzed to assess learner satisfaction.

Results: Total faculty time in the flipped classroom was reduced by 20% compared to the traditional curriculum. Mean pre- and post-test scores demonstrated significant increases across all modules. The majority of students rated the curriculum excellent or outstanding overall (89.2%) and with respect to individual curriculum components (online video content, 89.2%; in-class clinical reasoning, 86.5%; in-class surgical skills sessions, 97.3%). Students reported the interactive in-class sessions as the most effective components. The vast majority (94.6%) agreed or strongly agreed that the new curriculum should be continued. There was no statistically significant difference between mean NBME exam score of the study cohort compared to the historical cohort (73.8 vs 76.3, p=0.19). Medical student interest in pursuing surgery as a career increased significantly after completion of the new curriculum (3.29 to 6.13, p=0.0003, 10-point Likert), with 85.7% reporting the flipped classroom contributed to this increase.

Conclusions: This study demonstrates that implementation of a flipped classroom in the core clerkship is not only feasible, but results in high learner satisfaction and knowledge acquisition. Students reported increased interest in surgery as a result of the curriculum. The flipped classroom reduced faculty time commitment, with similar NBME exam scores compared to the traditional didactic model.

A Qualitative Analysis of a Novel Curriculum on Mistreatment for the Surgery Core Clerkship

Mazer, Laura MD MS; Liebert, Cara MD; Lin, Dana MD; Lau, James MD FACS., Stanford University, Stanford, CA

Background: Almost half of medical students report having experienced mistreatment during clinical rotations. This problem is particularly prevalent on surgical clerkships. Previously described interventions have not resulted in a significant decrease in reports of mistreatment.

Method: We initiated a pilot program to address mistreatment and change the culture of the surgical clerkship at our institution. The program consists of two sessions during the first and last week of the 8-week core clerkship. In the first session, students discuss their definitions of and experiences with mistreatment. They watch video vignettes that challenge their definitions and set expectations for real world clinical teaching, followed by discussion. The second session is a debrief, allowing students to analyze and reflect on experiences. The clerkship director proctors both sessions, emphasizing the administration's support for students facing mistreatment. Rather than focus on policies or procedures, this program focuses on engaging students in the dialogue. Students are empowered to define mistreatment, given a better understanding of the perspective of attendings and residents, and shown that the administration is committed to their wellbeing. Following the debrief, students completed anonymous open-ended written reflections regarding their impressions of the program. A qualitative analysis of student responses was performed using a modified grounded theory approach.

Results: We identified several recurring themes in student evaluations of the pilot program (n=17). All respondents felt the program had a positive impact. They noted the value of the program in its ability to: establish expectations (53%); allow for shared experiences (29%); raise awareness of resources (35%); and provide emotional support (24%). As a result of the program, students felt that the learning environment and rotation culture was improved (59%), and that they or their peers would be more likely to consider surgery as a specialty (18%).

Conclusions: Mistreatment is an intensely personal experience, and its potential deleterious effects on emotional wellbeing and clinical learning often go unseen. A rotation-specific mistreatment program focused on engaging students and creating a shared understanding was universally well received among core clerkship students in surgery, and may help to establish expectations, provide emotional support, and improve the clinical learning environment.

Competency-Based Medical Education: Can Both Junior and Senior Residents Achieve Competence After A Module?

Tim Dwyer, University of Toronto

Background: The implementation of competency-based medical education (CBME) as the orthopedic resident training format at the University of Toronto has moved post-graduate training away from a time-based training model, to a model based upon observable and measurable outcomes. We hypothesized that with intensive training to a set curriculum that junior and senior residents would be able to demonstrate and apply medical knowledge to an equivalent level.

Method: OSCE All residents undertaking a three-month sports medicine rotation were expected to pass a six station OSCE prior to moving to the next rotation. All stations were marked with a binary station-specific checklist and a 5-point overall global rating scale (Drefus). A global rating was also given for each domain of knowledge (history-taking, examination, image interpretation, clinical decision-making, consent, surgical technique).

Results: OSCE Over 18 months, 39 residents (21 junior, 18 senior) sat the end of rotation examination. A further six fellows also participated, for a total of 45 participants. The Cronbach's alpha of the six stations was 0.87. Analysis using a two-tail t test demonstrated a significant difference between junior and senior residents in both total checklist scores (%) and overall global ratings ($p < 0.01$). With regards to individual knowledge domains, there was also a significant difference between junior and senior residents for each (history $p < 0.05$, remainder $p < 0.01$).

Conclusion: Despite identical intensive teaching to a set curriculum within a CBME model, junior residents were not able to demonstrate or apply knowledge as well as senior residents, suggesting that overall experience is a critical factor.

Research priorities in surgical simulation: what has been achieved during the 21st century and what remains?

Maximilian J Johnston, John T Paige, Dimitrios Stefanidis, Rajesh Aggarwal, Shawn Tsuda, Ankur Khajuria, Sonal Arora on behalf of the Association for Surgical Education Simulation Committee.

Background: Ten research priorities for surgical simulation have been identified in recent years¹. The aim of this study was to establish what progress has been made regarding exploration and implementation of these research priorities and what still remains to be achieved.

Methods: Members of the Association for Surgical Education Simulation Committee conducted individualized literature reviews for each of the ten research priorities. An expert panel then collated the results to give a comprehensive overview of the current landscape for each simulation research priority.

Results: Excellent progress has been made in the development of metrics to assess technical, non-technical and teamwork skills in simulation. Whilst some of these have been translated into surgical curricula, this is not uniform across training and implementation differs within the surgical subspecialties. Furthermore, the best methods of feedback and debriefing to trainees and trainers have not yet been established. Arguably the three most important research questions are: 1. Does simulator competence equal clinical competence? 2. Does training on simulators transfer to improved clinical performance? 3. Does simulator training lead to improved patient outcomes? Unfortunately, these questions are also the most difficult to answer. Progress in answering these complex questions is comparably slower than other questions. Direct linkage of simulation training to improved patient outcome is seen as the holy grail of simulation research and this remains elusive at the present time.

Conclusions: Significant progress has been made in the areas of skills assessment and curriculum development using simulation. Further research into feedback, debriefing and decision-making is required to continue the promising progress in these areas. Whilst some progress has been made in analyzing the transfer of simulation competence to the clinical environment and the impact of simulation training on patient outcomes, there is much work to still be done in these areas. They should be the focus of simulation research in the years to come.

References: 1. Stefanidis D, Arora S, Parrack DM, et al. Research priorities in surgical simulation for the 21st century. *Am J Surg* 2012;203:49-53.

Young Surgeons on Speaking Up: When and How Surgical Trainees Voice Concerns About Supervisors' Clinical Decisions

Malini D. Sur MD^{1,2}, Nancy Schindler MD, MHPE^{1,3,4}, Puneet Singh MD¹, Peter Angelos MD, PhD^{1,2,5}, Alexander Langerman MD^{1,2,5,6,7}, ¹Department of Surgery, University of Chicago Medicine, Chicago, IL, ²MacLean Center for Clinical Medical Ethics, University of Chicago, Chicago, IL, ³Pritzker School of Medicine, University of Chicago, Chicago, IL, ⁴NorthShore University HealthSystem, Evanston, IL, ⁵Bucksbaum Institute for Clinical Excellence, University of Chicago, Chicago, IL, ⁶Department of Otolaryngology-Head and Neck Surgery, University of Chicago Medicine, Chicago, IL, ⁷Operative Performance Research Institute, University of Chicago Medicine, Chicago, IL

Introduction: Poor team communication is recognized as a major contributor to disasters within the aviation and health care industries. Flight crewmembers are trained to identify, verbalize, and escalate concerns about superordinates' plans, but literature exploring surgical trainees' responses to analogous clinical concerns is sparse. We sought to characterize how surgical trainees manage concerns about supervisors' clinical decisions.

Methods: Senior surgical residents at our institution were interviewed between April and July 2014 using a semi-structured interview guide. Residents' sense of responsibility for patient outcomes, factors affecting willingness to raise concerns, types of situations triggering concerns, and techniques utilized to verbalize concerns were elucidated. Transcripts initially generated by Dragon Naturally Speaking dictation software were reviewed for accuracy. Emerging themes were organized using NVivo 10 qualitative analysis software. Interviews were conducted until saturation was reached. A sample was independently coded for validation.

Results: 18 residents (11 female, 7 male) across 4 subspecialties participated: general surgery (8), urology (4), obstetrics and gynecology (3), and otolaryngology (3). The reported sense of responsibility for patient outcomes varied, with some residents conceding ultimate decision-making authority to supervisors ("it's their patient") and others emphasizing their own obligations to the patient ("the patient comes first"). Willingness to voice concerns was influenced by factors at multiple levels: systemic (i.e. departmental culture, degree of resident autonomy), supervisor (i.e. personality), trainee (i.e. knowledge base), and clinical context (i.e. severity of potential harm, strength of relevant evidence). Reported concerns commonly pertained to pre-operative workup, informed consent, and management of complications. Residents typically described verbalizing concerns in question-form to appear respectful and motivated by educational interests. Some favored stating concerns more directly, with a few even refusing to execute plans they disagreed with. Most felt they could escalate concerns to a program director or chairperson if necessary. None reported having had formal instruction in how to manage clinical concerns.

Conclusion: Several factors influence how surgical trainees voice concerns about supervisors' clinical decisions and no consistent approach is used. Given the success of didactics in other hazardous industries, a formal communications curriculum addressing surgical trainees' approach to raising concerns appears warranted to improve patient safety.

GOOGLE GLASS VIDEO FOR BEDSIDE PROCEDURAL SKILL ASSESSMENT IS INFERIOR TO THIRD PERSON VIDEO EVALUATION

Dylan O'Shea, Amy Morris, Kari Keys, Andrew Wright, Jonathan Ilgen, Heather Evans; Departments of Surgery, Medicine and Emergency Medicine, University of Washington, Seattle, WA

Introduction: Google Glass (GG) is a hands-free wearable computer that can rapidly obtain videos from an operator's first-person perspective. Such technologies may offer novel opportunities for trainee procedural assessment. The purpose of this pilot study was to assess the feasibility of using Glass video recording to evaluate trainee procedural skills, as compared to conventional third-person recording. We hypothesized that observers reviewing videos from this unobstructed perspective would rate procedural completion and trainee skill level with greater consistency.

Methods: Six surgical residents and one attending surgeon performed simulated internal jugular (IJ) catheter insertions using a partial task trainer. Procedures were recorded using GG and by a head-mounted camera on an observer, generating both a first-person (1P) and third-person (3P) recording of each procedure. Anonymized video recordings were assessed independently by four trained expert raters using a task-specific checklist and a global rating scale (GRS) described previously (Ma et al. *Adv Health Sci Educ Theory Pract*, 2012). Inter-rater reliabilities were calculated using intraclass correlation coefficients (ICC) for each item and for the checklist and GRS overall. Differences in inter-rater reliabilities using 1P and 3P perspectives were calculated (1P-3P) for each item to illustrate the relative performance of ratings under each video recording condition.

Results: Mean and median ICC differences for the checklist were -0.18 and -0.22, respectively. Mean and median ICC differences for the GRS were -0.08 and -0.10, respectively. Inter-rater reliabilities were consistently higher when assessments were completed under the 3P condition, and the greatest differences in reliability between the 1P and 3P recordings concerned items on the task-specific checklist, such as assessing the needle angle (-0.49) and visualizing the needle tip on ultrasound (-0.62).

Conclusion: Despite the allure of assessing procedural skills using first-perspective GG videos, inter-rater reliabilities were consistently lower than when assessments were completed using traditional video recordings. These pilot findings suggest limitations of these first-perspective recordings (such as trainee head movements, limited peripheral vision, and camera angles) and/or rater unfamiliarity with this novel assessment perspective

Feasibility of Force Feedback Vessel Ligation Simulator in Knot-tying Proficiency Training

Justin Hsu BS¹, James R. Korndorffer MD, MHPE, FACS², Kimberly M. Brown MD, FACS¹; ¹University Of Texas Medical Branch, Department of Surgery, Galveston, TX, USA; ²Tulane University School Of Medicine, Department of Surgery, New Orleans, LA, USA

Introduction: Current best practice for surgical skills acquisition entails deliberate practice to proficiency in the skills lab in preparation for clinical cases. Tying secure knots without exerting excessive force is an important skill for open procedures in which excessive bleeding may result from rough handling of tissues. We have developed a force feedback simulator that measures instantaneous force exerted on a blood vessel analog during knot tying and displays visual feedback of force exerted. This pilot study examines the benefits of using such simulator as part of a deliberate practice curriculum.

Methods: The vessel ligation simulator consists of a Silastic tubing suspended horizontally, and a Vernier Dual-Range Force Sensor perpendicularly attached to the tubing. Knot quality was assessed using a digital caliper measurement of knot diameter. Two-handed vessel ligations were each made with 2 square knots using 3-0 Silk sutures on the tubing at a fixed offset to the sensor to measure the vertical force exerted. Third-year medical students performed 10 vessel ligations as pre-test. Then, a display with instantaneous force data and maximum force threshold ($\pm 0.6\text{N}$, pre-determined by expert performance) was shown while the participant continued vessel ligations until proficiency, defined as 3 consecutive ligations made below the maximum force threshold and below the knot diameter cutoff (1.15mm). The display was removed and 10 ligations were performed as post-test. Average peak forces for pre-tests and post-tests were compared using Student's t-test.

Results: Students exerted significantly lower average peak forces on the post-test compared to pre-test ($0.55 \pm 0.17\text{ N}$ vs. $1.33 \pm 0.51\text{ N}$, $p < 0.01$). There were no significant differences between pre-test and post-test ligations that failed the knot diameter cutoff (3.6 ± 2.1 failures vs. 1.4 ± 0.5 failures, $p = 0.13$). Mean ligations made until proficiency was 12 ± 5 knots and mean time taken was 17 ± 7 minutes.

Conclusion: This pilot study demonstrates feasibility of real-time feedback in the acquisition of expert level skills of minimal force exertion during quality vessel ligations with visual feedback. The curriculum can be completed in a reasonable amount of time, and may also work as a warm-up exercise prior to a surgical case.

Development, implementation, and outcomes of an evidence-based surgery clerkship clinical skills exam

Jesse Moore, MD FACS, Cate Nicholas, MS PA EdD University of VT College of Medicine

Background: End of clerkship multiple-choice exams and oral exams allow assessment of a student's cognitive knowledge and clinical reasoning. Simulation provides a unique opportunity to assess the clinical skills acquired during a surgery clerkship. We developed an evidence based clinical skills exam (CSE) for the end of the surgery clerkship using standardized patients.

Methods: A literature search identified clinical skills and content that surgeons and primary care physicians felt should be learned by the end of a surgery clerkship. Five content areas showed significant overlap: breast mass, acute abdomen, GI bleeding, shock, and bowel obstruction. Five skills overlapped: abdominal exam, resuscitation, breast exam, writing progress notes, and writing admission orders. Existing CSEs or OSCEs available from the ACS Division of Education, MedEd Portal, and internal oral exams were modified to create 5 standardized patient (SP) encounters. One patient encounter includes a faculty administered oral exam. SPs completed checklists for the history taking, physical exam, professionalism, and communication skills domains. Faculty provide feedback on presentation skills, oral exam, written notes, and orders.

Results: Sixty two students have completed the CSE. Scores in each domain were calculated across cases. Mean scores were 98% for professionalism, 97% for communication, 87% for physical exam, and 81% for history taking. 6% (n = 4) of students failed to meet the passing standard in history taking (n=3) or physical exam (n=1) and remediated. 51-72% of students missed elements of the social history. Heart and lung exams were not satisfactorily performed in 55-83% of cases.

Conclusions: An evidence based CSE utilizing SPs can be developed for a surgical clerkship. The high pass rate indicates students are learning the necessary skills and content. Most exam failures are related to history taking; with elements of the social history being the most frequently neglected.

Defining “Honors” in the Surgery Clerkship

Jeremy M. Lipman MD¹, Kimberly D. Schenarts PhD², ¹MetroHealth Medical Center, Case Western Reserve University School of Medicine, Cleveland, OH, ²University of Nebraska Medical Center, Omaha, NE

Background: Although highly influential, no published criteria exist defining who should receive the highest grade in the surgery clerkship (“Honors”). Therefore, significant variability exists in how this evaluation is assigned. Identifying the critical characteristics of the student receiving this grade may improve its usefulness in residency selection, class standing and focusing students’ efforts. The purpose of this study was to attain expert consensus on the characteristics of an honors student in the third year surgery clerkship.

Methods: A three round modified Delphi technique was used in two parallel cycles to obtain expert consensus from the two major stakeholders; program and clerkship directors in surgery. Experts were recruited from across the United States, though not from the same institutions. Program directors first listed the characteristics they felt critical for a student to demonstrate to receive an honors grade which would not be affected by time of year or prior clinical experience. The experts then ranked the characteristics in order of importance. This list was then revised in the third round to achieve consensus. The clerkship directors then repeated this entire process.

Results: All 15 of the invited clerkship directors and 14/15 of invited program directors participated. A total of 65 unique characteristics were submitted by program directors who achieved consensus on 23. Clerkship directors submitted 62 and achieved agreement on 22 characteristics. Ten of the final characteristics were identical between the two groups. These were: communication skills, “shelf” exam score, synthetic ability (organizing data into meaningful care plans), absence of professionalism issues, outstanding work ethic, takes advantage of learning opportunities, accurate and complete history and physicals, enthusiasm, becomes an essential member of the care team and outstanding clinical acumen.

Conclusion: Expert consensus on the characteristics of an honors student in the surgery clerkship was achieved, accounting for the opinions of both clerkship and program directors in surgery. By using these criteria to define an honors student, the grade becomes emblematic of these ten characteristics deemed critical by both major stakeholders. This may reduce grade inflation within and between institutions and provide program directors with a consistent and reliable assessment of excellence.

Needlestick Injuries among Medical Students, Surgical Residents, Faculty, and Surgical Staff at One Institution

Lynn Y. Choi M.D., Rosalicia Torres, Soheil Syed, Ashar Ata, Carl Rosati M.D., Albany Medical Center, Albany, NY

Surgical employees are at high risk for needlestick injuries and are at fault for underreporting these exposures, due to time. In this study, we determine factors that lead to a lack of reporting needlestick injuries. This is the first study to compare the incidence of sharps injuries among all staff involved in the operating room at one institution— medical students, surgical residents, faculty, and surgical staff. IRB approval was obtained and an anonymous electronic survey was distributed to third and fourth year medical students, surgical residents (general and subspecialties), general surgery attendings, surgical technicians and nurses at Albany Medical Center; assessing needlestick injury history, whether they were reported, and why they were not reported. The overall survey response rate was 37% (196/528). Among the four groups 55% (107/195) had a needlestick injury. Survey of action taken after first sharps injury showed 85% (29/34) of surgical technicians and nurses, 64% (16/25) of attendings, 50% (16/32) of residents and 44% (7/16) of medical students reported the incident; overall 64% (68/107) reported their injuries. Surgical staff had an incidence rate ratio (IRR) of 1.33 ($p = 0.085$) when compared to attendings. Survey of fear of admitting a needlestick injury was observed by 69% (11/16) of medical students, 53% (17/32) of residents, 24% (6/25) of attendings and 26% (9/34) of surgical staff. Compared to attendings, medical students had an IRR of 2.86 ($p=0.008$) and residents had an IRR of 2.21 ($p=0.04$). 64% and 66% of respondents attributed not reporting their exposure due to the time consuming process and/or the patient was low risk, respectively. To our knowledge this is the first study to include surgical ancillary staff and furthermore we demonstrated that the surgical ancillary staff are 1.33 times more likely to report their needlestick injuries compared to attendings. The prevailing hypothesis for this is that there are sufficient resources to substitute ancillary staff, whereas it is more difficult to find replacement staffing for surgical residents and attendings. The key factors in underreporting of needlestick injuries at our institution are due to time and fear. Further research is necessary to mitigate these factors.