The Use of Error Analysis to Assess Resident Performance

Anne-Lise D. D'Angelo, MD[1], **Katherine E. Law, MS**[2], **Elaine R. Cohen, MEd**[1], **Jacob A. Greenberg, MD, EdM**[1], **Calvin Kwan, BS**[1], **Caprice Greenberg, MD, MPH**[1], **Douglas A. Wiegmann, PhD**[2], and **Carla M. Pugh, MD, PhD**[1]

[1]University of Wisconsin-Madison, School of Medicine and Public Health, Department of Surgery

[2]University of Wisconsin-Madison, School of Engineering, Department of Industrial and Systems Engineering

Abstract

Background—The aim of this study is to assess validity of a human factors error assessment method for evaluating resident performance during a simulated operative procedure.

Methods—Seven PGY4-5 residents had 30 minutes to complete a simulated laparoscopic ventral hernia (LVH) repair on Day 1 of a national, advanced laparoscopic course. Faculty provided immediate feedback on operative errors and residents participated in a final product analysis of their repairs. Residents then received didactic and hands-on training regarding several advanced laparoscopic procedures during a lecture session and animate lab. On Day 2, residents performed a nonequivalent LVH repair using a simulator. Three investigators reviewed and coded videos of the repairs using previously developed human error classification systems.

Results—Residents committed 121 total errors on Day 1 compared to 146 on Day 2. One of seven residents successfully completed the LVH repair on Day 1 compared to all seven residents on Day 2 (p=.001). The majority of errors (85%) committed on Day 2 were technical and occurred during the last two steps of the procedure. There were significant differences in error type (p=<.001) and level (p=.019) from Day 1 to Day 2. The proportion of omission errors decreased from Day 1 (33%) to Day 2 (14%). In addition, there were more technical and commission errors on Day 2.

Conclusion—The error assessment tool was successful in categorizing performance errors, supporting known-groups validity evidence. Evaluating resident performance through error classification has great potential in facilitating our understanding of operative readiness.

Keywords

Assessment; Simulation; Human Factors; Error; Surgery Education

Corresponding Author: Anne-Lise D. D'Angelo, Department of Surgery, University of Wisconsin Hospital and Clinics, 600 Highland Ave., H4-395 CSC, Madison, WI 53792, 608-262-0446, amaag@uwhealth.org.

Background

Recent literature has called into question whether graduating general surgery residents are ready for operative independence.[1,2] This question continues to fuel the debate regarding the restructuring of general surgery curricula.[3] However, the majority of this data is based on anecdotal evidence and opinion surveys that can be subject to bias.[1–2,4–6] Before we can make crucial decisions regarding changes to the surgical residency curricula, we must develop appropriate assessment tools to explicitly and objectively evaluate residents' ability to operate independently and define knowledge gaps and skill deficiencies.

Current assessment tools include task-specific and global rating scales for evaluating technical[7]; operative[8] and non-technical[9,10] skills. The majority of assessment tools for surgical performance are based on procedure time, and subjective and objective assessments of technical skill.[11] Although rating scales provide a framework for assessing operative ability, they do not provide specific information regarding the causes and characteristics of failures in performance. Operative errors committed by residents may result from deficits in knowledge and skill; however, errors can also result from lapses in attention and difficulty with perception.[12] Further confounding the issue, the operating room is a complex environment thus, contributing to the difficulty in determining and evaluating the causes of intra-operative errors. An assessment tool that incorporates the evaluation of error characteristics may better help to identify specific deficits and complexities in operative training.

The field of human factors engineering (HFE) has a long history in developing error-based assessments to improve human performance in high-risk and complex industries including aviation[13] and nuclear energy.[14] HFE is the scientific discipline concerned with understanding interactions among humans and other elements of a system in order to optimize human well-being and overall system performance.[15] Work in this field led to the development of classification frameworks designed to analyze the causes of errors and define generic classifications for human error mechanisms.[12–13,16] The Cognitive Error Taxonomy is a framework that allows for in depth analyses identifying which decision-making steps were made in error or simply bypassed out of habit.[13,16] Using these error based frameworks in a high-risk and complex environment, such as the operating room, allows us to identify how failures are related to cognitive processes.[12]

Simulation allows for error analysis without the cost of morbidity and mortality to patients. Our prior work involves design and fabrication of decision-based operative simulators to analyze intraoperative performance and decision making.[17] This study implements HFE-based error classification methods to investigate the causes and types of errors that trainees commit while also evaluating how they recover from these errors. The study aim is to assess validity of a human factors error assessment method for evaluating resident performance during a simulated operative procedure. Our hypothesis is that the error assessment method will detect changes in performance based on procedure step and participant knowledge thus, supporting known groups validity.

Methods

Setting and Participants

This study was a post-hoc comparative analysis of resident performance during a 2008 Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) course on Advanced Laparoscopic Hernia Surgery. The course occurred over a two-day period and seven senior residents (PGY4-5) participated in the study. The University of Wisconsin Institutional Review Board granted this study exempt status.

Research Protocol

During the course, residents had 30 minutes to complete a simulated laparoscopic ventral hernia (LVH) repair using a previously developed box-trainer style simulator.[17] On Day 1 of the course, the simulator contained a midline 10×10cm ventral hernia located 5 cm above the umbilicus. Surgical faculty acted as surgical assistants during the procedure and provided feedback upon completion. The feedback was not standardized, but catered to individual resident performance during the procedure. As part of the feedback, residents participated in a final product analysis of their repairs with faculty. Residents then received didactic and hands-on training through lectures and an animate lab. On Day 2, residents performed a nonequivalent LVH repair on a similar trainer. The simulator contained a 10×10cm ventral hernia located in the right upper quadrant. Residents also received individualized feedback for this repair. Information on resident demographics and laparoscopic experience were collected. In addition, all simulated procedures were video-recorded using both external and laparoscopic views. Months after the course, a blinded faculty member analyzed and individually graded resident's final hernia repair for completeness.

Laparoscopic Ventral Hernia (LVH) Simulator

The simulator is a box-style trainer with a plastic base covered by a fabric skin that contains a ventral hernia defect. It was created with the intent to induce decision making while performing a LVH repair. The simulator was designed to look like the abdominal cavity, specifically from the diaphragm to the upper plane of the pelvic cavity. The simulator is fabricated to represent the layers of the abdominal wall including: skin, subcuticular tissue, and peritoneum. Inside, simulated organs including bowel and omentum are layered within the cavity to provide the appearance of gross anatomy. The simulator allows for the use of both laparoscopic and open surgical instruments, except cautery, during each step of the procedure. Examples at each procedure step are shown below in Figure 1.

Post-hoc analysis

Three investigators reviewed the previously recorded LVH repairs from the SAGES course in order to identify, classify and categorize residents' errors. According to the Bellagio Conference on Human Error, errors were defined as "…something that has been done which was: (i) not intended by the actor, (ii) not desired by a set of rules or an external observer, (iii) that led the task or system outside acceptable limits".[18, p.25]

Using previously developed human error classifications, errors were categorized based on type (omission versus commission) and level (cognitive versus technical).[13] *Omission* errors were defined as failure to perform a step entirely. *Commission* errors represented failure to perform a step correctly. For example, failure to measure the hernia defect was categorized as an *omission* error, whereas measuring the hernia defect with an inaccurate method was categorized as a *commission* error. Errors in information, diagnosis, and strategy were categorized as *cognitive* and errors in action, procedure, or mechanics were classified as *technical* (Table 1). Day 1 and Day 2 comparisons were performed for error type and level as well as hernia repair completion rates. It was expected that many of the rule-based errors, which effect both level and type, would be corrected by day two. For this procedure, rule based errors include those relating to procedure steps including memory and execution.

Multimedia Video Task Analysis™

Investigators used commercially available software, Multimedia Video Task Analysis™ (MVTA™) (University of Wisconsin-Madison) to code and quantify errors during the repair. MVTA™ is a user interface designed to allow simultaneous video review and data annotation and categorization. Using this program, the investigators captured time of error occurrence, duration or error, and temporal relationship of errors to procedure step (Figure 2).

Data Analysis

Paired samples t-tests were used to compare the number of errors per participant from Day 1 and Day 2. Chi-square tests were used to compare Day 1 and Day 2 for dichotomous data: (1) error types, (2) error levels, and (3) completion rate. All analyses were performed using IBM SPSS Statistics Version 21.0 (IBM Corp, Armonk, NY) and a p-value < .05 was considered significant.

Results

Seven surgery residents completed the full two-day laparoscopic hernia course. Three residents were PGY-4s and four residents were PGY-5s. Six out of seven residents were male. On average, residents reported performing 6.86 (SD=5.01) LVH procedures prior to the course.

Residents committed 121 total errors on Day 1 compared to 146 on Day 2. Errors ranged from 11 to 23 errors per resident on Day 1 compared to 14 to 32 errors per resident on Day 2. One of seven residents successfully completed the LVH repair on Day 1. The six residents that did not complete the LVH repair on Day 1 only reached the mesh orientation step. On Day 2, all seven residents successfully completed the procedure (p=.001).The proportion of omission errors decreased from Day 1 (33%) to Day 2 (14%). There were significant differences in error type (p=<.001) and level (p=.019) from Day 1 to Day 2.

Residents committed fewer cognitive, but more technical errors on Day 2. Details of intraoperative errors on Day 1 and Day 2 are illustrated in Table 2. For example, common cognitive errors on Day 1 included failure to place anchoring sutures into the mesh, failure to properly visualize the hernia defect, and ineffective strategies to roll and insert the mesh.

On Day 2, common technical errors included multiple attempts to pass suture into the Endo Close™ device and tacking without visualization. These errors occurred during steps that the majority of residents did not reach on Day 1.

Figure 3 demonstrates the proportion of cognitive versus technical errors during each step of the procedure. Residents committed fewer technical errors on Day 2 during the steps previously reached on Day 1. Technical errors increased on Day 2 during the steps that were not previously reached on Day 1. The majority of errors (85%) committed on Day 2 were technical.

Discussion

Existing tools for assessing surgery resident performance largely focus on assessment of manual techniques and procedure time.[11] Most of these assessment tools fail to capture the causes of resident's performance failures, which may provide information regarding knowledge gaps and skill deficits. HFE error-based assessment methods are designed to evaluate the causes for performance failures in complex environments, such as the operating room. The aim of this study was to assess validity of a human factors error assessment method for evaluating resident performance during a simulated operative procedure, the LVH repair.

In this study, residents performed a simulated LVH repair on Day 1 and Day 2 of a SAGES course. A key finding was that the number of errors increased from Day 1 (N=121) to Day 2 (N=146 total errors), despite the fact that more residents were successful in completing the hernia repair. Further analysis revealed that 85% of the Day 2 errors happened during steps that the majority of residents never reached on Day 1. This finding underscores that operative success does not necessarily equate to less errors and that final product analysis alone may miss opportunities for improvement. This finding also supports our ability to use the error assessment method as a valid measure of performance as there is a strong correlation between error type and procedure step as predicted.

Additional analysis demonstrated that the type of errors committed also changed from Day 1 to Day 2. On Day 2, there were fewer omission errors (failure to perform a step entirely), indicating that residents remembered to perform more steps of the procedure. However, despite improvements in memory, there were more commission errors (failure to perform a step correctly). This pattern, decreased omission errors and increased commission errors, represents a logical progression in knowledge acquisition. On Day 1, residents may be unaware of particular steps of the procedure resulting in omission errors and poorer performance. After feedback and instruction, residents may be cognizant of the procedure steps required, but may not have had the opportunity to actually perform the technical aspects. Resident may have additional commission errors on Day 2 as they are attempting to complete the previously unperformed procedure steps. This progression of events supports our hypothesis that the error assessment method will detect changes in performance based on procedure step and participant knowledge (including memory). These findings support content and known groups validity for the HFE error assessment method.

Lastly, there was also a shift in the level of errors performed by residents. On Day 2, residents committed fewer cognitive errors. While we are unsure of the exact scientific and theoretical basis accounting for this observation it is likely to be multi-factorial. The improvement may reflect that residents were able to incorporate faculty feedback and self-assessment into their subsequent performance. Improved memory, better strategy and incorporation of new rules are all possibilities. It is also possible that the residents were more comfortable with the simulated setting and the simulator. However, it is unlikely that this accounted for all of the improvement as some of the most disruptive cognitive errors related to port-placement strategy (distance from hernia) and placement of anchoring sutures on the mesh. Rasmussen's skills, rules, knowledge-based error framework also helps to explain some of our observations regarding cognitive errors.[16] A number of the strategy type of cognitive errors resulted from forgetting or lack of prior knowledge of general principles and rules. Rule based errors are easier to correct than some skill and knowledge-based errors. As such, many rule based errors do not require long hours of repetitive practice for improvement. Having an understanding of which deficits require high volume repetition versus short focused interventions may allow us to better design resident training curricula.

Overall, the HFE error-based assessment method provided information about resident performance that was not previously quantifiable with checklists and surveys that focused on technical skill. This study demonstrates the feasibility of applying an error-based assessment method to evaluations of resident performance in simulated environments. In addition, we were able to detect changes in the nature of resident performance that would not have been evident through final product analysis. Moreover, our results show that it is necessary to take a broader look at performance beyond operative case logs, checklists, and surveys when assessing resident performance and evaluating deficits in intra-operative knowledge and skill.

Study limitations include those inherent in qualitative research methods. Deductive reasoning and categorization based on observations is prone to observer bias.[19] In addition, there are limitations in the use of simulation to assess operative performance. Despite the limitations, this study shows promising results and provides a basis for additional research categorizing errors committed during surgical procedures. The focus on error type (commission vs. omission) and error level (cognitive vs technical), helped to support our hypothesis; however, there are a number of additional classifications that may further define the nature of the errors committed. In addition, future work may help us to better understand the underlying reasons for residents' intraoperative errors. A closer analysis of cognitive errors and how this relates to performance over time is also warranted. Moreover, additional work is needed to evaluate error detection and management. Prior literature has demonstrated that it is not only the number of errors committed that correlates with poor patient outcomes, but also how surgeons detect and manage errors.[20]

As the debate about the length and structure of surgical residency curricula continues, we must design assessment tools that provide quantifiable data on resident performance. To assess resident readiness, we need to provide trainees with the opportunity to operate independently and commit errors thus, facilitating a better understanding of deficits in knowledge and skills. Our application of HFE-based error assessment in a simulated

Author Manuscript

environment has demonstrated measurable changes in resident performance not previously quantified.

Acknowledgments

References

1. Mattar SG, Alseidi AA, Jones DB, et al. General surgery residency inadequately prepares trainees for fellowship: results of a survey of fellowship program directors. Ann Surg. 2013; 258:440–449. [PubMed: 24022436]

2. Richardson JD. ACS transition to practice program offers residents additional opportunities to hone skils. The Bulletin of the American College of Surgeons. 2013 Available at http://bulletin.facs.org/2013/09/acs-transition-to-practice-program-offers-residents-additionalopportunities-to-hone-skills/.

3. Hoyt BD. Executive Director's report: Looking forward - February 2013. Bulletin of the American College of Surgeons. 2013 Available at http://bulletin.facs.org/2013/02/looking-forward-february-2013/.

4. Bucholz EM, Sue GR, Yeo H, Roman SA, Bell RH Jr, Sosa JA. Our trainees' confidence: results from a national survey of 4136 US general surgery residents. Arch Surg. 2011; 146:907–914. [PubMed: 21844434]

5. Coleman JJ, Esposito TJ, Rozycki GS, Feliciano DV. Early subspecialization and perceived competence in surgical training: are residents ready? J Am Coll Surg. 2013; 216:764–771. discussion 71–3. [PubMed: 23521960]

6. Nakayama DK, Taylor SM. SESC Practice Committee survey: surgical practice in the duty-hour restriction era. Am Surg. 2013; 79:711–715. [PubMed: 23816005]

7. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997; 84:273–278. [PubMed: 9052454]

8. Larson JL, Williams RG, Ketchum J, Boehler ML, Dunnington GL. Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. Surgery. 2005; 138:640–647. discussion 7–9. [PubMed: 16269292]

9. Sharma B, Mishra A, Aggarwal R, Grantcharov TP. Non-technical skills assessment in surgery. Surg Oncol. 2011; 20:169–177. [PubMed: 21129950]

10. Yule S, Paterson-Brown S. Surgeons' non-technical skills. Surg Clin North Am. 2012; 92:37–50. [PubMed: 22269259]

11. Schmitz CC, DaRosa D, Sullivan ME, Meyerson S, Yoshida K, Korndorffer JR Jr. Development and verification of a taxonomy of assessment metrics for surgical technical skills. Acad Med. 2014; 89:153–161. [PubMed: 24280844]

12. Reason, J. The human contribution: Unsafe acts, accidents and heroic recoveries. Burlington, VT: Ashgate; 2008.

13. Hooper BJ, O'Hare DPA. Exploring human error in military aviation flight safety events using post-incident classification. Aviation, Space, and Environmental Medicine. 2013; 84:803–813.

14. O'Connor T, Papanikolaou V, Keogh I. Safe surgery, the human factors approach. Surgeon. 2010; 8:93–95. [PubMed: 20303890]

15. [Accessed March 6th 2014] Human Factors and Ergonomics Society. About HFES. Available at https://www.hfes.org//Web/AboutHFES/about.html.

16. Rasmussen J. Human errors: A taxonomy for describing human malfunction in industrial installations. Journal of Occupational Accidents. 1982; 4:311–333.

17. Pugh C, Plachta S, Auyang E, Pryor A, Hungness E. Outcome measures for surgical simulators: is the focus on technical skills the best approach? Surgery. 2010; 147:646–654. [PubMed: 20403517]

18. Senders, JW.; Moray, N. Organization NAT. Human Error: Cause, Prediction, and Reduction. Taylor & Francis; 1991.

19. Angrosino, Michael V. Observer Bias. In: Lewis-Beck, MS.; Bryman, A.; Liao, TF., editors. Encyclopedia of Social Science Research Methods. Thousand Oaks, CA: SAGE Publications, Inc.; 2004. p. 758-759.

20. de Leval MR, Carthey J, Wright DJ, Farewell VT, Reason JT. Human factors and cardiac surgery: a multicenter study. J Thorac Cardiovasc Surg. 2000; 119:661–672. [PubMed: 10733754]
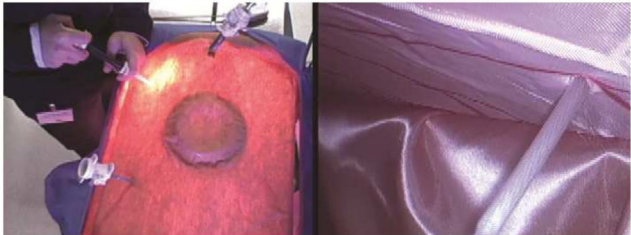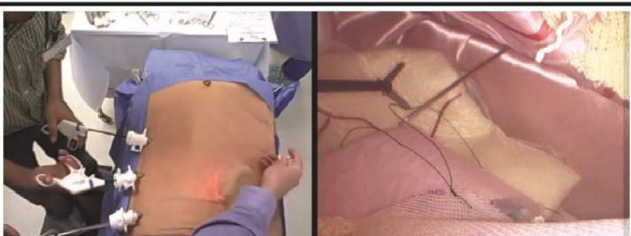
Procedure Step	Image
Port Placement	
Lysis of Adhesions	
Measurement of Hernia Defect	
Securing Anchoring Sutures to Abdominal Wall	
Tacking	

Figure 1.
Examples of procedure steps performed during the simulated LVH with corresponding images
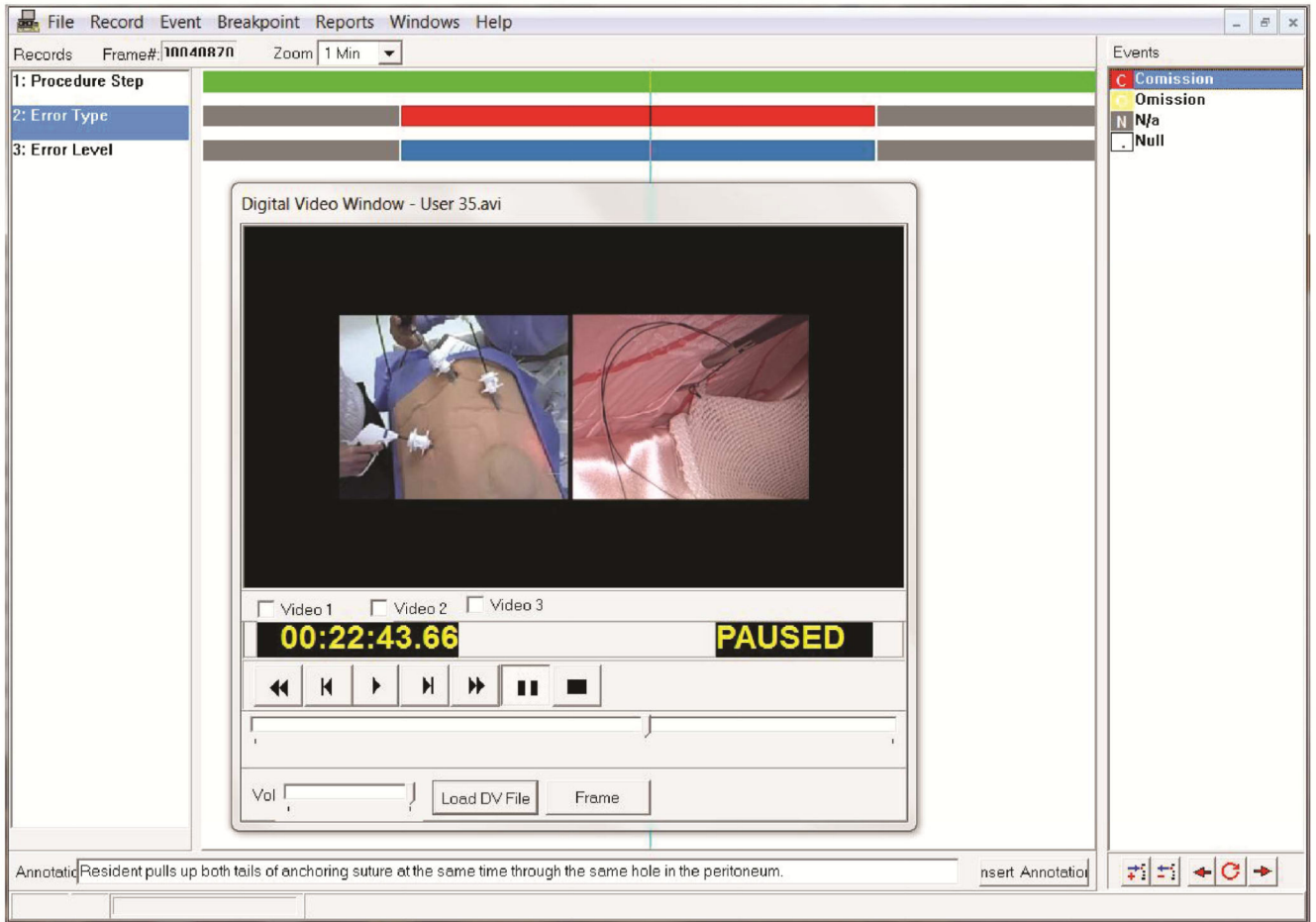
Figure 2.
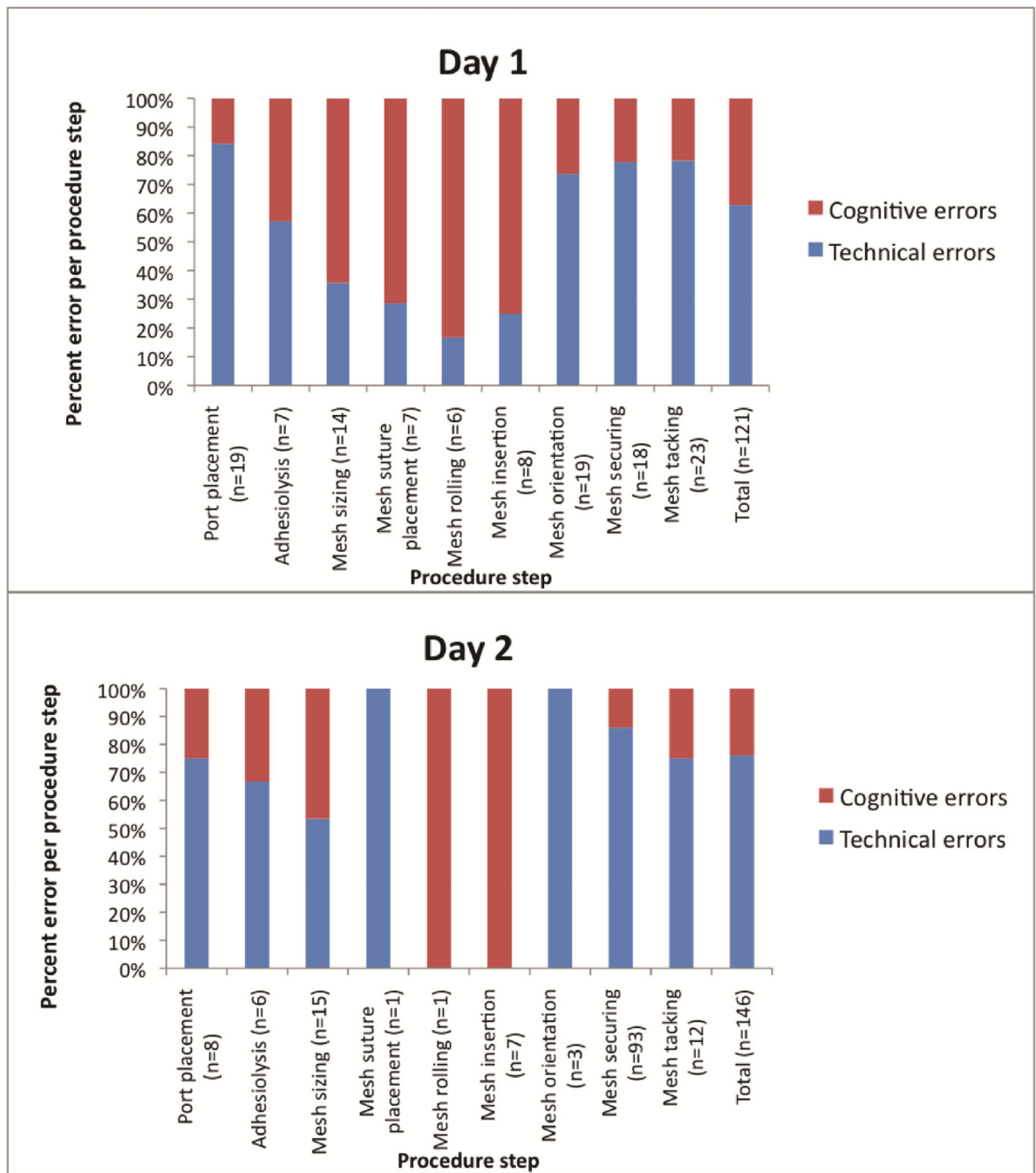Multimedia Video Task Analysis™ (MVTA™) user interface

Figure 3.
Proportion of cognitive versus technical errors during each step of the procedure

Table 1

Error classification definitions and examples adapted from Rasmussen's Cognitive Error Taxonomy.[16]

Error Level		Definition	Example
Cognitive			
	Information	Inability to detect cues arising from the change in system state.	Resident continues to use tacker multiple times despite the fact that it is out of tacks. Does not realize that no tacks are going into mesh.
	Diagnostic	Inaccurate diagnosis of the system state on the basis of the information available.	Resident recognizes mesh is slipping and incorrectly associates the issue with the size of the mesh when it is actually due to the suture slipping through the abdomen. Later correctly realizes the extent of the problem and how far it had slipped.
	Strategy	Selection of a strategy that fails to achieve the intended goal.	Resident does not put anchoring sutures in mesh.
Technical			
	Mechanical	Occurrence of structural or mechanical failures that do not provide an opportunity for resident intervention.	Suture slips through Endo Close™ device.
	Procedure	Execution of a procedure inconsistent with the strategy selected.	Resident puts an inappropriate number of anchoring sutures into mesh.
	Action	Failure to properly execute the procedure.	Resident makes multiple attempts at snaring the suture with the Endo Close™ device.

Table 2

Details of intraoperative errors on Day 1 and Day 2.

	Day 1	Day 2	p-value
Total LVH Completion			
No. (%) of residents with complete repairs	1/7 (14%)	7/7 (100%)	.001
Total number of errors	121	146	
Mean (SD) Participant Errors	17.3 (4.3)	20.9 (5.8)	.26
Error Type			
No. (%) of *omission errors*	40 (33%)	20 (14%)	<.001
No. (%) of *commission errors*	81 (67%)	126 (86%)	
Error Level			
No. (%) of *cognitive errors*	45 (37%)	35 (24%)	.019
No. (%) of *technical errors*	76 (63%)	111 (76%)	