Validation of the *Omni*: A Novel, Multimodality, and Longitudinal Surgical Skills Assessment

Morgan L. Cox, MD[*], **Donald A. Risucci, PhD**[†], **Brian F. Gilmore, MD**[*], **Uttara P. Nag, MD**[*], **Megan C. Turner, MD**[*], **Shanna R. Sprinkle, MD**[*], **John Migaly, MD**[*], **Ranjan Sudan, MD**[*]

[*]Department of Surgery, Duke University, Durham, North Carolina

[†]New York Medical College, Valhalla, New York

Abstract

OBJECTIVE: The breadth of technical skills included in general surgery training continues to expand. The current competency-based training model requires assessment tools to measure acquisition, learning, and mastery of technical skill longitudinally in a reliable and valid manner. This study describes a novel skills assessment tool, the *Omni*, which evaluates performance in a broad range of skills over time.

DESIGN: The 5 *Omni* tasks, consisting of open bowel anastomosis, knot tying, laparoscopic clover pattern cut, robotic needle drive, and endoscopic bubble pop, were developed by general surgery faculty. Component performance metrics assessed speed, accuracy, and quality, which were scaled into an overall score ranging from 0 to 10 for each task. For each task, ANOVAs with Scheffe's post hoc comparisons and Pearson's chi-squared tests compared performance between 6 resident cohorts (clinical years (CY1–5) and research fellows (RF)). Paired samples t-tests evaluated changes in performance across academic years. Cronbach's alpha coefficient determined the internal consistency of the *Omni* as an overall assessment.

SETTING: The *Omni* was developed by the Department of Surgery at Duke University. Annual assessment and this research study took place in the Surgical Education and Activities Lab.

PARTICIPANTS: All active general surgery residents in 2 consecutive academic years spanning 2015 to 2017.

RESULTS: A total of 62 general surgery residents completed the *Omni* and 39 (67.2%) of those residents completed the assessment in 2 consecutive years. Based on data from all residents' first assessment, statistically significant differences ($p < 0.05$) were observed among CY cohorts for bowel anastomosis, robotic, and laparoscopic task metrics. By pair-wise comparisons, mean bowel anastomosis scores distinguished CY1 from CY3–5 and CY2 from CY5. Mean robotic scores distinguished CY1 from RF, and mean laparoscopic scores distinguished CY1 from RF, CY3, and CY5 in addition to CY2 from CY3. Mean scores in performance on the knot tying and endoscopic tasks were not significantly different. Statistically significant improvement in mean scores was observed for all tasks from year 1 to year 2 (all $p < 0.02$). The internal consistency analysis revealed an alpha coefficient of 0.656.

Correspondence: Inquiries to Morgan L. Cox, MD, Duke University Medical Center, DUMC 3443, 2301 Erwin Road, Durham, NC 27710; morgan. cox@duke.edu.

CONCLUSIONS: The *Omni* is a novel composite assessment tool for surgical technical skill that utilizes objective measures and scoring algorithms to evaluate performance. In this pilot study, 3 tasks demonstrated discriminative ability of performance by CY, and all 5 tasks demonstrated construct validity by showing longitudinal improvement in performance. Additionally, the *Omni* has adequate internal consistency for a formative assessment. These results suggest the *Omni* holds promise for the evaluation of resident technical skill and early identification of outliers requiring intervention. (J Surg Ed 75:e218e228. 2018 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

INTRODUCTION

With technological advances and increasing surgical complexity, the breadth of technical skill required of general surgery trainees continues to expand from open to minimally invasive operations. Competency-based training models have been developed to ensure general surgery residents have a baseline level of technical skill to decrease patient risk and, ideally, increase trainee autonomy in the operating room.[1–4] However, effective simulation curricula require reliable and valid objective assessment tools that measure of acquisition, learning, and mastery of technical skill longitudinally.[5–7]

Given this need for objective assessments, many studies have developed technical skill models that apply new technologies such as motion analysis, accelerometer data, crowd sourcing, and machine learning.[8–11] These studies target a singular area of surgical skill, with laparoscopy being the most popular model, leaving a persistent need for a novel, global instrument to assess technical skill.[12–14]

Historically, the assessment methods of these novel simulators and models utilize surveys, participant questionnaires, video review, and global rating scores to evaluate performance.[15] The Objective Structured Assessment of Technical Skills (OSATS) is the most commonly used assessment.[16–18] OSATS is often modified to include task or procedure specific checklists for assessment in the simulation lab or operating room.[19–21] However, completion of the OSATS or modified OSATS includes the subjective opinion of an evaluator requiring multiple experts to review performance and ensure reliability.[22,23]

The need to develop and perfect a broad set of technical skills during general surgery training continues to be of the utmost importance to ensure graduates have a baseline competency across surgical modalities regardless of future subspecialization. While the American Board of Surgery requires residents to complete one-time certifications in the simulation-based Fundamentals of Laparoscopic Surgery (FLS) and Fundamentals of Endoscopic Surgery (FES), there is no objective test that measures technical skills of a surgical resident across various platforms such as open, laparoscopic, robotic, and endoscopic surgery in a longitudinal manner. Therefore, a need exists to develop a single test that can be implemented easily and quickly to assess progression of surgical skills through the course of a surgical residency.[24,25] This study aimed to address this gap and develop and

validate a composite assessment tool of open, laparoscopic, robotic, and endoscopic surgical skill that can evaluate the progression of performance over time.

MATERIAL AND METHODS

Study Population

This study was approved by the Duke University Institutional Review Board. All general surgery residents at a single, academic institution across 2 consecutive academic years were included. Assessments occurred annually and included data collected from June 2016 to September 2017. Interns completed the *Omni* during boot camp prior to starting residency while residents completed the assessment as time allowed through the fall and winter of each academic year. Residents were stratified by clinical year (CY) 1 to 5 and research fellows (RF). Of note, at our institution residents complete 2 clinical years prior to starting their research fellowships. Preliminary and categorical interns were included for analysis.

Omni Tasks

Each of the 5 *Omni* tasks was developed by a general surgery faculty member at our institution with the goal of establishing a longitudinal assessment of all core technical modalities necessary to be a competent general surgeon. We aimed to include a set of fundamental skills essential for completion of both basic and complex surgical procedures and should be developed prior to program completion. Individual tasks needed to be difficult enough that residents would not excel at the task early in training so improvement in performance could be assessed on an annual basis. No instructional videos for successful completion of the *Omni* tasks were provided to residents of any level. Proctors read a standardized script for each task explaining how to successfully accomplish the skill without penalty. The 5 tasks included open bowel anastomosis, knot tying, laparoscopic pattern cut, robotic needle drive, and endoscopic bubble pop. A snapshot of all tasks can be seen in Figure 1.

The bowel suturing task was designed to target isolated suturing skills by requiring the resident to complete an end-to-end bowel anastomosis in less than 10 minutes. Each resident is provided a 20 mm double layer bowel model (Limbs & Things Ltd., Savannah, GA), needle driver, forceps, and a 3–0 Maxon double-armed suture. No instructions are provided on the proper or ideal technique to perform the bowel anastomosis. Time starts when the bowel is touched, and time ends at 10 minutes or when the suture is cut upon completion. Quality and accuracy of performance is based on the amount of anastomosis completed, spacing of the stitches, and the likelihood of leakage.

Knot tying is assessed via an apparatus consisting of a wooden dowel sitting within shallow notches on the top of a plastic cup. The dowel is marked in the center with 2 separate black lines. The resident is provided two 0 silk sutures. First, the resident uses a single-handed technique with the nondominant hand to tie 5 square knots (10 separate throws) around the dowel. The resident must make sure their knot is between the black lines while not lifting the dowel off the cup or out of the notches. This process is repeated with the second suture although the resident uses a 2-handed technique to tie the 5 square knots. Time starts

when the suture is crossed prior to the first knot and ends when the suture is released following the final throw. The times for both tasks are combined, and all 10 square knots (20 separate throws total) have to be completed in less than 5 minutes. Accuracy is determined by placement of the knots between the 2 black lines. Quality of performance is measured by the number of times the dowel is lifted out of the shallow notches.

The laparoscopic task is an advanced pattern cut exercise using a unique clover pattern developed at our institution that is stamped onto a regular, multilayer 4-by-4 gauze. The gauze is placed within an FLS box trainer (Limbs & Things Ltd., Savannah, GA) and left free floating (i.e., not secured with the large clamp superiorly or small alligator clamps inferiorly). The resident is given a Maryland dissector and an endoscopic scissors in order to cut out the clover with as little deviation outside the black lines as possible. Each resident is instructed to start cutting from the edge of the gauze, and only the top layer with the stamped clover is scored. Time starts upon touching the gauze and ends when the clover is cut free. There is a maximum of 5 minutes allotted for this task. Accuracy and quality of the pattern cut task is determined by deviation outside the black lines and percent of pattern completely cut free, respectively.

The da Vinci Skills Simulator (Intuitive Surgical Inc., Sunnyvale, CA) is utilized for the robotic *Omni* task. Each resident completes the Needle Driving 2 module. This module consists of a free needle that has to be grasped with the highlighted instrument and driven through sequentially highlighted circles on the simulated sponge. Time starts when the instruments begin to move, and time ends with completion of the tenth needle drive. There is a maximum completion time of 5 minutes for this task. Accuracy and quality are determined by the simulator scores including number of needle drives completed and an overall performance score.

The endoscopic assessment takes place on the GI Mentor (Simbionix Ltd., Airport City, Israel) via the EndoBubble 1 module. The tip of the scope is navigated through the tunnel while avoiding the walls and touching each subsequent balloon with the end of a needle in order to pierce it. Time to completion is calculated by the simulator in addition to the errors accrued by touching the tunnel walls. The maximum time for completion is 5 minutes. The number of wall hits and overall simulator score determine the accuracy and quality of performance on this task.

Data Acquisition

All active residents completed the 6 *Omni* tasks during a single session on an annual basis starting in the 2016–2017 academic year as a residency program requirement. Proctors were trained simulation laboratory staff members instead of surgical faculty who may have biases resulting from prior experiences working with residents in the clinical arena. Proctors were not blinded to residents' training levels. The first annual administration of the *Omni* required proctors to complete raw data collection on paper that was later entered into an electronic database for analysis. Prior to the second annual administration for the *Omni*, an electronic survey was created for individual task data entry using Research Electronic Data Capture tools hosted at Duke University.[26] These data were automatically compiled for each participating resident based on academic year.

Using an institutional educational database, additional resident data were collected including demographics and whether the resident had completed a month-long endoscopy rotation at the time of each *Omni* performance. Unfortunately, the database was unable to automate individual resident case logs broken down into categories including technical skills in each of the 5 Omni task domains prior to the current analysis. The demographic and prior experience data were linked with the *Omni* Research Electronic Data Capture data via blinded, unique identifiers for statistical analysis.

Scoring System

At the time of *Omni* completion, proctors assessed each performance based on time to completion and predetermined parameters consisting of checklist items and objective measures specific to the task at hand. Raw data were converted into a calculated score via an algorithm developed a priori. To assess the *Omni* construct, each raw data point was categorized as a time, accuracy, or quality measure. The 3 components were scaled to produce an overall score ranging from 0 to 10 for each individual task. The points awarded for time ranged 0 to 2 while 0 to 4 points were awarded for each of the quality and accuracy categories. An overall *Omni* score, ranging 0 to 50, was calculated giving equal weight to each of the 5 tasks.

Outcomes

The primary outcome of interest was the extent to which individual task performance metrics differentiated amongst residents at different levels of clinical training (i.e., CY). Secondary outcome measures included changes in individual skills metrics across test administrations, the correlation of the task metrics to each other, and overall internal consistency of the *Omni* assessment. *Omni* performance metrics were also compared between residents who had completed a prior dedicated endoscopy rotation and those that had not.

Statistical Analysis

Scores from residents' first *Omni* assessment were used to compare performance across CY. Descriptive statistics are reported for overall scores and component scores broken down by task and CY. Categorical variables are reported as number and percentage while continuous variables are reported as mean and standard deviation (SD).

Construct validity was evaluated in several ways. One-way ANOVA was used to analyze differences between CY1–5 and RF in overall mean task scores and actual time (in seconds) to complete each task. Scheffe's post hoc comparisons determined which resident levels differed significantly ($p < 0.05$) from each other. Pearson's chi-squared tests examined the association between CY and categorical metrics from each task (i.e., speed, accuracy, and quality). Paired samples t-tests were applied to data from those residents who completed the *Omni* twice in order to assess construct validity with respect to progression of skill over time (i.e., an additional year of residency training).

The cohort completing the *Omni* for the first time was then stratified by those who had and had not completed the month-long endoscopy rotation at the time of their *Omni* assessment.

These subgroups were compared utilizing independent samples t-test and chi-Pearson's chi-square as appropriate.

Finally, Spearman's rho assessed for correlations between tasks, and internal consistency of the overall scores from the 5 *Omni* tasks was analyzed via Cronbach's alpha coefficient. Comparisons were 2-tailed for all analyses. A p value of <0.05 was considered statistical significance. Statistical analysis was completed using SPSS v23 (IBM, Armonk, NY) (Table 1).

RESULTS

A total of 62 (100%) general surgery residents completed the *Omni* with 39 (67.2%) residents completing the assessment in consecutive years. There were 3 missing scores for the robotic task due to data entry error. Therefore, the primary analysis of the robotic task scores included 59 residents. Only 37 residents were included in the robotic analysis evaluating improvement between the first and second *Omni* assessments.

Based on data from all residents' first assessment, significant differences were observed among CY cohorts for the bowel anastomosis, robotic needle drive, and laparoscopic pattern cut tasks. The *Omni* also identified performance outliers in both positive and negative directions. Mean scores stratified by resident level and *Omni* task can be seen in Figure 2 while the frequency of overall scores on the various *Omni* tasks are listed in Table 2.

Bowel Anastomosis Task

The bowel anastomosis task had a mean overall score of 5.6 (SD 2.3). There were statistically significant differences observed among the CY and RF cohorts based on the overall score (F 9.67; degrees of freedom (df) = 5, 56; $p < 0.001$). These overall scores were able to discriminate a CY1 from CY3, CY4, and CY5 residents in addition to CY2 from CY5 (all $p < 0.05$). This ability to differentiate was based on the accuracy ($p = 0.001$) and quality ($p = 0.005$) components of the task.

Knot Tying Task

The knot tying task had a mean overall score of 8.5 (SD 1.2). Mean differences in performance on the knot tying task were not significant (F 0.52; df = 5, 56; $p = 0.762$). Additionally, the task component scores of speed, accuracy and quality were also unable to differentiate based on resident level (all $p > 0.05$).

Laparoscopic Pattern Cut Task

The laparoscopic pattern cut task had a mean overall score of 4.7 (SD 2.6). There were statistically significant differences observed among the CY and RF cohorts based on the overall task score (F 6.71; df = 5, 56; $p < 0.001$). These overall scores were able to discriminate a CY1 from RF, CY3, and CY5 in addition to CY2 from CY3 (all $p < 0.05$). This ability to differentiate was based on all 3 of the task's component scores including speed ($p = 0.000$), accuracy ($p = 0.003$), and quality ($p = 0.021$).

Robotic Needle Drive Task

The robotic needle driving task had a mean overall score of 3.4 (SD 2.8). There were statistically significant differences observed among the CY and RF cohorts based on the overall task score (F 4.15; df = 5, 53; p = 0.003). These overall scores were able to discriminate a CY1 from RF (p = 0.049). This ability to differentiate was based on the quality component of the task score (p = 0.002). Although pairwise comparisons did not reveal a significant difference based on the calculated time component (p = 0.085), analyzing speed with time to completion as a continuous variable was able to discriminate between clinical year residents (p = 0.004).

Endoscopic Bubble Pop Task

The endoscopic bubble pop task had a mean score of 8.3 (SD 1.7). Mean differences in performance on the knot tying task were not significant (F 1.43; df = 5, 56; p = 0.230). However, significant mean differences were observed for the speed (p = 0.008) and accuracy (p = 0.049) components of the overall score.

A subgroup analysis divided residents' first-time assessments based on completion of a month-long endoscopy rotation at the time of *Omni* assessment. Twenty-nine residents (46.8%) had completed the rotation while 33 residents (53.2%) had not. The mean overall scores of residents who had and had not completed an endoscopy rotation were 8.2 (SD 2.0) and 8.3 (SD 1.3), respectively. This difference was not significantly different (t = 0.14, p = 0.888).

Construct Validity

The first and second average overall scores for those residents completing the *Omni* twice over 2 academic years can be seen in Figure 3. Evaluation of performance between the first and second *Omni* assessments revealed statistically significant improvement in the mean difference of overall scores for all 5 tasks (all p < 0.02).

Internal Consistency

There was a steady increase in overall *Omni* scores across CYs ranging from 26.1 to 39.8 out of 50 (Table 3). Overall scores for individual tasks all demonstrated significant correlations with each other (range 0.32–0.62; all p < 0.015) with the exception of the endoscopic bubble pop task (r = 0.22 or less with other measures). Evaluation of internal consistency utilizing overall *Omni* scores revealed a Cronbach's alpha coefficient of 0.669.

DISCUSSION

Our pilot study implemented a novel surgical skills assessment tool, the *Omni*, with an objective scoring algorithm to assess the longitudinal development of technical skills in a variety of surgical modalities. The *Omni* establishes a baseline level of skill without pretraining and allows for the repeated evaluation of technical performance throughout residency training and the identification of performance outliers within each domain of surgical approaches. Open bowel anastomosis, laparoscopic pattern cut, and robotic needle drive tasks were able to differentiate performance based on level of the trainee. The *Omni*

also revealed resident performance significantly improves on all 5 tasks from one clinical year to next providing promising data on the construct validity of the assessment and suggesting its potential to adequately assess residents on an annual basis.

Interestingly, the endoscopic task did not correlate with the other 4 assessment metrics within the *Omni*. This suggests the endoscopic bubble pop task could be of a different construct and assessing a different component of surgical technical skill, and this aspect of technical skill has been deemed important by the American Board of Surgery indicated by the FES assessment requirement. The endoscopy task is similar to the laparoscopic task in that it requires interpretation of a 3-dimensional skill via a 2-dimensional display which requires visual perception aptitude. The difference between these 2 tasks may lie in the fact that general surgery residents have significantly more experience with laparoscopic surgery throughout residency compared to endoscopic procedures. Additionally, the endoscopic task was unable to differentiate residents by clinical level or by completion of a month-long endoscopy rotation. Altogether, the construct validity of the endoscopic task remains in question and further investigation of is warranted including analysis based on the passage of FES.

Prior literature has described the immense value of surgical simulation in developing and assessing residents' technical skill. Simulation allows for the development of basic and complex skills in a low-risk setting, but assessment requires dedicated effort from experts either in real-time or as video review in order to obtain adequate reliability.[15,23,27] Additionally, the transferability of skills demonstrated in the simulation lab to the operating room and its impact on patient outcomes has not been well defined.[28,29] This is likely due to the fact that assessment tools targeting a single skills task or procedure under-represent a resident's technical proficiency, whereas the *Omni* aims to more accurately reflect technical skill ability by providing a global assessment in multiple surgical modalities including open, laparoscopic, robotic, and endoscopic surgery.[30]

Abdelrahman et al.[12] developed an inverted peg transfer designed to be more difficult than the FLS peg transfer in order to assess advanced laparoscopic skills. Their study team found the inverted peg transfer better discriminated between novices and experts compared to the regular peg transfer. This model was similar to the advanced laparoscopic pattern cut included in the *Omni* which also advances a standardized FLS task. However, the prior study addressed only a laparoscopic task for assessment while the *Omni* is all encompassing with 5 different surgical tasks providing a better global assessment of residents' technical skill. Additionally, the inverted peg transfer study included subjects categorically different from each other in terms of skill level. Novices consisted of medical students and surgical interns, and experts had 3 years of laparoscopic surgery experience. Therefore, their results in the ability to discriminate performance based on experience is not generalizable to general surgery residency where the difference in skill level between interns and chief residents is less than the spectrum of skills between medical students and faculty.

An ideal skills assessment tool not only discriminates performance based on clinical level, but it should also be able to identify outliers of both high and low performance. The *Omni* identified these outlying residents in all 5 tasks. While this is another promising feature of

the *Omni*, we have to await further reliability and validity studies to adequately interpret these outlying results. Ideally, threshold levels of minimum and maximum performance markers can be created to help identify residents lagging behind or those who are competent and ready to be challenged with more complex skills. Early detection of these outliers via the annual *Omni* assessment potentially allows for individualized and targeted remediation. Based on an individual's performance across the 5 *Omni* tasks, intervention could be designed to target a global deficiency in skills or a particular domain that is lacking.

Our institution has previously published on the implementation of an in-depth and broad simulation curriculum with 35 modules utilizing the American College of Surgeons/ Association of Program Directors in Surgery National Technical Skills Curriculum.[31] Implementation increased the amount of cadaveric and animal tissue labs, both of which increased resident satisfaction. As we have worked to integrate competencies into this simulation curriculum, the development of the *Omni* has not only allowed for an objective assessment of skill progression but also provided a metric to evaluate the curriculum itself by defining areas that require additional or alternate timing of training.

As with all research studies, this investigation is not without its limitations. Most importantly, this is a pilot study with a limited sample size. The breakdown by clinical year remains small with repeated measures over just 2 years. The inability of the knot tying and endoscopic tasks to discriminate performance levels is likely due to a lack of variability in mean scores that mostly fell on the higher end of the scoring scale (i.e., a "ceiling effect"). This pilot study did not contain the power to discriminate between these bunched scores which could be addressed with a change in the scoring algorithm and increased sample size or modifications to the tasks themselves. It is also feasible that knot tying skills in particular are the earliest surgical skills taught at the medical student level during the surgery clerkship which could lead to an improved set of technical skills in this area even at the intern level. Further investigation including a more difficult modification to the knot tying task will help determine the cause of the lack of construct validity of this task.

Additionally, the *Omni* was administered to the incoming intern class of the 2016 to 2017 academic year, but it took several months to complete assessment of the remaining residents. The following academic year, 2017 to 2018, residents of all levels were assessed near the beginning of the academic year. Therefore, the timeframe between assessments was longer and likely more accurate for those residents progressing from CY1 to CY2 compared to the more senior residents with just a few months between assessments.

Finally, simulation lab staff were used as proctors in order to reduce bias that clinical faculty could introduce based on clinical experience with residents. However, the simulation staff do have hands-on experience gathering data on residents throughout the simulation curriculum including CY1, CY2, and RFs. Therefore, there could be unmeasured bias by not blinding the proctors as raters. However, administering the *Omni* without utilizing the simulation lab staff is not feasible, and a future inter-rater reliability analysis will shed light on the quality and standardization of the raters.

Given the promising results in this pilot study of the *Omni's* potential ability to adequately assess and discriminate global technical performance over time, future studies are in development. Scoring algorithms for the knot tying and endoscopic tasks will be adjusted to better stratify residents across the calculated scores (0–10) which will likely increase the ability to discriminate performance by resident level with future analyses. Ideally, a reliability analysis, including inter-rater reliability on the laparoscopic and bowel anastomosis quality measures, and further validation of all 5 tasks will occur in a multiinstitutional study. Not only will this increase the sample size and generalizability of the results, but it should provide enough variability in the timing of FES and FLS assessments to allow for subgroup analyses based on completion of these required certification and how that correlates or impacts *Omni* performance. Furthermore, we would like to correlate performance in the different areas of the *Omni* with traditional training measures including board scores, in-service training exams, clinical milestones, monthly operative case logs, and patient outcomes.

CONCLUSIONS

The *Omni* is a novel assessment tool for surgical technical skill with 3 tasks (open bowel anastomosis, laparoscopic pattern cut, and robotic needle drive) that reliably discriminate by clinical year. Resident technical skill level improved over time on all 5 tasks, and the *Omni* has adequate internal consistency for a formative assessment. These results suggest the *Omni* holds promise for the evaluation of resident technical skill longitudinally for early identification of performance outliers requiring intervention.

ACKNOWLEDGMENTS

Disclosures:

MLC is supported by a National Institutes of Health T32 Training Grant with grant number T32HL069749. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ABBREVIATIONS:

OSATS	Objective Structured Assessment of Technical Skills
ABS	American Board of Surgery
FLS	Fundamentals of Laparoscopic Surgery
FES	Fundamentals of Endoscopic Surgery

CY	clinical year
RF	research fellow
REDCap	Research Electronic Data Capture
SD	standard deviation
df	degrees of freedom
ACS	American College of Surgeons
APDS	Association of Program Directors in Surgery

REFERENCES

1. Angelo RL, Ryu RKN, Pedowitz RA, et al. A proficiency-based progression training curriculum coupled with a model simulator results in the acquisition of a superior arthroscopic bankart skill set. Arthroscopy. 2015;31:1854–1871. 10.1016/j.arthro.2015.07.001. [PubMed: 26341047]

2. Beard JD. Assessment of surgical skills of trainees in the UK. Ann R Coll Surg Engl. 2008;90:282–285. 10.1308/003588408X286017. [PubMed: 18492389]

3. Griswold-Theodorson S, Ponnuru S, Dong C, Szyld D, Reed T, McGaghie WC. Beyond the simulation laboratory: a realist synthesis review of clinical outcomes of simulation-based mastery learning. Acad Med. 2015;90:1553–1560. 10.1097/ACM.0000000000000938. [PubMed: 26375271]

4. de Montbrun S, Louridas M, Grantcharov T. Passing a technical skills examination in the first year of surgical residency can predict future performance. J Grad Med Educ. 2017;9:324–329. 10.4300/JGME-D-16-00517.1. [PubMed: 28638511]

5. Mitchell EL, Arora S, Moneta GL, et al. A systematic review of assessment of skill acquisition and operative competency in vascular surgical training. J Vasc Surg. 2014;59:1440–1455. 10.1016/j.jvs.2014.02.018. [PubMed: 24655750]

6. Buckley CE, Kavanagh DO, Traynor O, Neary PC. Is the skillset obtained in surgical simulation transferable to the operating theatre. Am J Surg. 2014;207:146–157. 10.1016/j.amjsurg.2013.06.017. [PubMed: 24238602]

7. 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;138:640–649. 10.1016/j.surg.2005.07.017. [PubMed: 16269292]

8. Zia A, Sharma Y, Bettadapura V, Sarin EL, Essa I. Video and accelerometer-based motion analysis for automated surgical skills assessment. Int J Comput Assist Radiol Surg. 2018;13:443–455. 10.1007/s11548-018-1704-z. [PubMed: 29380122]

9. Azari DP, Frasier LL, Quamme SRP, et al. Modeling surgical technical skill using expert assessment for automated computer rating. Ann Surg. 2017. 10.1097/SLA.0000000000002478.1.

10. Lendvay TS, White L, Kowalewski T. Crowdsourcing to assess surgical skill. JAMA Surg. 2015;150:1086. 10.1001/jamasurg.2015.2405. [PubMed: 26421369]

11. Mason JD, Ansell J, Warren N, Torkington J. Is motion analysis a valid tool for assessing laparoscopic skill. Surg Endosc. 2013;27:1468–1477. 10.1007/s00464-012-2631-7. [PubMed: 23233011]

12. Abdelrahman AM, Yu D, Lowndes BR, et al. Validation of a novel inverted peg transfer task: advancing beyond the regular peg transfer task for surgical simulation-based assessment. J Surg Educ. 2017. 10.1016/j.jsurg.2017.09.028.

13. Poolton JM, Zhu FF, Malhotra N, et al. Multitask training promotes automaticity of a fundamental laparoscopic skill without compromising the rate of skill learning. Surg Endosc. 2016;30:4011–4018. [PubMed: 26743112]

14. Fonseca AL, Evans LV, Gusberg RJ. Open surgical simulation in residency training: a review of its status and a case for its incorporation. J Surg Educ. 2013;70:129–137. 10.1016/j.jsurg.2012.08.007. [PubMed: 23337682]

15. Atesok K, Satava RM, Marsh JL, Hurwitz SR. Measuring surgical skills in simulation-based training. J Am Acad Orthop Surg. 2017;25:665–672. 10.5435/JAAOS-D-16-00253. [PubMed: 28953081]

16. 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. 10.1002/bjs.1800840237. [PubMed: 9052454]

17. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. Am J Surg. 1997;173:226–230. [PubMed: 9124632]

18. van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J. Objective assessment of technical surgical skills. Br J Surg. 2010;97:972–987. 10.1002/bjs.7115. [PubMed: 20632260]

19. Hopmans CJ, den Hoed PT, van der Laan L, et al. Assessment of surgery residents' operative skills in the operating theater using a modified Objective Structured Assessment of Technical Skills (OSATS): a prospective multicenter study. Surgery. 2014;156:1078–1088. 10.1016/j.surg.2014.04.052. [PubMed: 25231747]

20. Niitsu H, Hirabayashi N, Yoshimitsu M, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. Surg Today. 2013;43:271–275. 10.1007/s00595-012-0313-7. [PubMed: 22941345]

21. Shah J, Munz Y, Manson J, Moorthy K, Darzi A. Objective assessment of small bowel anastomosis skill in trainee general surgeons and urologists. World J Surg. 2006;30:248–251. 10.1007/s00268-005-0074-1. [PubMed: 16425076]

22. Chang OH, King LP, Modest AM. Hur H-C. Developing an objective structured assessment of technical skills for laparoscopic suturing and intracorporeal knot tying. J Surg Educ. 2016;73:258–263. 10.1016/j.jsurg.2015.10.006. [PubMed: 26597729]

23. Sawyer JM, Anton NE, Korndorffer JR, DuCoin CG, Stefanidis D. Time crunch: increasing the efficiency of assessment of technical surgical skill via brief video clips. Surgery. 2018. 10.1016/j.surg.2017.11.011.

24. Vassiliou MC, Dunkin BJ, Marks JM, Fried GM. FLS and FES: comprehensive models of training and assessment. Surg Clin N Am. 2010;90:535–558. 10.1016/j.suc.2010.02.012. [PubMed: 20497825]

25. Peters JH, Fried GM, Swanstrom LL, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. Surgery. 2004;135:21–27. 10.1016/S0039. [PubMed: 14694297]

26. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42:377–381. 10.1016/j.jbi.2008.08.010. [PubMed: 18929686]

27. Dath D, Regehr G, Birch D, et al. Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills. Surg Endosc. 2004;18:1800–1804. 10.1007/s00464-003-8157-2. [PubMed: 15809794]

28. Anastakis DJ, Regehr G, Reznick RK, et al. Assessment of technical skills transfer from the bench training model to the human model. Am J Surg. 1999;177:167–170. [PubMed: 10204564]

29. Anderson DD, Long S, Thomas GW, Putnam MD, Bechtold JE, Karam MD. Objective Structured Assessments of Technical Skills (OSATS) does not assess the quality of the surgical result effectively. Clin Orthop Relat Res. 2016;474:874–881. 10.1007/s11999-015-4603-4. [PubMed: 26502107]

30. Van Nortwick SS, Lendvay TS, Jensen AR, Wright AS, Horvath KD, Kim S. Methodologies for establishing validity in surgical simulation studies. Surgery. 2010;147:622–630. 10.1016/j.surg.2009.10.068. [PubMed: 20015529]

31. Henry B, Clark P, Sudan R. Cost and logistics of implementing a tissue-based American College of Surgeons/Association of Program Directors in Surgery surgical skills curriculum

for general surgery residents of all clinical years. Am J Surg. 2014;207:201–208. 10.1016/j.amj-surg.2013.08.025. [PubMed: 24239527]

COMPETENCIES:

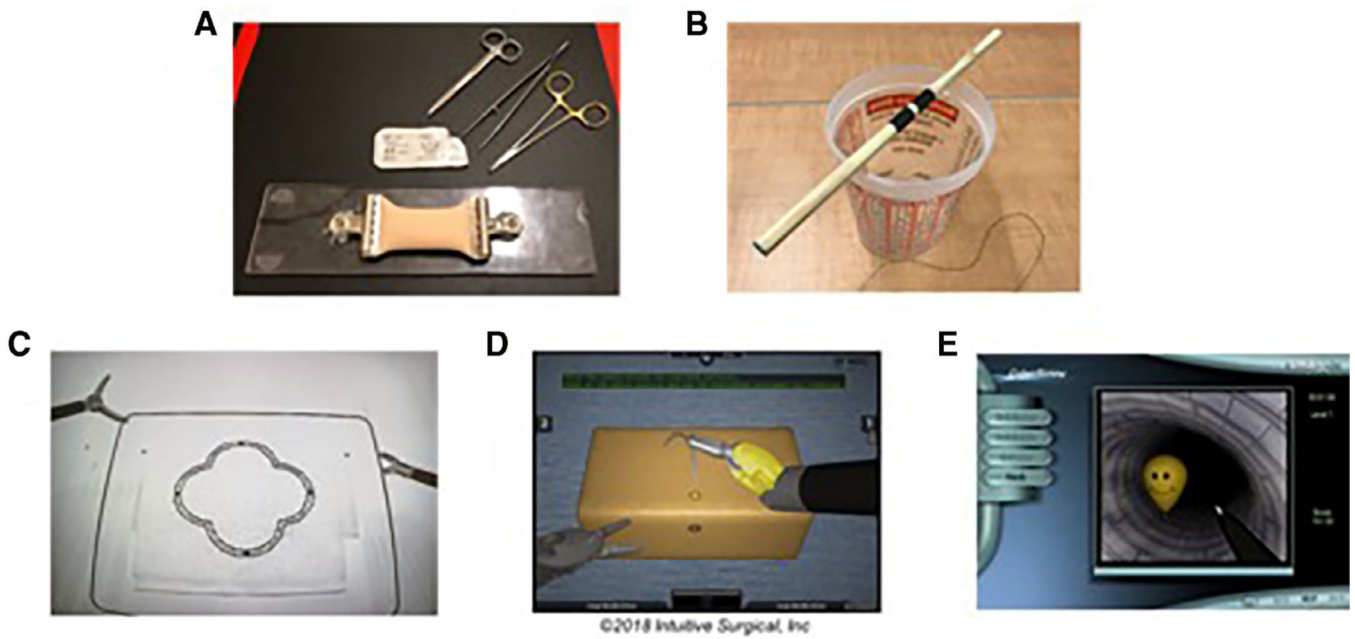Practice-Based Learning and Improvement, Medical Knowledge, Patient Care

FIGURE 1.

Snapshots of the 5 *Omni* tasks: (A) bowel anastomosis, (B) knot tying, (C) laparoscopic pattern cut, (D) robotic needle drive, and (E) endoscopic bubble pop.
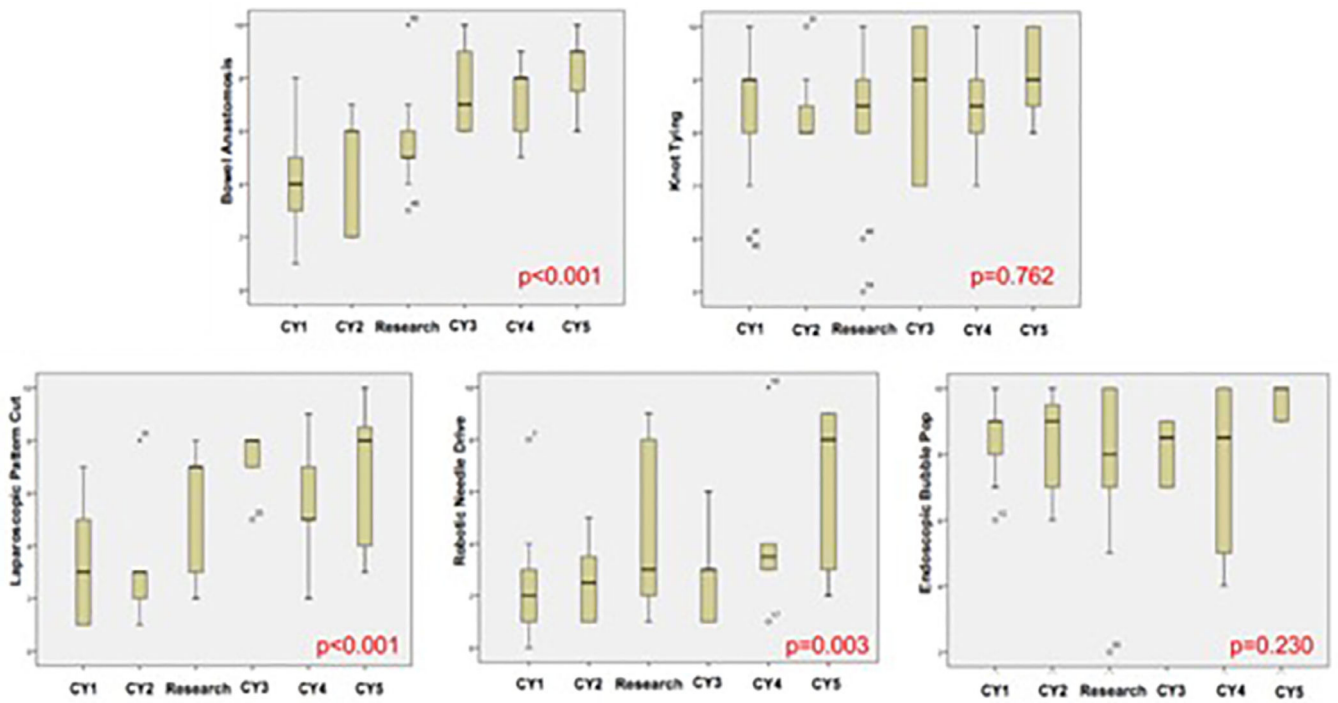
FIGURE 2.
Box plots with mean overall scores on the first attempt by clinical year for each *Omni* task.
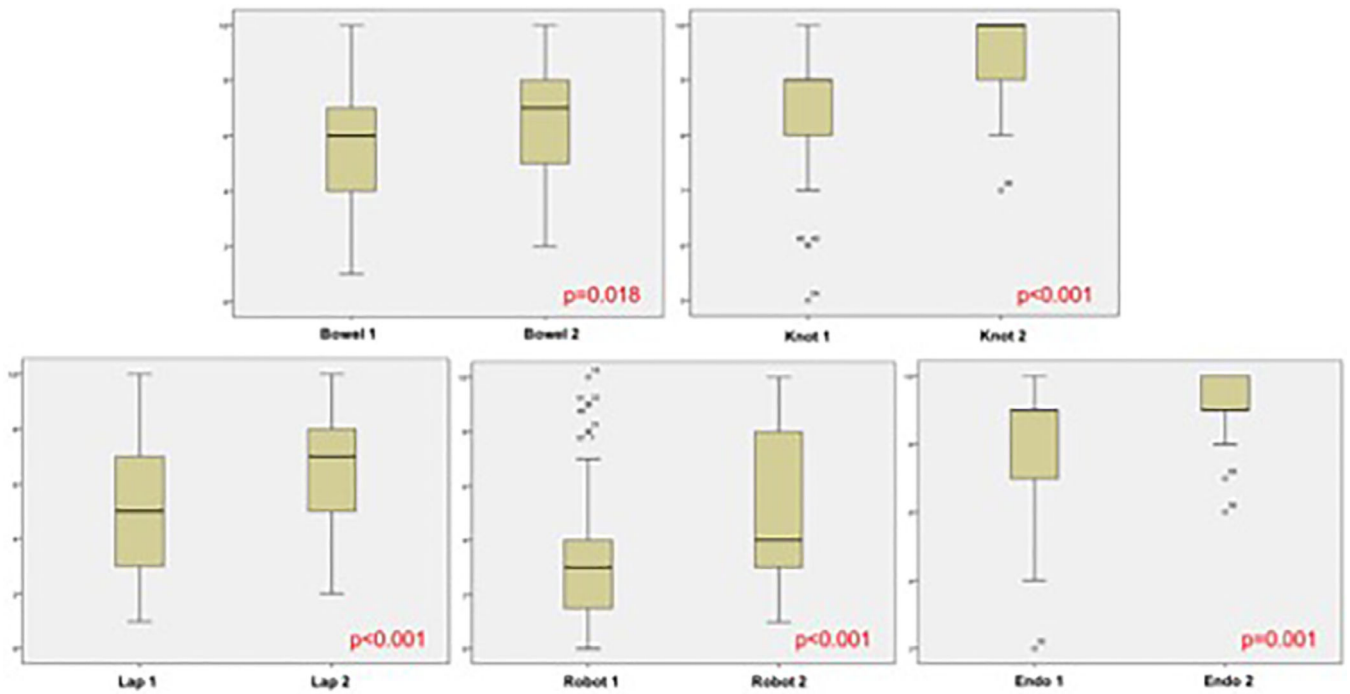
FIGURE 3.
Box plots with mean overall scores of the first and second *Omni* assessments broken down by task.

TABLE 1.

Baseline Resident Characteristics

	2016 to 2017	2017 to 2018	1st Completion	2nd Completion
Residents	49	52	62	39
CY1	10 (20.4%)	11 (21.2%)	21 (33.9%)	-
CY2	8 (16.3%)	8 (15.4%)	8 (12.9%)	8 (20.5%)
RF	14 (28.6%)	16 (30.8%)	14 (22.5%)	16 (41.0%)
CY3	4 (8.2%)	6 (11.5%)	6 (9.7%)	4 (10.3%)
CY4	6 (12.2%)	5 (9.6%)	6 (9.7%)	5 (12.8%)
CY5	7 (14.3%)	6 (11.5%)	7 (11.3%)	6 (15.4%)
CY1				
CY1 Categorical	6 (60%)	7 (63.6%)	13 (61.9%)	-
Preliminary	4 (40%)	4 (36.4%)	8 (38.1%)	-
Sex				
Male	34 (69.4%)	36 (69.2%)	43 (69.4%)	27 (69.2%)
Female	15 (30.6%)	16 (30.8%)	19 (30.6%)	12 (30.8%)

*
All data reported as count (percent).

TABLE 2.

Frequencies of Overall *Omni* Scores by Task

	Bowel N=62	Knot N=62	Laparoscopic N=62	Robotic N=59	Endoscopic N=62
Overall score					
0	0 (0%)	0 (0%)	0 (0%)	3 (4.8%)	0 (0%)
1	1 (1.6%)	0 (0%)	7 (11.3%)	12 (19.4%)	0 (0%)
2	6 (9.7%)	0 (0%)	7 (11.3%)	12 (19.4%)	1 (1.6%)
3	7 (11.3%)	0 (0%)	14 (22.6%)	14 (22.6%)	0 (0%)
4	5 (8.1%)	0 (0%)	2 (3.2%)	5 (8.1%)	1 (1.6%)
5	11 (17.7%)	1 (1.6%)	8 (12.9%)	1 (1.6%)	2 (3.2%)
6	12 (19.4%)	3 (4.8%)	2 (3.2%)	1 (1.6%)	4 (6.5%)
7	6 (9.7%)	5 (8.1%)	9 (14.5%)	1 (1.6%)	9 (14.5%)
8	6 (9.7%)	21 (33.9%)	10 (16.1%)	4 (6.5%)	8 (12.9%)
9	5 (8.1%)	17 (27.4%)	2 (3.2%)	5 (8.1%)	24 (38.7%)
10	3 (4.8%)	15 (24.2%)	1 (1.6%)	1 (1.6%)	13 (21.0%)

*
All data reported as count (percent).

TABLE 3.

Overall *Omni* Scores by Clinical Year

	N	Overall Score (0–50)
CY1	21	26.1 (5.4)
CY2	8	27.0 (5.1)
RF	13	32.2 (6.0)
CY3	6	34.5 (3.8)
CY4	6	33.2 (7.6)
CY5	5	39.8 (8.6)
Total	59	30.3 (7.2)

*All data reported as mean (SD).