Validation of the AUA BLUS Tasks

Timothy M. Kowalewski,* Robert Sweet,†,‡ Thomas S. Lendvay,§ Ashleigh Menhadji,† Timothy Averch,† Geoffrey Box,† Timothy Brand,† Michael Ferrandino,|| Jihad Kaouk,† Bodo Knudsen,† Jaime Landman,† Benjamin Lee,¶ Bradley F. Schwartz† and Elspeth McDougall†

From the Departments of Mechanical Engineering (TMK) and Urology (RS), University of Minnesota, Minneapolis, Minnesota, Departments of Urology, University of Washington and Seattle Children's Hospital, Seattle (TSL) and Madigan Army Medical Center, Uniformed Services University of the Health Sciences (TB), Tacoma, Washington, Boston University School of Medicine (AM), Boston, Massachusetts, University of Pittsburgh Medical Center (TA), Pittsburgh, Pennsylvania, Department of Urology, Ohio State University (GB, BK), Columbus and Cleveland Clinic (JK), Cleveland, Ohio, Departments of Urology, Duke University (MF), Durham, North Carolina, University of California-Irvine (JL), Orange, California, Department of Urology and Oncology, Tulane University (BL), New Orleans, Louisiana, Division of Urology, Southern Illinois University (BFS), Carbondale, Illinois, and Department of Urologic Sciences, University of British Colombia (EM), Vancouver, British Columbia, Canada

Abbreviations and Acronyms

ATT = attending surgeon

C-SATS = Crowd-Sourced Assessment of Technical Skills

FLS = Fundamentals of Laparoscopic Surgery

MS = medical student

OSATS = Objective Structured Assessment of Technical Skills

R1 = year 1 postgraduate residents

R2 = year 2 postgraduate residents

R5 = year 5 postgraduate residents

Purpose: Standardized assessment of laparoscopic skill in urology is lacking. We investigated whether the AUA (American Urological Association) BLUS (Basic Laparoscopic Urologic Skills) skill tasks are valid to address this need.

Materials and Methods: This institutional review board approved study included 27 medical students, 42 urology residents, 18 fellows and 37 faculty urologists across 8 sites. Using the EDGE (Electronic Data Generation and Evaluation) device (Simulab, Seattle, Washington) 454 recordings were collected on peg transfer, pattern cutting, suturing and clip applying tasks, which together comprise the expert determined BLUS tasks. We collected synchronized video and tool motion data for each trial. For each task errors, time, path length, economy of motion, peak grasp force and EDGE score were collected. An expert panel of 5 faculty members performed GOALS (Global Objective Assessment of Laparoscopic Skills) evaluations on a representative subset of peg transfer and suturing skill tasks performed by 24 participants (IRR = 0.95).

Results: Demographically derived skill levels proved unsuitable to evaluate construct validity. Separation of mean scores by grouped skill levels was strongest for the suturing task. Objective motion metrics and errors supported construct validity vis-à-vis correlation with blinded expert video ratings (motion metrics $R^2 = 0.95$, $p < 0.01$). Expert scores appeared to reward errors in suturing but not in block transfer.

Accepted for publication October 6, 2015.

The corresponding author certifies that, when applicable, a statement(s) has been included in the manuscript documenting institutional review board, ethics committee or ethical review board study approval; principles of Helsinki Declaration were followed in lieu of formal ethics committee approval; institutional animal care and use committee approval; all human subjects provided written informed consent with guarantees of confidentiality; IRB approved protocol number; animal approved project number.

Supported by the American Urological Association.

* Financial interest and/or other relationship with C-SATS and Simulab.

† No direct or indirect commercial incentive associated with publishing this article.

‡ Correspondence: Department of Urology, Medical School, University of Minnesota, 420 Delaware St. Southeast, MMC 394, Minneapolis, Minnesota 55455 (telephone: 612-625-7486; FAX: 612-626-0428; e-mail: rsweet@umn.edu).

§ Financial interest and/or other relationship with C-SATS.

|| Financial interest and/or other relationship with TransEnterix.

¶ Financial interest and/or other relationship with *Journal of Endourology*, Argos and Porges™.

0022-5347/16/1954-0998/0
THE JOURNAL OF UROLOGY®

http://dx.doi.org/10.1016/j.juro.2015.10.087
Vol. 195, 998-1005, April 2016
Printed in U.S.A.

Conclusions: BLUS skill task performance scoring can discriminate among basic laparoscopic technical skill levels. Self-reported demographics are an unreliable source of determining laparoscopic technical skill.

Key Words: urologic surgical procedures/education; laparoscopy/standards; clinical competence; education, medical; validation studies

THE introduction of laparoscopy demanded that surgeons in practice and training acquire a new set of surgical skills. The experience of laparoscopic cholecystectomy in the late 1980s in general surgery was fraught with unacceptable complications and outcomes, which ultimately drove SAGES (Society of American Gastrointestinal and Endoscopic Surgeons) to create a curriculum in 1997 known as FLS.[1–3] FLS underwent a series of validation studies and was ultimately endorsed by ACS (American College of Surgeons) in 2005. All general surgery residents are now required to pass the FLS examination to sit for the written board examination.

In 2009 the AUA examined the FLS curriculum and determined that it would be prudent to also pursue a basic laparoscopic training credentialing initiative. AUA created the BLUS curriculum based on FLS. BLUS includes 4 psychomotor task components, which are nonoperative, patient-free psychomotor skill tasks mirroring FLS (peg transfer, pattern cutting, needle driving and knot tying), and a fifth task consisting of a custom clip applying exercise. The tasks are performed with real laparoscopic tools on physical objects demanding psychomotor and visual-spatial laparoscopic skill with the objectives of minimizing completion time and avoiding specific errors.

Metrics and errors for these tasks were further determined by the BLUS working group and ultimately incorporated into a computerized box trainer to help objectify assessment.[4] Sweet et al found good acceptability and construct validity evidence for all psychomotor exercises in a cohort of 81 urologists.[5] More recently Mandava et al provided additional supporting evidence in a cohort of 28 participants.[6] BLUS also includes a completely rewritten cognitive component based on the handbook of laparoscopy.[7] The multiple choice cognitive examination required a passing score of 80 points.

The goal of this study was to assess the ability of BLUS and its associated automated metrics to objectively discriminate laparoscopic skill in a cohort of urologists and trainees. We investigated evidence of construct validity for the psychomotor skill portion of this curriculum using objective performance metrics and structured assessment by video review.

MATERIALS AND METHODS

Subject Pool and Data Collection

Our institutional review board approved study involved 124 subjects across 8 academic urology training programs in the United States. All subjects filled out a pretask questionnaire to self-report anonymous demographic information (training, age, laparoscopic experience level, nonrobotic laparoscopic caseload, etc). At each site trained examination proctors enforced consistent data collection procedures.

Subjects created an AUA account, logged into the system and performed the online cognitive skill examination. Each subject watched a demonstration video of each task and then completed each of the 4 BLUS tasks as described by Sweet et al[5] but on the EDGE platform. Time was limited to 6 minutes for all tasks. Subjects were allowed up to 5 practice attempts on the peg transfer task and 1 on all subsequent tasks. The supplementary table (http://jurology.com/) shows tasks and errors. Each completed task included a video recording of the task synchronized with streaming tool motion and force data derived from the EDGE device. Data were sampled at 30 Hz and included the Cartesian (xyz) position of the tool tip in cm, the roll of the tool shaft in degrees, the angle of the grasper handle in degrees and the force induced at the grasper handle in N. For each skill task the summary motion metrics (total time, path length, economy of motion, peak grasp force, etc) were automatically computed by the system.

Immediately after each task any nonautomated errors that required a human observer (number of drops, instrument tips out of field of view, inaccurate cutting or needle placement, etc) were manually calculated and entered by the trained proctor administering the testing session. Table 1 lists the errors. Weighted penalties for

Table 1. *Demographic spread of skill grouping among participating subjects*

Grouping	No. Subjects (%)	No. Tasks (%)
Original:		
MS	26 (22)	106 (23)
R1	14 (12)	60 (13)
R2	11 (9.4)	46 (10)
Yr 3 residents	6 (5.1)	25 (5.5)
Yr 4 residents	6 (5.1)	19 (4.2)
R5	4 (3.4)	16 (3.5)
Fellows	17 (15)	62 (14)
ATTs	33 (38)	120 (26)
Combined (laparoscopic experience):		
MS (0)	26 (22)	106 (31)
R1-R2 (0)	25 (21)	106 (31)
Yr 3 residents-R5 (0)	16 (14)	60 (17)
ATTs (3+ cases/mo)	20 (17)	71 (21)
Other*	30 (26)	111 (24)

* Not used in demographic analyses.

errors that contributed to the EDGE score were determined a priori through a Delphi consensus process among the AUA BLUS committee. Relative severity was determined by providing each member with 100 severity points to distribute accordingly across the list of errors. They were tallied, presented, discussed and tallied a second time. Errors were weighted accordingly.

Data files were excluded if 1) a subject never completed the online multiple choice examination on the AUA website, 2) any data were marked invalid by the proctor or showed obvious signs of incompletion, or 3) an automated program flagged files as demonstrating sensor malfunctions. Figure 1 shows a sample of the typical methods to identify and remediate data integrity issues. Beyond these exclusions the data were not modified in any way.

Construct Validity

Skill Categories via Self-Reported Demographics. Self-reported demographics were used as grouping variables to establish representative skill sets according to traditional techniques for establishing categories of skills based on prior work.[5] Training level or present status groups consisted of MS, R1-R5, fellows and ATTs. Self-reported laparoscopic caseloads were also solicited in 2 ways. Participants were asked to report the number of nonrobotic laparoscopic procedures that they had done per week in the last year. Participants were also asked to report the number of total nonrobotic laparoscopic procedures performed per month. These values were cross-tabulated to examine the consistency of self-reporting. Reports of a higher weekly caseload compared to the monthly caseload were taken as evidence of unreliable self-reporting.

The present status and monthly laparoscopic caseload variables were combined into groups, including MS, R1-R2, R1-R5 (each group with zero laparoscopic experience) and ATTs with 3 or more laparoscopic cases per month (table 1). This grouping allowed comparisons of groups with at least 15 subjects and 50 trials. Because virtually no urological residents reported laparoscopic experience, the group was excluded from demographic based statistical analysis. A multicomparison test using 1-way ANOVA and the Tukey honestly significant difference criterion at $p < 0.05$ were used to determine the ability to separate task metric means across the combined demographic skill categories.

Skill Level via Faculty Evaluations of Videos. A panel of 5 expert faculty urologists, including 1 each from 5 of the 8 training programs, evaluated a representative subset of task videos. Only peg transfer and suturing were selected for the representative subset. The videos showing the fastest, slowest and median task time in each of the combined demographic groups (MS and R1-R2 with zero laparoscopic experience, etc) were then selected to represent the spread in and across groups. This resulted in 24 videos (12 of peg transfer and 12 of suturing).

Faculty raters used the well validated GOALS survey tool[8] domains appended with the pass/fail final question originally appearing in OSATS[9] via a custom-built web based video display and survey tool (Zoho, Pleasanton, California). OSATS is a validated standard tool of technical skill assessment for open surgery, which uses structured surveys and a panel of faculty surgeons with expert domain knowledge. It includes a pass/fail checkbox. GOALS, which is OSATS applied to laparoscopic surgery, does not include the pass/fail checkbox. Construct validity was considered evidence that a test measured the intended construct. For example, our test (the BLUS task and scoring method) correctly distinguished and classified the skill level of a subject compared to the ground truth.

The order of the 24 videos was randomized for all raters. Reviewers were blinded to the identity of the subjects whose videos they reviewed and to the scores of other reviewers. The Cronbach α was considered to indicate the measure of concordance of faculty panel scores

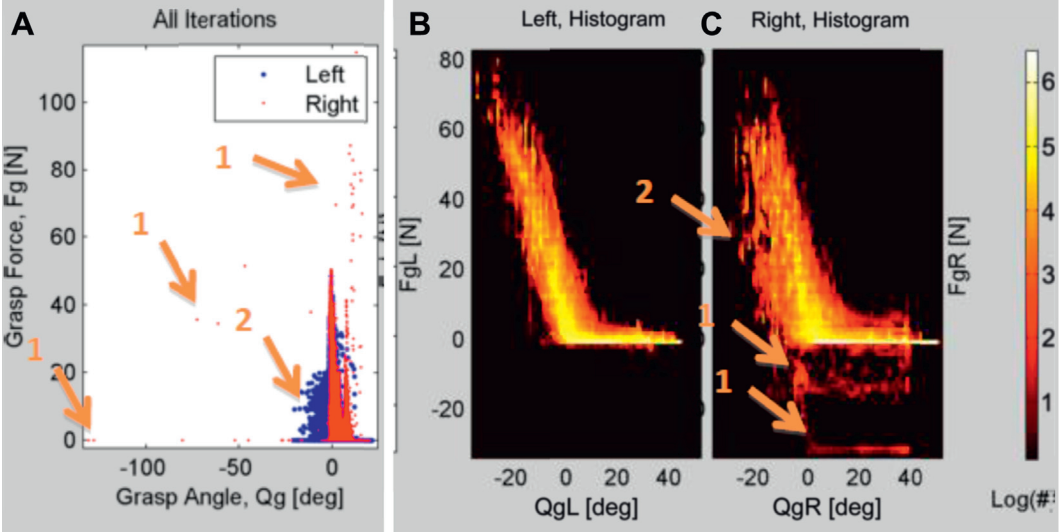

Figure 1. Cumulative plots of grasper angle vs grasper force for left and right hands on peg transfer (*A*) and suturing (*B* and *C*) tasks. Visible anomalies include outliers (*1*) and incorrectly signed data (*2*) in some recordings. Key shows frequency of occurrence heat map. Yellow indicates highly frequent observations. Dark red indicates extremely rare observations.

with a threshold of 0.9 indicating excellent agreement, 0.7 indicating good agreement and levels at or below 0.5 indicating poor, unacceptable levels of agreement.[10,11] The mean combined global GOALS ratings of the panel were considered the ground truth measure of the technical skill shown in each video.

RESULTS

A total of 454 complete trials from 117 participants were used for this study, including 110 of peg transfer, 110 of pattern cutting, 115 of suturing and 119 of clip applying. Table 1 shows the self-reported demographic spread among all tasks included in the study. For 7 subjects a total of 52 trials were excluded from the original 506 recordings of 124 subjects, including 37 (7.3%) due to proctor error during recording and 15 (3.0%) due to corrupted sensor data.

Self-reported demographics revealed internal inconsistencies. Cross-tabulation of self-reported caseloads showed disagreement between weekly and monthly categories for at least 25% of all subjects (table 2). For example, it is impossible to perform 1 or 2 cases per week and 1 or 2 cases per month. Only monthly caseloads were used on subsequent analyses. Cross-tabulation of present status and self-reported monthly caseloads indicated a dearth of laparoscopic experience among the urologists in our subject pool, even among some ATTs (table 3).

The separation of means and overall trends for task metrics across the combined skill groups was strongest for task time (fig. 2). Given the multicomparison test results at $p < 0.05$, only the suturing task provided significantly different means between the expert group of ATTs with 3 or more laparoscopic cases per month and all other skill levels. Peg transfer and pattern cutting did so for 2 other skill levels (R1-R2 and MS with zero laparoscopic experience). Clip applying only differed significantly for MS with zero laparoscopic experience.

The interrater reliability of ratings among the 5 faculty experts was excellent ($\alpha = 0.95$). Correlations between faculty scores and task metrics showed good to excellent agreement in the majority of cases (table 4).

DISCUSSION

Overall the data from this study provides moderate construct validity evidence for the ability of the BLUS technical skill examination to discriminate among laparoscopic basic skills. Two approaches were used, including self-reported demographic information and a blinded video review. In the demographic approach our choice of monthly caseload and present status as grouping variables aimed to conservatively separate skill levels and provide a balanced distribution in each group (table 1). This resulted in clear trends in favor of construct validity between these skill categories based on combined groupings and task metrics, most notably task time (fig. 2). However, only the suturing task provided a complete statistical separation of means for all skill levels with this approach.

The use of validated structured survey instruments such as GOALS and OSATS on individual task videos provided an alternative means to establish a ground truth skill rating. This approach obviated many challenges of demographic based skill groupings and provided excellent reliability ($\alpha = 0.95$). The drawback of this approach is that it is more resource intensive and challenging to scale to large data sets such as all 454 skill task videos in our existing database. Our analysis was limited to only 24 videos of the peg transfer and suturing tasks. Derived from the minimum, median and maximum task times in each demographic grouping, these tasks and videos were assumed to be most representative of the core skill set (fig. 3).

Given these assumptions, BLUS task metrics provided strong evidence of construct validity via good to excellent correlations between task metrics and faculty video ratings (table 4). Based on this finding the complete set of 454 skill task video recordings for this study is currently undergoing blinded video review using C-SATS.[12–15] C-SATS is a crowd sourced version of OSATS that was shown to be statistically concordant in skill evaluation when replacing a small panel of authoritative experts (faculty surgeons) with a large anonymous crowd in certain situations.

Our results also provide 2 surprising observations. Faculty scores showed a moderate but significant negative correlation with errors (better

Table 2. *Self-reported nonrobotic laparoscopic caseload*

No. Cases/Wk	No. Cases/Mo. (No. subjects): 0	1–2	3–5	Greater than 5
0	72	3	0	0
1–2	16	4	10	4
3–5	2	0	1	4
Greater than 5	0	0	0	1

Table 3. *Self-reported nonrobotic experience by current status*

	No. Cases/Mo				No. FLS Experience			
	0	1–2	3–5	Greater than 5	None	Heard of It	Tried Parts of It	Passed
MS	26	0	0	0	12	13	1	0
Residents:								
R1	14	0	0	0	7	4	3	0
R2	11	0	0	0	1	6	4	0
Yr 3	6	0	0	0	0	2	3	1
Yr 4	6	0	0	0	1	1	2	2
R5	4	0	0	0	0	3	0	1
Fellow	17	0	0	0	4	10	3	0
ATT	6	7	11	9	5	15	8	5

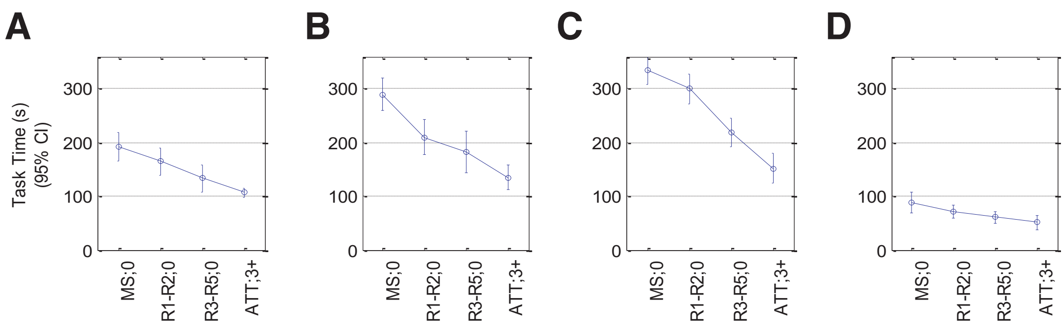

Figure 2. Construct validity results of self-reported demographic based skill groupings and task time, including peg transfer (*A*), cutting (*B*), suturing (*C*) and clip application (*D*).

scores had fewer errors) for the peg transfer task but the opposite was true for suturing. Blinded faculty video ratings effectively favored speed and rewarded errors in suturing tasks (table 4). Additionally the correlation between task time was stronger for suturing than for peg transfer. Of the 4 BLUS tasks suturing is perhaps the most clinically relevant and complex one. If this task is indeed valid, these data suggest that attending surgeons may in fact reward speed over accuracy de facto during a blinded video review despite the explicit emphasis on accuracy in GOALS or OSATS domains. This may highlight a fundamental limitation in the suturing task or in the widely used blinded video review methods. Without improving the integrity of accurately tracking errors (eg by instrumented means vs proctor observation) or considering improved suturing tasks our existing data cannot fully clarify this surprising finding.

This study revealed many limitations in our methodology of performing validation research. Self-reported data about caseload, training level, status or age are often relied on to establish categories of skill, eg the FLS program.[2,3] Such demographically derived categories of skill proved unsuitable to evaluate the construct validity of our data. This is further confounded by confusion and different protocols regarding what represents counting a case during residency. The current study indicates a significant rate of self-contradictory reporting as evidenced by the mismatch between weekly and monthly laparoscopic caseloads in more than 25% of subjects (table 2). The specific wording of the weekly caseload questionnaire may have added to this discrepancy. Such observations underscore the need for hard metrics (eg those extracted from medical records or direct measurements) in favor of self-reported counterparts.

Finally, despite coordinated efforts across 8 geographically distinct centers to recruit urological laparoscopic surgeons our data indicate that few urology residents and fellows have much laparoscopic experience (table 3). This presents yet another challenge for demographic based construct validity. More importantly it underscores the need for the BLUS tasks, examination and curriculum. Urology residents continue to experience fewer laparoscopic cases due to various factors (eg caps on resident working hours and the increasing popularity of surgical robots). Yet laparoscopic surgery remains an important part of the urological surgical skill set and it is the basis of fundamental robotic surgery.

Human error among examination proctors contributed to the majority of data exclusions despite uniform proctor training and experimental protocols. Moreover the accuracy of proctor reported task errors may be questionable, given several observed inconsistencies during a review of approximately 50 videos and comparison to tallied errors. This may undermine the observation noted that faculty reviewers may reward errors for the suturing task.

However, the stronger correlation between faculty scores and task time remains unchallenged. While the concordance among faculty was high ($\alpha = 0.95$), it was not perfect. Faculty raters remain

Table 4. *Correlation of faculty scores to task metrics in 12 videos each*

	Peg Transfer		Suturing	
	Pearson R	p Value	Pearson R	p Value
EDGE overall score	0.95	0.00	0.95	0.00
Time	−0.91	0.00	−0.95*	0.00*
Tool path	−0.95	0.00	−0.81	0.00
Movement count:				
By speed	−0.94	0.00	−0.72	0.01
By grasps	−0.87	0.00	−0.87	0.00
Jerk cost	−0.63	0.03	−0.82	0.00
Motion economy (mean speed)	0.62	0.03	Not significant*	
Grasp force variance + max		Not significant*		
Errors	−0.68	0.02	0.62*	0.03*

* Notable results for discussion.

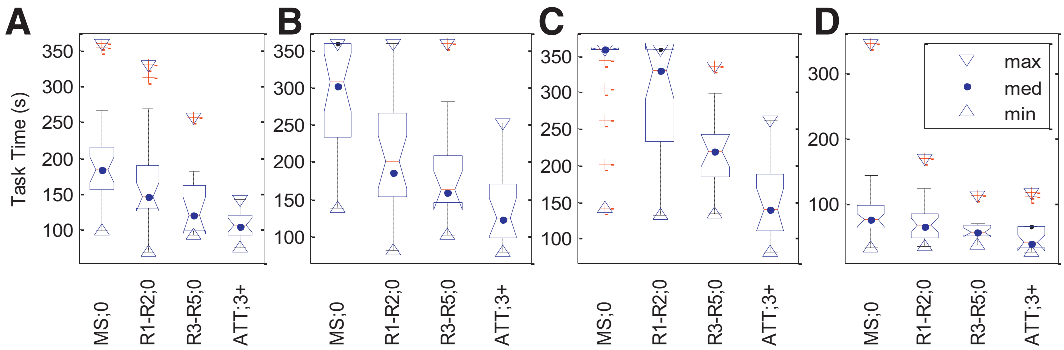

Figure 3. Selection of representative videos to establish ground truth skill categories as rated by faculty panel, including peg transfer (*A*), cutting (*B*), suturing (*C*) and clip application (*D*). Imposed time limit for all tasks was 360 seconds. Maximum (*max*), median (*med*) and minimum (*min*) task times were taken from each combined skill category. Only peg transfer and suturing were adopted for faculty review.

prone to human limitations, including finite attentional resources (eg our web interface recorded that not all faculty raters viewed all videos in their entirety) and inconsistencies in subjective evaluations (eg different scores were given for the same video by the same rater). Collectively all of these observations motivate the use of more objective means of assaying skill metrics. One possible alternative would apply instrumented tissues to provide consistent, objective, automated measurements and obviate the need for human proctors to enter errors.

The lack of any standardized assessment to certify even basic laparoscopic skills puts urology behind other similar specialties, which are all trending toward standardized means of assessment. As with any other high stakes examination (ie written boards) we expect BLUS to evolve as technical requirements, trends and continuous data from the examination help shape it with time. We believe that these data and prior results[5,6] provide enough evidence to justify the launch of a pilot program that is not tied to credentialing across residency programs so that we can answer the core question of how performance correlates with proficiency across graduating residents or other providers who are performing laparoscopic procedures. Then and only then would we advocate its use for actual high stakes testing and credentialing.

CONCLUSIONS

The BLUS peg transfer and suturing skill tasks showed good construct validity based on a consensus of established objective metrics and blinded video review by the expert faculty panel. All 4 skill tasks showed evidence of construct validity using demographically derived skill categories. However, physician self-reported demographics such as case volume proved inaccurate. Performance benchmarks based on such self-reported demographic data are questionable. When possible, the use of automated systems and sensors may improve the quality of data collection and analyses for stronger validation science in surgical simulation. Data from this work and prior results justify a pilot study across residency programs to determine whether the examination can predict the proficiency of graduating residents.

REFERENCES

1. Derossis AM, Fried GM, Abrahamowicz M et al: Development of a model for training and evaluation of laparoscopic skills. Am J Surg 1998; **175:** 482.
2. Fried GM: FLS assessment of competency using simulated laparoscopic tasks. J Gastrointest Surg 2008; **12:** 210.
3. 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.
4. Kowalewski TM, White LW, Lendvay TS et al: Beyond task time: automated measurement augments fundamentals of laparoscopic skills methodology. J Surg Res 2014; **192:** 329.
5. Sweet RM, Beach R, Sainfort F et al: Introduction and validation of the American Urological Association Basic Laparoscopic Urologic Surgery Skills curriculum. J Endourol 2012; **26:** 190.
6. Mandava SH, Liu J, Maddox MM et al: Stratification of expert vs novice laparoscopists using the Basic Laparoscopic Urologic Surgery (BLUS) curriculum at a single institution. J Surg Educ 2015; **72:** 964.
7. Collins S, Lehman DS, McDougall EM et al: AUA BLUS Handbook of Laparoscopic and Robotic Fundamentals. Linthicum: American Urological Association 2015.
8. Vassiliou MC, Feldman LS, Andrew CG et al: A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005; **190:** 107.
9. Martin J, Regehr G, Reznick R et al: Objective Structured Assessment of Technical Skill (OSATS) for surgical residents. Br J Surg 1997; **84:** 273.

10. Cronbach LJ and Shavelson RJ: My current thoughts on coefficient alpha and successor procedures. Educ Psychol Meas 2004; **64:** 391.

11. Hayes AF and Krippendorff K: Answering the call for a standard reliability measure for coding data. Commun Methods Meas 2007; **1:** 77.

12. Chen C, White L, Kowalewski T et al: Crowd-Sourced Assessment of Technical Skills: a novel method to evaluate surgical performance. J Surg Res 2013; **187:** 65.

13. Holst D, Kowalewski TM, White LW et al: Crowd-Sourced Assessment of Technical Skills: an adjunct to urology resident surgical simulation training. J Endourol 2014; **29:** 604.

14. Holst D, Kowalewski TM, White LW et al: Crowd-Sourced Assessment of Technical Skills: differentiating animate surgical skill through the wisdom of crowds. J Endourol 2015; **29:** 1183.

15. White LW, Kowalewski TM, Dockter RL et al: Crowd-sourced assessment of technical skill: a valid method for discriminating basic robotic surgery skills. J Endourol 2015; **29:** 1295.

EDITORIAL COMMENTS

The authors present a labor intensive study validating the AUA BLUS tasks, which together with a cognitive examination represent a curriculum comparable to FLS administered by ACS. The focus of this study was on validation of the 5 psychomotor tasks, which were able to discriminate basic laparoscopic skill levels among students, residents and faculty.

The ultimate goal of the BLUS program is not clear. Currently a passing score on FLS is required to sit for the written board examination in general surgery. Will the ABU (American Board of Urology) be receptive to a similar requirement? Issues such as cost will need to be addressed. Most importantly the reality is that conventional laparoscopy has generally been abandoned in favor of robotic surgery. Even technically simpler procedures such as adrenalectomy and nephrectomy are being performed robotically at many centers. As such how relevant is or will BLUS be going forward for our trainees? The real question, one that I hope the investigators will pursue, is whether the psychomotor tasks of BLUS can be used to assess robotic urological skills. I believe that the cognitive examination should be so used.

Jeffrey A. Cadeddu
Department of Urology
University of Texas Southwestern Medical Center
Dallas, Texas

The AUA funded this study with the purpose of creating a validated curriculum including psychomotor skills in traditional laparoscopy similar to the FLS curriculum, which is widely used by general surgeons. The BLUS project is the first step toward that goal. Four exercises were selected, including peg transfer, pattern cutting, needle driving and knot tying. A panel of 5 expert faculty urologists drawn from 5 of the 8 training programs evaluated a representative subset of 24 task videos (peg transfer and suturing) from the total of 454 skill task video recordings. The BLUS peg transfer and suturing skill tasks showed good construct validity. C-SATS is currently used for blinded review of all 454 skill task video recordings. Reporting these results would have added to the usefulness of this study. There was a significant lack of self-reported laparoscopic experience even among the attending urologists in this study.

The increasing use of robotic assistance in the United States for minimally invasive urological surgery will result in a further decrease in laparoscopic expertise without robotic assistance among urologists. However, basic laparoscopic skills are an important prerequisite for safe robotic surgery. We must continue to develop a validated psychomotor skill curriculum for laparoscopic and robotic urological surgery.

Chandru P. Sundaram
Department of Urology
Indiana University School of Medicine
Indianapolis, Indiana

REPLY BY AUTHORS

The shift from laparoscopy to robotics is undeniable. However, the development and demonstration of basic laparoscopic skills remain important. The upcoming multispecialty fundamentals of robotic surgery curriculum that is currently undergoing validation trials will be the next obvious step.

C-SATS will make studies such as this less labor intensive and enable our specialty to rapidly establish validity evidence as well as assessment for these and other technical skill sets for urology. Public demand for documenting technical skill is inevitable. Hopefully our relative governing bodies will maintain control and oversight of our specialty and be proactive through the adoption and development of a series of technical skill programs with validity evidence such as the AUA BLUS.