EDUCATION CORNER

From the Society for Vascular Surgery

The model for Fundamentals of Endovascular Surgery (FEVS) successfully defines the competent endovascular surgeon

Cassidy Duran, MD,[a] Sean Estrada, PhD,[b] Marcia O'Malley, PhD,[b] Malachi G. Sheahan, MD,[c] Murray L. Shames, MD,[d] Jason T. Lee, MD,[e] and **Jean Bismuth, MD,[a]** *Houston, Tex; New Orleans, La; Tampa, Fla; and Stanford, Calif*

Objective: **Fundamental skills testing is now required for certification in general surgery. No model for assessing fundamental endovascular skills exists. Our objective was to develop a model that tests the fundamental endovascular skills and differentiates competent from noncompetent performance.**

Methods: **The Fundamentals of Endovascular Surgery model was developed in silicon and virtual-reality versions. Twenty individuals (with a range of experience) performed four tasks on each model in three separate sessions. Tasks on the silicon model were performed under fluoroscopic guidance, and electromagnetic tracking captured motion metrics for catheter tip position. Image processing captured tool tip position and motion on the virtual model. Performance was evaluated using a global rating scale, blinded video assessment of error metrics, and catheter tip movement and position. Motion analysis was based on derivations of speed and position that define proficiency of movement (spectral arc length, duration of submovement, and number of submovements).**

Results: **Performance was significantly different between competent and noncompetent interventionalists for the three performance measures of motion metrics, error metrics, and global rating scale. The mean error metric score was 6.83 for noncompetent individuals and 2.51 for the competent group ($P < .0001$). Median global rating scores were 2.25 for the noncompetent group and 4.75 for the competent users ($P < .0001$).**

Conclusions: **The Fundamentals of Endovascular Surgery model successfully differentiates competent and noncompetent performance of fundamental endovascular skills based on a series of objective performance measures. This model could serve as a platform for skills testing for all trainees. (J Vasc Surg 2015;62:1660-6.)**

The idea behind fundamentals testing is not novel. In response to the core competency mandates for practice-based learning by the Accreditation Council for Graduate Medical Education (ACGME) as well as public demand for changes in surgical education and the role of learners in patient care, there has been rapid development of tools designed to train and assess surgical resident performance. In conjunction with the Society of American Gastrointestinal and Endoscopic Surgeons, the ACGME has now successfully validated two fundamental skills assessment models for laparoscopic and endoscopic procedures. The goal of these fundamentals programs was to "teach a standard set of [skills] in order to help ensure a minimal standard of care for all patients undergoing" laparoscopic and endoscopic procedures.[1] The Fundamentals of Laparoscopic Surgery (FLS) program has undergone robust validation and is now the most broadly adopted simulation program for assessing technical skill. FLS certification, a prerequisite for certification through the American Board of Surgery, contains both cognitive and skills components, and has been correlated with clinical performance.[2] Similarly, as of 2015, Fundamentals of Endoscopic Surgery (FES) is also required for board certification, after robust validation of the model for endoscopic assessment.[3] These fundamentals models serve as a template for how to successfully develop and validate a model for standard evaluation of practitioners in a given field.

From the Methodist DeBakey Heart & Vascular Center, Department of Cardiovascular Surgery, Houston Methodist Hospital,[a] and the Department of Mechanical Engineering, Mechatronics and Haptics Interfaces Laboratory, Rice University,[b] Houston; the Department of Vascular Surgery, Louisiana State University School of Medicine, New Orleans[c]; the Department of Surgery, University of South Florida, Tampa[d]; and the Department of Vascular Surgery, Stanford University School of Medicine, Stanford.[e]

Author conflict of interest: none.

Presented at the 2014 Vascular Annual Meeting of the Society for Vascular Surgery, Boston, Mass, June 5-7, 2014.

Additional material for this article may be found online at www.jvascsurg.org.

Correspondence: Cassidy Duran, MD, Department of Cardiovascular Surgery, 6550 Fannin St, Ste 1401, Houston, TX 77030 (e-mail: caduran@houstonmethodist.org).

The editors and reviewers of this article have no relevant financial relationships to disclose per the JVS policy that requires reviewers to decline review of any manuscript for which they may have a conflict of interest.

0741-5214

http://dx.doi.org/10.1016/j.jvs.2015.09.026

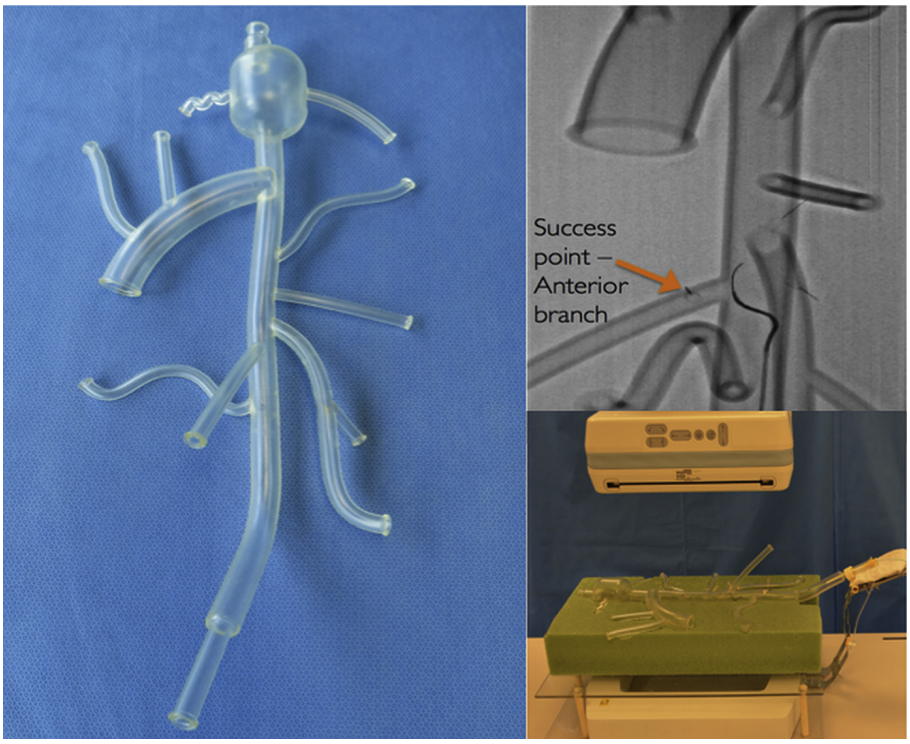

Fig 1. Fundamentals of Endovascular Surgery (FEVS) model with eight tasks designed to test basic endovascular skill.

To date, endovascular simulation has primarily been promoted as a tool for vendors to promote and train clients in their new devices. Although the developers of the simulation technology have created generic simulations that have been validated to varying degrees, including carotid artery stenting, renal interventions, and lower extremity angioplasty,[4-6] there is no precedent in the literature for this procedure-based approach to skills assessment on a broad scale. In fact, the validated fundamentals models have intentionally sought to avoid the procedure-specific approach to better identify requisite technical skills that apply to a broad range of laparoscopic/endoscopic procedures for the purpose of trainee assessments.

Given the ever-expanding procedures being performed by endovascular means, facilitated by the new devices appearing on the market, expecting each trainee to be tested on individual procedures would be extremely challenging if not impossible. In reality, endovascular procedures, as with laparoscopic and endoscopic procedures, can be broken into a series of tasks, and competency builds on a foundation of basic skills that apply to most interventions. The subtle refinements of those skills (informed by sound clinical judgment) are the makings of an endovascular expert. The aim of this study was to develop and beta-test a simulation platform for its validity for fundamental skills assessments of endovascular procedures.

Although many global rating devices for endovascular procedures have been cited in the literature, at present, no universally accepted device has been adopted by vascular educators. This is partly because the studies that have used global rating scales to evaluate endovascular skill have been weakened by a lack of reliability testing, were limited to testing a single simulated procedure (ie, carotid artery stenting), or failed to evaluate performance in the operating room as a standard, rather, basing all performance findings on virtual-reality simulation cases.[4-6] Before initiating this study, we therefore developed and internally validated a global assessment tool, applicable to all endovascular procedures, for technical endovascular skill in a live operating room setting. We then used this tool for assessment purposes in the current study (Appendix, online only).

In response to the standard criticism that assessment of procedural-based performance relies too heavily on subjective outcomes measures, electromagnetic tool tip tracking was performed to allow for motion analysis and, therefore, objective measurements of performance. These metrics characterize the quality of movement by smoothness and efficiency using validated algorithms that have successfully identified "expert movement" across a broad range of domains. Motion metrics derived from motor control-inspired metrics have proven useful for quantifying movement for different motor control tasks performed by healthy individuals and people recovering from various movement-inhibiting injuries and diseases.[7-11]

METHODS

Development of the experimental platform. The Fundamentals of Endovascular Skills (FEVS) model is nonanatomic (although anatomically inspired) and designed to enable the assessment of basic endovascular skills (Fig 1).

Table I. The fundamental tasks list, as developed and refined by consensus of members of the education committee of the Association of Program Directors in Vascular Surgery (APDVS)

Task	*Description*
1	Navigate up and over a bifurcation
2	Reshape a reverse curve catheter (task requires anteriorly oriented, downsloping branch cannulation.)
3	Imaging using oblique C-arm angulation (navigate into a third order vessel with posterior takeoff)
4	Cannulate right angle branch
5	Cannulate a branch vessel extending from an aneurysm
6	Maintain wire position during catheter/device exchange
7	Gate cannulation (tests accurate positioning and spatial awareness in aneurysmal space)
8	Cannulate branch off of type 3 arch anatomy (reshape catheter in the arch)

The process of developing the model started with querying members of the Association of Program Directors in Vascular Surgery (APDVS) education committee to make a list of skills that represent a fundamental skills set needed to perform basic endovascular procedures. During a retreat held for vascular surgery educators and members of the APDVS committee for education/simulation, participants were briefly introduced to the concept of fundamental skills assessment by one of the developers of both FLS and FES. After that, all participants in the group listed the skills they believed to meet the definition of a "fundamental task." Thereafter, each item was discussed individually to assess for redundancy and under-representation of key skills and then refined to a list of eight tasks (Table I). The additional item of ultrasound-guided vascular access was also included; however, it was not included in the final list due to limitations of the proposed platforms as well as pre-existing models and guidelines for the skill.[12]

After the initial retreat, we developed a silicone model that enabled testing of the agreed upon tasks. The group then reconvened, and the initial prototype was evaluated informally for face validity and adequate representation of the agreed upon tasks, after which two subsequent prototypes were created in response to expert feedback. A follow-up meeting was convened for final approval of the model before initiation of the testing phase.

Testing of the model. For the beta-testing phase of the study, four of the eight tasks were chosen, based on our ability to perform motion-based metric analysis without disrupting the setup for electromagnetic device tracking. Enrollment was performed consecutively and participation was voluntary.

Study participants. Twenty individuals (18 men, 2 women) consented to participate in this Institutional Review Board approved study performed at Houston Methodist Hospital. Participants included residents with minimal to moderate endovascular experience, fellows with moderate to extensive experience, and vascular surgeons with extensive endovascular experience. For the purpose of analysis, participants were categorized into a noncompetent group (those with <30 cases performed as the primary operating physician) and a competent group (≥ 30 cases performed as primary surgeon).

Tasks. The participants performed the first four fundamental endovascular tasks on each platform repeated over three sessions in a randomized fashion. The tasks included for beta-testing were cannulation of an anterior branch, an up-and-over maneuver, cannulation of a third order, posteriorly oriented branch, and a right angle branch. The first and second left lateral branches were cannulated while the C-arm was at 0° (or anterior/posterior), the anterior branch was cannulated at 75° left anterior oblique, and the posterior branch was cannulated at anterior/posterior until the catheter was sufficiently inside of the first left lateral branch, where the rest of the task was performed at 75° left anterior oblique. A preselected catheter and guidewire were used during each task to navigate to the branch of interest.

At the time of orientation to the model, participants were instructed to use proper technique as defined through expert consensus. This included elements agreed upon by the group to represent accepted standards, which emphasize or impose safe and effective practice to the users. These standards included ALARA (as low as reasonably achievable) principles of radiation exposure, maintaining safe and sufficient wire access for catheter advancement and exchanges, and use of proper reshaping techniques for maximizing benefit of preshaped catheters. The goal was to successfully advance the catheter tip to an identified success point 2 cm inside of the branch of interest in the FEVS model ≤ 5 minutes.

Performance assessment. The Global Rating Assessment Device for Endovascular Skill (GRADES, Table II) is a structured grading tool that was developed to assess individuals based on their performance in endovascular skills. One GRADES evaluation was completed for each session by a trained observer. For the purposes of this study, only two of the domains—efficiency and wire/catheter manipulation—were applicable to the tasks. A review of validated assessment tools for endovascular skills revealed that similar limitations would apply, and we therefore proceeded with the tool validated within our own institution.[5,13,14] A combined score of efficiency plus wire/catheter manipulation (score from 0 to 10) was computed for tasks performed.

Motion analysis. The kinematic movement of the catheter tip was recorded using the Aurora electromagnetic tracking device (NDI, Waterloo, ON, Canada) on the model. The NDI Window Field Generator with five degrees of freedom electromagnetic sensors was used to track the position and orientation of sensors placed in the catheter tip when navigating inside of the model. Analysis of motion metrics examines the quality of movement by deriving and quantifying number and duration submovements as well as accelerational changes (spectral arch length) that have been shown to correlate to expertise in nonsurgical domains. Applying these principles to catheter

Table II. Global Rating Device for Endovascular Surgery (GRADES) for domains of "efficiency," "wire and catheter manipulation," and "use of device" were included for analysis of robotic task performance

	Score		
Category	*1*	*3*	*5*
Efficiency	Constantly changing focus of operation or persisting at a task without progress	Slow but planned and reasonably organized	Confidently conducts operation, maintaining focus on component of the procedure until better done by another approach
Wire and catheter manipulation	Often unaware of the wire position, frequent loss of wire without losing position	Maintains awareness of wire position with occasional loss of wire access; can exchange a catheter over wire but slowly and with hesitation; occasional back and forth motion of wire	Always aware of wire position, no loss of wire access, efficient exchange of catheters over wire without hesitation
Use of the device	Inappropriate position, pressure and deployment	With effort can position the device; seems to understand appropriate pressures and deployment procedures but is hesitant.	Effortlessly positions the device in the appropriate position, and accurately uses the correct pressure and deployment strategies
Image quality	Unable to clearly capture relevant anatomy and does not understand which views are necessary for the case	Clearly captures relevant anatomy after several attempts; uses different views to do so; doesn't capture all required views for the procedure.	Clearly captures relevant anatomy within the first few attempt; understands and utilizes all required views for optimal imaging
Image safety (fluoroscopy, contrast use)	Uses much more fluoroscopy and/or contrast than is required to capture good quality images: seems unaware of the fluoroscopy or contras use	Makes an effort to minimize fluoroscopy and contrast use but uses more than is absolutely necessary to capture good quality images	Clearly understands the importance of minimizing radiation exposure and contrast use and dies so while simultaneously capturing high quality images
Autonomy	Unable to complete the entire procedure, even with extensive verbal guidance	Able to complete procedure with moderate verbal prompting	Able to complete procedure independently without verbal prompting

tip motion analysis objectively characterized the quality of motion and correlated it with performance.[15]

Data analysis. Data were obtained from the movement of the catheter tip and the global rating scale while participants executed each of the experimental tasks. After the combined global rating score per session for each participant was obtained, a Pearson correlation coefficient was computed to assess the relationship between each repeated measure from each metric to the global rating scores.

RESULTS

Overall, the success rate for task completion was 99% (100% for the competent group). Global rating scale results for the efficiency and catheter/wire manipulation domains were significantly different between the two groups. The mean total (of a possible 10 points) for was 8.6 for the competent users and 4.3 in the noncompetent group ($P = .039$). Completion times were not significantly different between the two groups for any task (Table III). Analysis of the GRADES score and its correlation to motion metrics, R values (obtained by averaging the data for each metric across all tasks and all sessions) were strong (Table IV).

Motion analysis revealed that for three of the four tasks, there were significant differences in performance level of the competent and noncompetent participants, the exception being the anterior branch (Fig 2). We believe that this was because the model allows for cannulation of the branch without correctly reshaping the catheter. This error was captured in the postprocedure video review, but in this instance, the shorter (more efficient) pathway was the incorrect maneuver, namely, failing to reshape a reverse curve catheter.

Table III. Correlation coefficient s (r) and P values comparing motion metrics and global rating scores for manual catheterization of physical model

	Model	
Metric	r	P *value*
Number of submovements	0.80	.001
Average submovement duration, seconds	0.79	.001
Spectral arc length	0.77	.001

Learning curve. In per-session analysis, we noted no improvement in performance across sessions (Fig 2), indicating that no learning occurred across the three sessions. This is important, because it indicates that proper performance on the model cannot be "learned" with limited practice.

Table IV. Completion times for competent and noncompetent participants on the Fundamentals of Endovascular Surgery (FEVS) model

Variable	*Competent, seconds*	*Noncompetent, seconds*	P *value*
Right angle	35.4	31	.74
Up and over	41.7	44.4	.35
Anterior	79.6	81.4	.89
Third order; posterior	112.4	121.9	.59

DISCUSSION

The fundamentals model has been well validated in laparoscopy and endoscopy as a means of assessing competence. The process used for successfully developing these models has likewise been well described. In light of the growing interest on the part of the surgery and vascular surgery boards as well as ACGME to include technical skills assessments in the certification process, that the model adopted is valid, accessible, and yields reproducible results is crucial. To develop a model that would test the fundamental technical skills of endovascular interventions, we relied on expert consensus from a number of experienced vascular surgeons, including several members of the APDVS education committee.

For the validation process, we explored objective motion-based metrics of performance[16] and correlated these findings with subjective assessments in the form of a global rating scale for endovascular procedures and based on previous endovascular experience. Movement smoothness is widely regarded as a hallmark of skilled, coordinated movement and has been used as a measure of motor performance in basic motor control tasks and rehabilitation applications.[13,14] Given the previous successes of objective assessment of laparoscopic surgical skill using motion-based performance metrics, we sought to explore the applicability of such measures for evaluating fundamental endovascular tasks. We included motion-based metrics inspired from principles of human motor control. These measures, such as smoothness of movement and movement efficiency, which we derived from the kinematic data of tool tip motion, give insight into the quality of movement.[10] This study is the first to explore the potential application of these metrics in the evaluation of the tasks performed during endovascular procedures.

The submovement extraction algorithms are designed to identify the number of discrete units of movements and key characteristics of those movements for motion data. They quantify the absolute number of discrete movements within a given motion aimed at completing a task and the duration of each of these movements. Motion analysis has shown that expertise for the performance of any given task is inversely proportional to number of movements and directly proportional to movement duration. Simply put, submovements are fewer in number and longer in duration for experts. These characteristics contribute to the smoothness of the pattern.

Spectral arch length examines the changes in acceleration over time by transforming them onto the frequency domain, as opposed to the time domain, and is a representation of the Fourier transformation of acceleration changes. Because performance was not correlated to time, assessment on the frequency domain proved to be a valuable alternative approach for quantifying the complexity of the speed trajectory and, therefore, to demonstrate an alternative approach for objective measurement of performance. Although these metrics all quantify smoothness, they are mathematically distinct. As such, they are not redundant, but rather enhance the overall validity of the approach.

Surprisingly, task completion time did not correlate with experience. These results were contrary to results from studies assessing the performance of laparoscopic surgery,[11,17] where task completion time was a useful metric for determining skill level. Although the novices rapidly completed some of the tasks, the process for doing so revealed maneuvers that would not be considered safe in a clinical setting (ie, advancing a catheter without a wire in place). The concept of inculcating fundamental skills in endovascular surgery is to promote safe trainees. What they ultimately end up doing as they progress into more advanced stages of their careers is not the benchmark here. Rather, this model will provide a training and assessment opportunity for a good foundation in endovascular skills.

Validation of the FEVS is contingent on not only the validity of the model itself but also the assessment tools. Motion metrics have been previously validated, as has the GRADES score. For high-stakes testing, however, there must be a more detailed and robust means of evaluating performance. This assessment should be done by a trained grader and must be reliable, valid, and easy to use. We are in the process of validating an error metric scoring tool (through analysis of video from the fluoroscopic images obtained during data collection for the current study) that will capture specific errors. Ideally, this tool would not require grading to be performed by an expert in the field, but rather, a well-trained layperson would be able to reliably assess users. Similarly validated procedural checklists and global rating scales require that the assessor have high degree of knowledge/experience,[13,14] and avoiding this requirement is crucial to the viability of the platform for widespread acceptance. Like the model itself, development of this tool requires expert consensus, must be thoroughly vetted through validity testing, and likely requires the additional input of psychometricians trained in the field of testing development, because it will serve as the basis for standardized evaluation. In its current format, the error metric scoring tool developed for the model effectively differentiates performance between the two groups. It nonetheless requires further validation steps to ensure it is sufficiently robust and reliable in this setting.

Among the primary limitations to this study, which we plan to address as we move forward with the validation

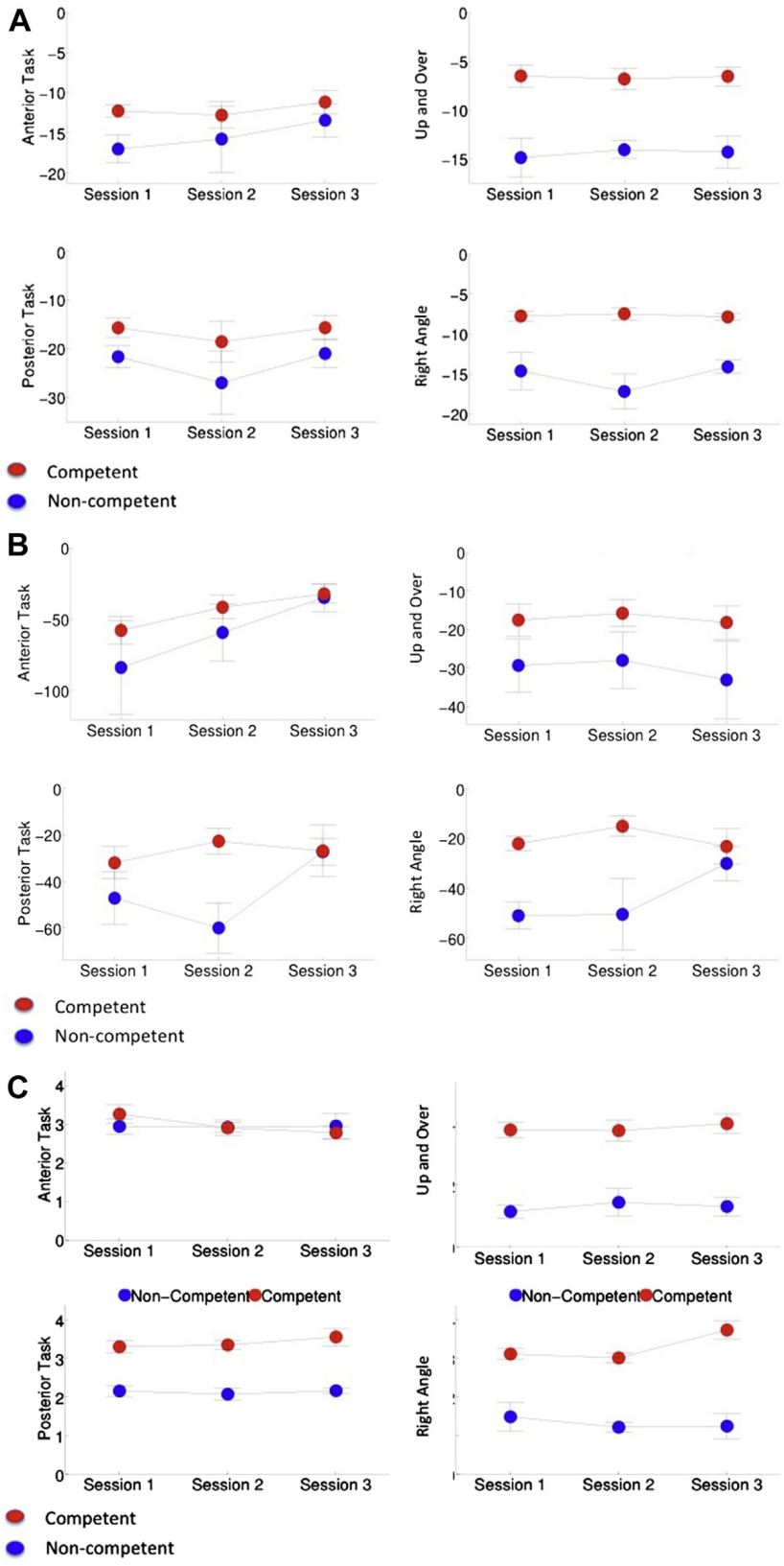

Fig 2. Mean metric scores for competent (*red*) and noncompetent (*blue*) participants performing fundamentals tasks: **(A)** spectral arc length metric scores, **(B)** number of submovements metric, and **(C)** duration of submovements metric. The *error bars* show the standard deviation.

process for this model, are the limited number of participants (especially experts) enrolled. This pilot series of data was collected from a small number of individuals at a single institution. We plan to expand our study to include multiple training institutions and increase the number of experienced individuals in our pool of participants. We will also test all eight tasks on the next round of validation testing for the model.

The silicone model assessed here offers an inexpensive model for the development of a training or testing platform that could be easily accessible to all interventional trainees. We believe model accessibility is critical for such a training program. At a cost of <$8000 for the FLS system and complete accessory package, this has likely played a major role in its success.

For the high-stakes testing platform, it is likely that a virtual model (made available at designated testing centers) will serve as a more reliable and precisely measurable platform. The FEVS model has been developed in an identical virtual format through a collaboration with Simbionix (Cleveland, Ohio). In the next phase of validity testing, the virtual model on the Simbionix simulator will serve as an additional platform for evaluation, and the larger, multi-institutional studies now in their infancy will provide the important answers to these questions.

CONCLUSIONS

The FEVS model successfully differentiates competent and noncompetent performance of fundamental endovascular skills based on a series of objective performance measures. This model could serve as a platform for skills testing for all trainees.

AUTHOR CONTRIBUTIONS

Conception and design: CD, MO, MGS, MLS, JL, JB
Analysis and interpretation: CD, SE, MO, MLS
Data collection: CD, SE
Writing the article: CD, SE
Critical revision of the article: MO, MGS, MLS, JL
Final approval of the article: CD, SE, MO, MGS, MLS, JL, JB
Statistical analysis: SE
Obtained funding: CD, JB
Overall responsibility: CD

REFERENCES

1. Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin LF, Schirmer B, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. Surgery 2004;135:21-7.
2. Barden CB, Specht MC, McCarter MD, Daly JM, Fahey TJ. Effects of limited work hours on surgical training. J Am Coll Surg 2002;195:531-8.
3. Fried GM, Marks JM, Mellinger JD, Trus TL, Vassiliou MC, Dunkin BJ. ASGE's assessment of competency in endoscopy evaluation tools for colonoscopy and EGD. Gastrointest Endosc 2014;80:366-7.
4. Dayal R, Faries PL, Lin SC, Bernheim J, Hollenbeck S, DeRubertis B, et al. Computer simulation as a component of catheter-based training. J Vasc Surg 2004;40:1112-7.
5. Berry M, Lystig T, Beard J, Klingestierna H, Reznick R, Lonn L. Porcine transfer study: virtual reality simulator training compared with porcine training in endovascular novices. Cardiovasc Intervent Radiol 2007;30:455-61.
6. Chaer RA, Derubertis BG, Lin SC, Bush HL, Karwowski JK, Birk D, et al. Simulation improves resident performance in catheter-based intervention: results of a randomized, controlled study. Ann Surg 2006;244:343-52.
7. Lin HC, Shafran I, Yuh D, Hager GD. Towards automatic skill evaluation: detection and segmentation of robot-assisted surgical motions. Comput Aided Surg 2006;11:220-30.
8. Van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. Br J Surg 2010;97:972-87.
9. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. BMJ 2003;327:1032-7.
10. Hernandez JD, Bann SD, Munz Y, Moorthy K, Datta V, Martin S, et al. Qualitative and quantitative analysis of the learning curve of a simulated surgical task on the da Vinci system. Surg Endosc 2004;18:372-8.
11. Rosen J, Brown JD, Barreca M, Chang L, Hannaford B, Sinanan M. The Blue DRAGON—a system for monitoring the kinematics and the dynamics of endoscopic tools in minimally invasive surgery for objective laparoscopic skill assessment. Stud Health Technol Inform 2002;85:412-8.
12. Lamperti M, Bodenham AR, Pittiruti M, Blaivas M, Augoustides JG, Elbarbary M, et al. International evidence-based recommendations on ultrasound-guided vascular access. Intensive Care Med 2012;38:1105-17.
13. Bech B, Lönn L, Falkenberg M, Bartholdy MJ, Rader SB, Schroeder TV, et al. Construct validity and reliability of structured assessment of endovascular expertise in a simulated setting. Eur J Vasc Endovasc Surg 2011;42:539-48.
14. Riga CV, Bicknell CD, Hamady MS, Cheshire NJ. Evaluation of robotic endovascular catheters for arch vessel cannulation. J Vasc Surg 2011;54:799-809.
15. Duran C, Estrada S, O'Malley MK, Lumsden AB, Bismuth J. Kinematics effectively delineates accomplished users of endovascular robotic with a physical training model. J Vasc Surg 2015;61:535-41.
16. Estrada S, O'Malley MK, Duran C, Schulz DG, Bismuth J. On the development of objective metrics for surgical skills evaluation based on tool motion. In: Proceedings of the 2014 IEEE International Conference on Systems, Man and Cybernetics. San Diego, CA; Oct 5-8, 2014.
17. Hofstad EF, Våpenstad C, Chmarra MK, Langø T, Kuhry E, Mårvik R. A study of psychomotor skills in minimally invasive surgery: what differentiates expert and nonexpert performance. Surg Endosc 2013;27:854-63.

Submitted Nov 4, 2014; accepted Sep 23, 2015.

Additional material for this article may be found online at www.jvascsurg.org.

APPENDIX (online only).

Intraoperative validation of a Global Rating Assessment Device for Endovascular Skill (GRADES)

Cassidy Duran, MD,[a] **Brian Dunkin, MD,**[b] **Vid Fikfak, MD,**[b] and **Jean Bismuth, MD,**[a] *Houston, Tex*

Objective: **Endovascular skills assessments have been widely implemented in simulation studies without rigorous validation. In the manner of the validation process performed for the widely endorsed Objective Structured Assessment of Technical Skills and Global Objective Assessment of Laparoscopic Skills assessments, we sought to develop a tool based on modifications from the previously used global rating scales of Reznick, Chaer, and Lee, for endovascular procedures. The aims of the study were to assess inter-rater reliability, assess internal consistency, and evaluate the tool's ability to ascertain construct validity of endovascular performance.**

Methods: **The Global Rating Assessment Device for Endovascular Skill tool was developed based on modifications from previously published work. The tool was piloted at simulation courses (n = 30) and anchors further refined to optimize accuracy of assessments. Two trained assessors used the finalized tool to complete 30 evaluations in a wide range of live endovascular cases.**

Results: **Trained assessors achieved good inter-rater reliability scores in 5 of 6 domains and the total score (*r* value range, 0.69-0.77). Internal consistency was high, with a Cronbach α score of 0.89. Experienced trainees outperformed novices across all domains.**

Conclusions: **The Global Rating Assessment Device for Endovascular Skill tool is a valid device for assessing endovascular performance and should be used as the standard global rating device for endovascular skills assessment going forward.**

Surgical skills assessment is rapidly becoming an essential component of residency and fellowship training and requires an objective, standardized toolset for evaluating trainees. This is especially true when developing a curriculum for widespread credentialing of trainees, because those being evaluated must be reliably and consistently evaluated by multiple assessors. General surgery programs have already begun incorporating these types of standard assessments, such as the Fundamentals of Laparoscopic Surgery, which is now prerequisite to sitting for the boards. Governing bodies in vascular surgery are likewise in the process of setting milestones for cognitive, nontechnical, and technical skills that are on track to be incorporated into the credentialing process in the near future. These requirements necessitate the creation of generally applicable, consistent assessment tools for endovascular procedures.

Global rating scales have been widely shown to be valid, user-friendly tools for assessing surgical skill. They have uniformly been shown to be preferable to procedural checklists throughout a variety of surgical subspecialties. Reznick et al[1] first validated a global rating scale in their description of a novel Objective Structured Assessment of Technical Skill for general surgical procedures. This methodology has since been widely applied to other disciplines and procedures with good results.[2-4]

In the endovascular community, modified versions, however, have often not been rigorously validated before being used to validate the proposed simulation tasks themselves. The studies evaluating endovascular skill using global rating scales have been weakened by a lack of reliability testing, testing a single simulated procedure (ie, carotid artery stenting), or failing to evaluate performance in the operating room as a standard, rather basing all performance findings on virtual-reality simulation cases.[5-7] As the stakes become higher and skills assessment moves from the realm of research to a required step towards certification, it is crucial that our assessment methods use reliable, reproducible, and valid tools. As such, there remains no universally agreed upon rating system for these types of procedures. The goal of this study was therefore to develop and validate a global assessment tool, applicable to all endovascular

From The Methodist DeBakey Heart & Vascular Center, Department of Cardiovascular Surgery[a]; and The Methodist Institute for Technology, Research and Education, The Houston Methodist Hospital.[b]

Author conflict of interest: none.

Presented at the 2013 Vascular Annual Meeting of the Society for Vascular Surgery, San Francisco, Calif, May 30-June 1, 2013.

Correspondence: Cassidy Duran, MD, Department of Cardiovascular Surgery, 6550 Fannin St, Ste 1401, Houston, TX 77030 (e-mail: caduran@houstonmethodist.org).

The editors and reviewers of this article have no relevant financial relationships to disclose per the JVS policy that requires reviewers to decline review of any manuscript for which they may have a conflict of interest.

procedures, for technical endovascular skill in a live operating room setting.

METHODS

Participants and evaluators. Participants were recruited under an Institutional Review Board-approved protocol. During the initial development of the global rating scale, vascular fellows were recruited at endovascular skills courses to test and refine the scale. The scale was piloted for use on simulators at two such courses and further refined based on findings from 60 participants. Thereafter, trainees at a single institution were recruited for participation in the study. Participants were categorized into an experienced group (postgraduate year 7 in the final 2 months of training) with >200 endovascular cases performed and a novice group (postgraduate year 6 within the first 2 months of training) with <30 cases performed.

Evaluators with significant endovascular experience (vascular surgeons with ≥5 years of post-training experience and a technologist in the animal laboratory for catheterization with >25 years of experience as a catheterization laboratory technologist and interventional animal laboratory researcher) were instructed on the appropriate use of the tool in a series of training sessions. Evaluations were then performed in 30 live cases.

Global Rating Assessment Device for Endovascular Skill. The Global Rating Assessment Device for Endovascular Skill (GRADES) rating scale consisting of six domains, each graded on a 5-point Likert scale, was developed through a process of expert consensus, pilot use at endovascular skills training events, and serial refinement of anchor descriptors for scores of 1, 3, and 5.

Evaluation. The trained evaluators assessed performance in the operating room in a variety of endovascular cases. These consisted of carotid artery stenting (n = 1), endovascular aortic aneurysm repair (n = 5), peripheral and renal interventions (n = 18), and central venous lesions (n = 6). Performance was scored using the GRADES tool immediately postoperatively. Trainees were then asked to use the same rating scale to self-assess their own performance. Completed assessments were separately entered into a database by a blinded third party. Correlation scores were calculated between the faculty vascular surgeon performing the case (trained assessor) and the trainee (self-assessment).

Statistical analysis. Data were collected and analyzed in Excel (Microsoft Corp, Redmond, Wash). Internal consistency was calculated using the Cronbach α, and interobserver reliability was assessed with the intraclass correlation coefficient. Construct validity was evaluated by the Student *t*-test, correlating scores with experience level. A P value of <.05 was considered statistically significant.

RESULTS

A total of 30 assessments were performed on a variety of venous and arterial endovascular procedures, including lower extremity and visceral angioplasty and stenting, endovascular aortic repair, and central venous interventions. Total performance scores correlated with trainee experience ($P = .0008$), and each domain displayed a clear difference between experienced and novice trainees. Internal consistency was high (Cronbach $\alpha = 0.84$). Good interobserver correlation was demonstrated for the total score ($r = 0.83$) and five of the six domains. Scores on the domain that correlated poorly between assessors (use of the device) correlated well ($r = 0.71$) between the two hands on participants in the case (faculty assessor and trainee); however, trained assessors demonstrated more reliable agreement in the other five domains than the assessors-trainees.

DISCUSSION

In previously published endovascular global rating scales, there is often no analysis of interobserver reliability.[5-7] A small number of studies have successfully demonstrated the validity of a scale, but only in a single procedure type and only in the simulated setting.[8-10] A 2010 review of assessments for endovascular skills found seven studies that had used global rating scales as part of their evaluation process.[10] Five of the investigators used global rating scales to assess performance on virtual-reality simulators with the aim of assessing construct validity of the simulated task, one study assessed transference of skill acquired on simulators to the operating room setting, and one used a porcine model.

Evaluation tools and simulation tasks are often lumped into a single concept for studies examining technical skills assessment. The validation steps for these two important components of developing educational models must, however, be considered independently. One must first have a well-validated means of performing assessments, which can then be applied to the validation process for the tasks themselves. By starting in a live operating room, a setting in which there can be no doubt that the task is representative of the skill being assessed, we eliminated the potential confounding variable of the simulator.

Inter-rater reliability speaks to the degree to which two trained observers agree about the performance rating awarded for a given task being assessed. Reliability in scoring is crucial for an assessment that could affect passage of milestones for certification. In arguably the most rigorous validation work performed in endovascular assessment to date, Bech et al[11,12] addressed the lack of well-validated tools available by developing the Structured Assessment of Endovascular Expertise, a comprehensive assessment tool for evaluating technical and nontechnical skills. Although the work successfully validated the tool for iliac angioplasty in a virtual environment, only three points on the scale specifically assessed technical skill.[11,12] Their work reinforced the utility of global rating scales, but the comprehensive 29-point assessment tool is too broad for use in a strict technical assessment.

The Van Heerzel et al[10] validation of simulation for carotid artery stent procedures included reliability testing of the tool, but broad applicability to other endovascular procedures is unclear. In most of the studies, modified versions of Reznick's Objective Structured Assessment of Technical

Skill tool were used without prestudy validation and often by a single user, meaning tasks were being assessed for construct validity by a single rater's score.

Although such use of a process is perhaps acceptable for introducing new concepts in virtual simulation, higher-stakes assessments require more rigorous reliability testing. Both components of reliability (ie, quality of the tool and training of the assessor) were thoroughly addressed in this study through multiple tool refinements as well as extensive discussion guided by case observation to agree on appropriate performance rating. The descriptive anchors used for scores of 1, 3, and 5 on the tool were modified after piloting of GRADES in simulation settings to minimize variance based on individual interpretation of the scale.

Internal consistency measures the degree of correlation among different items on a test that is designed to measure performance on an overall construct and is an additional means of supporting validity. Internal consistency was high (Cronbach $\alpha = 0.84$), indicating that the items tested correlate. This would be an expected result in a sound tool, because one would anticipate that as training proceeds, performance should improve equally across the individual components that comprise overall endovascular technical skill.

Finally, the construct validity—the capacity of the tool to differentiate trainees at the novice level from those with more robust endovascular experience—represents a necessary component of tool validation. Again, we believe that by assessing performance in the operating room, we have eliminated the potential confounding element of the simulator itself. Furthermore, the case variety represented in the current study ensures the broad applicability of the tool to all endovascular procedures.

This work has some notable limitations. We chose to evaluate a live case, despite the entailed bias, because consensus among our study group was that video-based performance evaluation would potentially limit our ability to fully evaluate all domains. By choosing to evaluate live case performance, blinding was not possible. This limitation was mitigated somewhat by the second assessor, who was not familiar with the participants before enrollment in the study. The strong correlations between that unbiased assessor and the faculty surgeon also participating in the case suggest that our results were not strongly impacted by bias.

CONCLUSIONS

As demand grows for including technical skills assessments in both the training and credentialing process for vascular surgeons, the need for a single reliable, well-validated, and widely applicable tool is growing. The GRADES tool is a valid device for assessing endovascular performance and should be used as the standard global rating device for endovascular skills assessment going forward.

AUTHOR CONTRIBUTIONS

Conception and design: CD, MO, MGS, MLS, JL, JB
Analysis and interpretation: CD, SE, MO, MLS
Data collection: CD, VF
Writing the article: CD, VF
Critical revision of the article: CD, VF, BD, JB
Final approval of the article: CD, VF, BD, JB
Statistical analysis: VF
Obtained funding: Not applicable
Overall responsibility: CD

REFERENCES

1. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. Am J Surg 1997;173:226-30.
2. Curry M, Malpan A, Li R, Tantillo T, Jog A, Blanco R, et al. Objective assessment in residency-based training for transoral robotic surgery. Laryngoscope 2012;122:2184-92.
3. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondre K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005;190:107-13.
4. Duran CA, Shames M, Bismuth J, Lee JT; Simulation Committee of the Association of Proram Directors in Vascular Surgery. Validated assessment tool paves the way for standardized evaluation of trainees on anastomotic models. Ann Vasc Surg 2014;28:115-21.
5. Berry M, Lystig T, Beard J, Klingestierna H, Reznick R, Lonn L. Porcine transfer study: virtual reality simulator training compared with porcine training in endovascular novices. Cardiovasc Intervent Radiol 2007;30:455-61.
6. Chaer RA, Derubertis BG, Lin SC, Bush HL, Karwowski JK, Birk D, et al. Simulation improves resident performance in catheter-based intervention: results of a randomized, controlled study. Ann Surg 2006;244:343-52.
7. Dayal R, Faries PL, Lin SC, Bernheim J, Hollenbeck S, DeRubertis B, et al. Computer simulation as a component of catheter-based training. J Vasc Surg 2004;40:1112-7.
8. Hislop SJ, Hsu JH, Narins CR, Gillespie BT, Jain RA, Schippert DW, et al. Simulator assessment of innate endovascular aptitude versus empirically correct performance. J Vasc Surg 2006;43:47-55.
9. Tedesco MM, Pak JJ, Harris EJ Jr, Krummel TM, Dalman RL, Lee JT. Simulation-based endovascular skills assessment: the future of credentialing? J Vasc Surg 2008;47:1008-13; discussion: 1014.
10. Van Herzeele I, Aggarwal R, Malik I, Gaines P, Hamady M, Darzi A, et al. Validation of video-based skill assessment in carotid artery stenting. Eur J Vasc Endovasc Surg 2009;38:1-9.
11. Bech B, Lonn L, Schroeder TV, Rader SB, Ringsted C. Capturing the essence of developing endovascular expertise for the construction of a global assessment instrument. Eur J Vasc Endovasc Surg 2010;40:292-302.
12. Bech B, Lonn L, Schroeder TV, Rader SB, Ringsted C. Construct validity and reliability of structured assessment of endovascular expertise in a simulated setting. Eur J Vasc Endovasc Surg 2011;42:539-48.