Proving the Effectiveness of the Fundamentals of Robotic Surgery (FRS) Skills Curriculum: A Single-Blinded, Multi-Specialty, Multi-Institutional Randomized Control Trial

Richard M. Satava, MD[1], Dimitrios Stefanidis, MD PhD[2], Jeffrey S. Levy, MD[3], Roger Smith, PhD[4], John R. Martin, MD[2], Sara Monfared, MD[2], Lava R. Timsina, PhD[2], Ara Wardkes Darzi, MD[5], Andrea Moglia, PhD[6], Timothy C. Brand, MD[7], Ryan P. Dorin, MD[8], Kristoffel R. Dumon, MD[9], Todd D. Francone, MD[10], Evangelos Georgiou, MD PhD[11], Alvin C. Goh, MD[12], Jorge E. Marcet, MD[13], Martin A. Martino, MD[14], Ranjan Sudan, MD[15], Justin Vale, MBBS[5], Anthony G. Gallagher, PhD[16]

1 Department of Surgery, University of Washington Medical Center, Seattle, WA

2 Department of Surgery, Indiana University School of Medicine, Indianapolis, IN

3 Department of Ob/Gyn, Drexel University College of Medicine, Institute of Surgical Excellence, Philadelphia, PA

4 Florida Hospital Nicholson Center, University of Central Florida College of Medicine, Celebration, FL

5 Department of Surgery, St. Mary's Hospital, Imperial College, London, United Kingdom

6 EndoCAS Simulation Center, University of Pisa, Pisa, Italy

7 Andersen Simulation Center, Madigan Army Medical Center, Tacoma, WA

8 Center for Education, Simulation and Innovation, Hartford Hospital, Hartford, CT.

9 Penn Medicine Clinical Simulation Center, Philadelphia, PA

10 Department of Colon and Rectal Surgery, Lahey Health and Medical Center, Burlington, MA

11, National and Kapodistrian University of Athens, Athens, Greece

12 Houston Methodist Hospital, Methodist Institute for Technology, Innovation, and Education, Houston, TX

13 USF Health Center for Advanced Medical Learning and Simulation, Tampa, FL

14 Lehigh Valley Health Network, Allentown PA.

15 Department of Surgery, Surgical Education and Activities Lab, Duke University Medical Center, Durham, NC

16 Director of Technology Enhanced Learning, ASSERT Centre, College of Medicine and Health, Brookfield Health Sciences Complex, University College Cork, Cork, Ireland

CORRESPONDENCE Richard M. Satava, MD FACS
Department of Surgery
University of Washington Medical Center
Office: 2884 Rancho Road
Pebble Beach, CA 93953 USA
Tel: (425) 765-0730
email: rsatava@uw.edu

This is the author's manuscript of the article published in final edited form as:

Satava, R. M., Stefanidis, D., Levy, J. S., Smith, R., Martin, J. R., Monfared, S., Timsina, L. R., Darzi, A. W., Moglia, A., Brand, T. C., Dorin, R. P., Dumon, K. R., Francone, T. D., Georgiou, E., Goh, A. C., Marcet, J. E., Martino, M. A., Sudan, R., Vale, J., & Gallagher, A. G. (2020). Proving the Effectiveness of the Fundamentals of Robotic Surgery (FRS) Skills Curriculum: A Single-blinded, Multispecialty, Multi-institutional Randomized Control Trial. Annals of Surgery, 272(2), 384–392. https://doi.org/10.1097/SLA.0000000000003220

Key Points

Question: Is the Fundamentals of Robotic Surgery (FRS) proficiency-based progression curriculum effective for teaching basic robotic surgery skills?

Findings: In an international multi-institutional, multi-specialty, blinded, randomized control trial, implementation of the FRS skills curriculum using various simulation platforms led to improved performance of surgical trainees on a transfer test compared with controls.

Meaning: The FRS is an effective simulation-based course for training to proficiency on basic robotic surgery skills before surgeons apply those skills clinically.

ABSTRACT

OBJECTIVE: To demonstrate the non-inferiority of the fundamentals of robotic surgery (FRS) skills curriculum over current training paradigms and identify an ideal training platform.

SUMMARY BACKGROUND DATA: There is currently no validated, uniformly accepted curriculum for training in robotic surgery skills.

METHODS: Single-blinded parallel-group randomized trial at 12 international American College of Surgeons (ACS) Accredited Education Institutes (AEI). Thirty-three robotic surgery experts and 123 inexperienced surgical trainees were enrolled between April 2015 and November 2016. Benchmarks (proficiency levels) on the 7 FRS Dome tasks were established based on expert performance. Participants were then randomly assigned to 4 training groups: Dome (n=29), dV-Trainer (n=30) and DVSS (n=32) that trained to benchmarks and control (n=32) that trained using locally available robotic skills curricula. The primary outcome was participant performance after training based on task errors and duration on five basic robotic tasks (knot tying, continuous suturing, cutting, dissection, and vessel coagulation) using an avian tissue model (transfer-test). Secondary outcomes included cognitive test scores, GEARS ratings, and robot familiarity checklist scores.

RESULTS: All groups demonstrated significant performance improvement after skills training ($p<0.01$). Participating residents and fellows performed tasks faster (DOME and DVSS groups) and with fewer errors than controls (DOME group; $p<0.01$). Inter-rater reliability was high for the checklist scores (0.82–0.97) but moderate for GEARS ratings (0.40–0.67).

CONCLUSIONS: We provide evidence of effectiveness for the FRS curriculum by demonstrating better performance of those trained following FRS compared with controls on a transfer test. We therefore argue for its implementation across training programs before surgeons apply these skills clinically.

1. BACKGROUND

Robotic surgery was introduced in clinical practice in 1995 and has seen tremendous growth since. Like other new technologies, robotic surgery introduced new skills for surgeons to master. While robotic surgery skills may be easier to acquire than laparoscopy[1], and industry sponsored training courses are already in existence and required before clinical use of existing robotic platforms[2], concerns have been raised about the competency of surgeons using this technology[3,4]. Recognizing the need for training in robotic surgery, numerous investigators have developed robotic skills curricula[5-9], but none has received universal acceptance and wide-spread adoption. There is currently no available comprehensive, industry agnostic, and uniformly accepted program to train and assess fundamental skills needed by any surgeon to perform robotic surgery.

Fundamentals of Robotic Surgery (FRS) is a basic robotic surgery skills training course for surgeons wanting to use this technology in their practice. This program was developed using a Full Life-cycle Curriculum Development (FLCD)[10] process during 4 sequential consensus conferences by 66 subject matter experts who were official representatives of all surgical societies whose members perform robotic surgery. The FRS development process was comprehensive and also included members from several surgical education and regulatory organizations[11] (Appendices 1 & 2). In short, following a needs assessment and gap analysis, the FRS development process began in December 2011 with a consensus conference focusing on defining the outcome measures of the program; seven unique robotic tasks were developed for training based on expert input that address 25 technical skills deemed necessary by experts to perform basic robotic surgery. Two additional consensus conferences focused on course development (cognitive, psychomotor, and team training components), the development of the simulator model to be used for skills training, and the metrics to be used for performance assessment in all tasks. A final conference outlined the design of the study to assess program effectiveness. Importantly, program development was "robotic-system agnostic" (i.e., not specific to one system), independent from industry influence, and limited to systems requiring total control by the surgeon (i.e., teleoperation). In addition, the FRS was designed to cover skills required by the surgeon and surgical team from the moment the patient enters the OR until they exit the OR.

The objective of this study was to demonstrate the effectiveness of this newly developed training program for basic robotic skill acquisition. We hypothesized that inexperienced robotic surgeons who would

successfully complete the proficiency-based-progression FRS curriculum would significantly improve their performance and outperform traditionally trained controls. We further sought to assess the optimal simulator for training in FRS by comparing the effectiveness of two virtual reality simulators and one physical model simulator to the controls.

2. METHODS

This multi-institutional, multi-specialty, single-blinded, parallel group randomized control trial was conducted between April 2015 and November 2016. Participating institutions were selected based on a competitive process. To be eligible, institutions had to have an American College of Surgeons' Accredited Education Institute (ACS-AEI), a minimum of 3 separate surgical specialties that were performing robotic surgery, availability of participants with variable experience in robotic surgery, and easy access to a robotic surgical system both in a simulated and clinical environment for training and testing. Participants included both robotic experts and novices. Experts were selected by the participating institute's primary investigators based on their experience, reputation, and academic publications and were required to have performed a minimum of 50 robotic cases as primary surgeons and be actively performing at least 2 robotic cases per month (averaged over 6 months). Novices included surgical residents, fellows, and faculty (Attendings) who had participated in less than 5 robotic cases. The study protocol was approved by the Institutional Review Board (IRB) of every participating institution (Appendix 3).

Study protocol

The first step of the protocol was to establish the benchmarks (proficiency levels) for training on FRS. To accomplish this, 30 experts were asked to complete the seven FRS tasks on the three simulation platforms used in the study until no performance improvement was seen on two consecutive attempts. Expert performance was then averaged for each task, outliers were removed, and the new average was used to set the benchmarks for the main study.

After FRS task proficiency benchmarks were defined, novices were voluntarily enrolled in the study protocol at each institution. Trainees completed a baseline questionnaire detailing demographic information and

prior surgical training and robotic experience. These data were deidentified and submitted to the study coordination center. In order to maintain allocation concealment and rater blindness to pre- and post-test status, the study coordination center assigned participants by simple randomization to the four study groups at the beginning of the study: the FRS physical Dome[12], the da Vinci Simulation System (DVSS), and the dV-Trainer training groups, and a control group.

The study algorithm is shown in Figure 1. Control group participants were offered the Intuitive da Vinci curriculum, the Robotic Training Network (RTN) curriculum, or participated in existing local training via their home institution. Of the 12 institutions with control group participants, 11 (92%) used all or part of the Intuitive da Vinci curriculum. One group used a local comprehensive didactic plus a local skills course. Of the 11 institutions which used the Intuitive curriculum, 4 (36%) used the entire Intuitive curriculum, 3 (28%) used the didactic portion of the Intuitive curriculum with modifications to all or some of the skills, and 4 (36%) used the RTN curriculum which includes elements of the Intuitive curriculum.

Participants of the three experimental groups first had to successfully complete the online FRS didactic course[13], followed by proficiency-based psychomotor skills training[14] on one of three platforms to which they had been randomly assigned. The Dome group trained on the FRS physical Dome (Florida Hospital Nicholson Center, Celebration, FL), the DVSS group trained using the da Vinci backpack virtual reality (VR) simulator (3-D Systems/Simbionix, Tel Aviv, Israel), and the dV-Trainer group trained using the Mimic robotic surgery VR simulator (Mimic Technologies, Inc., Seattle, WA). The three experimental groups were all asked to achieve expert-derived proficiency on the same 7 tasks (docking, ring transfer, knot tying, suturing, 4th arm cutting, puzzle piece dissection, and vessel energy dissection). To ensure consistency across training platforms, the seven tasks were first created in a VR model using computer-aided design/computer-aided manufacture (CAD/CAM) software. The model was adapted for the VR platforms used by the DVSS and dV-Trainer groups and was 3-D printed as described by Truong et al[15] to create the physical model for the Dome group (Figure 2).

Performance assessment

Robotic surgery knowledge was assessed in all groups (including the experts) via a 44-question multiple choice test developed by the investigator team. This cognitive test was first administered to the study experts

(n=30) who also provided feedback on its appropriateness. Mean expert score on the test was 39.4 correct out of 44 (89.55%) questions. Thus, the passing score for participants on this test was set at 39 or more correct answers.

Experimental groups were required to surpass this expert threshold before they could progress to the next step, which was psychomotor skills; otherwise they had to restudy the online FRS didactic course and retake the test. The performance of the baseline psychomotor skills of all groups was then assessed (pre-test) using the da Vinci Si surgical system (latest model available for robotic surgery at that point). The familiarity with the operation of the robotic system of all participants, including docking of the robotic arms and instrument insertion, was scored by the proctors using a standardized "familiarity checklist" developed by the investigators. Participants were then required to complete 5 robotic tasks on an avian tissue model (Figure 2). These tasks very closely resemble the knot tying, suturing, 4th arm cutting, puzzle piece dissection, and vessel energy dissection tasks of the training model. Participant performance was video recorded by SimCapture Ultraportable recording system with LiveCapture cloud-based video storage (B-Line Medical Inc., Washington, D.C.) and archived for later review. Upon training completion, participants' performance was reassessed on the 5 avian tissue model tasks (post-test). As a source of content validity evidence and to ensure representativeness to the domain of robotic surgery, avian tissue model tasks were designed to resemble the FRS tasks that were developed based on consensus of experienced robotic surgeons. Task format, instructions, and expert performance descriptions were the same for the training platforms used in the experimental groups and in the avian tissue model to ensure response process validity across pre-test, training, and post-test.[16,17] The similarity of the tasks between the avian tissue model and the FRS training model was ideal to capture skill transfer.[18,19]

After study completion, each video was reviewed by 2 blinded raters who recorded task duration in seconds and task errors using a 32-criteria task-specific checklist (numeric psychomotor metric test). Raters also completed the Global Evaluative Assessment of Robotic Skills (GEARS) rating scale[20]. Raters were required to participate in calibration scoring sessions prior to video assessment proper. During these sessions they reviewed and independently scored non-study videos (n = 5), and then discussed their ratings of the videos to calibrate and harmonize their approach to performance scoring.

The primary outcome of the study was participant performance based on errors and task time during the post-test on the avian tissue model. Secondary outcomes included cognitive test scores, GEARS ratings, and robot

familiarity checklist scores. We also assessed the inter-rater reliability for the assessment tools used in this project and expected it to be higher for the numeric psychomotor metric test versus the GEARS rating scale.

Sample size calculation and statistical analysis

Sample size calculation was based on the mean and standard deviation scores from the study by Goh et al[20] where a 16% difference was observed between robotic experts and novices (score of 28.3 vs. 23.9, respectively, with a standard deviation of 6). Using these data and an alpha error level of 0.05 and a power of 50% (1 – Beta), at least 10 subjects were required in each group in order to detect a difference of a similar or greater magnitude in our study. Analysis of variance (ANOVA) was used for group comparisons unless otherwise noted. We used multivariable generalized linear model (GLM) accounting for the clustering effect at institutional levels and for the baseline imbalances between the groups by controlling for the variables such as age, gender, dominant hand, and past experiences (such as using robotic simulator, DV trainer or DVSS Backpack). We used GLM to estimate the mean time and error differences in completeing all tasks at post-test from pre-test for the experimental groups compared to the control group. This analysis was repeated for each group of participants (attendings, fellows, and residents). To assess inter-rater reliability (IRR) for time measurements, we counted as agreement if both raters scored within 10 seconds of each other for time; a difference >10 seconds was disagreement. To assess IRR for the numeric psychomotor metric test scores, percent-agreement for each metric for each of the five tasks were calculated. IRR for GEARS scores was calculated similarly (percent-agreement for each metric). Of note, IRR was assessed both for the pre- and post-test scores. Because our initial data analysis revealed important baseline differences between groups despite the randomization, we also performed and report multivariate analyses that controlled for these differences.

Deidentification, randomization, blinding, and administrative management was conducted by the study coordination center, the independent Institute for Surgical Excellence (ISE)[21]; due to incompatibility of data acquisition systems at different ACS-AEIs (USA=9, International=3), data was stored locally. Video-reviewers were blinded as to institution, pre- and post-test, and trainee.

3. RESULTS

Of 14 institutions selected to participate in this trial, 2 were unable to obtain IRB approval within 1 year and withdrew. Of 123 participants enrolled, 99 (80%) completed the protocol and were included in the initial bivariate analysis (Figure 3). Surgical specialties represented included Urology, Gynecology, Colorectal Surgery, General Surgery, and Thoracic Surgery.

Participant baseline characteristics are shown in Table 1. Despite randomization, there were significant gender, age, level of training, and prior robotic and laparoscopic experience differences among groups at baseline. All groups had a 2-3x more men, except the dV-Trainer group which had twice as many women as men. In addition, while the distribution of residents, fellows, and attendings was consistent among groups, the control group had significantly more attendings (10 vs. 3-4 in other groups). Participants in the DVSS and control groups were also older and had more prior laparoscopic experience. On the other hand, all participants had minimal prior robotic experience as primary surgeons (0-4), except for the attendings in the control group who had 11 ± 18 cases.

Despite these imbalances, all groups successfully passed the cognitive test with scores > 90%, successfully passed the familiarity checklist, and performed similarly on the pre-test without statistically significant differences among groups in task completion time ($p=0.25$) or errors ($p=0.12$) (Figure 4). After training, all groups (including control) performed significantly better at post-test compared to baseline ($p<0.01$). Performance improvement for task time ranged between 32.4% and 39.5% and for task errors between 47.4% and 55.3%. Nevertheless, there were no statistically significant differences between groups at post-test for either time ($p=0.55$) or error ($p=0.59$) (Figure 4).

On subgroup analysis, attending surgeons performed significantly better at baseline compared with residents both based on task completion time ($p<0.01$) and errors ($p=0.03$), with smaller nonsignificant differences noted between attendings and fellows. However, no significant differences were noted between these subgroups during the post-test (Figure 5).

Of the 99 participants who completed the protocol, 84 participants had complete demographic data and were included in the multivariate analysis. This analysis revealed that, compared with the control group's performance during the post-test, attendings in the Dome group performed the tasks faster ($p=0.01$), fellows in the Dome and DVSS groups were also faster ($p<0.01$ for both), while residents in the Dome and DVSS groups were

faster than control but differences did not reach statistical significance (p=0.34 and p=0.23, respectively). Similarly, compared with the control group's performance during the post-test, fellows and residents in the Dome group committed fewer task errors (p<0.01) (Figure 6).

Two pairs of trained raters assessed the videos. The first pair assessed 69 videos and the second pair assessed 30 videos. Mean IRR across metrics of the numeric psychomotor metric test was high (0.82–0.97) for both rater pairs during the pre- and post-tests. However, mean IRR across GEARS metrics was low (0.40–0.67) during the pre- and post-test for both rater paris. The differences in mean IRR between the numeric psychomotor metric test and GEARS were statistically significant in all comparisons (p < 0.01).

4. DISCUSSION

In this study, we report the comprehensive development of the first multidisciplinary proficiency-based progression curriculum for basic robotic surgical skills and provide evidence for its effectiveness using a multi-institutional randomized control study design. Our hypothesis was in part proven as all training groups significantly improved their performance after training on one of the study's simulation platforms. Nevertheless, despite these improvements, training group performance at post-test did not differ from that of the control group that also improved. This surprising-at-first-glance finding is likely explained by the fact that all control group participants also participated in a "local standard" robotic skills curriculum at their local institution. Further, despite initial randomization, due to attrition the control group had 3 times as many attendings complete the study compared with the experimental groups. Our subgroup analysis demonstrated that attendings had significantly better performance at baseline compared with other less experienced participants. Therefore, this necessitated the use of multivariate analysis to control for differences in the number of attendings among groups.

Our analysis confirmed overall non-inferiorrity of the FRS to control and better performance for one or more training subgroups at post-test compared to control. The Dome group demonstrated faster and more accurate performance than control for fellow and resident participants, but not for attendings. Even though attendings got faster, they had similar errors as control at post-test. The DVSS group's residents and fellows also performed faster than controls at post-test but had similar errors. The dV-Trainer group demonstrated the smallest performance improvements after training, and in some metrics performed worse than the control group

(not statistically significant). These findings suggest that the simulation platform used for training had an impact on the transfer of acquired skills to the avian tissue model, with the physical Dome platform maximizing training effectiveness and with the DVSS VR platform close behind. The two VR platforms used in this study seemed to differ in their effectiveness as the DVSS group outperformed the dV-Trainer group based on several metrics. A recent randomized control study corroborates our findings as it demonstrated that training on the DVSS led to imporved participant performance on a urethrovesical anastomosis compared with training on the dV-Trainer[22]. However, it should also be noted that participants of the dV-Trainer group in our study reported problems with the Mimic software. Some of the 7 training tasks malfunctioned which required repeat attempts and may have resulted in unreliable assessments on the dV -Trainer before the post-test. This may have affected the post-test results. While the Mimic software for the dV-Trainer has been recently updated to correct the encountered issues, this new software was not available during our study period and therefore we cannot comment on its effectiveness.

Several systematic reviews have demonstrated the value of physical and virtual reality simulator training for laparoscopic skill acquisition[23-25]. Evidence for the effectiveness of robotic surgery simulators has been lagging behind, however[26]. Our study provides support for the effectiveness of simulation for robotic skill acquisition. The skills curriculum implemented in all platforms was based on Gallagher and O'Sullivan's proficiency-based-progression (PBP) methodology[14,27,28], which has been previously shown to effectively improve performance[29-34]. This training paradigm has been shown to be superior to traditional time-based approach[23] as it tailors training to individual needs and ensures uniform trainee performance at training completion[35].

In this study, we therefore present an effective skills curriculum for basic robotic surgery that is based on best training methodology and prepares the trainee. Importantly, the FRS curriculum has been developed following a state-of-the art curriculum development process that has included multiple consensus conferences to establish the tasks, metrics, and curriculum content based on expert robotic surgeon input from multiple surgical specialties. This comprehensive approach is unique in the literature. Furthermore, our study provides validity evidence for use of the avian tissue model in performance assessment.[16,17] Attending surgeons out-performed residents and fellows at baseline and on post-test, indicating that the internal structure of the avian tissue model was able to discriminate between more experienced and less experienced surgeons.

It should be mentioned that FRS was developed to be agnostic to any particular platform and therefore, this curriculum applies to basic robotic skills independent to the platform used. Moving forward newer platforms would have to be tested using the FRS tasks to obtain validity evidence and confirm our assumption about their universal application. In addition, knowledge specific to the use of each platform will have to be incorporated into the curriculum.

A noteworthy finding of our study is that we also found excellent IRR when our raters used task-specific checklist of psychomotor metrics for assessment, while their IRR when using the GEARS tool was insufficient despite prior rater training. A potential explanation may be that GEARS was initially designed to evaluate complex tasks and full robotic procedures as opposed to very precise simple skills. Nevertheless, this finding raises questions about the previously reported validity and reliability of the GEARS tool[12]. A growing body of research has indeed demonstrated superior reliability levels of quantitative checklist scoring systems in comparison to assessment tools that are based on Likert scales[36,37].

Limitations of the study include the observed group imbalances at baseline as described and attending attrition in the experimental groups. Nevertheless, we were able to control for these imbalances through our multivariate analysis and identify the true effect of training. In addition, video reviewers noted that despite having very specific instructions for participating institutions and frequent communication, there was some variability in the quality of the avian tissue model that was used for the assessments across sites. Furthermore, while participants in the control group received local training as expected that included some or all of the Intuitive curriculum, robotic training curricula varied across institutions, and we did not record the amount of robotic training each control group participant received. To account for these potential confounders, we controlled in our multivariate analysis for institutions. As mentioned previously, the dV-Trainer group results were likely influenced by platform issues at the time the study was performed. Since the conclusion of the study period, the the dV-Trainer software has been updated, so this group may now have had different training outcomes. Finally, we did not assess group performance during robotic procedures in the operating room. While intraoperative assessments were originally considered in this study, logistical, ethical, and financial issues did not allow its inclusion.

5. CONCLUSION

Using FLCD methodology and PBP process models, the FRS curriculum and simulator models have been created as an "open source" course (freely available at www.frsurgery.org) to be adopted, adapted, or reconfigured to suit requirements of those in need of a resource for robotic surgery training and assessment of basic robotic skills. In this international multi-institutional, non-inferiority blinded, randomized control trial evidence is provided of effectiveness of this curriculum by demonstrating better performance of those trained using FRS compared with controls and argues for FRS implementation across training programs. It is noted that IRR assessment with psychomotor performance metrics was high but was insufficient with GEARS.

6. ACKNOWLEDGEMENTS

Intuitive Surgical and the U.S. Department of Defense provided funding to support the four consensus conferences. Intuitive Surgical additionally provided funding to conduct the validation trial, but did not participate in curriculum development or trial design. The authors would like to acknowledge the significant contributions of the primary investigators and local cooridnators at each site who made this study possible.

Appendix 1. Organizations participating in the 4 consensus conferences

American Association Gynecologic Laparoscopy (AAGL) **
American College of Surgeons (ACS)
American Congress of OB-Gyn (ACOG)
American Urologic Association (AUA)**
American Academy of Orthopedic Surgeons (AAOA)**
American Association of Thoracic Surgeons (AATS)**
American Association of Colo-Rectal Surgeons (ASCRS)**
European Urology Association (EUA)
Florida Hospital Nicholson Center
Intuitive Surgical *
U.S. Department of Defense (DoD) *
U.S. Department of Veterans Health Affairs (VHA)
Minimally Invasive Robotic Association (MIRA)
Society for Robotic Surgery (SRS)
Society of American Gastrointestinal and Endoscopic Surgeons (SAGES)**
American Board of Surgery (ABS)
Accreditation Council of Graduate Med Education (ACGME)
Association of Surgical Educators (ASE)
Residency Review Committee (RRC) – Surgery
Residency Review Committee - Urology
Royal Australasian College of Surgeons (RACS)
Royal College of Surgeons-Ireland (RCSI)
Royal College of Surgeons-England (RCS)

*– Funding organizations
**– FRS Executive Committee

Appendix 2. Individuals participating in the 4 consensus conferences

Arnold Advincula, MD, FACS
Rajesh Aggarwal, MBBS
Abdulla Ali Al Ansari, MD, FRCS
David M. Albala, MD
Richard L. Angelo, MD
Mehran Anvari, MD
John Armstrong, MD, FACS
Garth Ballantyne, MD, MBA
Michele Billia, MD
James F. Borin, MD
David M. Bouchier-Hayes, MD
Timothy C. Brand, MD, FACS
Jan Cannon-Bowers, PhD
Sanket Chauhan, MD
Rafael F. Coelho, MD
Geoff Coughlin, MD
Alfred Cuschieri, MD
Prokar Dasgupta, MD
Ellen Deutsch, MD
Gerard Doherty, MD
Brian J. Dunkin, MD, FACS
Susan G. Dunlow, MD
Gary Dunnington, MD
Ricardo Estape, MD
Peter Fabri, MD
Vicenzo Ficarra, MD
Marvin Fried, MD
Gerald Fried, MD
Vicenzo Ficarra, MD
Anthony G. Gallagher, PhD
Larry R. Glazerman, MD, MBA
Teodor Grantcharov, MD, PhD, FACS
Piero Giulianotti, MD
David Hananel
James C. Hebert, MD, FACS
Robert Holloway, MD
Santiago Horgan, MD
Jacques Hubert, MD
Wallace Judd, PhD
Lenworth Jacobs, MD
Arby Kahn, MD
Keith Kim, MD, FACS
Sara Kim, PhD
Michael Koch, MD, FACS
Timothy Kowalewski, PhD
Rajesh Kumar, PhD
Kevin Kunkler, MD
Gyunsung Lee, PhD
Thomas S. Lendvay, MD
Raymond J. Leveillee, MD
Jeffrey S. Levy, MD
C.Y. Liu, MD
Fred Loffer, MD
Guy Maddern, FRACS
Scott Magnuson, MD
Javier Magrina, MD
Michael Marohn, MD
David Maron, MD
Martin A. Martino, MD, FACOG
W. Scott Melvin, MD
Francesco Montorsi, MD
Alex Mottrie, MD
Paul Neary, MD, FRCSI
Kenneth Palmer, MD
Eduardo Parra-Davila, MD, FACS
Ceana Nezhat, MD
Manuela Perez, MD, PhD
Cyril Perrenot, MD
Gary Poehling, MD
Vipul R. Patel, MD
Sonia L. Ramamoorthy, MD, FACS
Koon Ho Rha, MD, FACS, PhD
Judith Riess, PhD
Bernardo M. Rocco, MD
COL Robert Rush, MD
Richard Satava, MD, FACS
Brendan Sayers, MD
Daniel J. Scott, MD
Steve Schwaitzberg, MD
Neal Seymour, MD
Nazema Siddiqui, MD
Mika Sinanan, MD, PhD, FACS
Roger D. Smith, PhD
Hooman Soltanian, MD
Dimitrios Stefanidis, MD, PhD, FACS
Chandru Sundaram, MBBS
Robert Sweet, MD, FACS
Amir Szold, MD
Raju Thomas, MD
Oscar Traynor, MD
Edward Verrier, MD, FACS
Gregory S. Weinstein, MD
Thomas Whalen, MD

Appendix 3. ACS-Accredited Educational Institues participating in the FRS validation trial

Andersen Simulation Center at Madigan Army Medical Center, Tacoma, WA
Carolinas Simulation Center, Carolinas Medical Center, Charlotte, NC
Center for Education, Simulation & Innovation at Hartford Hospital, Hartford, CT
Centro EndoCAS, University of Pisa, Pisa, Italy
Duke University Medical Center, Durham, NC
Imperial College London, London, UK
Lahey Hospital & Medical Center, Burlington, MA
Lehigh Valley Health Network, Allentown, PA
Methodist Institute for Technology, Innovation, & Education, Houston, TX
University of Athens Medical School, Athens, Greece
Penn Medicine Clinical Simulation Center, University of Pennsylvania, Philadelphia, PA
University of South Florida Center for Advanced Medical Learning and Simulation, Tampa, FL

Figure 1. Protocol for FRS validation trial

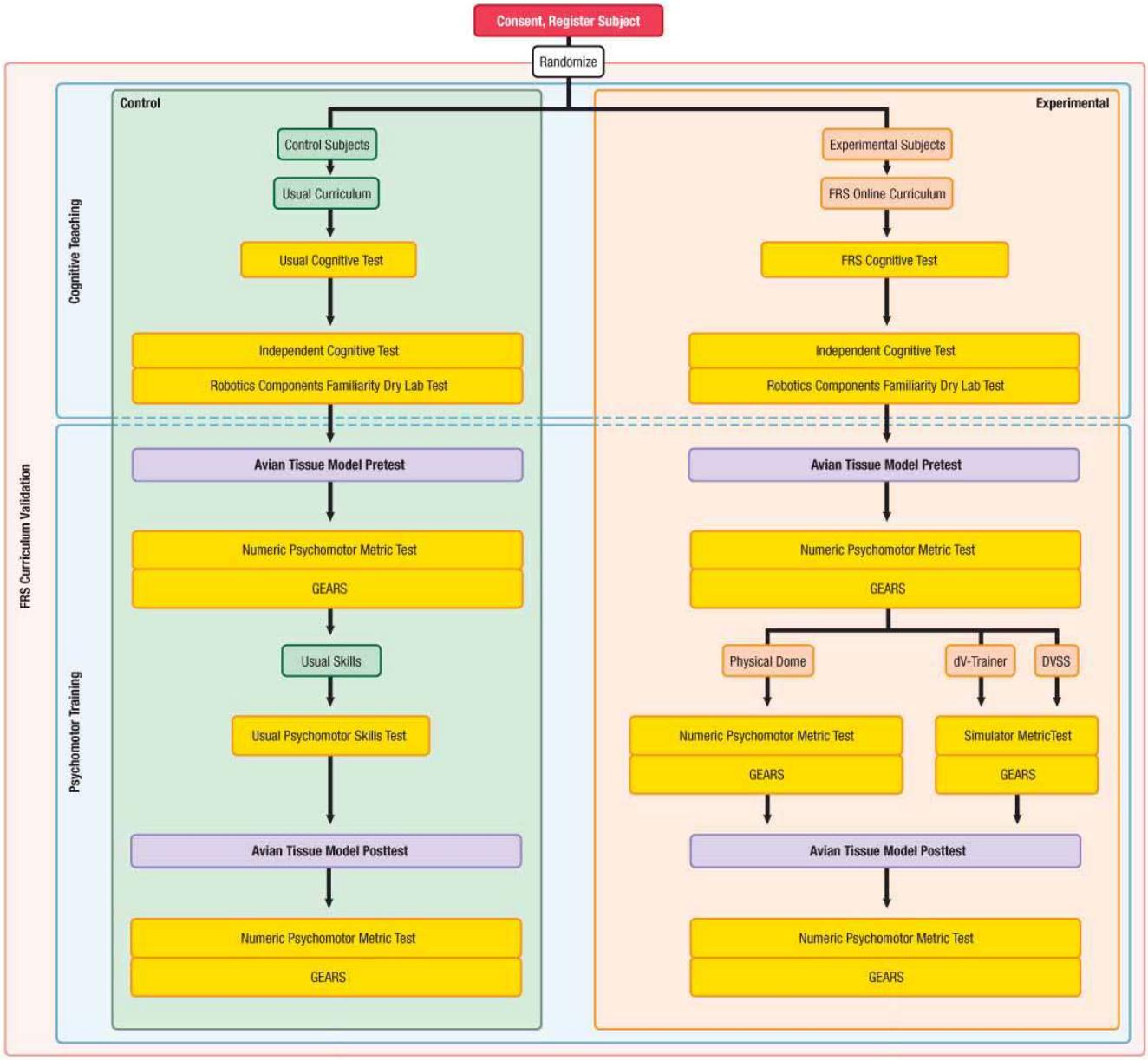


Figure 2. The FRS VR Dome (upper left), 3-D printed FRS physical Dome (upper right), and avian tissue model (lower left, lower right)

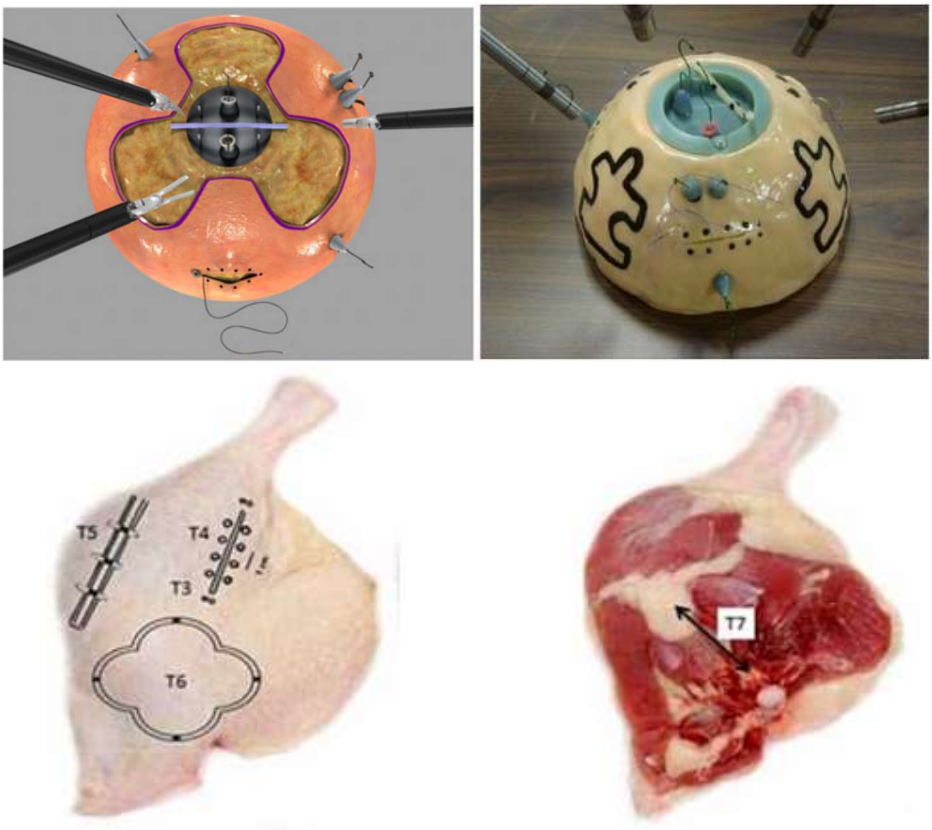


Figure 3. Consort diagram

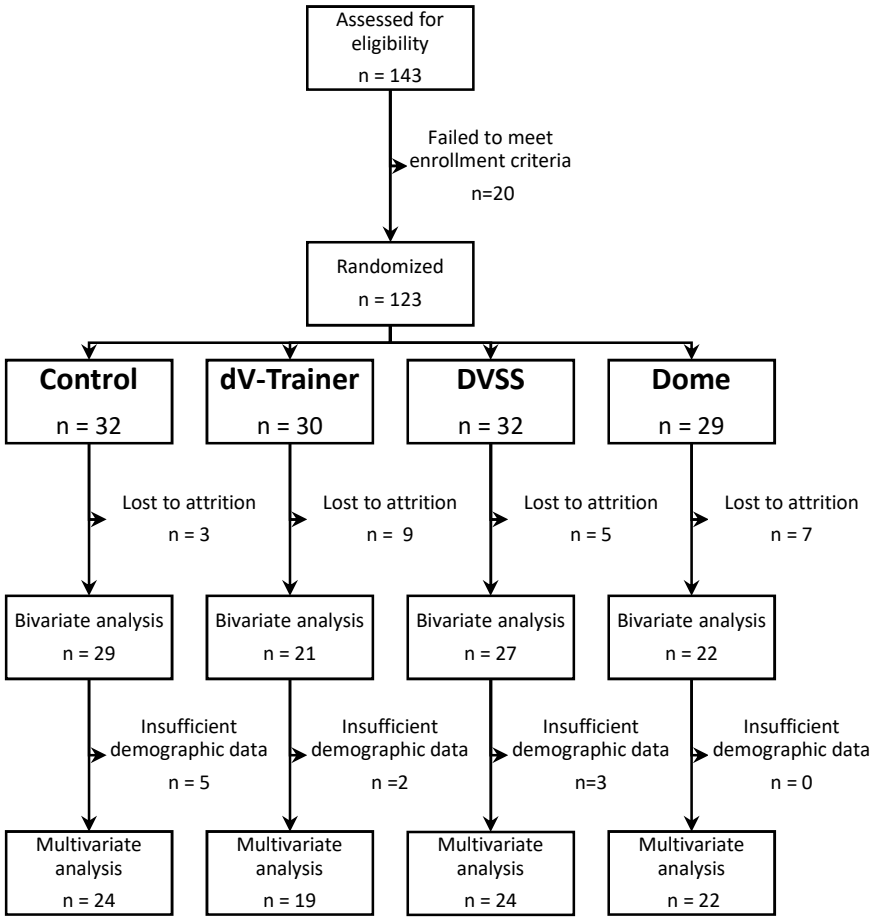
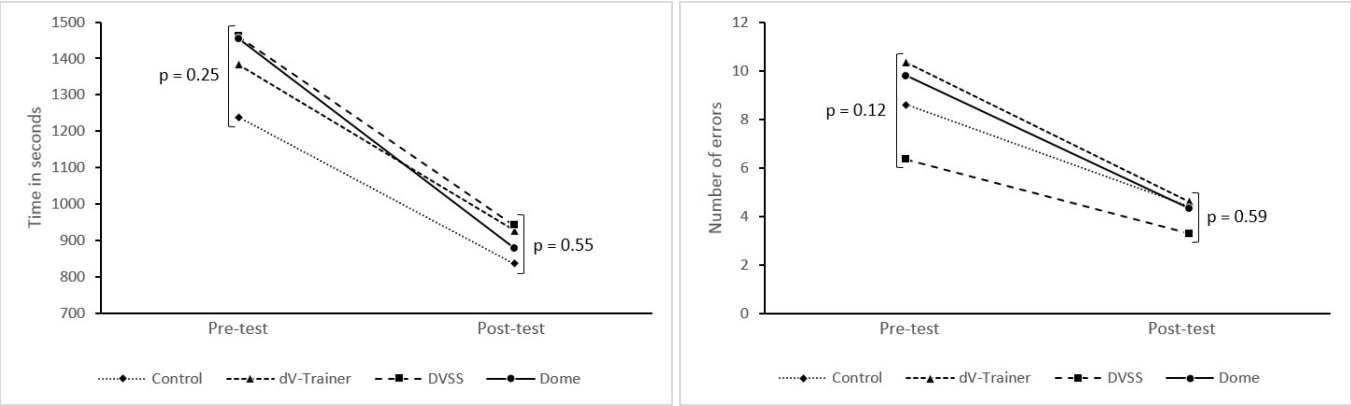
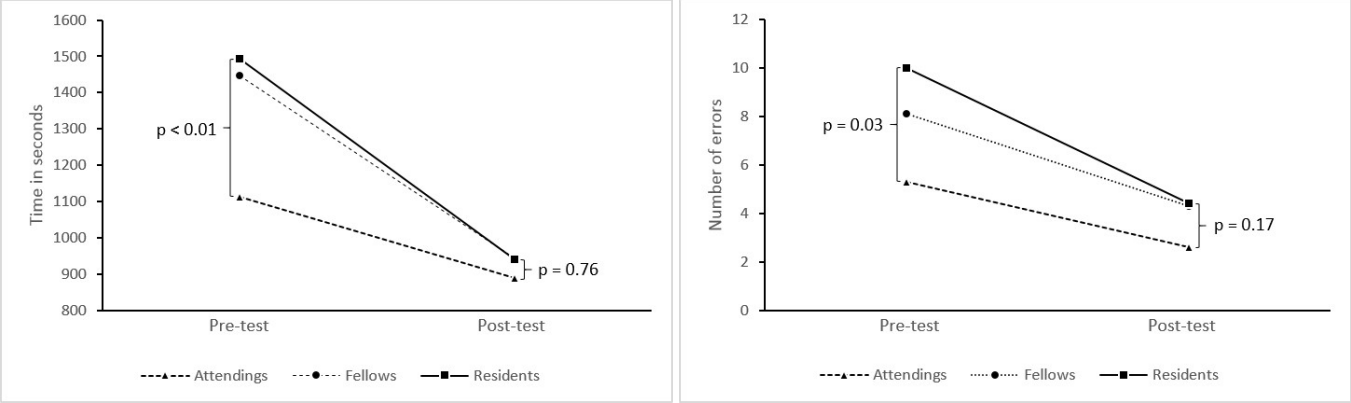


Figure 4. Study group performance comparison (task time & errors)



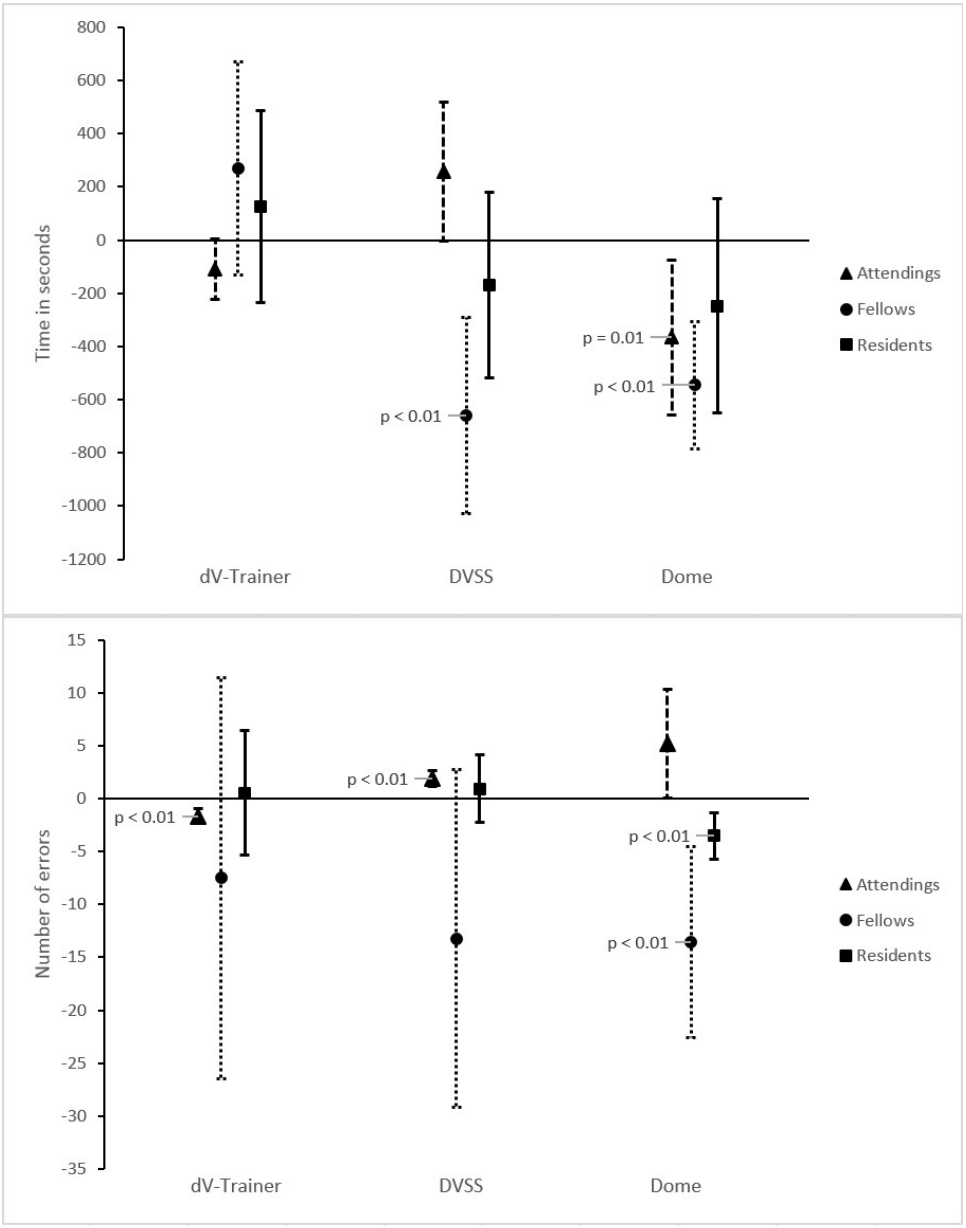
Symbols represent average group performance in all tasks in seconds (left) and average number of errrors in all tasks for each group (right). Change from pre- to post-test was significant ($p < 0.01$) for all groups for both time and errors.

Figure 5. Performance comparison of participants based on training level (task time & errors)



Symbols represent average group performance in all tasks in seconds (left) and average number of errors in all tasks for each group (right). Change from pre- to post-test was significant ($p < 0.01$) for all groups for both time and errors.

Figure 6. Group performance comparison with control group as reference (task time & errors)



Symbols represent estimated time (top) and error (bottom) difference from control group in seconds in all tasks adjusted for baseline differences among groups. Error bars represent standard error of the estimates. Control group performance has been set as 0 in the Y-axis. P-values refer to comparison of each group's performance with control on the same tasks.

Table 1. Participant demographic details

Group numbers		**Resident**	**Fellow**	**Attending**	**Total***
	Control	11 (38%)	2 (7%)	10 (34%)	29
	dV-Trainer	11 (52%)	4 (19%)	4 (19%)	21
	Dome	13 (59%)	4 (18%)	3 (14%)	22
	DVSS	13 (48%)	6 (22%)	3 (11%)	27
Group gender (Male:Female)		**Resident**	**Fellow**	**Attending**	**Total****
	Control	9:2	0:2	9:1	18:6
	dV-Trainer	6:5	1:3	0:4	7:12
	Dome	9:4	2:2	2:1	15:7
	DVSS	8:5	6:0	2:1	18:6
Age of participants		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
	Control	30 (3)	33 (0)	42 (8)	36 (8)
	dV-Trainer	29 (9)	34 (3)	35 (5)	31 (8)
	Dome	30 (3)	37 (8)	35 (4)	33 (6)
	DVSS	33 (9)	38 (3)	44 (14)	37 (10)
Practice Years		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
	Control	4.5 (2.8)	7.5 (2.1)	14.4 (9.0)	8.9 (7.9)
	dV-Trainer	4.5 (2.5)	6.4 (0.8)	8.3 (4.6)	5.8 (3.1)
	Dome	3.6 (1.9)	8.0 (4.2)	8.7 (1.5)	6.0 (4.9)
	DVSS	5.5 (8.1)	10.0 (4.0)	21.0 (15.6)	9.7 (9.8)
Years of robotic experience		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
	Control	0.8 (1.5)	2.5 (3.5)	2.6 (3.0)	1.7 (2.4)
	dV-Trainer	1.3 (1.4)	2.3 (3.3)	5.7 (1.5)	2.4 (2.6)
	Dome	0.4 (0.5)	3.0 (2.2)	1.6 (1.5)	1.2 (1.6)
	DVSS	1.0 (1.9)	0.8 (1.2)	4.0 (3.6)	1.3 (2.1)
Number of Robotic cases		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
	Control	25.6 (78.9)	46.5 (65.8)	33.9 (81.7)	29.7 (74.1)
	dV-Trainer	17.3 (31.4)	17.0 (22.9)	66.7 (76.4)	25.9 (41.9)
	Dome	9.5 (15.8)	52.5 (68.5)	10.7 (16.8)	18.2 (35.9)
	DVSS	6.7 (15.4)	11.8 (20.1)	0 (0)	7.2 (15.7)
Robotic cases as a primary surgeon		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
	Control	0.1 (0.3)	0 (0)	11.2 (17.7)	4.4 (12.0)
	dV-Trainer	0 (0)	0 (0)	3.8 (4.8)	0.8 (2.5)
	Dome	1.2 (2.1)	2.8 (4.9)	0 (0)	1.3 (2.7)
	DVSS	0 (0)	1.0 (1.7)	0 (0)	0.3 (0.9)
Total laparoscopic experience		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
	Control	80.5 (113.4)	342.5 (293.4)	322.5 (304.1)	220.7 (253.2)
	dV-Trainer	104.4 (165.8)	308.0 (246.0)	176.7 (215.5)	165.1 (200.1)
	Dome	140.0 (298.6)	175.5 (160.1)	343.3 (280.4)	165.8 (263.5)
	DVSS	183.5 (361.9)	598.0 (580.5)	750.0 (1060.7)	337.3 (497.0)
Primary surgeon laparoscopic experience		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
	Control	23.3 (45.2)	282.5 (378.3)	194.5 (185.9)	121.5 (172.8)
	dV-Trainer	35.6 (81.5)	93.0 (139.4)	41.3 (68.2)	50.1 (92.6)
	Dome	49.8 (95.0)	75.0 (86.6)	176.7 (194.0)	72.5 (113.4)
	DVSS	84.2 (229.1)	326.0 (390.4)	250.0 (353.6)	161.2 (278.4)

*****There were a number of participants whose training level was not specified

**There were a number of participants whose gender was not specified

References

1. Stefanidis D, Wang F, Korndorffer JR, Jr., Dunne JB, Scott DJ. Robotic assistance improves intracorporeal suturing performance and safety in the operating room while decreasing operator workload. *Surg Endosc.* 2010;24(2):377-382.
2. da Vinci Training. 2018; https://www.intuitivesurgical.com/training/. Accessed April 23, 2018.
3. Chen MM, Holsinger FC. Morbidity and Mortality Associated With Robotic Head and Neck Surgery: An Inquiry of the Food and Drug Administration Manufacturer and User Facility Device Experience Database. *JAMA Otolaryngol Head Neck Surg.* 2016;142(4):405-406.
4. Lee YL, Kilic GS, Phelps JY. Medicolegal review of liability risks for gynecologists stemming from lack of training in robot-assisted surgery. *J Minim Invasive Gynecol.* 2011;18(4):512-515.
5. Dulan G, Rege RV, Hogg DC, et al. Developing a comprehensive, proficiency-based training program for robotic surgery. *Surgery.* 2012;152(3):477-488.
6. Dulan G, Rege RV, Hogg DC, et al. Proficiency-based training for robotic surgery: construct validity, workload, and expert levels for nine inanimate exercises. *Surg Endosc.* 2012;26(6):1516-1521.
7. Arain NA, Dulan G, Hogg DC, et al. Comprehensive proficiency-based inanimate training for robotic surgery: reliability, feasibility, and educational benefit. *Surg Endosc.* 2012;26(10):2740-2745.
8. Green CA, Chern H, O'Sullivan PS. Current robotic curricula for surgery residents: A need for additional cognitive and psychomotor focus. *Am J Surg.* 2018;215(2):277-281.
9. Connolly M, Seligman J, Kastenmeier A, Goldblatt M, Gould JC. Validation of a virtual reality-based robotic surgical skills curriculum. *Surg Endosc.* 2014;28(5):1691-1694.
10. Zevin B, Levy JS, Satava RM, Grantcharov TP. A consensus-based framework for design, validation, and implementation of simulation-based training curricula in surgery. *J Am Coll Surg.* 2012;215(4):580-586 e583.
11. Smith R, Patel V, Satava R. Fundamentals of robotic surgery: a course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development. *Int J Med Robot.* 2014;10(3):379-384.
12. FRS Dome. http://frsdome.wordpress.com. Accessed April 23, 2018, 2018.
13. Fundamentals of Robotic Surgery Curriculum. 2018; http://www.frssurgery.org/frs-curriculum/. Accessed April 23, 2018.
14. Gallagher AG, O'Sullivan GC. *Fundamentals of Surgical Simulation: Principles & Practice.* New York, NY: Springer; 2011.
15. Truong M, Tanaka A, Perez M, Simpson K, Hearn G, Smith R. From Design to Conception: An Assessment Device for Robotic Surgeons. Interservice/Industry Training, Simulation, and Education Conference; 2014.
16. Messick S. VALIDITY. *ETS Research Report Series.* 1987;1987(2):i-208.
17. Downing SM, Yudkowsky R. *Assessment in health professions education.* Routledge; 2009.
18. Blume BD, Ford JK, Baldwin TT, Huang JL. Transfer of training: A meta-analytic review. *J Manage.* 2010;36(4):1065-1105.
19. Arthur Jr W, Bennett Jr W, Stanush PL, McNelly TL. Factors that influence skill decay and retention: A quantitative review and analysis. *Hum Perform.* 1998;11(1):57-101.
20. Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *J Urol.* 2012;187(1):247-252.
21. Consensus Conference Overview. 2018; http://frssurgery.org/curriculum-development/. Accessed April 23, 2018, 2018.
22. Hoogenes J, Wong N, Al-Harbi B, et al. A Randomized Comparison of 2 Robotic Virtual Reality Simulators and Evaluation of Trainees' Skills Transfer to a Simulated Robotic Urethrovesical Anastomosis Task. *Urology.* 2018;111:110-115.
23. Larsen CR, Oestergaard J, Ottesen BS, Soerensen JL. The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. *Acta Obstet Gynecol Scand.* 2012;91(9):1015-1028.

24. Nagendran M, Gurusamy KS, Aggarwal R, Loizidou M, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database Syst Rev.* 2013(8):CD006575.
25. Nagendran M, Toon CD, Davidson BR, Gurusamy KS. Laparoscopic surgical box model training for surgical trainees with no prior laparoscopic experience. *Cochrane Database Syst Rev.* 2014(1):CD010479.
26. Moglia A, Ferrari V, Morelli L, Ferrari M, Mosca F, Cuschieri A. A Systematic Review of Virtual Reality Simulators for Robot-assisted Surgery. *Eur Urol.* 2016;69(6):1065-1080.
27. Gallagher AG. Metric-based simulation training to proficiency in medical education:- what it is and how to do it. *Ulster Med J.* 2012;81(3):107-113.
28. Gallagher AG, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg.* 2005;241(2):364-372.
29. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg.* 2002;236(4):458-463; discussion 463-454.
30. Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Am J Surg.* 2007;193(6):797-804.
31. Van Sickle KR, Ritter EM, Baghai M, et al. Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying. *J Am Coll Surg.* 2008;207(4):560-568.
32. Gallagher AG, Seymour NE, Jordan-Black JA, Bunting BP, McGlade K, Satava RM. Prospective, randomized assessment of transfer of training (ToT) and transfer effectiveness ratio (TER) of virtual reality simulation training for laparoscopic skill acquisition. *Ann Surg.* 2013;257(6):1025-1031.
33. Angelo RL, Ryu RK, 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(10):1854-1871.
34. Cates CU, Lönn L, Gallagher AG. Prospective, randomised and blinded comparison of proficiency-based progression full-physics virtual reality simulator training versus invasive vascular experience for learning carotid artery angiography by very experienced operators. *BMJ Simulation and Technology Enhanced Learning.* 2016.
35. Stefanidis D. Optimal acquisition and assessment of proficiency on simulators in surgery. *Surg Clin North Am.* 2010;90(3):475-489.
36. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc.* 2003;17(10):1525-1529.
37. Gallagher AG, O'Sullivan GC, Leonard G, Bunting BP, McGlade KJ. Objective structured assessment of technical skills and checklist scales reliability compared for high stakes assessments. *ANZ J Surg.* 2014;84(7-8):568-573.