Impact of Physician Expertise on Probe Trajectory During Obstetric Ultrasound A Quantitative Approach for Skill Assessment

Maela Le Lous, Fabien Despinoy, Margaux Klein, Elisa Fustec, Vincent Lavoué, Pierre Jannin

Impact of Physician Expertise on Probe Trajectory During Obstetric Ultrasound A Quantitative Approach for Skill Assessment

Maela Le Lous, Fabien Despinoy, Margaux Klein, Elisa Fustec, Vincent Lavoué, Pierre Jannin

Impact of physician expertise on probe trajectory during obstetric ultrasound: a quantitative approach for skill assessment.

-Obstetrical Ultrasound Trajectory Metrics-

Maela LE LOUS, MD [1,2], Fabien DESPINOY, PhD [1], Margaux KLEIN, MD [2], Elisa FUSTEC MD [2], Vincent LAVOUE MD phD [2, 3], Pierre JANNIN MD PhD[1].

1. Univ Rennes, INSERM, LTSI - UMR 1099, F35000, Rennes, France.
2. Department of Obstetrics and Gynecology, University Hospital of Rennes, France.
3. INSERM 1242, Chemistry, Oncogenesis, Stress and Signaling, Rennes, France.

Corresponding author:

Maela Le Lous, MD

LTSI-INSERM , Université de Rennes 1, UMR 1099, 35000, Rennes, France.

Phone : Mobile +33 6 95 02 38 05 Work + 33 99 26 31 07

Fax

E-mail : maela.le.lous@chu-rennes.fr

Conflict of interest: The authors report no conflict of interest.

Funding: None

Word count 3298 ; Figure count : 2 ; Table count : 2

Abstract

Introduction The objective of the study was to identify objective metrics to evaluate the significance of a sonographer's expertise on trajectories of ultrasound probe during obstetric ultrasound training procedures.

Methods This Prospective observational study was conducted at Rennes University Hospital, department of obstetrics and gynecology. We evaluated a panel of sonographers (expert, intermediate and novice) in performing three tasks (Brain, Heart and Spine) with an obstetric ultrasound simulator (Scantrainer, Medaphor™, Cardiff, UK). The trajectories of the probe were logged and recorded by a custom data acquisition software. We computed metrics on the trajectories (duration, path length, average velocity, average acceleration, jerk, working volume) to compare the three groups and identify discriminating metrics.

Results A total of 33 participants were enrolled: 5 experts, 12 intermediates, and 16 novices. Discriminatory metrics were observed among the three levels of expertise for duration, velocity, acceleration and jerk for Brain and Spine tasks. Working volume was discriminatory for the Brain and the Heart task. Path length was discriminatory for the Brain task.

Conclusions Our results suggest a relationship between the sonographer's level of expertise and probe trajectory metrics. Such measurements could be used as an indicator of sonographer proficiency and contribute to automatic analysis of probe trajectory to evaluate the quality of sonography and the sonographer.

Funding None.

Keywords Ultrasound; Obstetrics; Simulation training; Education; Standard setting.

Introduction

Education in the field of obstetric ultrasound has shifted from an exclusively observation-based training to simulation-based training over the past decade[1]. Multiple factors may have contributed to the more widespread use of simulation: concerns for diagnostic errors, patient volunteers' or teachers' availability, and technological advances based on virtual reality (VR)[23]. Obstetric ultrasound simulators (OUS) may revolutionize the Obstetric and Gynecologic (OB/GYN) residents' curriculum for both training and evaluation[4].

The utility of OUS for ultrasound training is now well supported by evidence [5,6,7,8]. In a series of eight Danish studies, Toslgaard et al. successively explored the learning curves for novices, examined how to improve the efficiency of training with the use of dyad practice, and explored if improvements were sustained over time[9]. They also demonstrated skill transfer to subsequent clinical training[10]. Finally, they demonstrated an improvement for the patients after simulation training, such as a decrease in discomfort and improvements in their perception of safety and their confidence[10].

Evaluation of competency in obstetric ultrasound is a time-consuming process and requires pregnant volunteers who are willing to be used for training. One study has shown the potential of OUS as a substitute for evaluating trainees[11]. In this study, dexterity and quality of images obtained during evaluation were assessed by two independent examiners. However, dexterity was subjectively scored between 0 and 10 based on observation.

A fundamental issue for evaluation with OUS is the need for metrics. VR based simulators allow automatic recording of probe trajectories during training and evaluation. OUS could be a suitable tool to extract data from these probe trajectories for objective evaluation, and to identify metrics that may discriminate the level of expertise of participants during their training. The aim of our study was to assess the potential of OUS to identify objective metrics to measure expertise in ultrasonography.

Methods

Participants

The experimental set-up was conducted at the Department of Gynecology and Obstetrics at the University Hospital of Rennes. All training and assessments were carried out in an undisturbed environment. The participants were recruited in June 2018 and the study was conducted from July to September 2018. Approval was obtained from our local institutional review board.

The participants were divided into three groups based on their levels of experience: medical students (novice), OB/GYN residents (intermediate), and OB/GYN consultants (experts). All participants were recruited locally at the department of Gynecology and Obstetrics, and provided oral and written informed consent. The experts were professional who had every day practice on genuine patients (about 25 morphologic scans a week). The intermediates were interns who are familiar with routine examination scans (fetal movements, amniotic fluid, estimation of fetal weight, about 5 scans a week) but not with morphologic examination. The novices had absolutely no ultrasound experience. The participants in the 3 groups never practiced on the simulator before the study, as the simulator was bought right before study beginning.

The medical program at the University of Rennes is a 6-year traditional curriculum, and the gynecology rotations are completed during the 2 final years. The Novices were recruited during their gynecology rotations. During their prior medical training, the students had completed courses in pelvic anatomy, and ultrasound theory, but had no hands-on training. The Intermediates included OB/GYN residents who had not graduated in ultrasound but who were familiar with ultrasound equipment. The Experts were OB/GYN consultants who used ultrasound on a daily basis.

Virtual reality simulator

Training and assessments were performed using a high-fidelity VR based simulator (Scantrainer, Medaphor™, Cardiff, UK) designed for obstetric ultrasound training. It is composed of a monitor, an abdominal probe similar to a real one, docked into a Sigma haptic device, two screens and a computer (**Figure 1**). The haptic device provides realistic force feedback, when operator applies pressure on the abdominal probe. One monitor displays B-Mode 2D ultrasound pictures provided by the system and obtained from real patients. The second monitor displays an animated illustration of the probe position in a virtual patient.

Tasks performed

Step 1: Identification of tasks

Most fetal malformations are detected during the second trimester scan. This scan examines the fetus in 11 planes as requested by the National Committee for Prenatal Ultrasound[12].

Tasks reflecting the standardized second trimester examination of the fetus were selected to obtain a large sample of fetal anatomy screening, in particular tasks which involved switching from transverse to coronal or sagittal planes. We selected the tasks which would potentially reflect differences in ultrasound competences and provided the most realistic view of the fetus in three consecutive standardized planes. For Task 1 (Brain at 24 weeks of gestation), participants were asked to successively obtain the standardized view of the fetal head circumference, then the standardized plane of the cerebellum, and finally a coronal view of the brain through the cavum of septum pellucidum. For Task 2 (Heart at 22 weeks of gestation), participants were asked to successively obtain the standardized view of the four chambers, the view of the left ventricle and the aorta, and the view of the right ventricle and the pulmonary artery. For Task 3 (Spine at 22 weeks of gestation), participants were asked to obtain successively the standardized frontal view of the kidneys, a sagittal view of the spine, and the parasagittal view of the left diaphragm. **Figure 2** shows a pictorial representation of the three tasks. **The participant judged himself that the view was exact and he froze the image to finish exercise.**

All participants were asked to provide their age, gender, and years of clinical experience, and were assigned an identifier to anonymize the data. They then received a short introduction to the simulated setting, including how to operate the simulator and its functions. They also received a course about the recommended fetal planes. All trainees practiced 10 minutes in a "probe manipulation exercise" to familiarize themselves with the device, and to check the expertise difference between the groups. The score of this exercise was calculated by adding the scores of six multiple choice questions (pass: 1; fail: 0) designed to test recognition of six different 3-dimensional geometric solids (cone, pyramid, spiral, spheres, other complex geometric solids) by scanning them with the 2-D probe. Technical assistance was provided during the simulator test, but no instructions, feedback or time limit were provided. Each participant had to perform the three tasks three times.

Metric evaluation

The data were logged and recorded by a custom data acquisition software developed with Scantrainer, Medaphor^{TM}, Cardiff, UK. The probe position frequency sampling was 20 Hz. Computation of all metrics and trajectory analysis were run on a dedicated computer (Xeon E5-1650V4 @3.60GHz with 32Go RAM) using Matlab_R2017b. The analysis was done following a previously published technique[13]. A total of 6 selected metrics was analyzed:

1. **Duration (D)** corresponds to the execution time between the first time the hand moves the probe, until it has been released at the end of the task. It is measured in seconds (s).
2. **Path Length (PL)** represents the total distance travelled by the probe during the execution of the task. It is measured in millimeters (mm).
3. **Average Velocity (AV)** corresponds to the average linear speed of the probe during the task. It is measured in $mm.s^{-1}$.
4. **Average Acceleration (AA)** corresponds to the average instantaneous acceleration of the probe during the task. It is measured in $mm.s^{-2}$.
5. **Average Jerk (AJ)** corresponds to the average jerk (derivative of the acceleration) during the task, also known as "smoothness" measure. It is measured in $mm.s^{-3}$.
6. **Working volume (WV)** represents the volume of the convex hull for each trajectory. The convex hull of a trajectory is the smallest convex volume among those which contain it. It is measured in mm^3.

One-way Anova-3 tests were conducted to detect differences between the three groups of participants. **The Bonferroni correction was applied to the results to**

control alpha inflation ($p<0.08$). The participants were absolutely not told about what the metrics measured were.

Results

Demographics

Participant demographics according to expertise level are presented in **Table 1**.

Manipulation probe exercise

For the manipulation probe exercise, out of a total of six points (six multiple choice questions, pass: 1; fail: 0), the Experts' mean score was 5.6±0.5, the Intermediates' mean score was4.7±1.2, and the Novices' mean score was3.4±1.2 points ($p=0.002$). **Table 1**.

Trajectory Metrics

Analysis of the probe trajectory metrics revealed significant differences for Tasks 1, 2 and 3 between the Expert, Intermediate and Novice groups (**Table 2**).

Duration of the exercise(s)

The mean Duration(s) of the exercise differed between Novices, Intermediates, and Experts for Task 1 (respectively 157.5s ±176.2 versus 111.2s ±110.3 versus 35.6s ±16.3, $p<0.001$), Task 3 (112.8s ±64.7 vs. 82.3s ±50.6 vs. 35.5s ±11.4 $p<0.001$). For Task 2 (109.9s ±69.7 vs. 103.7s ±61.1 vs. 51.7s ±47.4 $p=0.009$) it was not significant.

Path length (mm)

The Path Length (mm) was shorter with increasing level of expertise for Task 1 (Novices: 1 442.3±1 134.4 vs. Intermediates: 1 795.4±1071.9 vs. Experts: 1 005.7±409.7, $p<0.001$) and for Task 3 (respectively, 1 515.3±569.4 vs. 1 590.8 ± 589.0 vs. 1 071.2 ±638.3, $p=0.013$).

For Task 2 (the "Heart" exercise), the Path Length was not statistically different between the three groups (1 428.7±508.2 vs. 1 266.5 ± 517.5 vs. 1 027.0 ± 1 138.3, NS).

Average velocity (mm/s)

The average velocity (mm/s) was significantly higher with increasing level of expertise for Task 1 and 3. For Task 1, average velocity was 19.2mm/s ±5.7 for Novices vs. 19.5mm/s ±13.8 for Intermediate and 29.5 mm/s ±8.7 for Experts ($p<0.001$), and for Task 3 it was respectively 15.6mm/s ± 5.2 vs. 22.2 ±7.5 vs. 29.1±8.6 ($p<0.001$). For Task 2, it was respectively 15.7 mm/s ±6.4 vs. 14.5±7.0 vs. 20.1 ± 6.4 ($p= 0.024$, NS).

Average Acceleration ($mm.s^{-2}$)

The average acceleration was statistically different between the three groups for Task 1 (137.0 $mm.s^{-2}$ ±48.2 vs. 138.6 ± 105.1 vs. 228.9 ±59.8, $p<0.001$), and for Task 3 (109.5 ± 37.0 vs. 157.3 ± 54.6 vs. 225.7 ± 80.6, $p<0.001$), but not for Task 2 (115.8±62.8 vs. 101.7±40.1 vs. 149.6±37.5, NS)

Average Jerk ($mm.s^{-3}$)

Jerk was statistically different between the three groups for Task 1 (2 755.7±1 357.0 vs. 2 728.5 ± 2 283.5 vs. 4 528.0±1 358.4, $p<0.001$) and for Task 3 (2 154.7 ± 845.5 vs. 3 099.7 ± 1 300.4 vs. 4 378.1 ± 1 605.7, $p<0.001$) increasing with level of expertise. For Task 2, there was no significant difference (2 426.8±1 886.4 vs. 2 012.1±716.7 vs. 2 929.7±761.1, NS).

Working Volume (mm^3)

The total Working Volume covered was statistically different between the three groups for Task 1 (1 118 661.6 mm^3 ±622487.1 vs. 1 438 677.7±716 948.1vs. 761 105.7±603 917.3 $p=0.002$).

For Task 2 (1 755 821.8 ± 1 318 436.4 vs. 1 275 117.0±674 373.0 vs. 825 782.0 ± 962 957.2, NS) and Task 3, the difference was not significant (1 618 668.3±624 129.9 vs. 1 759 287.7 ± 1 250 355.0 vs. 867 504.4±717 940.5, NS).

Discussion

Main Findings

This paper presents for the first time the use of metrics computed on probe trajectory during simulated fetal ultrasound with OUS to objectively assess the expertise of the user and his/her dexterity. The main result was that objective metrics (duration, acceleration, velocity and jerk) differed statistically according to the level of expertise in two of the three tasks.

The time for each task (Duration) decreased as the level of expertise increased ranging from 35s to 52s for Experts and from 109s to 157s for Novices. The Path Length, although significantly shorter for Experts in the “Brain” and “Spine” exercise, was not significantly different for the “Heart” exercise. This might be explained by the fact that the planes in the heart exercise are much more tightly bunched in a very small volume. There must have been a task effect. The “heart” exercise must have been less discriminant using trajectory metrics, because it implied tiny probe movements. The same could have been argued for the “Brain” exercise because there isn’t a lot a movement between the head circumference and the transcerebellar view, but a lot a movement is necessary to obtain the coronal view. It is to notice that some amount of initial movement among experts, must be tied to translating positioning and movements with the model to recall of positioning and movements with genuine patients.

Velocity and Acceleration were significantly higher with the increase in level of expertise suggesting that experienced sonographers move from one plane to another with a faster movement. Jerk, which is a derivative of Acceleration, may be interpreted as the variation of acceleration, or how sudden the variations are. We could have assumed that jerk would have been smaller and the gesture smoother for the experts, as it has been shown for surgical procedures[14]. However, during ultrasound examination, jerk was higher for experts because they most often combined a quick translation followed by a quick 90° rotation of the probe to switch, for example, from a transverse view of the kidney to a sagittal view of the spine. For the "Heart" exercise, average Jerk did not differ between the 3 groups, with may be explained again with the small quantity of movements between the different views. The Working Volume was smaller for the Experts than for the Intermediates and Novices, implying that the Experts remained within the area of interest. These results are similar to those observed by Zago et al about FAST (Focused Assessment with Sonography in Trauma) examination. In this study, they also used hand motion analysis to discriminate expertise, and found similar results : longer hand path and higher working volume for the novices[15]. In our study, intermediates performed more like novices than experts on many measures because they may not feel comfortable with morphologic examination.

These data are all the more important in France where, for the first time next year, the practical exam for students will be by OUS rather than on volunteers. Other developed countries are following the same trend, i.e., relying more on assessment by OUS. Objective metrics are thus required to respond to this move towards automatic and objective evaluation. Recording the trajectories of the probe, and comparing them to trajectories obtained with experts is an interesting way of evaluating students' level. In a clinical diagnostic perspective, trajectory metrics, taken separately, may not be relevant to measure performance. However, these metrics are a first step. We tested some metrics, and not all of them will have clinical meaning, but they do have a kinematic meaning. Some are obvious (duration) and some are not but may be discriminant (jerk, volume). It is part of the method to go from the identification of differences to explanation of the differences observed.

Madsen et al. also analyzed simulator-generated metrics on a high fidelity ultrasound transvaginal simulator[6]. They evaluated a group of 16 ultrasound novices along with a group of 12 OB/GYN consultants. The score was calculated by adding the scores of the seven modules (0, fail; 1, pass) for each participant. Of the 153 metrics, 48 reliably discriminated between levels of competence and demonstrated evidence of construct validity. However, in that study, simulator-generated metrics were dichotomic, marked either pass or fail, which is different from the present study with a continuous variable.

Few other studies have analyzed simulator-generated metrics, and none have analyzed trajectories. Furthermore, most publications on simulator-generated metrics focus on laparoscopy training[16,17]. Jones D et al. conducted a study extracting data form a laparoscopic simulator. They suggested a relationship between the training level of the surgeon and the forces imparted on the tissue during a laparoscopic simulation[18]. In another study about laparoscopic training, Rivard et al. selected 36 individual metrics on four tasks, including speed, motion path length, respect for tissue, accuracy, task-specific errors, and successful task completion. Time and motion path length were significantly different for all four tasks, and the other metrics for some of the tasks. They then used the validated metrics to create summary equations for each task, which successfully distinguished between the different experience levels[16]. Lastly, another study about laparoscopic procedures explored the correlation between path length or smoothness and outcome measures such as accuracy error, knot slippage, leakage, tissue damage, and operating time[19]. In that study, no correlation was found between the metrics and surgery outcomes, except for operative time. Finally, in another study by Sánchez-Margallo, the suturing performance was successfully assessed by the motion analysis method. They demonstrated construct validity for the execution time and path length[20].

To integrate a simulator in a training and assessment program, it is necessary to demonstrate face, content and construct validity of that simulator. The construct validity means the ability to discriminate between different levels of expertise. By opposition, face and content validity means how convincing or realistic the simulator is according to experts, in a more subjective way[21]. It is interesting to notice how objective measures are used to inform construct validity. A study by Van Dongen et al. which aimed to demonstrate construct validity for a laparoscopic virtual reality simulator, used the clinical experience as a definition of expertise. Indeed, simulator metrics were tested in 16 novices, 16 residents and 16 experts to construct content validity of a laparoscopic simulator[22]. It appeared that performance of the various tasks on the simulator corresponded to the respective level of laparoscopic clinical experience.

In a study by Ramos et al., participants completed three virtual reality (VR) exercises using the Da Vinci Skills Simulator, as well as corresponding dry lab versions of each exercise. Simulator performance was assessed by metrics measured on the simulator. Dry lab performance was blindly video-evaluated by expert review using the six-metric GEARS tool. This study is interesting because their definition of expertise was based on an exercise and not just the years of clinical experience. Additionally, participants were ask to complete a questionnaire to evaluate face and content validity. [23] In another study by Kenney et al, construct validity of a robotic surgery virtual-reality trainer was assessed[24]. The performance was recorded using a built-in scoring algorithm including total task time, total instrument motion, and number of instrument collisions.

Experienced robotic surgeons outperformed novices in nearly all variables. Again, each subject completed a questionnaire after finishing the modules to assess face and content validity. All experienced surgeons ranked the simulator as useful for training and agreed with incorporating the simulator into a residency curriculum.

Strengths and Limitations

One limitation of our study is linked with the definition of expertise. The three expertise levels were defined by the clinical experience, not the quality of the scans or images that the participants actually produced. However, clinical experience levels were clearly defined (experts who had everyday practice of fetal ultrasound, residents who were much less experienced, and novices who were medical students with no experience at all). Moreover, to address the potential bias on how to define an expert, every participant achieved a probe manipulation exercise, which confirmed levels in the three groups. Another limitation is that we did not assess face and content validity, because we focused on objective metrics.

Implications-Interpretation

One challenge when teaching ultrasound is to explain how to obtain the view of the fetus and how to switch from one plane to another, especially to sagittal views. Future explorations are required to approach the optimal trajectory of the probe. This could help to better teach the technique, optimize scan duration, and assess the dynamic quality of the exploration. Future works should also assess skill transfer to clinical practice and trajectories on actual patients.

Conclusion

This study shows that objective trajectory metrics differ according to level of expertise in two OUS tasks. The connected OUS interface between the operator's hand and the patient, provides numerical data that can help better understand and assess skill acquisition. It is the responsibility of the clinicians to let the developers know what data they are interested in that would make the simulators suited for training.

Acknowledgements

The authors would like to thank all participants in this study, Felicity Neilson for English editing, and Stephen Thucker for technical support on the simulation system.

Disclosure of interests:

The authors declare no conflict of interest.

Authorship contribution:

MLL carried out the study, analyzed the literature, and wrote up the paper.

FD conducted the analyze of the metrics.

MK and EF had a role in participants enrollment, helped with experiments and data collection, read and corrected the paper.

VL read and added correction to the draft and had a role in the conception and correction of the paper.

PJ had a role in the conception, planning, and correction of the paper.

Details of Ethics Approval :

Approval was obtained from our local institutional review board (University Hospital of Rennes, **IRB Approval Avis n° 19.29**, 03/03/2019).

Funding : None

References

1. Cook DA. How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Med Educ* 2014; 48: 750–760.
2. Scalese RJ, Obeso VT, Issenberg SB. Simulation technology for skills training and competency assessment in medical education. *J Gen Intern Med* 2008; 23 Suppl 1: 46–49.
3. Scott DJ, Dunnington GL. The new ACS/APDS Skills Curriculum: moving the learning curve out of the operating room. *J Gastrointest Surg* 2008; 12: 213–221.
4. Tolsgaard MG. A multiple-perspective approach for the assessment and learning of ultrasound skills. *Perspect Med Educ* 2018; 7: 211–213.
5. Chalouhi GE, Quibel T, Lamourdedieu C, et al. [Obstetrical ultrasound simulator as a tool for improving teaching strategies for beginners: Pilot study and review of the literature]. *J Gynecol Obstet Biol Reprod (Paris)* 2016; 45: 1107–1114.
6. Madsen ME, Konge L, Nørgaard LN, et al. Assessment of performance measures and learning curves for use of a virtual-reality ultrasound simulator in transvaginal ultrasound examination. *Ultrasound Obstet Gynecol* 2014; 44: 693–699.
7. Tolsgaard MG, Rasmussen MB, Tappert C, et al. Which factors are associated with trainees' confidence in performing obstetric and gynecological ultrasound examinations? *Ultrasound Obstet Gynecol* 2014; 43: 444–451.
8. Tolsgaard MG, Chalouhi GE. Use of simulators for the assessment of trainees' competence: trendy toys or valuable instruments? *Ultrasound Obstet Gynecol*. Epub ahead of print 17 April 2018. DOI: 10.1002/uog.19071.
9. Tolsgaard MG, Ringsted C, Dreisler E, et al. Sustained effect of simulation-based ultrasound training on clinical performance: a randomized trial. *Ultrasound Obstet Gynecol* 2015; 46: 312–318.
10. Tolsgaard MG, Ringsted C, Rosthøj S, et al. The Effects of Simulation-based Transvaginal Ultrasound Training on Quality and Efficiency of Care: A Multicenter Single-blind Randomized Trial. *Ann Surg* 2017; 265: 630–637.
11. Chalouhi GE, Bernardi V, Gueneuc A, et al. Evaluation of trainees' ability to perform obstetrical ultrasound using simulation: challenges and opportunities. *Am J Obstet Gynecol* 2016; 214: 525.e1-525.e8.
12. cneof 2016 - Recherche Google, https://www.google.com/search?client=firefox-b&q=cneof+2016 (accessed 7 January 2019).
13. Despinoy F, Zemiti N, Forestier G, et al. Evaluation of contactless human-machine interface for robotic surgical training. *Int J Comput Assist Radiol Surg* 2018; 13: 13–24.
14. Ghasemloonia A, Maddahi Y, Zareinia K, et al. Surgical Skill Assessment Using Motion Quality and Smoothness. *J Surg Educ* 2017; 74: 295–305.
15. Zago M, Sforza C, Mariani D, et al. Educational impact of hand motion analysis in the evaluation of FAST examination skills. *Eur J Trauma Emerg Surg*. Epub ahead of print 15 March 2019. DOI: 10.1007/s00068-019-01112-6.
16. Rivard JD, Vergis AS, Unger BJ, et al. Construct validity of individual and summary performance metrics associated with a computer-based laparoscopic simulator. *Surg Endosc* 2014; 28: 1921–1928.
17. Shanmugan S, Leblanc F, Senagore AJ, et al. Virtual reality simulator training for laparoscopic colectomy: what metrics have construct validity? *Dis Colon Rectum* 2014; 57: 210–214.

18. Jones D, Jaffer A, Nodeh AA, et al. Analysis of Mechanical Forces Used During Laparoscopic Training Procedures. *J Endourol* 2018; 32: 529–533.

19. Cesanek P, Uchal M, Uranues S, et al. Do hybrid simulator-generated metrics correlate with content-valid outcome measures? *Surg Endosc* 2008; 22: 2178–2183.

20. Sánchez-Margallo JA, Sánchez-Margallo FM, Oropesa I, et al. Objective assessment based on motion-related metrics and technical performance in laparoscopic suturing. *Int J Comput Assist Radiol Surg* 2017; 12: 307–314.

21. Koch AD, Buzink SN, Heemskerk J, et al. Expert and construct validity of the Simbionix GI Mentor II endoscopy simulator for colonoscopy. *Surg Endosc* 2008; 22: 158–162.

22. van Dongen KW, Tournoij E, van der Zee DC, et al. Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? *Surg Endosc* 2007; 21: 1413–1417.

23. Ramos P, Montez J, Tripp A, et al. Face, content, construct and concurrent validity of dry laboratory exercises for robotic training using a global assessment tool. *BJU Int* 2014; 113: 836–842.

24. Kenney PA, Wszolek MF, Gould JJ, et al. Face, content, and construct validity of dV-trainer, a novel virtual reality simulator for robotic surgery. *Urology* 2009; 73: 1288–1292.

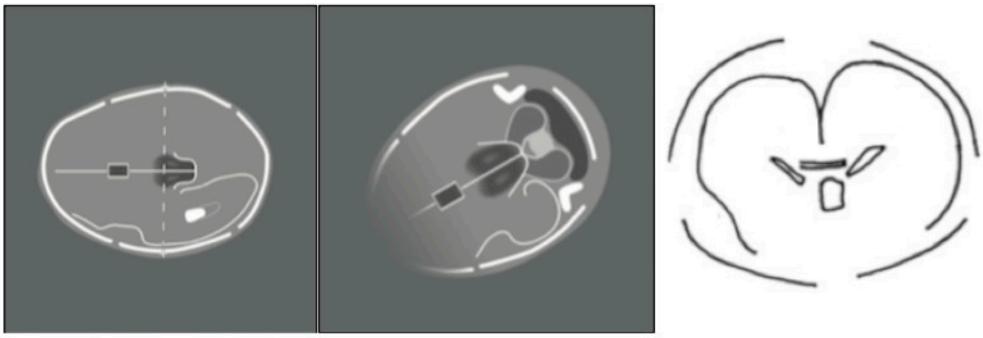
Legends

Figure 1. The virtual reality ultrasound simulator (Scantrainer, Medaphor^{TM}, Cardiff, UK).

Figure 2. Pictorial representation of the three tasks.

Figure 1. The virtual reality ultrasound simulator (Scantrainer, Medaphor^{TM}, Cardiff, UK).

Task 1: 3 Consecutive planes* (Brain examination: Head circumference/Cerebellum view/coronal view)



*The plane represented in white color is not part of National recommendations for standard examination.

Task 2: 3 Consecutive planes (Heart examination: Four chambers/ Left ventricular outflow tract view/ Right ventricular outflow view).

Task 3: 3 Consecutive planes (Kidneys/Spine/Left Diaphragm)

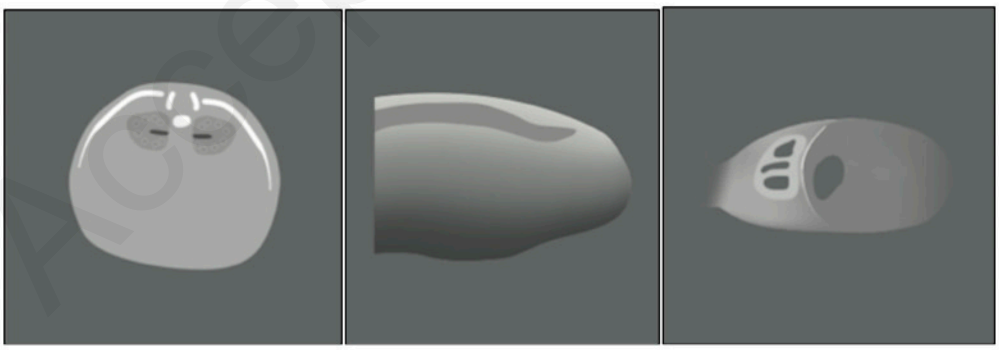


Table 1. Demographics and ultrasound experience of participants in the three groups

	Novices	Intermediates	Experts
No. of participants	16	12	5
Age (mean ±SD)	24.8±5.1	28.3±0.7	44.4±9.4
Gender			
F	12	12	4
M	4	0	1
Clinical experience (years)	0	3.3±0.7	19.8±10.2
Manipulation probe test score			
6 questions (1: pass; 0 : fail)	3.4±1.2	4.7±1.2	5.6±0.5
(mean ±SD)			

Table 2. 6 Metrics analyzed on probe trajectory during the three tasks according to level of expertise.

	Novices N=16	Intermediates N=12	Experts N=5	p
Duration (s)				
Brain	157.5±176.2	111.2±110.3	35.6±16.3	<0.001
Heart	109.9±69.7	103.7±61.1	51.7±47.4	0.009
Spine	112.8±64.7	82.3±50.6	35.5±11.4	<0.001
Path length (cm)				
Brain	144.2±113.4	179.5±107.1	100.5±40.9	<0.001
Heart	142.8±50.8	126.6±51.7	102.7±113.8	NS
Spine	151.5±56.9	159.0±58.9	107.1±63.8	0.013
Average Velocity (cm/s)				
Brain	1.9±0.5	1.9±1.3	2.9±0.8	<0.001
Heart	1.5±0.6	1.4±0.7	2.0±0.6	0.024
Spine	1.5±0.5	2.2±0.7	2.9±0.8	<0.001
Average Acceleration (cm/s^2)				
Brain	13.7±4.8	13.8±10.5	22.8±5.9	<0.001
Heart	11.5±6.2	10.1±4.0	14.9±3.7	0.014
Spine	10.9±3.7	15.7±5.4	22.5±8.0	<0.001
Average Jerk (cm/s^{-3})				
Brain	275.5±135.7	272.8±228.3	452.8±135.8	<0.001
Heart	242.6±188.6	201.2±71.6	292.3±76.1	NS
Spine	215.4±84.5	309.9±130.0	437.8±160.5	<0.001

Working Volume (cm^3)				
Brain	1 118.6±622.4	1 433.6±716.9	761.1±603.9	0.002
Heart	1 755.8±1318.4	1 275.1±674.3	825.7±962.9	0.009
Spine	1 618.6±624.1	1 759.2±1 250.3	867.5±717.5	NS