



**HAL**  
open science

# 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

► **To cite this version:**

Maela Le Lous, Fabien Despinoy, Margaux Klein, Elisa Fustec, Vincent Lavoué, et al.. Impact of Physician Expertise on Probe Trajectory During Obstetric Ultrasound A Quantitative Approach for Skill Assessment. *Simulation in Healthcare*, 2021, 16 (1), pp.67-72. 10.1097/SIH.000000000000465 . hal-02887320v2

**HAL Id: hal-02887320**

**<https://hal.science/hal-02887320v2>**

Submitted on 4 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 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

► **To cite this version:**

Maela Le Lous, Fabien Despinoy, Margaux Klein, Elisa Fustec, Vincent Lavoué, et al.. Impact of Physician Expertise on Probe Trajectory During Obstetric Ultrasound A Quantitative Approach for Skill Assessment. Simulation in Healthcare, Lippincott, Williams & Wilkins 2020, Publish Ahead of Print, 10.1097/SIH.0000000000000465 . hal-02887320

**HAL Id: hal-02887320**

**<https://hal.archives-ouvertes.fr/hal-02887320>**

Submitted on 22 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

3 -Obstetrical Ultrasound Trajectory Metrics-  
4

5 Maela LE LOUS, MD <sup>1,2</sup>, Fabien DESPINOY, PhD <sup>1</sup>, Margaux KLEIN, MD <sup>2</sup>, Elisa  
6 FUSTEC MD <sup>2</sup>, Vincent LAVOUE MD PhD <sup>2,3</sup>, Pierre JANNIN MD PhD<sup>1</sup>.

7

8 1. Univ Rennes, INSERM, LTSI - UMR 1099, F35000, Rennes, France.

9 2. Department of Obstetrics and Gynecology, University Hospital of Rennes, France.

10 3. INSERM 1242, Chemistry, Oncogenesis, Stress and Signaling, Rennes, France.

11

12 **Corresponding author:**

13 Maela Le Lous, MD

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

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

16 Fax

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

18

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

20 **Funding:** None

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

22 **Abstract**

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

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

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

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

43

44 **Funding** None.

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

46 **Introduction**

47 Education in the field of obstetric ultrasound has shifted from an exclusively  
48 observation-based training to simulation-based training over the past decade<sup>1</sup>. Multiple  
49 factors may have contributed to the more widespread use of simulation: concerns for  
50 diagnostic errors, patient volunteers' or teachers' availability, and technological advances  
51 based on virtual reality (VR)<sup>2,3</sup>. Obstetric ultrasound simulators (OUS) may revolutionize  
52 the Obstetric and Gynecologic (OB/GYN) residents' curriculum for both training and  
53 evaluation<sup>4</sup>.

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

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

67 A fundamental issue for evaluation with OUS is the need for metrics. VR based  
68 simulators allow automatic recording of probe trajectories during training and  
69 evaluation. OUS could be a suitable tool to extract data from these probe trajectories

70 for objective evaluation, and to identify metrics that may discriminate the level of  
71 expertise of participants during their training. The aim of our study was to assess the  
72 potential of OUS to identify objective metrics to measure expertise in  
73 ultrasonography.

74

Accepted manuscript

75 **Methods**

76

77 **Participants**

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

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

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

100

## 101 **Virtual reality simulator**

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

110

## 111 **Tasks performed**

### 112 *Step 1: Identification of tasks*

113 Most fetal malformations are detected during the second trimester scan. This  
114 scan examines the fetus in 11 planes as requested by the National Committee for  
115 Prenatal Ultrasound<sup>12</sup>.

116 Tasks reflecting the standardized second trimester examination of the fetus were  
117 selected to obtain a large sample of fetal anatomy screening, in particular tasks which  
118 involved switching from transverse to coronal or sagittal planes. We selected the tasks  
119 which would potentially reflect differences in ultrasound competences and provided the  
120 most realistic view of the fetus in three consecutive standardized planes. **For Task 1**  
121 **(Brain at 24 weeks of gestation)**, participants were asked to successively obtain the  
122 standardized view of the fetal head circumference, then the standardized plane of the  
123 cerebellum, and finally a coronal view of the brain through the cavum of septum



124 pellucidum. For Task 2 (Heart at 22 weeks of gestation), participants were asked to  
125 successively obtain the standardized view of the four chambers, the view of the left  
126 ventricle and the aorta, and the view of the right ventricle and the pulmonary artery. For  
127 Task 3 (Spine at 22 weeks of gestation), participants were asked to obtain successively  
128 the standardized frontal view of the kidneys, a sagittal view of the spine, and the  
129 parasagittal view of the left diaphragm. Figure 2 shows a pictorial representation of the  
130 three tasks. The participant judged himself that the view was exact and he froze the  
131 image to finish exercise.

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

144

145 **Metric evaluation**

146 The data were logged and recorded by a custom data acquisition software  
147 developed with Scantrainer, Medaphor™, Cardiff, UK. The probe position frequency  
148 sampling was 20 Hz. Computation of all metrics and trajectory analysis were run on a  
149 dedicated computer (Xeon E5-1650V4 @3.60GHz with 32Go RAM) using  
150 Matlab\_R2017b. The analysis was done following a previously published technique<sup>13</sup>. A  
151 total of 6 selected metrics was analyzed:

- 152 1. **Duration (D)** corresponds to the execution time between the first time the  
153 hand moves the probe, until it has been released at the end of the task. It is  
154 measured in seconds (s).
- 155 2. **Path Length (PL)** represents the total distance travelled by the probe during  
156 the execution of the task. It is measured in millimeters (mm).
- 157 3. **Average Velocity (AV)** corresponds to the average linear speed of the probe  
158 during the task. It is measured in mm.s<sup>-1</sup>.
- 159 4. **Average Acceleration (AA)** corresponds to the average instantaneous  
160 acceleration of the probe during the task. It is measured in mm.s<sup>-2</sup>.
- 161 5. **Average Jerk (AJ)** corresponds to the average jerk (derivative of the  
162 acceleration) during the task, also known as “smoothness” measure. It is  
163 measured in mm.s<sup>-3</sup>.
- 164 6. **Working volume (WV)** represents the volume of the convex hull for each  
165 trajectory. The convex hull of a trajectory is the smallest convex volume  
166 among those which contain it. It is measured in mm<sup>3</sup>.

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

169 control alpha inflation ( $p < 0.08$ ). The participants were absolutely not told about  
170 what the metrics measured were.  
171

Accepted manuscript

172 **Results**

173 **Demographics**

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

175 **Manipulation probe exercise**

176 For the manipulation probe exercise, out of a total of six points (six multiple  
177 choice questions, pass: 1; fail: 0), the Experts' mean score was  $5.6 \pm 0.5$ , the Intermediates'  
178 mean score was  $4.7 \pm 1.2$ , and the Novices' mean score was  $3.4 \pm 1.2$  points ( $p=0.002$ ).

179 **Table 1.**

180 **Trajectory Metrics**

181 Analysis of the probe trajectory metrics revealed significant differences for Tasks  
182 1, 2 and 3 between the Expert, Intermediate and Novice groups (**Table 2**).

183 *Duration of the exercise(s)*

184 The mean Duration(s) of the exercise differed between Novices, Intermediates,  
185 and Experts for Task 1 (respectively  $157.5s \pm 176.2$  versus  $111.2s \pm 110.3$  versus  $35.6s$   
186  $\pm 16.3$ ,  $p < 0.001$ ), Task 3 ( $112.8s \pm 64.7$  vs.  $82.3s \pm 50.6$  vs.  $35.5s \pm 11.4$   $p < 0.001$ ). For Task  
187 2 ( $109.9s \pm 69.7$  vs.  $103.7s \pm 61.1$  vs.  $51.7s \pm 47.4$   $p = 0.009$ ) it was not significant.

188 *Path length (mm)*

189 The Path Length (mm) was shorter with increasing level of expertise for Task 1  
190 (Novices:  $1442.3 \pm 1134.4$  vs. Intermediates:  $1795.4 \pm 1071.9$  vs. Experts:  $1005.7 \pm 409.7$ ,  
191  $p < 0.001$ ) and for Task 3 (respectively,  $1515.3 \pm 569.4$  vs.  $1590.8 \pm 589.0$  vs.  $1071.2$   
192  $\pm 638.3$ ,  $p = 0.013$ ).

193 For Task 2 (the "Heart" exercise), the Path Length was not statistically different  
194 between the three groups ( $1428.7 \pm 508.2$  vs.  $1266.5 \pm 517.5$  vs.  $1027.0 \pm 1138.3$ , NS).

195 *Average velocity (mm/s)*

196 The average velocity (mm/s) was significantly higher with increasing level of  
197 expertise for Task 1 and 3. For Task 1, average velocity was 19.2mm/s  $\pm$ 5.7 for Novices  
198 vs. 19.5mm/s  $\pm$ 13.8 for Intermediate and 29.5 mm/s  $\pm$ 8.7 for Experts ( $p < 0.001$ ), and for  
199 Task 3 it was respectively 15.6mm/s  $\pm$  5.2 vs. 22.2  $\pm$ 7.5 vs. 29.1  $\pm$ 8.6 ( $p < 0.001$ ). For Task  
200 2, it was respectively 15.7 mm/s  $\pm$ 6.4 vs. 14.5  $\pm$ 7.0 vs. 20.1  $\pm$  6.4 ( $p = 0.024$ , NS).

201 *Average Acceleration (mm.s<sup>-2</sup>)*

202 The average acceleration was statistically different between the three groups for  
203 Task 1 (137.0 mm.s<sup>-2</sup>  $\pm$ 48.2 vs. 138.6  $\pm$  105.1 vs. 228.9  $\pm$ 59.8,  $p < 0.001$ ), and for Task 3  
204 (109.5  $\pm$  37.0 vs. 157.3  $\pm$  54.6 vs. 225.7  $\pm$  80.6,  $p < 0.001$ ), but not for Task 2 (115.8  $\pm$ 62.8  
205 vs. 101.7  $\pm$ 40.1 vs. 149.6  $\pm$ 37.5, NS)

206 *Average Jerk (mm.s<sup>-3</sup>)*

207 Jerk was statistically different between the three groups for Task 1 (2 755.7  $\pm$ 1  
208 357.0 vs. 2 728.5  $\pm$  2 283.5 vs. 4 528.0  $\pm$ 1 358.4,  $p < 0.001$ ) and for Task 3 (2 154.7  $\pm$  845.5  
209 vs. 3 099.7  $\pm$  1 300.4 vs. 4 378.1  $\pm$  1 605.7,  $p < 0.001$ ) increasing with level of expertise.  
210 For Task 2, there was no significant difference (2 426.8  $\pm$ 1 886.4 vs. 2 012.1  $\pm$ 716.7 vs. 2  
211 929.7  $\pm$ 761.1, NS).

212 *Working Volume (mm<sup>3</sup>)*

213 The total Working Volume covered was statistically different between the three  
214 groups for Task 1 (1 118 661.6 mm<sup>3</sup>  $\pm$ 622487.1 vs. 1 438 677.7  $\pm$ 716 948.1 vs. 761  
215 105.7  $\pm$ 603 917.3  $p = 0.002$ ).

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

219 **Discussion**

220 *Main Findings*

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

226 **The time** for each task (Duration) decreased as the level of expertise increased  
227 ranging from 35s to 52s for Experts and from 109s to 157s for Novices. The Path Length,  
228 although significantly shorter for Experts in the “Brain” and “Spine” exercise, was not  
229 significantly different for the “Heart” exercise. This might be explained by the fact that  
230 the planes in the heart exercise are much more tightly bunched in a very small volume.

231 **There must have been a task effect. The “heart” exercise must have been less discriminant**  
232 **using trajectory metrics, because it implied tiny probe movements.** The same could have  
233 been argued for the “Brain” exercise because there isn’t a lot a movement between the  
234 head circumference and the transcerebellar view, but a lot a movement is necessary to  
235 obtain the coronal view. **It is to notice that some amount of initial movement among**  
236 **experts, must be tied to translating positioning and movements with the model to recall**  
237 **of positioning and movements with genuine patients.**

238 Velocity and Acceleration were significantly higher with the increase in level of  
239 expertise suggesting that experienced sonographers move from one plane to another with  
240 a faster movement. Jerk, which is a derivative of Acceleration, may be interpreted as the  
241 variation of acceleration, or how sudden the variations are. **We could have assumed that**  
242 **jerk would have been smaller and the gesture smoother for the experts, as it has been**

243 shown for surgical procedures<sup>14</sup>. However, during ultrasound examination, jerk was  
244 higher for experts because they most often combined a quick translation followed by a  
245 quick 90° rotation of the probe to switch, for example, from a transverse view of the  
246 kidney to a sagittal view of the spine. For the “Heart” exercise, average Jerk did not differ  
247 between the 3 groups, with may be explained again with the small quantity of movements  
248 between the different views. The Working Volume was smaller for the Experts than for  
249 the Intermediates and Novices, implying that the Experts remained within the area of  
250 interest. These results are similar to those observed by Zago et al about FAST (Focused  
251 Assessment with Sonography in Trauma) examination. In this study, they also used hand  
252 motion analysis to discriminate expertise, and found similar results : longer hand path and  
253 higher working volume for the novices<sup>15</sup>. In our study, intermediates performed more like  
254 novices than experts on many measures because they may not feel comfortable with  
255 morphologic examination.

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

266 is part of the method to go from the identification of differences to explanation of the  
267 differences observed.

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

276 Few other studies have analyzed simulator-generated metrics, and none have  
277 analyzed trajectories. Furthermore, most publications on simulator-generated metrics  
278 focus on laparoscopy training<sup>16,17</sup>. Jones D et al. conducted a study extracting data from a  
279 laparoscopic simulator. They suggested a relationship between the training level of the  
280 surgeon and the forces imparted on the tissue during a laparoscopic simulation<sup>18</sup>. In  
281 another study about laparoscopic training, Rivard et al. selected 36 individual metrics on  
282 four tasks, including speed, motion path length, respect for tissue, accuracy, task-specific  
283 errors, and successful task completion. Time and motion path length were significantly  
284 different for all four tasks, and the other metrics for some of the tasks. They then used the  
285 validated metrics to create summary equations for each task, which successfully  
286 distinguished between the different experience levels<sup>16</sup>. Lastly, another study about  
287 laparoscopic procedures explored the correlation between path length or smoothness and  
288 outcome measures such as accuracy error, knot slippage, leakage, tissue damage, and  
289 operating time<sup>19</sup>. In that study, no correlation was found between the metrics and surgery



290 outcomes, except for operative time. Finally, in another study by Sánchez-Margallo, the  
291 suturing performance was successfully assessed by the motion analysis method. They  
292 demonstrated construct validity for the execution time and path length<sup>20</sup>.

293 To integrate a simulator in a training and assessment program, it is necessary to  
294 demonstrate face, content and construct validity of that simulator. The construct validity  
295 means the ability to discriminate between different levels of expertise. By opposition,  
296 face and content validity means how convincing or realistic the simulator is according to  
297 experts, in a more subjective way<sup>21</sup>. It is interesting to notice how objective measures are  
298 used to inform construct validity. A study by Van Dongen et al. which aimed to  
299 demonstrate construct validity for a laparoscopic virtual reality simulator, used the  
300 clinical experience as a definition of expertise. Indeed, simulator metrics were tested in  
301 16 novices, 16 residents and 16 experts to construct content validity of a laparoscopic  
302 simulator<sup>22</sup>. It appeared that performance of the various tasks on the simulator  
303 corresponded to the respective level of laparoscopic clinical experience.

304 In a study by Ramos et al., participants completed three virtual reality (VR)  
305 exercises using the Da Vinci Skills Simulator, as well as corresponding dry lab versions  
306 of each exercise. Simulator performance was assessed by metrics measured on the  
307 simulator. Dry lab performance was blindly video-evaluated by expert review using the  
308 six-metric GEARS tool. This study is interesting because their definition of expertise was  
309 based on an exercise and not just the years of clinical experience. Additionally,  
310 participants were asked to complete a questionnaire to evaluate face and content validity.<sup>23</sup>

311 In another study by Kenney et al, construct validity of a robotic surgery virtual-reality  
312 trainer was assessed<sup>24</sup>. The performance was recorded using a built-in scoring algorithm  
313 including total task time, total instrument motion, and number of instrument collisions.

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

318

### 319 *Strengths and Limitations*

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

### 329 *Implications-Interpretation*

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

336

337

338 **Conclusion**

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

344

Accepted manuscript

345 **Acknowledgements**

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

348

349 **Disclosure of interests:**

350 The authors declare no conflict of interest.

351

352 **Authorship contribution:**

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

354 FD conducted the analyze of the metrics.

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

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

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

360

361 **Details of Ethics Approval :**

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

364

365 **Funding** : None

366 **References**

367

- 368 1. Cook DA. How much evidence does it take? A cumulative meta-analysis of  
369 outcomes of simulation-based education. *Med Educ* 2014; 48: 750–760.
- 370 2. Scalese RJ, Obeso VT, Issenberg SB. Simulation technology for skills training  
371 and competency assessment in medical education. *J Gen Intern Med* 2008; 23 Suppl 1:  
372 46–49.
- 373 3. Scott DJ, Dunnington GL. The new ACS/APDS Skills Curriculum: moving the  
374 learning curve out of the operating room. *J Gastrointest Surg* 2008; 12: 213–221.
- 375 4. Tolsgaard MG. A multiple-perspective approach for the assessment and learning  
376 of ultrasound skills. *Perspect Med Educ* 2018; 7: 211–213.
- 377 5. Chalouhi GE, Quibel T, Lamourdedieu C, et al. [Obstetrical ultrasound  
378 simulator as a tool for improving teaching strategies for beginners: Pilot study and  
379 review of the literature]. *J Gynecol Obstet Biol Reprod (Paris)* 2016; 45: 1107–1114.
- 380 6. Madsen ME, Konge L, Nørgaard LN, et al. Assessment of performance  
381 measures and learning curves for use of a virtual-reality ultrasound simulator in  
382 transvaginal ultrasound examination. *Ultrasound Obstet Gynecol* 2014; 44: 693–699.
- 383 7. Tolsgaard MG, Rasmussen MB, Tappert C, et al. Which factors are associated  
384 with trainees' confidence in performing obstetric and gynecological ultrasound  
385 examinations? *Ultrasound Obstet Gynecol* 2014; 43: 444–451.
- 386 8. Tolsgaard MG, Chalouhi GE. Use of simulators for the assessment of trainees'  
387 competence: trendy toys or valuable instruments? *Ultrasound Obstet Gynecol*. Epub  
388 ahead of print 17 April 2018. DOI: 10.1002/uog.19071.
- 389 9. Tolsgaard MG, Ringsted C, Dreisler E, et al. Sustained effect of simulation-  
390 based ultrasound training on clinical performance: a randomized trial. *Ultrasound*  
391 *Obstet Gynecol* 2015; 46: 312–318.
- 392 10. Tolsgaard MG, Ringsted C, Rosthøj S, et al. The Effects of Simulation-based  
393 Transvaginal Ultrasound Training on Quality and Efficiency of Care: A Multicenter  
394 Single-blind Randomized Trial. *Ann Surg* 2017; 265: 630–637.
- 395 11. Chalouhi GE, Bernardi V, Gueneuc A, et al. Evaluation of trainees' ability to  
396 perform obstetrical ultrasound using simulation: challenges and opportunities. *Am J*  
397 *Obstet Gynecol* 2016; 214: 525.e1-525.e8.
- 398 12. cneof 2016 - Recherche Google, [https://www.google.com/search?client=firefox-  
399 b&q=cneof+2016](https://www.google.com/search?client=firefox-b&q=cneof+2016) (accessed 7 January 2019).
- 400 13. Despinoy F, Zemiti N, Forestier G, et al. Evaluation of contactless human-  
401 machine interface for robotic surgical training. *Int J Comput Assist Radiol Surg* 2018;  
402 13: 13–24.
- 403 14. Ghasemloonia A, Maddahi Y, Zareinia K, et al. Surgical Skill Assessment Using  
404 Motion Quality and Smoothness. *J Surg Educ* 2017; 74: 295–305.
- 405 15. Zago M, Sforza C, Mariani D, et al. Educational impact of hand motion analysis  
406 in the evaluation of FAST examination skills. *Eur J Trauma Emerg Surg*. Epub ahead  
407 of print 15 March 2019. DOI: 10.1007/s00068-019-01112-6.
- 408 16. Rivard JD, Vergis AS, Unger BJ, et al. Construct validity of individual and  
409 summary performance metrics associated with a computer-based laparoscopic  
410 simulator. *Surg Endosc* 2014; 28: 1921–1928.
- 411 17. Shanmugan S, Leblanc F, Senagore AJ, et al. Virtual reality simulator training  
412 for laparoscopic colectomy: what metrics have construct validity? *Dis Colon Rectum*  
413 2014; 57: 210–214.

- 414 18. Jones D, Jaffer A, Nodeh AA, et al. Analysis of Mechanical Forces Used During  
415 Laparoscopic Training Procedures. *J Endourol* 2018; 32: 529–533.
- 416 19. Cesanek P, Uchal M, Uranues S, et al. Do hybrid simulator-generated metrics  
417 correlate with content-valid outcome measures? *Surg Endosc* 2008; 22: 2178–2183.
- 418 20. Sánchez-Margallo JA, Sánchez-Margallo FM, Oropesa I, et al. Objective  
419 assessment based on motion-related metrics and technical performance in laparoscopic  
420 suturing. *Int J Comput Assist Radiol Surg* 2017; 12: 307–314.
- 421 21. Koch AD, Buzink SN, Heemskerk J, et al. Expert and construct validity of the  
422 Symbionix GI Mentor II endoscopy simulator for colonoscopy. *Surg Endosc* 2008; 22:  
423 158–162.
- 424 22. van Dongen KW, Tournoij E, van der Zee DC, et al. Construct validity of the  
425 LapSim: can the LapSim virtual reality simulator distinguish between novices and  
426 experts? *Surg Endosc* 2007; 21: 1413–1417.
- 427 23. Ramos P, Montez J, Tripp A, et al. Face, content, construct and concurrent  
428 validity of dry laboratory exercises for robotic training using a global assessment tool.  
429 *BJU Int* 2014; 113: 836–842.
- 430 24. Kenney PA, Wszolek MF, Gould JJ, et al. Face, content, and construct validity  
431 of dV-trainer, a novel virtual reality simulator for robotic surgery. *Urology* 2009; 73:  
432 1288–1292.
- 433

434 **Legends**

435 **Figure 1.** The virtual reality ultrasound simulator (Scantrainer, Medaphor™, Cardiff,  
436 UK).

437 **Figure 2.** Pictorial representation of the three tasks.

Accepted manuscript

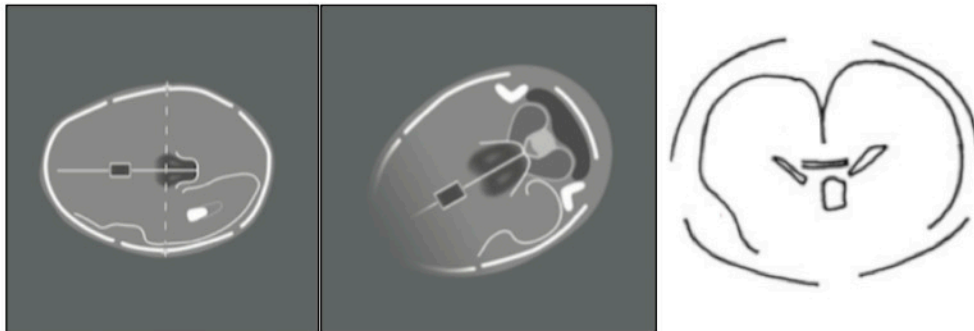
**Figure 1.** The virtual reality ultrasound simulator (Scantrainer, Medaphor™, Cardiff, UK).



Accepted manuscript



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 $\pm$ SD)	24.8 $\pm$ 5.1	28.3 $\pm$ 0.7	44.4 $\pm$ 9.4
Gender			
F	12	12	4
M	4	0	1
Clinical experience (years)	0	3.3 $\pm$ 0.7	19.8 $\pm$ 10.2
Manipulation probe test score			
6 questions (1: pass; 0: fail)	3.4 $\pm$ 1.2	4.7 $\pm$ 1.2	5.6 $\pm$ 0.5
(mean $\pm$ 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
<b>Duration (s)</b>				
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
<b>Path length (cm)</b>				
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
<b>Average Velocity (cm/s)</b>				
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
<b>Average Acceleration (cm/s<sup>2</sup>)</b>				
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
<b>Average Jerk (cm/s<sup>3</sup>)</b>				
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<sup>3</sup>)

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

---

Accepted manuscript

