



HAL
open science

Metrics used to evaluate obstetric ultrasound skills on simulators: A systematic review

Maela Le Lous, Margaux Klein, Caroline Tesson, Julien Berthelemy, Vincent Lavoue, Pierre Jannin

► To cite this version:

Maela Le Lous, Margaux Klein, Caroline Tesson, Julien Berthelemy, Vincent Lavoue, et al.. Metrics used to evaluate obstetric ultrasound skills on simulators: A systematic review. European Journal of Obstetrics & Gynecology and Reproductive Biology, 2021, 258, pp.16-22. 10.1016/j.ejogrb.2020.12.034 . hal-03101591

HAL Id: hal-03101591

<https://hal.science/hal-03101591>

Submitted on 2 Jan 2023

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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Metrics used to evaluate obstetric ultrasound skills on simulators: a systematic review.**

2 Maela LE LOUS^{1,2,3}, Margaux KLEIN², Caroline TESSON², Julien BERTHELEMY¹,

3 Vincent LAVOUE MD^{2,3}, Pierre JANNIN¹.

4

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

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

7 3. CIC Inserm 1414, University Hospital of Rennes, University of Rennes 1, Rennes, France.

8

9 **Corresponding author:**

10 Maela Le Lous, MD

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

12 Phone : Mobile +33 6 95 02 38 05 Work + 33 99 26 32 09

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

14

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

16 **Funding:** None

17 Words : 3636 ; Figure : 1 ; Tables 3 ; References : 24

18 **Abstract**

19 Obstetric ultrasound simulators are now used for training and evaluating OB/GYN
20 students but there is a lack of literature about evaluation metrics in this setting. In this
21 literature review, we searched MEDLINE and the COCHRANE database using the keywords:
22 (Obstetric OR Fetal) AND (Sonography OR Ultrasound) AND Simulation. Of a total of 263
23 studies screened, we selected nine articles from the title and the abstract in PubMed, in the
24 past 5 years. Two more article were added from bibliographies. A total of 11 articles were
25 therefore included. from which nine articles were selected from the title and the abstract in
26 PubMed. Two more articles were added from the bibliographies For each study, data about
27 the type of simulation, and the metrics (qualitative or quantitative) used for assessment were
28 collected. The selection of studies shows that evaluation criteria for ultrasound training were
29 qualitative metrics (binary success/fail exercise ; dexterity quoted by an external observer ;
30 Objective Structured Assessment of Ultrasound Skills (OSAUS) Score ; quality of images
31 according to Salomon's score) or quantitative criteria (Accuracy of Biometry – Simulator
32 generated metrics). Most studies used a combination of both. To date, simulator metrics used
33 to discriminate ultrasound skills are performance score quoted by external observers and
34 image quality scoring. Whether probe trajectory metrics can be used to discriminate skills is
35 unknown.

36 **Keywords :** Ultrasound; Obstetrics; Simulation training; Education; Evaluation ; Metrics.

37 **Introduction**

38 Ultrasound is a core skill required by all obstetrics and gynecology (OB/GYN)
39 students that have traditionally been taught during clinical time directly on patients[1]. In
40 France, it is also recommended to sit a national examination including theory and practice.
41 However, ultrasound training is associated with long-time learning curves and training
42 opportunities in clinical ultrasound are becoming scarcer [2]. Students require more than 24
43 months of clinical experience and 12-24 days of training in specialized ultrasound units
44 before feeling confident about performing transvaginal and transabdominal ultrasound scans
45 alone[2]. Furthermore, while ultrasound is usually considered safe for the patient, missed
46 diagnoses by unskilled operators could cause a loss of chance for both mother and fetus[3].
47 For example, the prenatal diagnosis of transposition of the great vessels – which requires
48 planned delivery and perinatal management – drastically influences postnatal outcome[4].
49 Obstetric ultrasound simulators (OUS) may facilitate access to ultrasound training and
50 evaluation as they do not need to be used in a clinical setting[5].

51 The use of OUS for ultrasound training has been largely evaluated [6]·[7]·[2]·[8]. In
52 a series of eight Danish studies, Toslgard et al. successively explored the learning curves for
53 students (undergraduate sonographer), examined how to improve the efficiency of training
54 with the use of dyad practice, and investigate whether improvements were sustained over
55 time[9]. They also demonstrated skill transfer to subsequent clinical training[10]. Finally, they
56 demonstrated an added value for the patients after simulation training, such as a decrease in
57 discomfort, and improvement in their perception of safety, and in their confidence[10].

58 Evaluating skills in obstetric ultrasound is a time-consuming process and requires
59 pregnant volunteers who are willing to be used for training as well as available teachers. OUS
60 are thus a good alternative for student evaluation[11]. However, a fundamental issue for
61 evaluation is the need for evaluation metrics which would limit subjectivity and increase

62 standardization and accuracy thereby producing meaningful comparisons. These can be either
63 qualitative, sometimes with a subjective assessment based on observation (dexterity score,
64 OSAUS score), or quantitative (image quality, time to complete exercise, probe trajectory).

65 Given the relative lack of data in the literature about metrics that are able to
66 distinguish between levels of expertise, we sought to review the literature and investigate
67 the current level of evidence about metrics used in ultrasound simulation training in the
68 field of obstetrics.

69

70

71 **Material and methods**

72 This review was planned, conducted, and reported in adherence with the Preferred
73 Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards of
74 quality[12].

75 A search of scientific publications was conducted in PubMed and the Cochrane
76 Library using the keywords: (Obstetric OR Fetal) AND (Sonography OR Ultrasound) AND
77 Simulation

78 Eligible studies had to be published in English or French within the past 5 years and to
79 investigate the use of ultrasound simulators in the specific field of obstetrics for health
80 profession learners at any stage (initial and continuous) for training or evaluation. Systematic
81 reviews were not included.

82 For each study we reported the type of metrics used for evaluation and distinguished
83 between articles addressing qualitative and quantitative metrics. We also noted whether
84 trajectory metrics were considered.

85 The date of last searched was January 11, 2020.

86

87 **Study Selection**

88 A total of 263 studies were screened, from which nine articles were selected from the
89 title and the abstract in PubMed. Two more articles were added from the bibliographies
90 (**Figure 1** – Article flow chart). No additional article was added from Google scholar.

91 Finally, 11 articles were analyzed (**Table 1**).

92

93

94

95 **Results**

96 The selection of studies shows that evaluation criteria for ultrasound training were
97 qualitative metrics (binary success/fail exercise – Dexterity – OSAUS score - Quality of
98 images) or quantitative criteria (Accuracy of Biometry – Simulator generated metrics). Most
99 studies used a combination of both. No study used self-evaluation criteria.

100

101 **Qualitative metrics**

102 *Binary Success / Fail exercise*

103 In 2019, Zimmermann et al. compared the results of a theoretical test before and after
104 a training program combining e-learning and simulation training for 33 midwives [13]. In
105 addition to the theoretical test, the evaluation metrics studied were: fetus localization, fetal
106 heartbeat detection, placenta localization, and cervix measurement. Participants were
107 evaluated three times: initial (T1), after e-learning (T2), and after the simulation training on a
108 simulator (T3). The notes obtained were 46.6 % [8.3–83.3 %] at T1. 87.2 % [66.6 à 100 %] at
109 T2, and 91.6 % [66.6 à100 %] at T3. Scores improved significantly between T1 and T2 ($p <$
110 0.001), and between T1 and T3 ($p < 0.001$). Placenta localization (success/fail) improved
111 considerably after training (100% versus 28%, $p= 0.02$) as did cervix measurement (90% after
112 versus 19% before training, $p= 0.03$). This descriptive cohort study used binary qualitative
113 evaluation (success/fail) criteria on a simulator.

114

115 *Dexterity score*

116 In a study by Chalouhi et al. in 2016 [14], OB/GYN residents, midwives or physicians
117 sitting the National OB/GYN ultrasound exam were asked to perform a standardized view of
118 the fetus (head circumference, four-chamber view of the heart, kidneys, spine, stomach, and
119 upper lip) first on a real patient volunteer and then on a simulator. The evaluation metrics

120 were dexterity (scored from 0 to 10 by an external observer) and image quality according to
121 Salomon's score[15]. Both the dexterity scores (6.5+-2 and 5.9+-2.3, respectively, p=0.31)
122 and the image quality were similar for the simulation and the real patient. The shortcomings
123 of this study were that dexterity was evaluated subjectively and no quantitative evaluation
124 metrics were used.

125 *The OSAUS score*

126 The Objective Structured Assessment of Ultrasound Skills (OSAUS) is, as the name
127 implies, an objective score for ultrasound evaluation which was validated by Tolsgaard et al.
128 in 2013[16]. The OSAUS score includes seven items, each of them scored from 1 to 5:
129 indication for the examination, knowledge about ultrasound equipment, image optimization,
130 systematic examination, interpretation of images, documentation of examination, and medical
131 decision making.

132 In another paper, Tolsgaard et al. reports on a multicenter randomized single-blinded
133 study of 26 students randomized between a simulation group and a clinical practice group.
134 During the first week of their residency, the interns received a 1-hour course on ultrasound
135 basics and pelvic anatomy. The first group (n=14) received a simulator training (Scantrainer,
136 MedaphorTM, Cardiff, UK) and a clinical training, whereas the control group received
137 traditional clinical training under supervision only (n=12). Students from the first group had
138 to complete seven training modules on the simulator (gynecology and first trimester
139 pregnancy modules) and achieve an expert (graduate sonographer) level of performance. The
140 mean time necessary to reach the expert performance level on the simulator was 3h 16
141 minutes (CI 95% [2h56; 3h36]) and after a mean of 30.3 attempts [27.6;32.9]. The metrics
142 recorded on the simulator were: the time spent on the simulator for each module; the resulting
143 scores; dichotomous variables caught by the simulator (for example, screening of the entire

144 uterus in a sagittal plane); and the number of attempts. The participants of the simulation
145 group also trained on a phantom mannikin (BluePhantom, CAE Healthcare, Sarasota, FL,
146 USA) for transvaginal ultrasound scan with their local sonographer. In the clinical training
147 group (control group), the interns started directly with a patient and could call his/her superior
148 if needed[9]. Competency was assessed by two independent external observers with the
149 OSAUS score after retrospectively viewing a video of a scan obtained on a real patient. The
150 OSAUS scores during the clinical evaluation were significantly higher in the simulation group
151 than in the control group (59.1 ± 9.3 vs 37.6 ± 11.8 ; $p < 0.001$) with significantly more successful
152 exercises in the simulation group (85.7% vs 8.3%; $p < 0.001$) and improvements in
153 optimization ($p < 0.001$), systematization ($p = 0.001$), decision making ($p = 0.005$),
154 documentation ($p < 0.001$), and image interpretation ($p < 0.001$). No difference was found for
155 equipment management ($p = 0.095$).

156 Another study, by Gueneuc et al. in 2019, used the modified OSAUS score (five
157 items, rated from 1 to 3) to assess the dexterity of 40 midwifery students randomized into two
158 groups with different schedules for ultrasound simulation training (Vimedix TM, CAE
159 Healthcare, Sarasota, USA)[17]. Group A (n=24) learned with the traditional academic
160 course: clinical ultrasound training followed by an evaluation on an OUS (E1A). Group B
161 (n=16) was evaluated on a simulator before the clinical ultrasound training (E1B). After these
162 initial assessments, both groups were provided with learning and training sessions on an OUS,
163 and re-evaluated in the same exercise (E2A and E2B). Group B then completed its clinical
164 ultrasound training and was evaluated 1 month later on the simulator (E3B). The evaluation
165 consisted in scoring the biometry images and of assessing their competence based on the
166 OSAUS score, by external observers. After the clinical training of Group B, their OSAUS
167 score (E3B) was significantly higher than that of Group A (E1A) after their clinical training

168 (3.45/5 vs. 2.375/5 [P=0.00017])[16]. The authors concluded that the best time to train on
169 simulators seems to be prior to clinical training on real patients.

170

171 **Quantitative metrics**

172 *Image quality scoring*

173 In the same study [17], Gueneuc et al. also compared the quality of the images
174 produced by the midwives before and after the simulation training, using the image-quality
175 score developed by Salomon et al.[15]. This score is used for quality control for routine
176 standardized fetal ultrasound images obtained in the second trimester of pregnancy. It is based
177 on expert analysis of the images and allows for good inter- and intra-reviewer reproducibility
178 (**Table 3**). The image quality score in Gueneuc et al.'s study was higher in group B (9.95
179 versus 12.67, p=0.003).

180 Chalouhi et al. published a prospective case control study in 2016 comparing a
181 simulation group A (N=10) and a clinical training group B (n=10)[6]. The metrics of
182 evaluation were image quality (again scored by external observers) according to Salomon's
183 score[15] and acquisition time. The total score in group A for image quality ($14.3/18 \pm 1.4$)
184 was significantly higher than that for group B (10.3 ± 2.75) ($p = 0.001$). The time needed to
185 accomplish the whole exam tended to be longer for group A (569 ± 174 s) compared to group B
186 (479 ± 104 s).

187 A Canadian prospective case control study by Rosen et al. in 2017 sought to assess the
188 impact of simulation training in 18 first or second year interns randomized in two groups:
189 simulation training and clinical training on a real patient [18]. The participants were asked to
190 perform four planes of the fetal brain on a real pregnant patient as a pre- and post-training
191 test: (1) head measurements (biparietal diameter and circumference), (2) cavum septum

192 pellucidum, (3) posterior horn of the lateral ventricle, and (4) posterior fossa. Each image was
193 scored from 0 to 5. The evaluation metrics assessed by external observers were the quality of
194 the image (focal adjustment, depth and zoom) caliper positioning, and landmark recognition.
195 They observed a significant improvement after the training in both the clinical (mean pre-
196 training score 13.3 versus mean post-training score 24.6; $p < 0.04$) and the simulation group
197 (mean pre-training score 15.9 versus mean post-training score 28.9, $p < 0.05$). They concluded
198 that initial training on a simulator was as useful as clinical training on a real patient.

199

200 *Fetal biometrics*

201 Burden et al. published a prospective study in 2013 using fetal biometrics to assess
202 participants' skills (18 students and 8 certified experts) in a virtual reality setting using an
203 ultrasound system on a full size mannikin (UltraSim™, MedSim Inc, Fort Lauderdale, FL,
204 USA)[19]. The planes studied were crown-rump-length of the fetus in the first trimester and
205 biparietal diameter, occipitofrontal diameter and femur length in the second trimester. The
206 metrics used for evaluation was the mean gap in percentage between the students'
207 measurements and the gold standard (experts' measurements). Over the five repetition scans,
208 the mean percentage difference fell from 7.1% to 4.1% for CRL measurement, and students
209 halved the time they took to make the measurement (141s to 70s $p < 0.001$). The time taken for
210 students to complete CRL and fetal biometry scans decreased significantly (all $P < 0.05$) with
211 repetitions.

212 In another prospective study by Le Lous et al. in 2017, the objective was to measure
213 the time needed to obtain the femur biometry plane before and after a 1-hour training in 30
214 students. The metrics used were the optimization of the image (gain, zoom, caliper
215 positioning) as well as the angle to the horizontal, and acquisition time. They showed a 30%

216 reduction in the time to obtain the right plane (86 sec versus 125 sec, $p=0.015$) after the
217 training, with a similar angle to the horizontal 10.08 versus 9.68, $p=0.84$ and similar femur
218 length (31.3mm versus 32mm, $p=0.15$). These criteria required a third person to measure the
219 length and angle on the images *a posteriori*.

220 Another study published in 2015 by Akoma et al. was unusual as they used a porcine
221 phantom simulator to assess if simulation training improved students' performance [20].
222 Local ultrasound machines were used and an examiner assessed the time to obtain the view,
223 and correct positioning of the calipers. After the training, each participant performed fetal
224 biometry three times on a real patient (head circumference, biparietal diameter, abdominal
225 circumference, and femur length), with the measurements having been previously made by an
226 expert as a reference. The participants were then randomized in two groups; the first group
227 ($n=12$) received 2 hours of training on a real patient and 1 hour on a simulator, and the other 3
228 hours in a clinical setting. New biometrics were obtained after the training on real patient. No
229 difference was observed between the two groups in terms of time to complete the
230 examination, fetal biometry errors, and number of satisfactory images ($p = 0.72$). This study
231 is interesting as it shows the potential of a low-cost simulator. However, once again, an
232 external observer was necessary.

233 Finally, in a multicenter prospective case control study conducted by Andreasen et al.
234 in 2019, the simulator training effect was measured using the exactitude of fetal weight
235 estimation before and after a simulation training [21]. The metrics of evaluation were first the
236 precision of fetal weight estimation, the quality of the images using Salomon's score, the time
237 spent on the simulator, and the time necessary to complete fetal weight estimation. Fetal
238 weight estimation was performed on a pregnant patient in the third trimester within 3 days
239 before delivery (elective c-section or labor induction).

240 Students were randomized into two comparable groups: “simulation training” or
241 “usual training”. The mean time spent on the simulator was 185.5 minutes. After completing
242 the training, the participants estimated the fetal weight once again and the images obtained
243 were scored by two experts using Salomon’s quality score [15]. The estimated fetal weight
244 was compared to the actual birth weight. After adjustment on the clinical experience, the
245 simulation group improved its diagnostic precision by 32% (CI 95 % [6.9, 50.1], (p=0.02).
246 Image quality scores improved by 1.2 points (CI 95 % [0.4, 20,1] p=0.0059) in the simulation
247 group, without any changes in the control group (IC à 95 %, [-0,8, 1,0] (p=0,78). In this study,
248 evaluation was based on a robust objective and quantitative criterion: the difference between
249 ultrasound estimation and actual birthweight. However, this is not a convenient criterion, first
250 because it is not easy to practice on pregnant volunteers just before delivery, and second
251 because we know that estimation of fetal weight late in the third trimester is challenging and
252 not the time when routine scans are performed. Overall, these observations highlight the
253 necessity to have other metrics.

254

255 *Simulator-generated metrics*

256 Simulator -generated metrics can also be used for the evaluation of the students, when
257 the simulator itself gives a feedback to the student at the end of the exercise.

258 Madsen et al. published a prospective study in 2017 comparing the learning curves of
259 20 midwives after two training sessions on two simulators: a virtual-reality simulator
260 (Scantrainer, Medaphor, Cardiff, UK) and a physical pelvic mannequin (BluePhantom, CAE,
261 Sarasota, FL) [22]. The participants first received 30 minutes of explanation about the
262 simulator and its functions, and then the number of repetitions necessary to reach a predefined
263 performance level and the time taken to complete the examination were recorded. Simulator
264 scores were based on automated measurements of movement accuracy, completeness of scan,

265 image optimization technique, and finally, the accuracy of different measurements. The
266 OSAUS score was also rated. The participants then measured cervical length on a real patient.
267 A good correlation was found between the time necessary to reach performance level
268 (Pearson correlation coefficient: 0.78, $p < 0.001$), and performance score in the clinical setting
269 and on simulators (Pearson correlation coefficient: 0.81, $p = 0.049$). This study showed that
270 performance of the different tasks was related and identified slower participants who needed
271 extra training.

272 Tolsgaard et al., in their randomized study in 2015, also collected simulator generated
273 metrics in the simulation group [9]. After completing a module, the simulator provided
274 automated feedback using dichotomous metrics: scanning through the entire uterus, and
275 sufficient optimization of the image. However, this automated feed-back was not sufficient to
276 evaluate the students and they mainly used the OSAUS score.

277

278 **Discussion**

279 This literature review describes the evaluation metrics currently used to assess ultrasound
280 simulation training in the field of obstetrics: they can be either quantitative or qualitative, and
281 assessed by an independent external observer or by the simulator itself. Simulation training is
282 being increasingly used to teach ultrasound skills that are crucial for learners in many medical
283 areas and especially for OB/GYNs and midwives. It is therefore important to define metrics
284 of evaluation.

285 Most of the studies in this review, focused on the quality of the images produced as a
286 means of assessing the participants' skills. This involves collecting the images produced in
287 the simulation task and scoring their quality *a posteriori* based on a score developed by
288 Salomon et al, with the help of one or two experts who look at all the images. The
289 Salomon score is an image quality score based on expert scoring and gives a good inter-
290 and intra-observer reproducibility[15]. To avoid the need for expert scoring, image quality
291 could also be assessed with deep learning image recognition[23].

292 However, beyond producing quality images, learners of ultrasound skills must also
293 know how to interpret them correctly and, of course, to diagnose major fetal defects. They
294 must also know how to manage the machine: most simulators include functions – focus,
295 depth, zoom - for image optimization. Thus another way of assessing students' skills is by
296 the OSAUS score, a valid and reliable tool developed by Tolgaard et al. to assess student's
297 ultrasound skills in all its aspects. It is composed of seven items and also scored by an
298 external examiner. The seven items assess the student's aptitude to: establish the
299 indication; manage the technical aspects of an ultrasound machine; optimize the image;
300 systematize the scan; interpret the images; and document the results (**Figure 2**).

301 Another metric which has been used is dexterity scored by an external observer.
302 Dexterity corresponds to the fluency in the execution of a task. For example, in Chalouhi et

303 al.'s study, dexterity was subjectively assessed by a score ranging from 0 to 10. While it is
304 relatively easy for an expert to correctly evaluate the skills of a student just by looking at
305 her/him and assessing ease of manipulation, this evaluation remains subjective. Future
306 work may be necessary to know if this parameter could be objectively and numerically
307 assessed comparing the smoothness of the trajectories according to expertise.

308 Of the 11 studies included in our review, only two used simulator generated metrics.
309 The automated feed-back provided by the simulator were: time spent on the simulator and
310 number of repetitions to achieve a certain level of performance, or dichotomous variables
311 such as optimization/no optimization, scanning of the entire uterus/or not. This shows the
312 volume examined, objectifies the systematization of the fetal scan, and flags up if a student
313 "forgot" to scan one part of the fetus. The time required to obtain the right plane is also
314 interesting, but the duration of the task might not be correlated with skill. The efficiency of
315 the movement and systemization of the examination could also be an interesting way of
316 automatically evaluating students. This could be assessed by recording the trajectory of the
317 probe. However, to the best of our knowledge, an automated analysis of the image quality
318 or of probe trajectory has not been studied to date.

319

320 All the studies included in this review agree that simulation may be of added value to
321 clinical training. Not only can the student be trained in an undisturbed environment outside
322 the clinical setting, but it may also allow an overview of the most frequent malformations,
323 especially for simulators which store a panel of pathological images which the student can
324 scan and record. The capacity to recognize a fetal defect may also be used for evaluation as
325 evaluation with volunteer patients typically involve normal fetuses. The criteria are
326 therefore the capacity to obtain the right plane with correct measurements, use the machine
327 correctly to position the calipers, and obtain a good image quality. The use of simulator

328 would allow evaluation of their capacity to make the right diagnosis – the ultimate aim of
329 ultrasound – especially of malformations for which the antenatal diagnosis impacts fetal
330 prognosis.

331 It is worth noting that none of these evaluation modalities assess non-technical skills.
332 Fetal ultrasound is not without psychological risks as it can generate uncertainty and
333 anxiety that can interfere with the mother-child bonding[24]. Telling a couple that their
334 fetus has a defect is difficult and this aspect cannot be taught or evaluated by a virtual
335 simulator. The best way to train for this may be through patient-actor simulations which
336 could be included in the training.

337 Finally, it has been demonstrated that clinical experience and time spent in specialized
338 ultrasound units are predictors of students' confidence in performing ultrasound
339 independently[2]. This highlights that simulation training can only ever be a complement
340 to clinical training, and not replace it.

341

342

343 **Conclusion**

344 Establishing clear metrics to evaluate ultrasound skills is challenging. Evaluation by
345 ultrasound simulation is interesting and different metrics of evaluation already exist,
346 though most involve an external observer. The main interest of the OUS would be the self-
347 teaching, if the technique did not require an external observer. The most common
348 evaluation criteria are the quality of images produced, the accuracy of measurement, and
349 the capacity to optimize the image. Additionally, the capacity to detect a fetal
350 malformation may be investigated as an evaluation criterion allowed by simulators.
351 Finally, dexterity could be an interesting way to automatically discriminate a level by
352 studying the trajectory of the probe.

353

354

355 **References**

- 356 [1] Patel H, Chandrasekaran D, Myriokefalitaki E, Gebeh A, Jones K, Jevé YB, et al. The
357 Role of Ultrasound Simulation in Obstetrics and Gynecology Training: A UK Trainees'
358 Perspective. *Simul Healthc* 2016;11:340–4. <https://doi.org/10.1097/SIH.000000000000176>.
- 359 [2] Tolsgaard MG, Rasmussen MB, Tappert C, Sundler M, Sorensen JL, Ottesen B, et al.
360 Which factors are associated with trainees' confidence in performing obstetric and
361 gynecological ultrasound examinations? *Ultrasound Obstet Gynecol* 2014;43:444–51.
362 <https://doi.org/10.1002/uog.13211>.
- 363 [3] Robyr R, Bernard J-P, Roume J, Ville Y. Familial diseases revealed by a fetal
364 anomaly. *Prenat Diagn* 2006;26:1224–34. <https://doi.org/10.1002/pd.1593>.
- 365 [4] Bravo-Valenzuela NJ, Peixoto AB, Araujo Júnior E. Prenatal diagnosis of
366 transposition of great arteries - an update review. *Ultrasonography* 2020.
367 <https://doi.org/10.14366/usg.20055>.
- 368 [5] Tolsgaard MG. A multiple-perspective approach for the assessment and learning of
369 ultrasound skills. *Perspect Med Educ* 2018;7:211–3. [https://doi.org/10.1007/s40037-018-](https://doi.org/10.1007/s40037-018-0419-8)
370 [0419-8](https://doi.org/10.1007/s40037-018-0419-8).
- 371 [6] Chalouhi GE, Quibel T, Lamourdedieu C, Hajal NJ, Gueneuc A, Benzina N, et al.
372 [Obstetrical ultrasound simulator as a tool for improving teaching strategies for beginners:
373 Pilot study and review of the literature]. *J Gynecol Obstet Biol Reprod (Paris)* 2016;45:1107–
374 14. <https://doi.org/10.1016/j.jgyn.2015.12.011>.
- 375 [7] Madsen ME, Konge L, Nørgaard LN, Tabor A, Ringsted C, Klemmensen AK, et al.
376 Assessment of performance measures and learning curves for use of a virtual-reality
377 ultrasound simulator in transvaginal ultrasound examination. *Ultrasound Obstet Gynecol*
378 2014;44:693–9. <https://doi.org/10.1002/uog.13400>.
- 379 [8] Tolsgaard MG, Chalouhi GE. Use of simulators for the assessment of trainees'
380 competence: trendy toys or valuable instruments? *Ultrasound Obstet Gynecol* 2018.
381 <https://doi.org/10.1002/uog.19071>.
- 382 [9] Tolsgaard MG, Ringsted C, Dreisler E, Nørgaard LN, Petersen JH, Madsen ME, et al.
383 Sustained effect of simulation-based ultrasound training on clinical performance: a
384 randomized trial. *Ultrasound Obstet Gynecol* 2015;46:312–8.
385 <https://doi.org/10.1002/uog.14780>.
- 386 [10] Tolsgaard MG, Ringsted C, Rosthøj S, Nørgaard L, Møller L, Freiesleben NLC, et al.
387 The Effects of Simulation-based Transvaginal Ultrasound Training on Quality and Efficiency
388 of Care: A Multicenter Single-blind Randomized Trial. *Ann Surg* 2017;265:630–7.
389 <https://doi.org/10.1097/SLA.0000000000001656>.
- 390 [11] Chalouhi GE, Bernardi V, Gueneuc A, Houssin I, Stirnemann JJ, Ville Y. Evaluation
391 of trainees' ability to perform obstetrical ultrasound using simulation: challenges and
392 opportunities. *Am J Obstet Gynecol* 2016;214:525.e1-525.e8.
393 <https://doi.org/10.1016/j.ajog.2015.10.932>.
- 394 [12] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred
395 reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015
396 statement. *Syst Rev* 2015;4:1. <https://doi.org/10.1186/2046-4053-4-1>.
- 397 [13] Zimmermann R, Mousty E, Mares P, Letouzey V, Huberlant S. E-learning et
398 simulation en échographie focalisée pour la formation continue des sages-femmes en salle de
399 naissance. *Gynécologie Obstétrique Fertilité & Sénologie* 2019;47:836–40.
400 <https://doi.org/10.1016/j.gofs.2019.10.010>.

401 [14] Chalouhi GE, Bernardi V, Gueneuc A, Houssin I, Stirnemann JJ, Ville Y. Evaluation
402 of trainees' ability to perform obstetrical ultrasound using simulation: challenges and
403 opportunities. *Am J Obstet Gynecol* 2016;214:525.e1-525.e8.
404 <https://doi.org/10.1016/j.ajog.2015.10.932>.

405 [15] Salomon LJ, Winer N, Bernard JP, Ville Y. A score-based method for quality control
406 of fetal images at routine second-trimester ultrasound examination. *Prenat Diagn*
407 2008;28:822–7. <https://doi.org/10.1002/pd.2016>.

408 [16] Tolsgaard MG, Ringsted C, Dreisler E, Klemmensen A, Loft A, Sorensen JL, et al.
409 Reliable and valid assessment of ultrasound operator competence in obstetrics and
410 gynecology. *Ultrasound Obstet Gynecol* 2014;43:437–43. <https://doi.org/10.1002/uog.13198>.

411 [17] Gueneuc A, De Garnier J, Dommergues M, Rivière M, Ville Y, Chalouhi GE. Impact
412 de l'intégration de la simulation en échographie obstétricale dans le cursus des étudiants
413 sages-femmes. *Gynécologie Obstétrique Fertilité & Sénologie* 2019;47:776–82.
414 <https://doi.org/10.1016/j.gofs.2019.07.004>.

415 [18] Rosen H, Windrim R, Lee YM, Gotha L, Perelman V, Ronzoni S. Simulator Based
416 Obstetric Ultrasound Training: A Prospective, Randomized Single-Blinded Study. *J Obstet*
417 *Gynaecol Can* 2017;39:166–73. <https://doi.org/10.1016/j.jogc.2016.10.009>.

418 [19] Burden C, Preshaw J, White P, Draycott TJ, Grant S, Fox R. Usability of virtual-
419 reality simulation training in obstetric ultrasonography: a prospective cohort study.
420 *Ultrasound Obstet Gynecol* 2013;42:213–7. <https://doi.org/10.1002/uog.12394>.

421 [20] Akoma UN, Shumard KM, Street L, Brost BC, Nitsche JF. Impact of an Inexpensive
422 Anatomy-Based Fetal Pig Simulator on Obstetric Ultrasound Training. *J Ultrasound Med*
423 2015;34:1793–9. <https://doi.org/10.7863/ultra.15.14.12004>.

424 [21] Andreasen LA, Tabor A, Nørgaard LN, Ringsted C, Sandager P, Rosthøj S, et al. Is
425 simulation training only for inexperienced trainees? A multicenter randomized trial exploring
426 the effects of simulation-based ultrasound training on obstetricians' diagnostic accuracy.
427 *Ultrasound in Obstetrics & Gynecology* n.d.;n/a. <https://doi.org/10.1002/uog.20362>.

428 [22] Madsen ME, Nørgaard LN, Tabor A, Konge L, Ringsted C, Tolsgaard MG. The
429 Predictive Value of Ultrasound Learning Curves Across Simulated and Clinical Settings:
430 Predictive Value of Learning Curves During Sonographic Simulation. *J Ultrasound Med*
431 2017;36:201–8. <https://doi.org/10.7863/ultra.16.01037>.

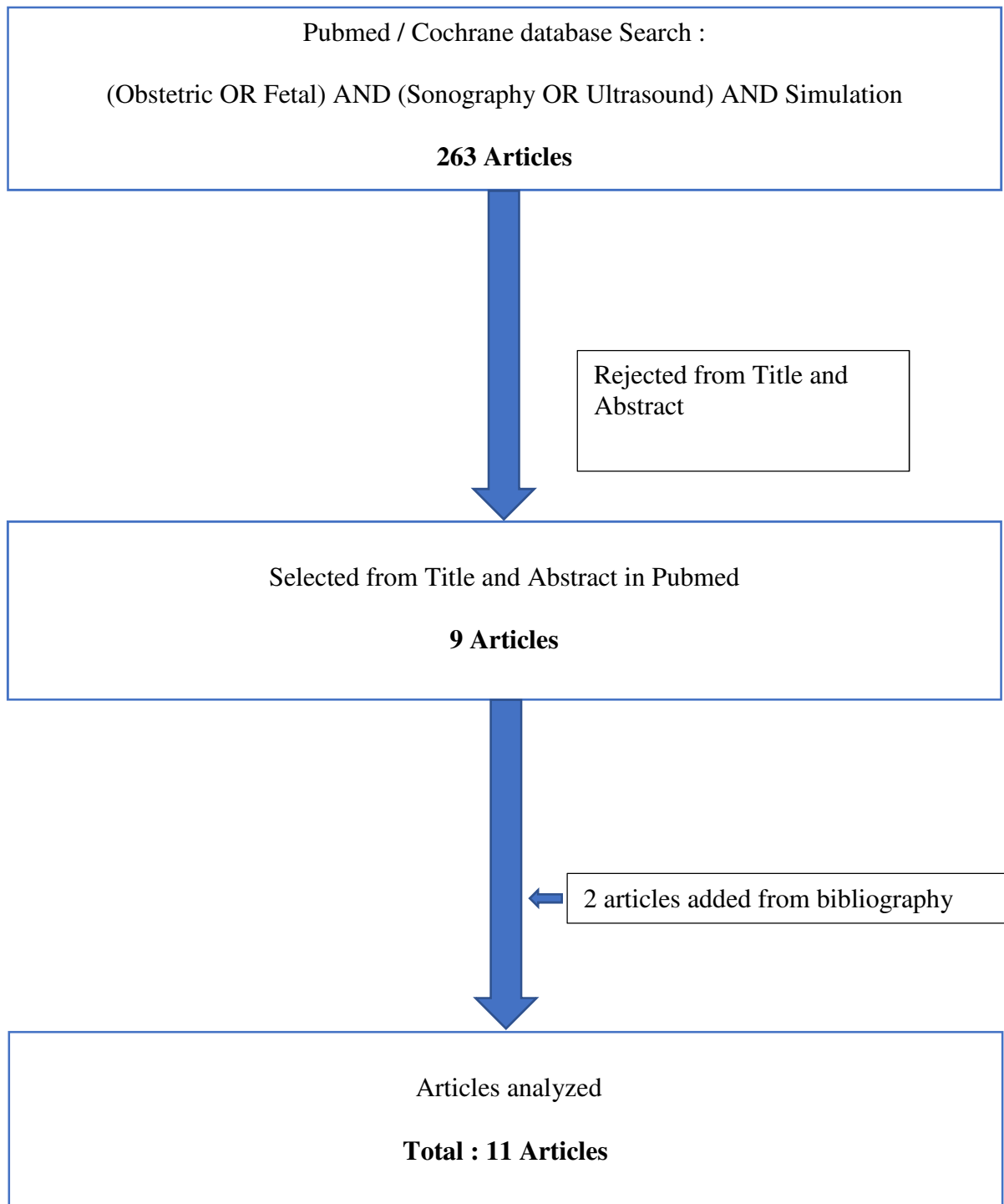
432 [23] Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in
433 ultrasound imaging in obstetrics and gynecology. *Ultrasound Obstet Gynecol* 2020.
434 <https://doi.org/10.1002/uog.22122>.

435 [24] Rothenberger SE, Moehler E, Reck C, Resch F. Prenatal stress: course and
436 interrelation of emotional and physiological stress measures. *Psychopathology* 2011;44:60–7.
437 <https://doi.org/10.1159/000319309>.

438

439

Figure 1. Flow Chart



Author	Journal	Design	N	Learning objective	System	Planes	Trajectory metrics	Metrics used
Andreasen et al. 2019	Ultrasound in obstetrics & gynecology	Prospective case control	70	Impact of simulation according to level experience	Medaphor Scantrainer (Cardiff, UK)	HC AC FL	No	Image quality by Salomon criteria Time on simulator Time to achieve task Precision of Fetal biometrics
Gueneuc et al. 2019	Gynecology Obstetrics Fertility and Sénology	Prospective	40	Second trimester scan Placenta localization Fetal presentation Fetal biometrics Amniotic fluid evaluation	Vimedix TM, CAE Healthcare, (Sarasota, USA)	IAF HC BIP AC FL	No	Image quality by Salomon criteria OSAUS score
Zimmerman et al. 2019	Gynecology Obstetrics Fertility and Sénology	Retrospective	33	Placenta localization Fetal presentation Fetal biometrics Amniotic fluid evaluation	ND	0	No	Fetal presentation, maniability of the probe, cervix position, placenta localization
Rosen et al. 2017	Journal of Obstetrics and Gynecology Canada	Prospective case control	18	Fetal brain examination	Vimedix Ob/Gyn CAE Healthcare, Montreal, Quebec	BIP HC Cerebellum Posterior ventricle Cavum of septum pellucidum	No	Image optimization (gain, zoom, calipers) Acquisition benchmark Calipers position
Lelous et al. 2017	Journal of Gynecology Obstetrics and Human Reproduction	Prospective		Femoral Length	Vimedix TM Simulator	FL	No	Image optimization (gain, zoom, calipers) Angle to the horizontal Time to achieve task
Madsen et al. 2017	Journal of Ultrasound in Medicine	Prospective	20	Cervical Length	Scantrainer, Medaphor, Cardiff, UK BluePhantom, CAE, Sarasota, USA	FL	No	OSAUS score Time to achieve task Time spent on simulator Number of repetitions Simulator generated metrics
Chalouhi et al. 2016	Journal of obstetrics Gynecology and Human Reproduction	Prospective case control	20	Second trimester scan	ND	HC BIP AC FL Placenta localization Fetal presentation	No	Image quality by Salomon criteria Time to achieve task

Chalouhi et al. 2016	American Journal of Obstetrics and Gynecology	Prospective	29	Fetal biometrics	Vimedix TM (CAE Healthcare, Sarasota, USA)	BIP/HC AC FL Upper lip Four chambers view Right/Left cardiac outlet Kidneys Diaphragm Stomach Spine	No	Image quality by Salomon criteria Dexterity (1 – 10)
Akoma et al. 2015	Journal of ultrasound in medicine	Prospective case-control	24	Fetal biometrics	0	BIP HC AC FL	No	Measurement Time to achieve task
Tolsgaard et al. 2016	Ultrasound in Obstetrics & Gynecology	Prospective case-control	26	Cervical Length	Scantrainer, Medaphor TM , Cardiff, UK BluePhantom, CAE Healthcare, Sarasota, FL, USA	NC	No	OSAUS Score Time spent on the simulator Simulator generated metrics Number of repetition
Burden et al. 2013	Ultrasound in Obstetrics & Gynecology	Prospective	26	Crown-Rump Length Fetal biometrics	UltraSim TM ; MedSim Inc., Fort Lauderdale, FL, USA	CRL BIP/HC FL	No	Mean gap with gold standard (%) Time to achieve task

HC : Head circumference ; BIP : Biparietal diameter ; AC : abdominal circumference ; FL : femur length ;AFI : Amniotic fluid index ; CRL Crown Rump Length

Table 1. Literature overview of ultrasound simulation metrics.

1 **Table 2. OSAUS Score**

2

	1	2	3	4	5
1. Indication for the examination If applicable. Reviewing patient history and knowing why the examination is indicated	Displays poor knowledge of the indication of the examination		Displays some knowledge of the indication of the examination		Displays ample knowledge of the indication of the examination
2. Applied knowledge of ultrasound equipment Familiarity with the equipment and its functions, selecting probe, using buttons and application of gel.	Unable to operate equipment		Operate the equipment with some experience		Familiar with operating the equipment
3. Image optimization Consistently ensuring optimize image quality by adjusting gain, depth, focus, frequency	Fails to optimize images		Competent image optimization but not done consistently		Consistent optimization of images
4. Systematic examination Consistently displaying systematic approach to the examination and presentation of relevant structures according to guidelines.	Unsystematic approach		Displays some systematic approach		Consistently displays systematic approach
5. Interpretation of images Recognition of image pattern and interpretation of findings.	Unable to interpret any findings		Does not consistently interpret findings correctly		Consistently interpret findings correctly
6. Documentation of examination Image recording and focused verbal/written documentation.	Does not document any images		Documents most relevant images		Consistently documents relevant images
7. Medical decision making If applicable. Ability to integrate scan results into the care of the patient ad medical decision making.	Unable to integrate findings into medical decision making		Able to integrate findings into a clinical context		Consistent integration of findings into medical decision making

3

4

5

6

7

8

Table 3. Image quality score by Salomon et al.

	Biparietal diameter	Abdominal circumference	Femur Length
1	Symetrical	Symetrical	Both femur extremities are visibles
2	Thalami present	Stomach present	Angle to the horizontal <45°
3	Cavum pellucidum present	Sinus porta present	Sufficient zoom
4	No Cerebellum	No kydney	Correct position of the calipers
5	Sufficient zoom	Sufficient zoom	
6	Correct position of the calipers and ellipse	Correct position of the calipers and ellipse	
Total Score	6	6	4