UCLA | CRESST NATIONAL CENTER FOR RESEARCH ON EVALUATION, STANDARDS, AND STUDENT TESTING

SIMULATION-BASED ASSESSMENT OF ULTRASOUND PROFICIENCY

Markus R. Iseli, John J. Lee, Katerina Schenke, Seth Leon, Deborah Lim, Barbara Jones, and Li Cai

OCTOBER 2019

CRESST REPORT **865**

The work herein reported was supported by grant number W81XWH-16-C-0146 from the U.S. Army Medical Research Acquisition Activity to Pélagique, LLC to National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

To cite from this report, please use the following as your APA reference: Iseli, M. R., Lee, J. J., Schenke, K., Leon, S., Lim, D., Jones, B., & Cai, L. (2019). *Simulation-based assessment of ultrasound proficiency* (CRESST Report 865). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Table of Contents

Simulation-Based Assessment of Ultrasound Proficiency

Markus R. Iseli, John J. Lee, Katerina Schenke, Seth Leon, Deborah Lim, Barbara Jones, and Li Cai

CRESST/University of California, Los Angeles

Abstract: Ultrasound proficiency is an important skill for military and nonmilitary personnel. This study looks at the development of a validated method to assess both declarative and procedural skills. The study consisted of a user model validation part and a skill decay part. Methods were developed for taking data from a SonoSim® ultrasound training solution hardware and software tool and processing the data for use in a Bayesian network (BN). A total of 86 participants were assessed on the knowledge and skills of fetal biometry in obstetrics (OB, $n = 58$) and of internal jugular vein cannulation (IJV, $n = 37$). We found that scanning experience, declarative knowledge, procedural skills, and model inferences strongly correlated with each other overall: Better performance (declarative and procedural) correlated with higher scores predicted by our model. Experienced participants performed better than inexperienced participants on many metrics and exhibited smaller variations in performance. We found small but not statistically significant skill decay which could be indicative of a testing effect and individual differences.

INTRODUCTION

Simulation technology is becoming widely used and accepted as a valid method of training for various medical skills, with ultrasound being one of the best studied (Mackay, Zhou, Lewis, Fraser, & Atkinson, 2018). The medical health sciences community (MHS) is adopting ultrasonography across greater numbers of personnel and across more professional disciplines as its benefits to the efficiency and effectiveness of military healthcare are appreciated (Nelson & Sanghvi, 2016). Medical schools are also seeing the need to integrate ultrasound training into their curriculum (Chiem et al., 2016; Dinh et al., 2015). Mobile health, a subset of telemedicine concerned with the use of mobile networks and devices for healthcare delivery, makes the possibility of medical diagnosis ubiquitously available through diagnostic hardware, such as ultrasound probes that can be directly connected to a provider’s smartphone (Selig, Collins, Church, & Zeman, 2019).

With use of ultrasound in medicine becoming more widespread, the accurate and valid modeling of student knowledge, skills, and attributes (KSA) is thus of high importance for the valid assessment of ultrasound proficiency and competence. This holds true for basic traditional procedures like fetal biometry (MacGregor & Sabbagha, 2009), and for ultrasound-guided procedures (Ahmed et al., 2016; American Institute of Ultrasound in Medicine, 2019; Chenkin, Lee, Huynh, & Bandiera, 2008) like internal jugular vein cannulation. Both these procedures were used in this study.

The fundamental principles of ultrasound include five goals (American Institute of Ultrasound in Medicine, 2019, p. 1952):

1. To select a transducer that will provide optimal penetration of the anatomic region of interest.
2. To maximize sound wave transmission to and echo acquisition from target structures.
3. To maximize image quality.
4. To accurately interpret images.
5. To keep the potential for bioeffects as lows as reasonably achievable (ALARA).

The focus for this study was on measuring the skills in obtaining a high-quality image (Goals 2 and 3) for taking measurements or determining location of a needle insertion and accurately interpreting the images (Goal 4), that is, successfully identifying anatomical structures.

This research builds on earlier CRESST efforts to examine ultrasound skills using a computer-based simulator with probe (Chung, Gyllenhammer, Baker, & Savitsky, 2013; Iseli & Savitsky, 2015). Chung et al. (2013) found that for focused assessment with sonography for trauma (FAST) exams, window interpretation performance was significantly better using the SonoSim® simulator-based training than classroom-based training. This has implications in terms of the use of sonography in trauma and combat hospitals in that training to improve proficiency can be provided using computer-based simulation versus live patients. Iseli and Savitsky (2015) described an ontology that was developed to model the domain of ultrasonography skills which included a procedural portion that broke out the six phases of ultrasound examination including indication, preparation, acquisition, interpretation, presentation, and integration of findings (see next section; similar phases to Todsen et al., 2015). This ontology was generated by four subject matter experts with a rater agreement of .55, and was composed of eight Level 1 entities, 24 Level 2 entities, and 119 Level 3 entities (Iseli & Savitsky, 2015).

In this study, CRESST researchers initially examined the validity of its ultrasound proficiency model for two tasks: fetal biometry (OB) and ultrasound-guided internal jugular vein cannulation (IJV), with physician, resident, and medical student participants. For a subset of these participants, skill decay was explored for the same two tasks with knowledge and skills assessed at approximately zero, two, six and 12 weeks following mastery training. An automated assessment approach was used in this study to determine its validity and reliability to reduce if not eliminate the subsequent need for examining interrater reliability as required by other scoring methods such as the Objective Structured Assessment of Ultrasound Skills (OSAUS, Todsen et al., 2015).

This report is organized as follows: The Models section describes the organization of the domain knowledge and skills into declarative and procedural ontologies. The subsection User Model describes the integration of the procedural ontology into a Bayesian network and the subsection Skill Decay gives a brief overview on skill decay that, given sufficient collected data, could be modeled using latent growth models. The Metrics section then outlines the procedural, declarative, and model-related metrics (including measures) used in this study. The Methods section describes the two parts of the study (user model validation and skill decay parts), the recruitment of participants, the administration protocols, and how the collected data were processed. The Results and Discussion section presents results by study part and by task and is followed by the Limitations section and the Summary section.

MODELS

To guide assessment and computational model design, we first created an ontology to outline and define the subdomains of ultrasound relevant to our study: fundamentals of ultrasound, obstetrics (i.e., fetal biometry), and ultrasound-guided procedures (i.e., internal jugular vein cannulation).

An ontology is composed of “explicit formal specifications of the terms in the domain and relations among them” (Noy & McGuiness, 2001, p. 1). These terms and relations can be represented in spreadsheets as well as graphically, and can be used to develop corresponding Bayesian network (BN) models. The existing Ontology for Fundamentals of Ultrasound (Iseli & Cai, 2016) proved to be an excellent framework to fit knowledge, skills, and attributes (KSAs) for new subdomains because it describes the procedural phases and steps required to be successful at performing an ultrasound examination or procedure. Its structure was slightly updated to contain the following six phases listed below with a subset of their associated performance-related questions:

- **Indication:** Is ultrasound examination or ultrasound-guided procedure indicated? Were the risks/benefits or indications/contraindications correctly determined?
- **Preparation:** Were the appropriate steps regarding safety, ergonomics, equipment, and planning undertaken to be well prepared for the examination/procedure? Were the appropriate procedures selected?
- **Acquisition:** What was the performance on psychomotor skills (i.e., inferred from hand motion data)? How appropriate was the choice of scanning windows and views?

Were the resulting images of anatomy and pathology optimally displayed, using imaging standards? Was confirmatory analysis of the results of a scan/procedure needed, and if yes, how well was it executed?

- **Interpretation:** How well were the generated images and eventual sonographic artifacts interpreted? Were normal and pathological findings accurately identified and interpreted?
- **Presentation:** Were findings reported and archived in accord with the institution's standards?
- **Integration:** Were reported findings integrated into medical decision making, patient management, and care?

Subdomain-related declarative knowledge that is needed to perform the procedural steps is stored in a "declarative ontology," which currently contains taxonomic relations between entities related to ultrasonography within and between subdomains or topics. For example, the declarative entity ANATOMY lists all the different anatomical parts discussed in the courses (e.g., heart, head and neck, musculoskeletal system, etc.), and the entity SONOGRAPHIC ARTIFACTS lists all the different types of artifacts (e.g., acoustic shadowing, edge artifacts, etc.).

Figure 1 gives a top-level overview of the two ontologies.

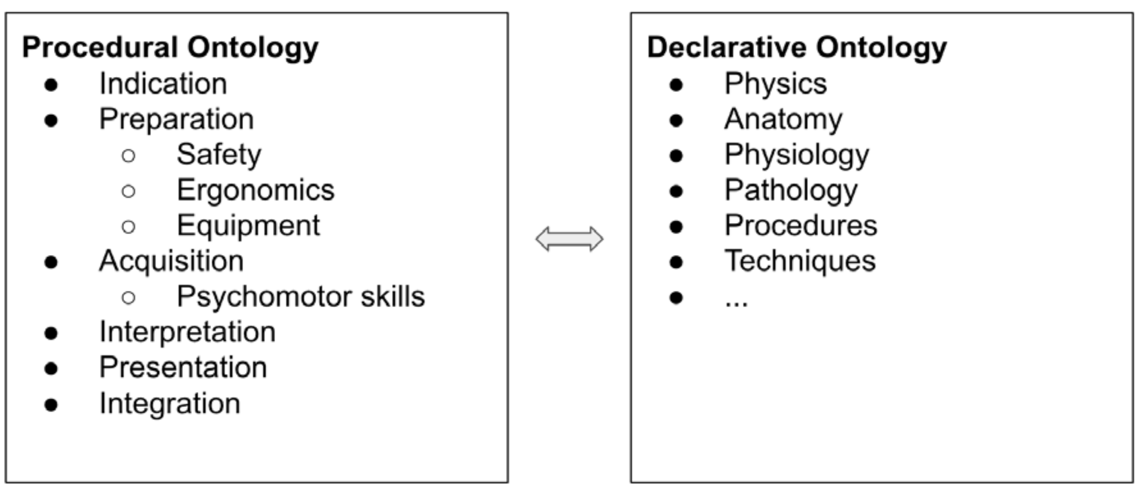


Figure 1. Ultrasound procedural and declarative ontologies.

To express the procedural steps in the context of their subdomains, the procedural ontology currently embeds references to the declarative ontology as uppercase text. For example, in the procedural ontology, in the Interpretation phase, the step "Differentiate SONOGRAPHIC ARTIFACTS from normal ANATOMY or PATHOLOGY" has three references written in uppercase script that link to the corresponding nodes in the declarative ontology.

Proficiency can thus be stated/reported in two different ways: (a) proficiency of procedural knowledge (high level), and (b) proficiency of procedural knowledge in the context of a specialty or subdomain (low level). The former will return a proficiency statement of "Differentiate SONOGRAPHIC ARTIFACTS from normal ANATOMY or PATHOLOGY"; the latter will detail the context as "Differentiate SONOGRAPHIC ARTIFACTS (e.g., acoustic shadowing) from normal ANATOMY (e.g., kidney) or PATHOLOGY (e.g., kidney stone)."

Given the above ontology, we then defined the computational models to be used to model a user's KSA (user model) and a user's skill decay (skill decay model) as outlined below.

User Model

To model ultrasound competency, we used Bayesian networks (Almond, Mislevy, Steinberg, Yan, & Williamson, 2015; Culbertson, 2015) because these models are versatile, are easily expandable and scalable, can easily be combined with other latent variable or probabilistic models, are able to make probabilistic inferences about observed data, and can deal with missing data. Bayesian networks (BNs) are probabilistic graphical models that represent joint probabilities on a set of variables by encoding the variables' independence assumptions. BNs can be used to infer a person's knowledge, skills, and attributes (KSAs), given observations of the person's behavior.

Our BN model incorporates the measures of knowledge, quality, and efficiency, and background and individual difference measures. The BN model developed for this study is based on the procedural ontology and is shown in Figure 2, where nodes representing latent (i.e., not directly measured) variables are depicted in light blue and nodes representing observable variables are orange. Evidence in the form of probability scores for each metric (see Table 1) is entered into the observable variables which will update the BN probabilities of mastery of the various latent skills. Independent of which topic evidence is entered, the same network is used and the top-level latent score from the node Procedural Skills is used in this study. However, to show from which topic the evidence was collected, we assign this score different labels: *BN probability of OB skills* and *BN probability of IJV skills*. See Appendix A for definitions of the BN nodes.

SME User Model Reviews

A panel of 14 subject matter experts, all medical doctors, provided feedback on our ontology design, KSA wording and terminology, our study design, and study materials. We also had a machine learning expert review our computational models as well as the metrics format and requirements. Reviews of ontologies and models were mostly positive, requiring only minor wording changes.

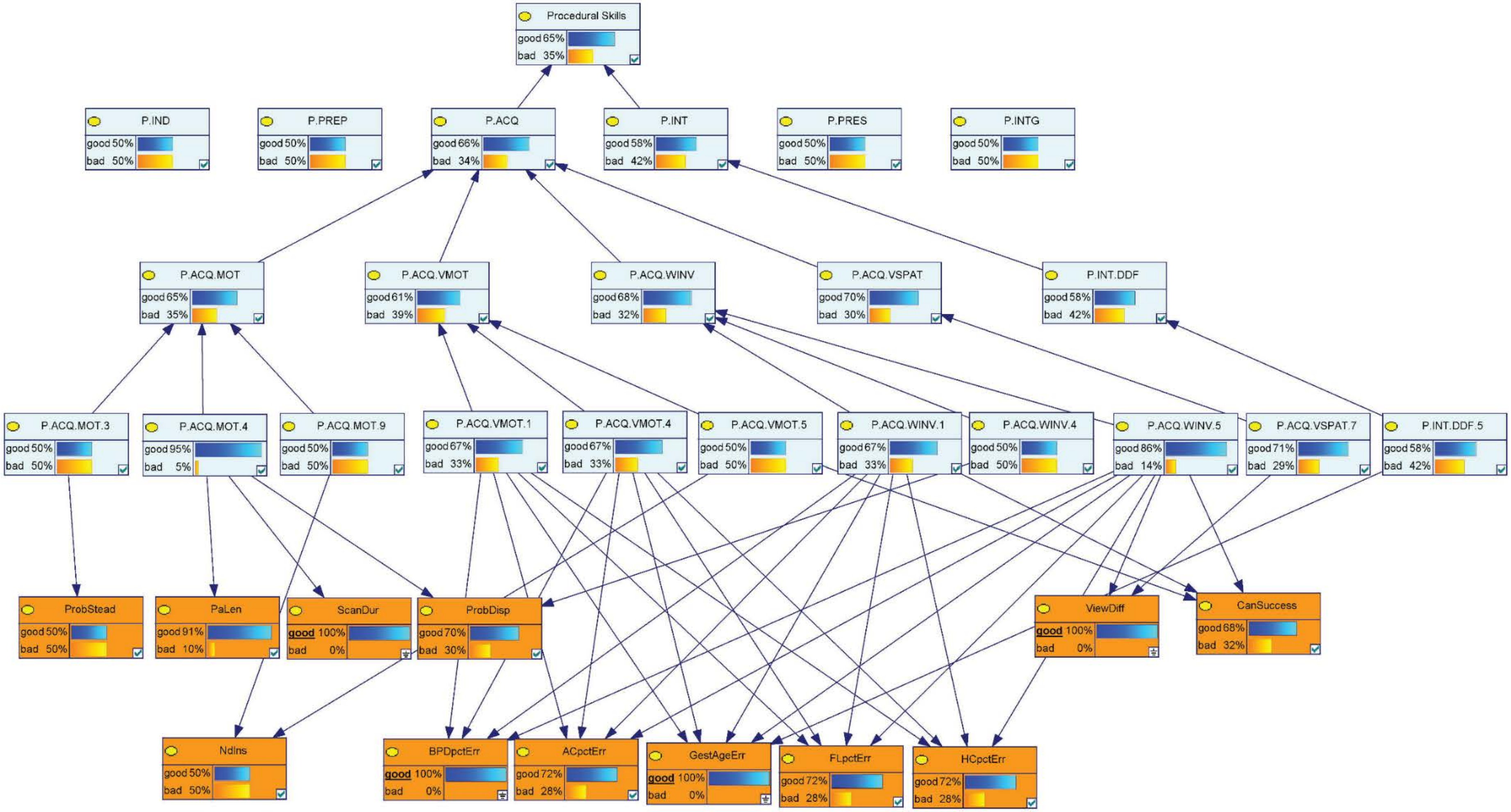


Figure 2. Ultrasound proficiency Bayesian network (BN). See Appendix A for node descriptions.

Skill Decay

The scientific literature on skill decay is limited, but what does exist suggests that complex skills, once mastered, degrade over several months. Cognitive aspects of the complex skills appear to degrade first, with notable degradation after approximately three to six months. Psychomotor elements persist longer but degrade as well, with notable decay occurring after 10 to 12 months (Arthur, Bennett, Stanush, & McNelly, 1998; Arthur et al., 2007; Driskell, Willis, & Copper, 1992; Wisher, Sabol, Ellis, & Ellis, 1999). Critically, among the first cognitive skills to show signs of degradation are the abilities to prioritize and execute actions and responses in the appropriate order (Wisher et al., 1999). Andersen (2012) describes this skill as "expert judgment," the cognitive skills required to go beyond routine, automated procedures to analytical behavior leading to situational awareness that allows better judgment on choice of actions and when and how they're performed.

To ameliorate skill decay and improve skill retention, refresher training is recommended. However, there are several factors and moderators that potentially influence skill decay or retention. Arthur et al. (1998) and Wisher et al. (1999) mention the following factors: (a) miscellaneous task characteristics such as psychomotor versus cognitive tasks, recall versus recognition tasks, number of steps involved, task integration, level of task organization, task structure and complexity, and task difficulty; (b) training characteristics such as distribution of practice (e.g., part vs. whole, massed vs. distributed), degree of overlearning, instructional strategies and training methods, programmed learning, memory aids, spacing of trials or sessions, feedback, and hypnosis during training; (c) retention interval and test characteristics such as rehearsal, test trial characteristics, kinds of rehearsal, relearning, practice during rehearsal, test taking during retention interval, and repetition of test trials; (d) individual differences such as motivation of trainee, amount of previous training, intelligence of trainee, ability of trainee, and trainee age; and (e) perceptual skills that involve the ability to discriminate between and to classify stimuli based on perceivable properties (e.g., visual-spatial cognition). Assuming that all participants in our study got trained and assessed on selected tasks only within the SonoSim® automated assessment system and that we were able to measure when each participant received which instruction and performed which task, we should have been able to measure the abovementioned factors (a), (b), and (c). Some of the factors in (d) can be assessed by survey self-evaluation questions, whereas factors in (e) can only be inferred from observed performance. For our assumptions to hold, we ideally would need to ensure that the skills and associated tasks selected for this study were not part of the participants' normal work and that all activity in the system would need to be logged. The latter condition is easily satisfied—the former condition, however, is harder to satisfy, especially if study training and testing are very closely related to or integrated into the existing participants' curriculum. In any case, we designed survey questions that collected information on the time, amount, and frequency of training on knowledge and skills related to our study received during the retention interval.

Because of the limited number of participants who completed the skill decay portion of the study, we were unable to estimate latent growth curve models (Bollen & Curran, 2006). Instead, to understand the extent to which skill decay occurred in participants' procedural skills across OB and IJV, we will illustrate the mean scores on tasks across time for each individual.

METRICS

Metrics that demonstrated sensitivity to expertise and most directly reflected ultrasonography knowledge and skill were included. We assessed declarative knowledge (answers to multiple-choice questions) and procedural skills (hand motion data, performance on ultrasound simulator test cases). To assess declarative KSAs, we defined multiple-choice questions for the following four subdomains: fundamentals of ultrasound, second- and third-trimester fetal biometry (OB), ultrasound-guided procedures, and internal jugular vein cannulation (IJV). See Appendix B for multiple-choice mastery question examples. To assess procedural KSAs, we defined simulation performance metrics as shown in Table 1. These metrics were calculated for each participant and task.

Table 1

Procedural Metrics Used for the Assessment of Procedural Skills

Metric type	Metric short name	Metric description	Metric interpretation
Procedural metrics			
OB&IJV	scan duration	Scan duration to achieve optimal view (in milliseconds)	Shorter duration is not necessarily desirable as it might indicate more thorough examination and identification of landmarks
OB&IJV	scan path length	Scan path length (in radians)	Shorter path lengths are desirable as participants were asked to use economy of probe motion
OB&IJV	probe acceleration	Scan path smoothness—or hand steadiness: calculated as probe acceleration (radians/second^2)	Smaller acceleration is preferable as it shows better steadiness
OB&IJV	scan view error	Alignment error to optimal scan view: Quaternion distance to optimal view quaternion (radians)	Smaller distance / difference to optimal view is better
OB	BPD score	Biparietal diameter (BPD): 1.0 minus percentage error relative to optimal value (float between 0.0 and 1.0)	Higher score is better
OB	HC score	Head circumference (HC): 1.0 minus percentage error relative to optimal value (float between 0.0 and 1.0)	Higher score is better
OB	AC score	Abdominal circumference (AC): 1.0 minus percentage error relative to optimal value (float between 0.0 and 1.0)	Higher score is better
OB	FL score	Femur length (FL) score: 1.0 minus percentage error relative to optimal value (float between 0.0 and 1.0)	Higher score is better
OB	gestational age score	Calculated gestational age: 1.0 minus percentage error relative to optimal value (float between 0.0 and 1.0)	Higher score is better
IJV	cannulation score	Cannulation success: successful (coded 0.95), unsuccessful (0.65), dangerous (0.05: e.g., puncture of carotid artery)	Higher score is better
IJV	needle insertions	Number of needle insertions (integer number)	Lower number is better

Metric type	Metric short name	Metric description	Metric interpretation
IJV	probe displacement	Probe displacement on pressure pad (millimeters)	Lower number is considered better
Declarative metrics			
OB&IJV	fundamentals score	Fundamentals of Ultrasonography mastery score (0 = none correct to 1 = 100% correct)	Higher value is better
IJV	USGP score	Ultrasound guided procedures (USGP) mastery score (0 = none correct to 1 = 100% correct)	Higher value is better
OB	OB score	2nd and 3rd Trimester mastery score (0 = none correct to 1 = 100% correct)	Higher value is better
IJV	IJV score	Internal Jugular Vein mastery score (0 = none correct to 1 = 100% correct)	Higher value is better
Bayes net model metrics			
OB	BN probability of OB skills	Bayesian network probability of fetal biometry skill mastery using above procedural metrics as evidence (0 = no skill to 1 = best skill)	Higher value is better
IJV	BN probability of IJV skills	Bayesian network probability of internal jugular vein cannulation skill mastery using above procedural metrics as evidence (0 = no skill to 1 = best skill)	Higher value is better

Note. Metric type defines the addressed assessment topic which can be either fetal biometry (OB), internal jugular vein cannulation (IJV), or metrics collected independent of topic (OB&IJV).

METHODS

Participants

Participants were drawn from physicians, residents, and medical students, mostly from the University of California, Los Angeles, and some from surrounding hospital facilities in the greater Los Angeles area. For the OB task, additional participants were drawn from OB/GYN students at the Uniformed Services University for the Health Sciences (USUHS) in Bethesda, Maryland. Our two-group experience criterion for both OB and IJV tasks was determined as follows: Participants who reported on the survey that they had performed the OB- or IJV-related procedures more than 10 times on real patients were considered experienced; all other participants are considered inexperienced. Participant recruiting and data collection followed

UCLA Institutional Review Board (IRB) approved guidelines and protocols. Detailed demographics are given in the results section of this report.

Materials

Two different setups were developed for the data collection, each utilizing SonoSim® hardware and software (U.S. Patent No. 8,480,404, 2013), one for the fetal biometry assessment (OB) and one for internal jugular vein cannulation task (IJV). Both were developed by Pélagique, Inc. All the setups included an ultrasound probe, an external monitor, and a laptop loaded with the SonoSimulator® software for each respective task. The fetal biometry assessment only included one station (see Figure 3), whereas IJV had two (see Figure 4)—one for the right neck and one for the left. This allowed for quicker administration of the tasks by not requiring the neck pad to be switched for left- and right-side scanning and needle cannulation. For IJV, we also had a mannikin and simulated needle (see Figure 5) at each of the two stations (Dong, Savitsky, & Osher, 2009). The OB assessment contained 15 tasks: three task types times five cases/patients. The task types were (a) measurement of head circumference (HC) and biparietal diameter (BPD); (b) measurement of abdominal circumference (AC); and (c) measurement of femur length of the fetus. The IJV assessment contained 12 tasks: four tasks (each presenting a different case) repeated three times.[1] The needle was not sharp, and retracted into the syringe to simulate insertion into the neck of the mannikin. External cameras in each station were used to take the video of the participant interacting with the virtual patient, particularly to show the position of the probe, hand, and for IJV, the simulated needle used for the cannulation. Proctor scripts were developed for both assessment topics and for both parts of the study: user model validation and skill decay. See Appendix C and Appendix D for the proctor scripts.

One known limitation of the fetal biometry task is that the lateral movement (translation) of the ultrasound probe did not show up on the dynamically updated ultrasound image on the monitor. This translation feature was added for the IJV task, and participants used ultrasound gel to allow for easy gliding of the probe over the neck pad surface.

[1]We focused on the analysis of left-sided IJV because of analysis time constraints and some missing data for right-sided tasks that still need to be explored.

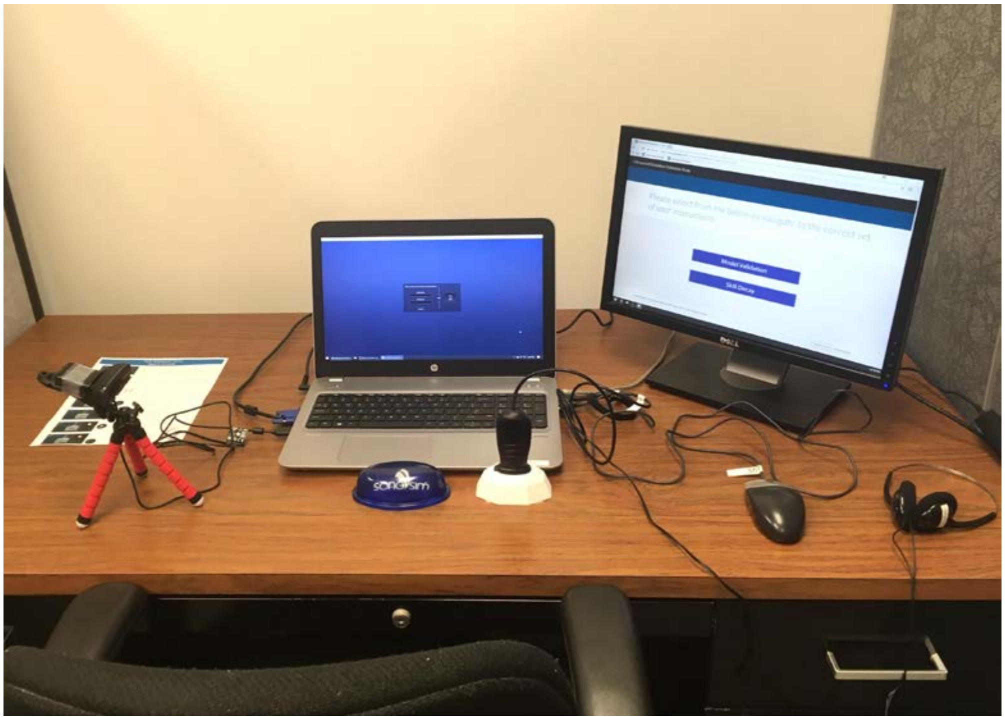


Figure 3. Setup for fetal biometry (OB task).

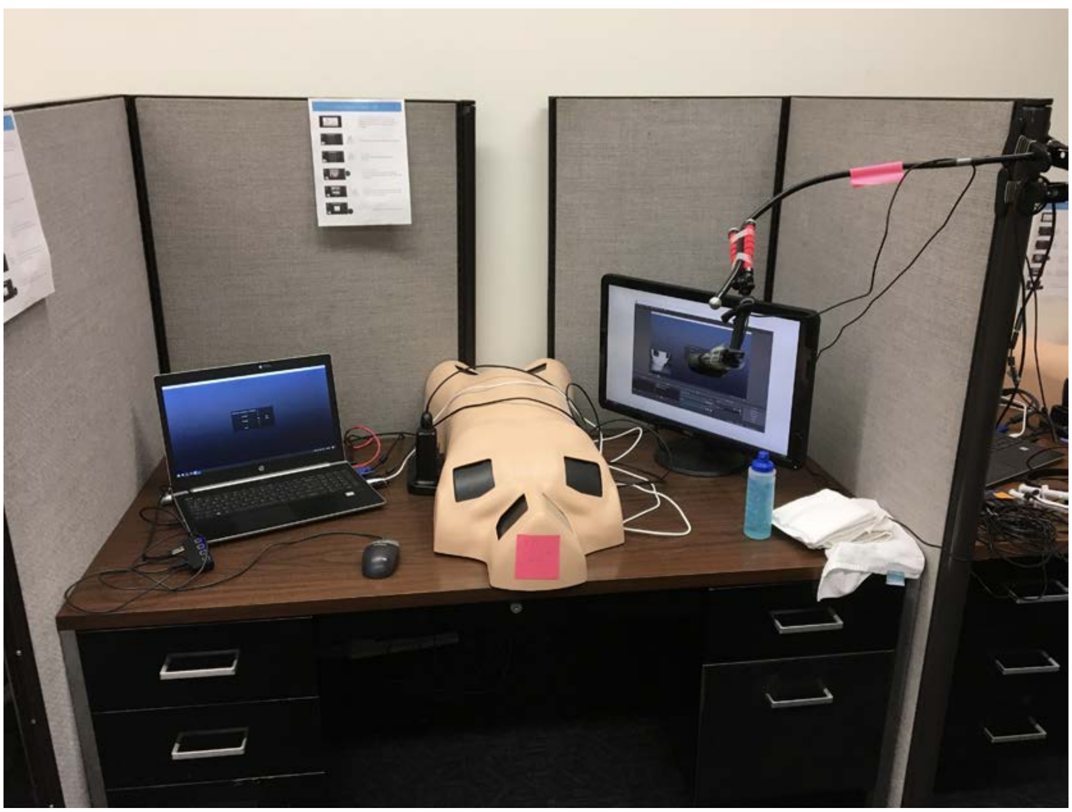


Figure 4. Right neck setup for internal jugular vein (IJV) cannulation task.

Figure 5. Close-up of needle used for the IJV tasks.

Study Design

We initially proposed a group study design where the participant groups would have represented different levels of experience (expert, medical doctor, resident, medical student). However, given recent results of a previous study, which had shown that categorization into groups of experience levels was subjective and could not be confirmed by collected data (Iseli, Savitsky, & Schenke, 2019), we decided instead to do a correlational analysis informed by participant background information that was closely related to the assessment questions and tasks, such as the number of scans performed by the participant on the same or similar topics.

The user model validation study was combined with the skill decay study, because we realized that the declarative knowledge and procedural skills trained and assessed in the user model study would be very similar to—if not the same as—the knowledge and skills assessed in the planned skill decay study. This combination of studies simplified study design and reduced the logistic efforts and planning required.

Study research questions and hypotheses were as follows:

- User model validation: To what extent do the model predictions of sonography understanding and skills correspond to external measures of sonography understanding and skill?
  - H1: Model predictions are correlated with other metrics (declarative or procedural).
  - H2: Experienced participants will perform better than inexperienced participants on most metrics. However, according to a prior study (Iseli et al., 2019), depending on task, the metrics scan duration and scan path length may not be reliable proficiency indicators.

- Skill decay: What is the decay of knowledge and skills of novices trained to criterion on the SonoSim® Ultrasound Training Solution over time?

With the combination of the two studies, we split the study into two consecutive parts, which are described below and illustrated in Figure 6. Part 1, the user model validation part, trained participants to basic ultrasound knowledge and skills criteria, collected pretraining data on a selected topic, trained participants in the same topic, and finally collected posttraining assessment data for the same topic. Data acquired in Part 1 were used to validate our models and to evaluate training effectiveness for the selected topic of training.

Part 2, the skill decay part, followed participants over a 12-week period. Participants were assessed four times on the same selected topic at 0 weeks, 2 weeks, 6 weeks, and 12 weeks. Data acquired in this part were used to evaluate potential skill decay.

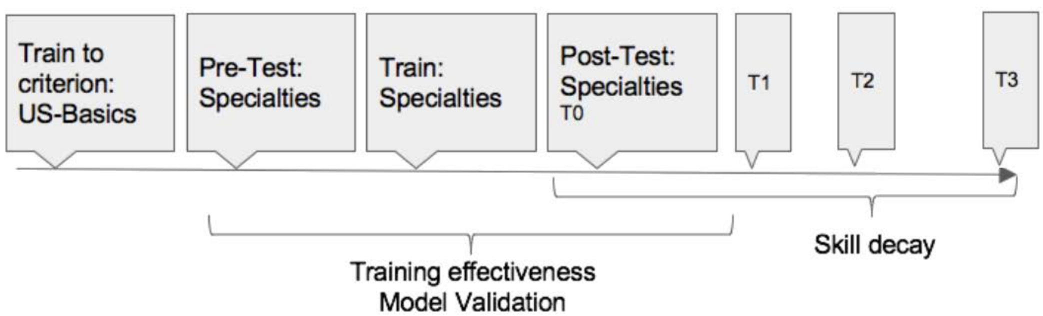


Figure 6. Study design for combination of user model validation and skill decay studies.

Procedure

Participant Recruitment

Participants were recruited from both a university hospital environment (medical students, medical residents, and practicing physicians) at UCLA and from the Uniformed Services University of the Health Sciences (USUHS) in Bethesda, Maryland. We sought to enroll a full-skill spectrum of participants, ranging from novice to expert ultrasonographers.

User Model Validation Part

At the start of the user model validation portion of the study, participants first logged into the SonoSim® Course Library, where they completed an introductory survey and then were tested on their declarative knowledge in fundamentals of ultrasound and—depending on their chosen study topic—in either second- and third-trimester ultrasound (OB topic) or ultrasound-guided procedures and internal jugular vein cannulation (IJV topic). During their onsite proctored session, participants watched a training simulation video on the use of the SonoSimulator® and practiced the related skills using the probe (and needle for the IJV tasks).

Note that for user model validation, no ultrasound-topic-related training was provided. After demonstrating the skills needed to utilize the simulator, participants were tested on their procedural (scanning) skills in their chosen topic and finally completed a postsurvey (see Appendix E). Training and assessment both utilized SonoSim® software and hardware.

Validation of the user model was partly based on validation measures proposed in Iseli and Jha (2016) as well as in Almond et al. (2015). The expected outcome of the user model validation study was a set of measures that discriminates among sonographers (users) of various experience levels and a model that incorporates these measures to yield predictions of competency on the modeled knowledge and skills. We used a two-step approach: (a) evaluation of measures and metrics, and (b) evaluation of model.

Step 1: Evaluation of measures and metrics. This step first evaluated descriptive statistics of our metrics and made sure they were within reasonable range. Then we compared highest performing to lowest performing metrics. For example, we expected that, compared to inexperienced participants, experienced participants showed better procedural performance (e.g., smaller distance to optimal quaternion view angles) and better knowledge (e.g., higher multiple-choice assessment scores).

Step 2: Evaluation of model. This step included the inspection of correlations between model inferences of performance and external measures and the evaluation of whether these inferences could discriminate, at the group level, between experts and novices. Experts should be predicted to have higher probabilities of understanding and performance across all the major constructs.

Skill Decay Part

The skill decay portion of the study was split into four sessions: The initial training and assessment of declarative as well as procedural skills at Week 0 (sd0), followed by three assessment-only retest sessions that evaluated skill decay at Week 2 (sd1), Week 6 (sd3), and Week 12 (sd6).[2]

In Week 0, all participants completed training courses in the online SonoSim® Course Library which provided declarative knowledge on second- and third-trimester ultrasound (OB topic) and—for participants partaking in the IJV study—on ultrasound-guided procedures and IJV cannulation. At the end of the courses, their declarative knowledge was assessed with multiple-choice mastery questions, and they scheduled their first proctored onsite visit where they were provided procedural skills training and assessment. During the onsite procedural skills training, videos guiding participants in the use of ultrasound for OB or IJV were shown, where they could pause the video to practice the skills being taught. On completion of the

[2]The digits in sd0, sd1, sd3, and sd6 initially represented the number of months when retests were planned to occur during the skill decay (sd) part of our study. Because of delays, the time intervals between retests had to be halved, resulting in Weeks 0, 2, 6, and 12 instead of Months 0, 1, 3, and 6 to accommodate the study time constraints.

onsite training, participants were assessed on tasks for their chosen study topic: fetal biometry tasks (OB) or IJV cannulation tasks (IJV).

In Weeks 2, 6, and 12, participants were retested on the same declarative knowledge mastery questions and on their procedural scanning skills for their chosen study topic. These weeks correspond with timepoints: sd0, sd1, sd3, and sd6. At the end of the final session, participants completed a postsurvey. See Appendix F for the specific online and onsite participant tasks for each session.

Data Preprocessing

Figure 7 illustrates the current design of our backend which extracts all the raw time series data from a Redis[3] database, extracts relevant measures, calculates the abovementioned metrics, scores those metrics according to scoring models described in the next section, and enters these scores into our Bayesian network model.

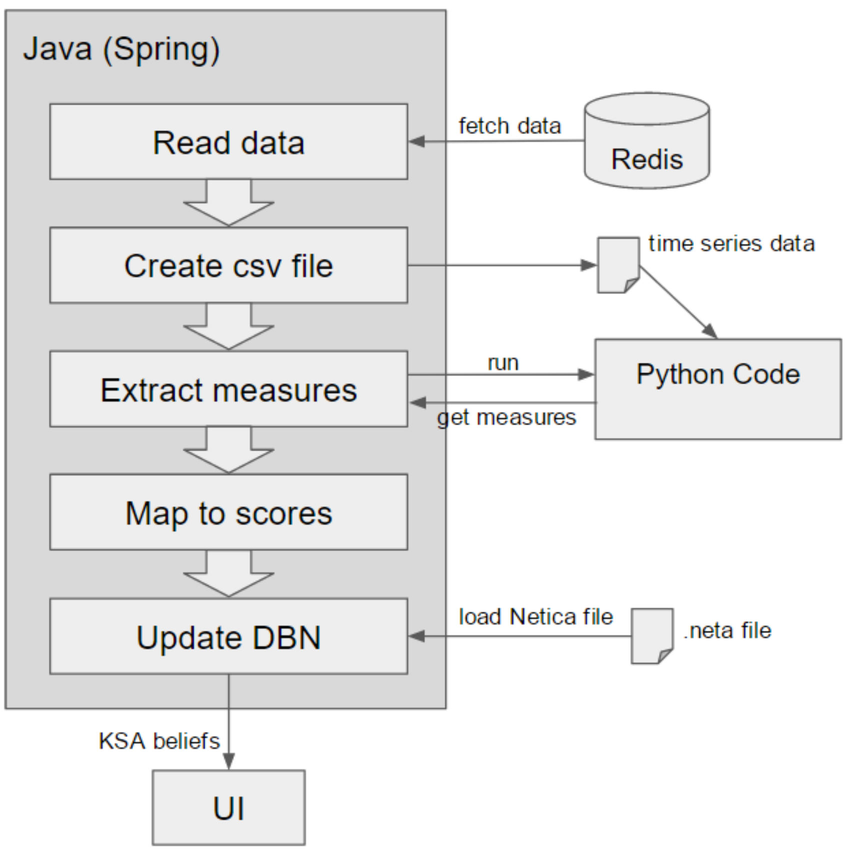


Figure 7. Model integration design.

[3]Redis is an open source (BSD licensed), in-memory data structure store, used as a database, cache and message broker (https://redis.io)

We extracted time series data, discrete event data, and biometry measurements stored in a Redis database in a compressed protocol buffer format. The time series data consisted of probe and needle positions and orientations in space at each timestamp, while the discrete event data consisted of software logs of user actions (such as probe calibration actions). As part of the data inspection, we omitted responses that fell beyond two standard deviations from the mean on path length. The biometry measurement data consisted of the measurement values when the user performed OB tasks. We used the discrete event data as a reference to indicate the beginning of a task (during the last needle calibration event) and the end of a task (in OB tasks, when the user freezes the image; in IJV tasks, when needle animation ends in the time series data). In some cases, after the calibration event, there was some lead time before the user started moving the probe. Hence, we truncated the start of this segment up to the point when the probe position changed. In addition, due to a glitch in the software, some time series data entries were recorded incorrectly and had to be interpolated.

Scoring of Procedural Metrics

We planned to use multimodal metrics distributions for the scoring functions ("Map to Scores" box in Figure 7) to distinguish between different levels of performance: for example, experienced, intermediate, and inexperienced performance. Figure 8 shows an example of the distribution of scan path length (in radians) for one specific task. For this example, a bimodal distribution can be seen with experts showing a distribution around smaller path lengths (μ_{expert} = 2.78 rad, σ_{expert} = 0.57 rad) and novices showing longer path lengths (μ_{novice} = 18.34 rad, σ_{novice} = 7.83 rad).

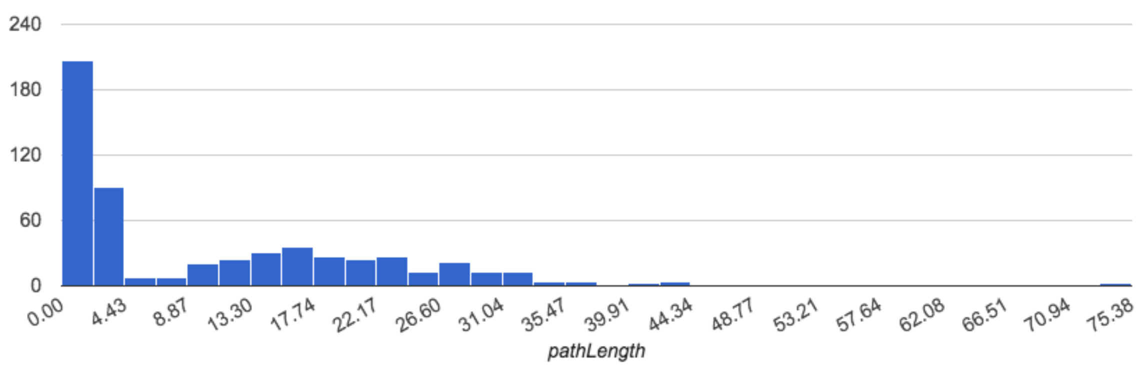


Figure 8. Histogram which shows the distribution of scan path length (in radians) for one specific task.

Because of a too small sample size, we had to simplify our planned approach and applied a somewhat heuristically determined two-group experience criterion (see Participants section) which we used to model expert performance metrics as normal distributions with means and standard deviations. Note that these models are task-dependent, meaning that depending on task, the optimal (expert) metric distributions will change. Table 2 below outlines the scoring function for each metric and indicates if the function is dependent on task or not. The scoring functions yield the probability of expertise (the “score”), calculated as the posterior probability using the specified Normal distribution parameters mean (mu) and standard deviation (sigma). This score is then entered as evidence into the corresponding BN node.

Table 2

Scoring Functions for Each Metric

Metric	BN variable	Scoring function	Task dependency
probe acceleration	ProbStead	N(mu,sigma), see table [N(var=mu)=1.0] (normal distribution)	yes
scan path length	PaLen	N(mu,sigma), see table	yes
scan duration	ScanDur	N(mu,sigma), see table	yes
probe displacement	ProbDisp	N(mu,sigma), see table	yes, IJV only
cannulation score	CanSuccess	CanSuccess = (cannulation score + 1) / 4	no, IJV only
needle insertions	NdIns	1:0.95; 2:0.65; 3:0.35; >3:0.05 [Number of insertions: score]	no, IJV only
scan view error	ViewDiff	N(mu,sigma), mu=0, sigma = pi/180	no, OB only
AC score	ACpctErr	= AC score	no, OB only
BPD score	BPDpctErr	= BPD score	no, OB only
FL score	FLpctErr	= FL score	no, OB only
HC score	HCpctErr	= HC score	no, OB only
gestational age score	GesAgeErr	= gestational age score	no, OB only

Note. N stands for Normal distribution.

RESULTS AND DISCUSSION

In this section, we first discuss results for the user model validation part of the study and then present the results of the skill decay part.

User Model Validation (MV)

This subsection presents the results for the user model validation part of this study. It is subdivided into the topics OB and IJV. Descriptive statistics will be presented first, followed by correlational analysis, and experienced vs. inexperienced user performance comparisons.

Descriptive Statistics (MV: OB&IJV)

Participants (MV: OB&IJV). Across both OB and IJV topics, 86 individuals participated in the user model validation study with nine participants taking part in both OB and IJV tasks. This included eight physicians (9.3%), 13 residents (15.1%), 61 medical students (70.9%), and four participants that specified "other" (4.7%). Participants came from a range of specialties, specifically 16 in obstetrics and gynecology (18.6%), seven in emergency medicine (8.1%), and seven dispersed across six other specialties—including two undecided (8.1%) and 56 (65.1%) with specialty listed as "not applicable." Twenty-seven worked in clinical settings (31.4%), seven in nonclinical settings (8.1%), and 18 worked in both (20.9%). Their ages ranged from 20 to 65, with the majority within the 23- to 29-year-old age range. Sixty percent of participants were female and 37% were male. The majority of participants had less than one hour of OB ultrasound experience and less than one hour of ultrasound-guided IJV experience. See further details below. Overall, more than one half of participants (52%) had less than six hours of ultrasound experience when starting the study. In participants' self-assessment, 54% stated that their skills before the study for image acquisition were very poor or below average; for image interpretation, 57% stated that their skills were very poor or below average; and for ultrasound-guided procedures, 69% considered themselves to be very poor or below average. Six participants overlapped between the topics for user model validation part of the study and nine participants overlapped between topics for the skill decay part.

Correlations (MV: OB&IJV). The correlations between the mastery question scores, OB experience, and Bayes net procedural skill scores are shown in Table 3. Performance on fundamentals of ultrasound mastery questions (*fundamentals score)* was significantly corelated with *USGP score*, $r = .38$, $p < .05$; *IJV score*, $r = .58$, $p < .001$; *OB score*, $r = .51$, $p < .001$; number of times having performed OB tasks on patients *(Number performed OB)*, $r = .20$, $p < .05$; and hours of performance on OB tasks *(hours performed OB),* $r = .23$, $p < .05$. Other significant correlations were: *USGP score* with *IJV score*, $r = .58$, $p < .001$; *OB score* with *Number performed OB,* $r = .53$, $p < .001$ and with *hours performed OB,* $r = .60$, $p < .001$. These correlations make sense in that the mastery test of declarative knowledge, Fundamentals of Ultrasonography, is expected to be strongly related to other declarative knowledge mastery tests on similar, yet more specific topics (Items 2–4 in the table) and to prior experience (Items 5 and 6 in the table).

There were also significant correlations between BN procedural scores and their corresponding declarative topic knowledge metrics: *BN probability of OB skills* with *OB score, r* =.47, *p* < .01; *BN probability of IJV skills* with *IJV score*, *r* =.54, *p* < .01. These correlations are also expected as topic-related declarative knowledge and procedural skills should be highly related.

Similarly, there were significant positive correlations between *Number performed OB* and *hours performed OB, r =.89, p* < .001. Finally, there was also a significant positive correlation between *Number performed OB* and *BN probability of OB skills*, *r* =.29, *p* < .05, and *hours performed OB* and *BN probability of OB skills, r = .37, p* < .01. The correlations are also what one might expect as they are related to prior-experience-related metrics.

Table 3

Correlation Matrix of Declarative Knowledge and Participant Background Information—OB

Variable	1	2	3	4	5	6	7
1. fundamentals score	—						
2. USGP score	.38*	—					
3. IJV score	.58***	.58***	—				
4. OB score	.51***	-.09	.41	—			
5. Number performed OB	.20*	.25	.32	.53***	—		
6. hours performed OB	.23*	.26	.31	.60***	.89***	—	
7. BN probability of IJV skills	.19	.20	.54**	.51	.25	.14	—
8. BN probability of OB skills	.16	.42	.48	.47**	.29*	.37**	.58

**p* < .05 (two-tailed). ***p* < .01 (two-tailed). ****p* < .001 (two-tailed).

User Model Validation Results (MV: OB)

Participants (MV: OB). Fifty-eight individuals participated including four physicians (6.9%), 12 residents (2.7%), and 38 medical students (65.5%). The most common specialty included 16 participants (27.8%) in Obstetrics and gynecology, and six participants (1.3%) dispersed across four other specialties (including two undecided). Twenty (34.5%) worked in clinical settings, seven (12.1%) in nonclinical settings, and 15 (25.9%) worked in both. Their ages ranged from 22 to 65, with the majority within the 24- to 29-year-old age range. Sixty percent of participants were female and 38% were male. The majority of participants had less than six hours of OB ultrasound experience and less than one hour of ultrasound-guided IJV experience. Overall, more than two thirds of participants (67.2%) had less than 21 hours of ultrasound experience when starting the study. In participants' self-assessment, 51.7% stated that their skills before the study for image acquisition were very poor or below average; for image interpretation 51.7% stated that their skills were very poor or below average, and for ultrasound-guided procedures, 65.5% considered themselves to be very poor or below average.

Correlations (MV: OB). Table 4 shows the Pearson correlations between the OB procedural metrics. For H1, there are many significant correlations that we were expecting to see. There was a significant positive correlation between *scan duration* and *probe acceleration*, $r = .62$, $p < .001$, between *scan duration* and *scan path length*, $r = .63$, $p < .001$, and between *scan duration* and *BPD score*, $r = .27$, $p < .05$. *Probe acceleration* was positively correlated with *scan path length*, $r = 1.00$, $p < .001$.

As expected, *scan view error* was negatively correlated with all the biometric measurement scores (*score_**), however, only *HC score* and *gestational age score* were significantly correlated, $r = -.29$, $p < .05$ and $r = -.26$, $p < .05$, respectively. There was also a significant positive correlation between the *HC score* and *AC score*, $r = .78$, $p < .001$, where both scores reflect measures of the circumference of an anatomical object, head or abdomen, respectively.

There were also significant positive correlations between *gestational age score* and *BPD score,* $r = .34$, $p < .05$; *HC score,* $r = .41$, $p < .01$; *AC score*, $r = .33$, $p < .05$; and *FL score*, $r = .85$, $p < .001$. For H1, these correlations were expected to be positive.

For procedural skills BN scores, *BN probability of OB skills* was significantly negatively correlated with *scan duration,* $r = -.34$, $p < .01$, and with *probe acceleration*, $r = -.48$, $p < .001$; *scan path length*, $r = -.48$, $p < .001$; and *scan view error,* $r = -.53$, $p < .001$. There were significant positive correlations between *BN probability of OB skills* and *HC score*, $r = .40$, $p < .01$; *FL score*, $r = .45$, $p < .001$; and with *gestational age score*, $r = .57$, $p < .001$. These significant correlations also match what we would expect.

Similarly, for OB mastery score, there were significant correlations between declarative and procedural metrics: *OB score* correlated with *probe acceleration,* $r = -.37$, $p < .05$; *scan path length*, $r = -.37$, $p < .05$; *FL score*, $r = .40$, $p < .05$; *gestational age score,* $r = .47$, $p < .01$; and *BN probability of OB skills*, $r = .47$, $p < .01$. These significant correlations match what we expected to see: Better performance on declarative knowledge means better performance on procedural skills.

The correlation of $r = 1.00$, $p < .001$ between *probe acceleration* and *scan path length* indicates that these two metrics are basically measuring the same dimension.

Table 4

Correlation Matrix of Procedural Skills (Scan Duration, …, Gestational Age Score) With Bayesian Network Model Prediction (BN Probability of OB Skills) and Declarative Knowledge (OB Score)

Variable	1	2	3	4	5	6	7	8	9	10
1. scan duration	—									
2. probe acceleration	.62***	—								
3. scan path length	.63***	1.00***	—							
4. scan view error	-.17	-.08	-.07	—						
5. BPD score	.27*	.01	.02	-.09	—					
6. HC score	-.11	-.12	-.13	-.29*	.1	—				
7. AC score	-.02	-.02	-.03	-.15	.06	.78***	—			
8. FL score	.06	-.12	-.11	-.2	.19	.12	.15	—		
9. gestational age score	.00	-.17	-.17	-.26*	.34*	.41**	.33*	.85***	—	
10. BN probability of OB skills	-.34**	-.48***	-.48***	-.53***	.14	.40**	.19	.45***	.57***	—
11. OB score	-.06	-.37*	-.37*	-.31	.04	.21	.14	.40*	.47**	.47**

*$p < .05$ (two-tailed). **$p < .01$ (two-tailed). ***$p < .001$ (two-tailed).

Experienced vs. inexperienced (MV: OB)

Declarative skills. No comparisons between experienced and inexperienced participants can be made, because experienced participants were not required to take the declarative mastery tests.

Procedural skills. Differences in procedural performance on OB tasks between experienced and inexperienced participants are shown in Table 5. Overall, experienced participants performed better than inexperienced participants. Statistically significant differences, at the $p < .001$ to .05 level, were found for the following metrics: *scan view error*, all biometric measurement scores except *HC score*, *BN probability of OB skills* score, and *OB score* on the mastery test.

For our hypothesis H2, the distance from optimal view *scan view error*, as hypothesized, showed a statistically significant difference between experienced and inexperienced participants, $p = .026$. However, *scan path length* ($p = .41$), *probe acceleration* ($p = .39$), and *HC score* ($p = .984$) were not statistically significantly different. Since *HC score* and *BPD score* are both originating from the same acquired image (thalamic plane view), the fact that performance on *HC score* was not a discriminator between experienced and inexperienced participants is unexpected. One possible explanation could be that when matching the fetal skull with a caliper ellipse, caliper placement errors on one axis of the ellipse might cancel out measuring errors on the other axis. That fact that *AC score*, which also uses a caliper ellipse for measurement, was able to distinguish between experienced and inexperienced participants, could be due to a higher difficulty of acquiring the optimal viewing plane for abdominal circumference measurement. This could be due to insufficient instruction or less visible ultrasound image landmarks.

Table 5

Metrics Comparison Between Experienced and Inexperienced Performance—OB

	Experienced (*n* = 13)		Inexperienced (*n* = 46)		Total (*n* = 59)		
	M (SD)	Range	*M (SD)*	Range	*M (SD)*	Range	*p* value
scan duration	40575.1 (16893.8)	13407.9- 78517.5	40778.0 (1674.4)	15852.9- 91707.7	40733.3 (16627.8)	13407.9- 91707.7	.942
probe acceleration	.2(.1)	.1 - .4	.2 (.1)	.0 - .4	.2 (.1)	.0 - .4	.390
scan path length	1.4 (5.5)	4.7 - 25.3	11.4 (5.6)	3.0 - 26.4	11.2 (5.6)	3.0 – 26.4	.410
scan view error	.6 (1.3)	.2 – 1.0	.8 (.5)	.2 - 3.8	.7 (.5)	.2 – 3.8	.026
HC score	1.0 (.0)	.9 - 1.0	1.0 (.1)	.1 - 1.0	1.0 (.0)	.8 - 1.0	.984
AC score	1.0 (.0)	.9 - 1.0	.9 (.1)	.2 - 1.0	.9 (.1)	.2 - 1.0	.03
FL score	1.0 (.1)	.8 - 1.0	.8 (.2)	.3 - 1.0	.9 (.2)	.3 - 1.0	< .001
BPD score	1.0 (.0)	.9 - 1.0	.9 (.1)	05 - 1.0	.9 (.1)	.5 - 1.0	< .001
gestational age score	1.0 (.0)	.1 - 1.0	.9 (.1)	.8 - 1.0	.9 (.0)	.8 - 1.0	< .001
BN probability of OB skills	.6 (.0)	.6 - .7	.6 (.0)	.5 - .7	.6 (.1)	.5- .7	.003
OB score	67.8 (8.0)	59.0 - 83.0	48.8 (13.5)	17.0 - 83.0	52.6 (14.7)	17.0 - 83.0	< .001

Graphs with kernel density estimates were plotted based on task type—Task Type 1: tasks involved the measurement of head circumference and biparietal diameter (Figure 9); Task Type 2: tasks involved the measurement of abdominal circumference (Figure 10); and Task Type 3: tasks involved the measurement of femur length (Figure 11). As observed for all task types, scores for experienced participants were largely higher than that of inexperienced participants, and this was further corroborated by small *p* values between experienced and inexperienced participants—*HC score* ($p = .127$), *AC score* ($p = .017$), *FL score* ($p < .001$), and *BPD score* ($p < .001$). The following figures show kernel density plots of the metrics by task type.

In general, the experienced participants performed better on the biometric measurement scores (except for head circumference, *HC score*), exhibiting a smaller standard deviation than the inexperienced participants.

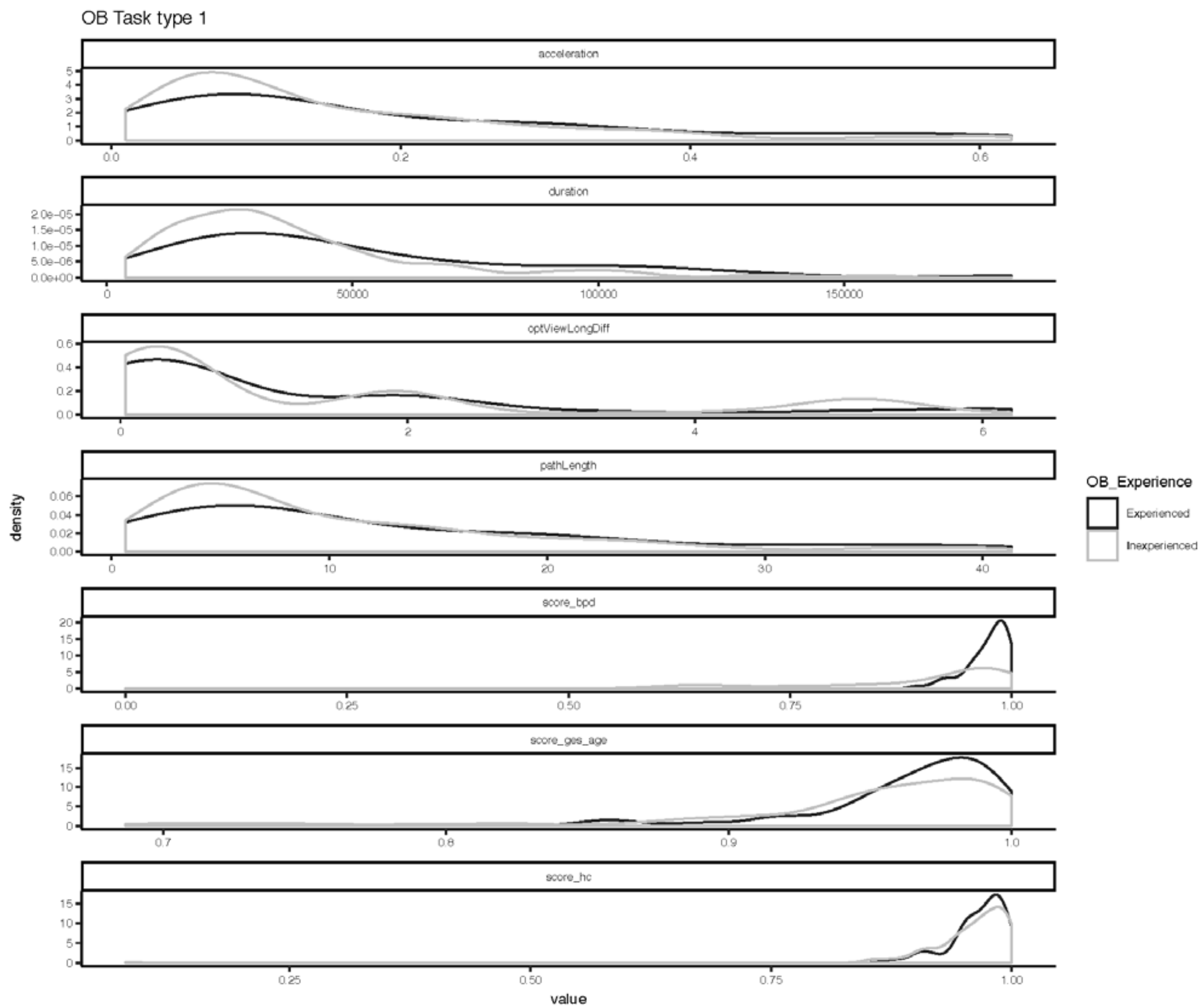


Figure 9. Kernel density plots for OB Task Type 1: Measurement of biparietal diameter and head circumference in the thalamic viewing plane.

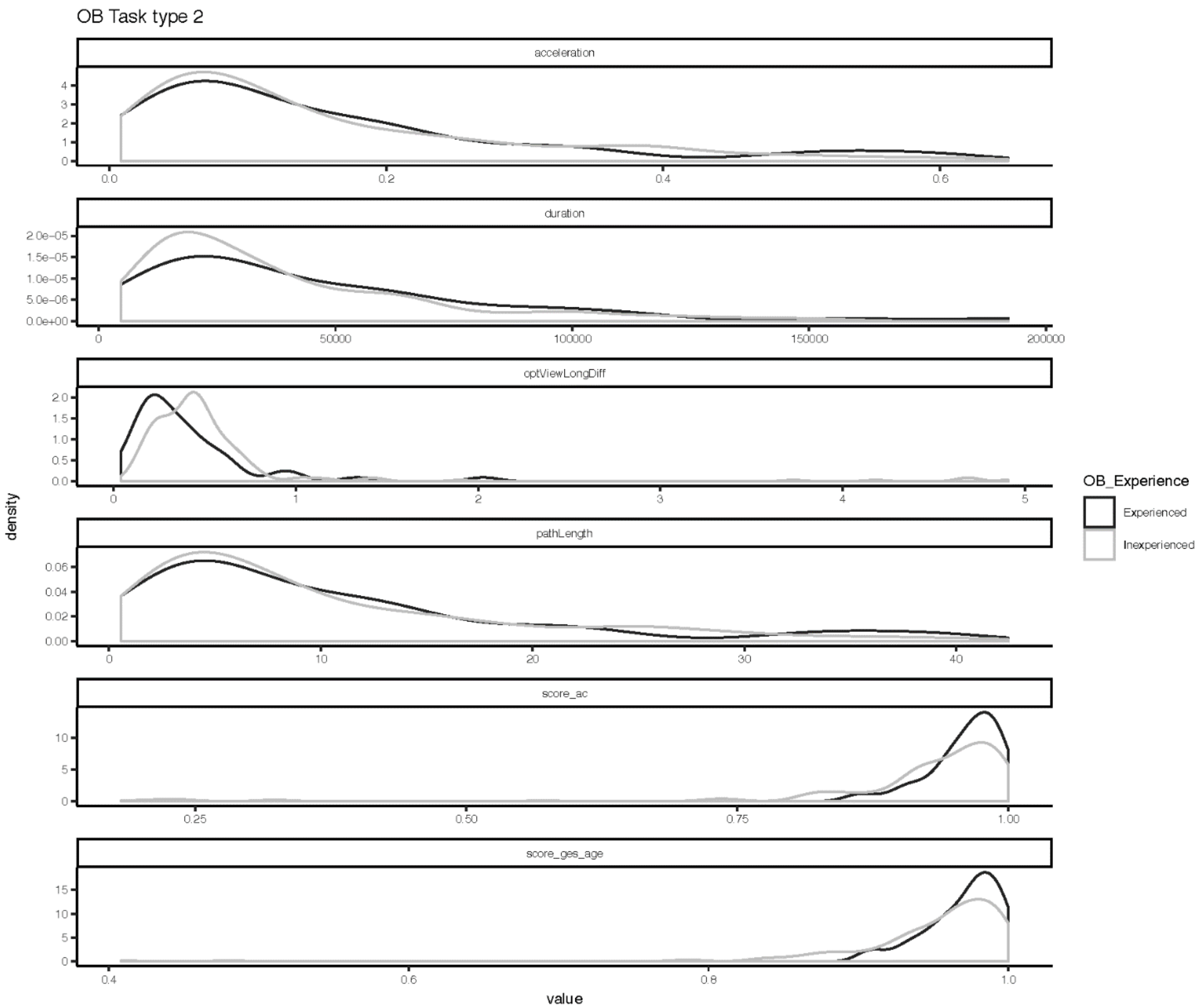


Figure 10. Kernel density plots for OB Task Type 2: Measurement of abdominal circumference.

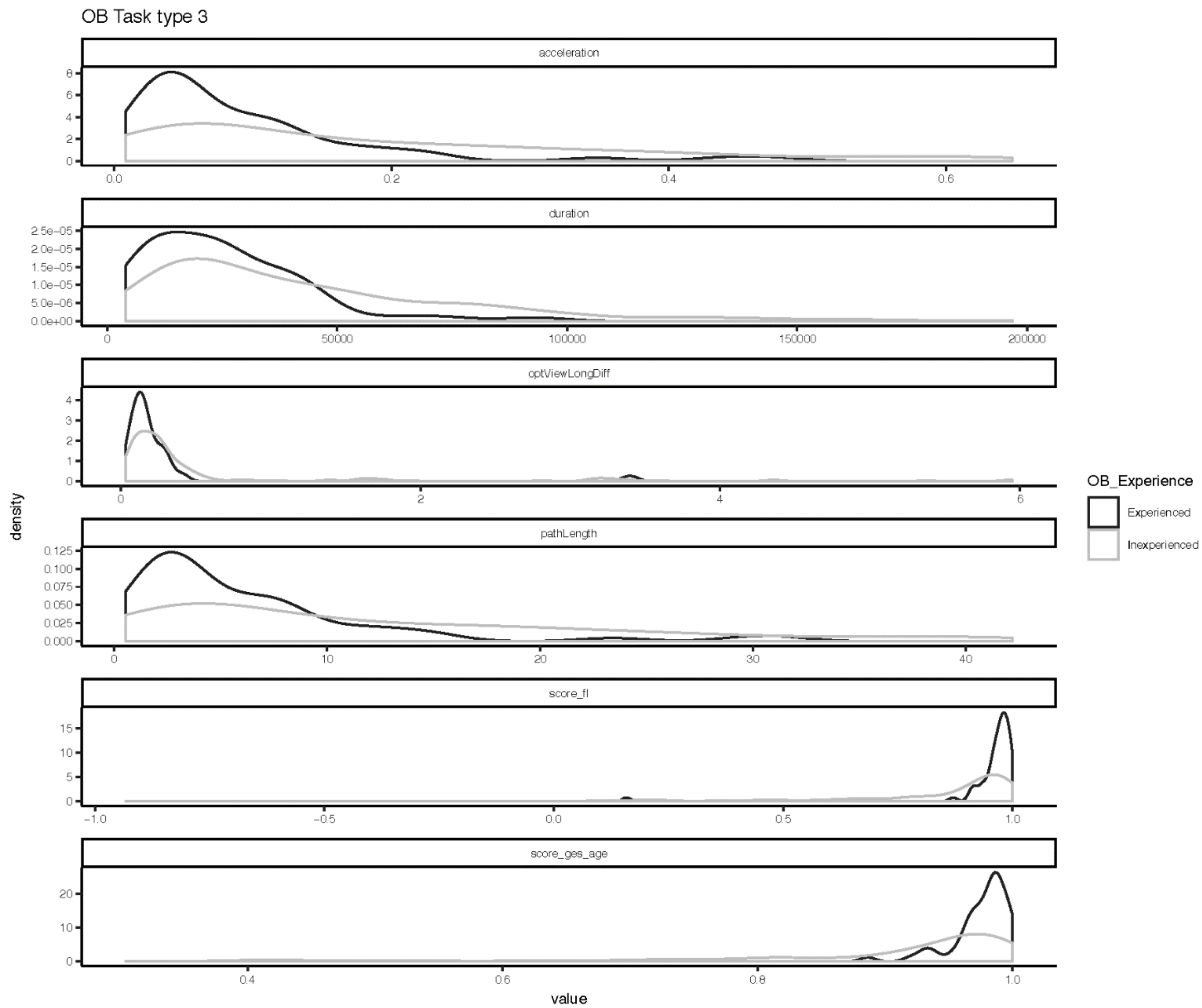


Figure 11. Kernel density plots for OB Task Type 3: Measurement of femur length.

User Model Validation Results (MV: IJV)

Participants (MV: IJV). Thirty-seven individuals participated including five physicians (13.5%), two residents (5.4%), and 30 medical students (81.1%). The most common specialty was emergency medicine with six participants (16.2%), and five participants (13.5%) were dispersed across four other specialties (including two undecided). Eight participants worked in clinical settings (21.6%), zero in nonclinical settings, and six worked in both settings (16.2%). Their ages ranged from 20 to 53, with the majority within the 22- to 26-year-old age range. Fifty-four percent of participants were female and 43% were male. The majority of participants had less than one hour of OB ultrasound experience and less than one hour of ultrasound-guided IJV experience. See further details below. Overall, more than three fourths of participants (75.7%), had less than six hours of ultrasound experience when starting the study. In participants' self-assessment, 56.8% stated that their skills before the study for image acquisition were very poor or below average; for image interpretation 62.1% stated that their skills were very poor or below average; and for ultrasound-guided procedures, 67.6% considered themselves to be very poor or below average.

Correlations (MV: IJV). The correlations for the procedural skill metrics for the IJV task are shown in Table 6. There were significant positive correlations between *scan duration* and *probe acceleration*, $r = .56$, $p = .01$; *scan duration* and *scan path length,* $r = .68$, $p = .001$; and *probe acceleration* and *scan path length,* $r = .91$, $p = .001$. In addition, there were also positive correlations between *probe displacement* and *scan duration*, $r = .68$, $p = .001$; between *probe displacement* and *probe acceleration,* $r = .68$, $p = .001$; and between *probe displacement* and scan path length, $r = .84$, $p = .001$. As expected, there were significant negative correlations between number of *needle insertions* and *cannulation score*, $r = -.48$, $p = .01$; *needle insertions* and *IJV score* on the mastery test, $r = -.90$, $p = .001$; and *needle insertions* and *IJV score* on the mastery test, $r = -.54$, $p = .01$. Cannulation success score (*cannulation score*) was also positively correlated with *BN probability of IJV skills,* $r = .59$, $p = .001$ and *IJV score* on the mastery test, $r = .64$, $p = .001$. Finally, there was a significant positive correlation between *BN probability of IJV skills* and *IJV score* on the mastery test, $r = .56$, $p = .01$. For H1: These significant correlations also provide additional validity support for how procedural skills relate to each other, to the *BN probability of IJV skills*, and to the declarative knowledge score (IJV score on the mastery test, *IJV score*).

Similarly, as stated in the OB section, the correlation of $r = .91$, $p < .001$ between *probe acceleration* and *scan path length* indicates that these two metrics are basically measuring similar dimensions.

Table 6

Correlation Matrix of IJV Metrics

	1	2	3	4	5	6	7
1. scan duration	—						
2. probe acceleration	.56**	—					
3. scan path length	.68***	.91***	—				
4. needle insertions	.05	-.03	-.02	—			
5. cannulation score	.21	-.11	-.07	-.48**	—		
6. probe displacement	.68***	.68***	.84***	.09	-.03	—	
7.ijv_proc_BN	-.14	-.22	-.17	-.90***	.59***	-.023	—
8. IJV score	.32	.11	.24	-.54**	.64***	.20	.56**

*p < .05 (two-tailed). **p < .01 (two-tailed). ***p < .001 (two-tailed).

Experienced vs. inexperienced (MV: IJV)

Declarative skills. No comparisons between experienced and inexperienced participants can be made, because experienced participants were not required to take the declarative mastery tests.

Procedural skills. The descriptive statistics for experienced versus inexperienced participants for the IJV tasks is shown in Table 7. Regarding our hypothesis H2: Experienced participants scored significantly higher than inexperienced participants only for *cannulation score*, $p < .05$. The *cannulation score* is the most important metric in terms of completing the cannulation task without errors which would bring harm to the patient (e.g., puncturing the vein through, or cannulating an artery by mistake). The metrics of *scan duration*, *probe acceleration*, *scan path length*, *needle insertions*, and *probe displacement* did not show significant differences between experienced and inexperienced participants. We would have expected that *probe displacement* would be a good discriminator between experienced and inexperienced participants. Possible reasons why could be, amongst others (a) inexperienced participants took greater risks and cannulated without moving the probe position a lot and without orienting it in short or long axes; and (b) the employed simulator version was not realistic enough so that experienced participants could not use their routine scanning skills.

Table 7

Metrics Comparison Between Experienced and Inexperienced Performance - IJV

	Experienced (*n* = 4)		Inexperienced (*n* = 33)		Total (*n* = 37)		
	M (SD)	Range	*M (SD)*	Range	*M (SD)*	Range	*p* value
scan duration	46601.3(20428.6)	28134.8- 65826.1	55821.1(2288.1)	17623.7- 106929.5	54824.4(2255.7)	17623.7- 106929.5	.523
probe acceleration	.2 (.0)	.1 - 02	.3 (.3)	.0 – 1.5	.2 (.2)	.0 – 1.5	.689
scan path length	8.9 (2.9)	6.5 - 13.1	1.3 (5.8)	2.1 – 31.2	1.1 (5.5)	2.1 - 31.2	.724
needle insertions	2.1 (1.0)	1.1 – 3.1	3.2 (2.6)	1.0 – 1.2	3.1 (2.5)	1.0 - 1.2	.642
cannulation score	.9 (.2)	.1 - .9	.5 (.2)	.1 - 1.0	.6 (.2)	.1 - 1.0	.016
probe displacement	.2 (.1)	.1 – .4	.2 (.1)	.1 – .5	.2 (.1)	.1 – .5	.463
BN probability of IJV skills	.6 (.0)	.6 - .6	.6 (.0)	.5 - .6	.6 (.0)	.5 - .6	.103

Figure 12 shows kernel density plots for one of the four IJV cannulation tasks. It confirms our findings from the correlation analysis and illustrates the larger standard deviation for inexperienced participants.

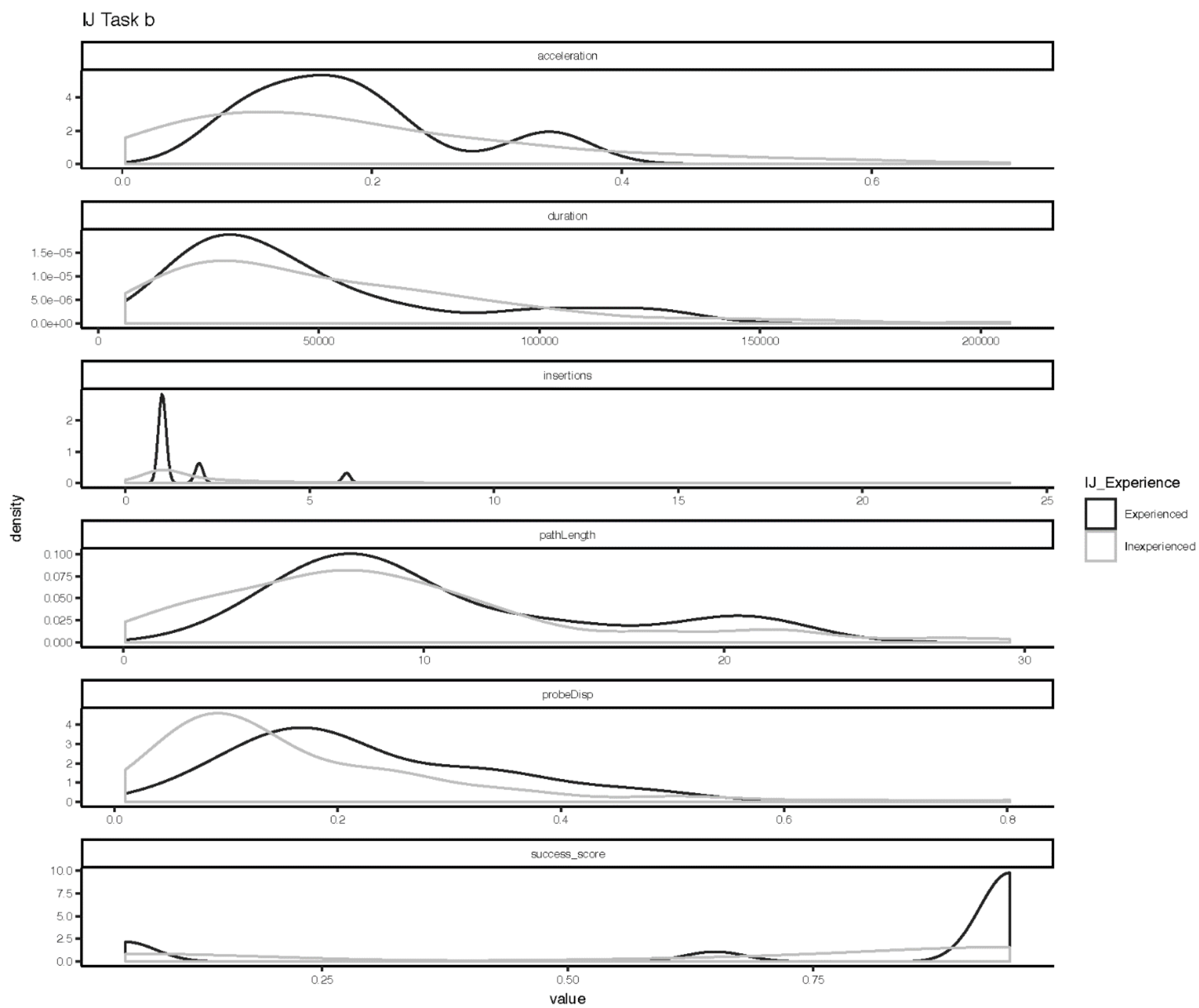


Figure 12. Kernel density estimates plot for an IJV task.

Skill Decay Part (SD)

This subsection presents the results for the skill decay part of this study. It is subdivided into the topics OB and IJV. Descriptive statistics will be presented first, followed by an analysis of declarative and procedural skill changes.

Changes in Declarative Skills (SD: OB&IJV)

The descriptive statistics for the mastery question scores for both topics can be found in Table 8. The means for the skill decay part (sd0) of the study are generally higher compared to the user model validation (mv) means, which is expected since at the mv timepoint participants had not yet received any instruction. In general, during the skill decay part (sd0 through sd6), while the performance decays a little, it is not statistically significant (p = .145),[4] and seems to level off which could be indicative of skill retention rather than decay. It is possible that there might be a testing effect (Karpicke & Aue, 2015) to some degree, where repeated testing actually has a practice effect of refreshing and retaining skills.

Table 8

Descriptive Statistics Mastery Questions for Model Validation and Skill Decay for OB and IJV Topics

Topic	Phase	*M*	*SD*	Min	Max	*n*
fund	mv	63.09	15.60	28	100	101
fund	sd0	78.88	13.66	21	100	80
OB	mv	52.27	14.36	17	83	49
OB	sd0	77.81	11.12	43	93	26
OB	sd1	71.12	1.91	52	87	16
OB	sd3	68.92	12.12	50	85	12
OB	sd6	69.50	9.56	52	80	12
usgp	mv	73.10	1.73	36	87	40
usgp	sd0	83.29	11.98	56	100	21
usgp	sd1	82.00	8.22	64	95	12
usgp	sd3	81.38	6.36	72	95	13
usgp	sd6	81.55	7.72	72	92	11
IJ	mv	61.18	15.99	17	90	39
IJ	sd0	8.24	15.17	48	98	21
IJ	sd1	76.85	7.87	64	90	13
IJ	sd3	73.54	11.59	50	90	13
IJ	sd6	74.45	1.63	57	88	11

Note. fund: fundamentals of ultrasonography; OB: 2nd and 3rd Trimester Fetal Biometry; usgp: Ultrasound-guided Procedures; IJ: Intrajugular Vein Cannulation.

The means for the mastery scores are graphed in Figure 13. The greatest decrease is from the sd0 timepoint to the sd1 timepoint for OB and IJ topics, while the greatest decrease for the ultrasound guided procedures (*usgp*) mastery topic was from the sd1 to sd3 timepoints. However, the decreases were not significant. Participants' *IJV score* and *USGP score* mean

[4] p value created with a fixed effect multilevel model with random intercepts, with time coded such that sd0 was 0, sd1 was 2, sd3 was 6, and sd6 was 12.

mastery scores started the highest and ended higher than the *OB score*. The average score for the fundamentals course started the lowest.

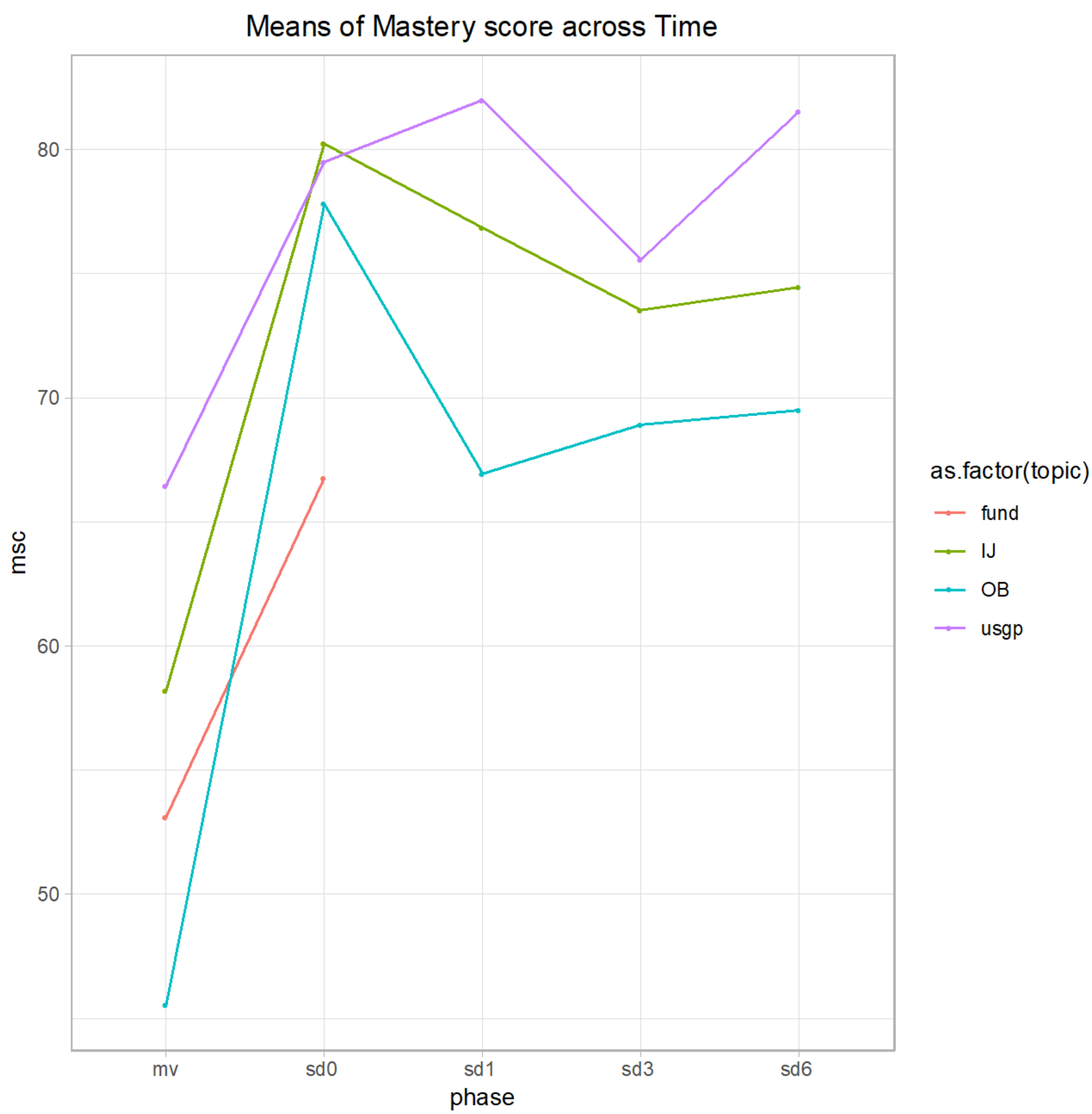


Figure 13. Change in declarative skills mean scores (msc) as a function of timepoints for ultrasound fundamentals knowledge (fund), ultrasound-guide procedures in general (usgp), internal jugular vein cannulation (IJ), and 2nd and 3rd trimester fetal biometry (OB).

Skill Decay Results (SD: OB)

Participants (SD: OB). Nineteen individuals participated including zero physicians, one resident (5.3%), and 17 medical students (89.5%). Two participants indicated a specialty (including one undecided), three worked in clinical settings (15.8%), two in nonclinical settings (1.5%), and six worked in both (31.6%). Their ages ranged from 22 to 35, with the majority within the 23- to 26-year-old age range. Sixty-eight percent of participants were female and 32% were male. The majority of participants had less than one hour of OB ultrasound experience and less than one hour of ultrasound-guided IJV experience. Overall, almost three fourths of participants (73.7%) had less than six hours of ultrasound experience when starting the study. In participants' self-assessment, 57.9% stated that their skills before the study for image acquisition were very poor or below average; for image interpretation, 63.2% stated that their skills were very poor or below average; and for ultrasound-guided procedures, 68.4% considered themselves to be very poor or below average.

To evaluate the skill decay of participants' procedural knowledge skills, we used data from participants who completed at least one session at either of the following timepoints: sd0, sd1, sd3, and sd6. We omitted participants who we characterized as experts from the skill decay analysis because their skill decay patterns were likely to differ from those of nonexperts.

Changes in procedural skills (SD: OB). We analyzed procedural data from two participants on the following metrics: scan duration, path length, steadiness, deviation from optimal (thalamic) view plane, relative errors for biometrics, and overall gestational age error. We ran random intercept multilevel models with time points nested within participant to understand the extent to which participants' slopes changed over time. In Table 9, the coefficient for time represents the slope parameter. The random effect variances are also depicted in the table. σ_2 is the within-participant residual variance, τ_{00} is the between-participant variance (the variation between the participant's intercept and the average intercept), and the ICC represents the intraclass correlation (the amount of variance in the outcome that the participant ID explains). Across the models for OB, it appears as though there is no statistically significant effect of time on participants' outcome suggesting that skill decay is not occurring. One exception is for *scan path length*, where the coefficient for time is -.25 suggesting that over time, path length actually decreases. However, because of the small sample, these results are to be taken with a grain of salt.

Table 9

Growth Curve Multilevel Model for OB

	scan duration (msec)		scan path length (rad)		scan view error (rad)		probe acceleration (rad/sec^2)		BN probability of OB skills	
Predictor	Estimates	*SE*	Estimates	*SE*	Estimates	*SE*	Estimates	*SE*	Estimates	*SE*
(Intercept)	61348.79**	21407.15	1.56***	1.35	.73***	.02	.16***	.02	.65***	.00
Time	-1545.21	3302.14	-.25 *	.12	-.00	.00	-.00 *	.00	-.00	.00
Random Effects										
σ^2	13502862812.90		16.56		.01		.00		.00	
τ_{00}	791664773.96$_{user}$		26.05$_{part.}$		.00 $_{part.}$		.01 part.		.00 part.	
ICC	.06 $_{user}$		.61 $_{part.}$		.13 $_{part.}$		.61 part.		.51 part.	
Observations	64		64		64		64		64	
Marginal R^2 / Conditional R^2	.003 / .059		.028 / .622		.000 / .127		.028 / .620		.000 / .512	

	gestational age score		BPD score		HC score		AC score		FL score	
Predictor	Estimates	*SE*	Estimates	*SE*	Estimates	*SE*	Estimates	*SE*	Estimates	*SE*
(Intercept)	.95***	.00	.90***	.02	.96***	.00	.93***	.02	.93***	.01
timenum	.00	.00	.00	.00	.00	.00	.00	.00	-.00	.00
Random Effects										
σ^2	.00		.01		.00		.01		.00	
τ_{00}	.00 part.		.00 part.		.00 part.		.00 part.		.00 part.	
ICC	.44 part.		.12 part.		.32 part.		.13 part.		.60 part.	
Observations	64		64		64		64		64	
Marginal R^2 / Conditional R^2	.012 / .444		.003 / .125		.018 / .335		.023 / .148		.002 / .602	

Note. The subscript "$_{part.}$" means relative to participant ID.
**p* < .05. ***p* < .01. ****p* < .001.

Figure 14 shows participants' performance on procedural knowledge across time for OB, averaged over all participants. The x-axis represents the number of days since sd0. We averaged participants' scores across all tasks. Note that we omitted scores that were higher than the 95th percentile to account for outliers in the data. We then applied a smoothing spline to produce a mean trend line across all participants to get a better understanding of trends in procedural knowledge over time.

Overall, performance across time on many of the variables reveal a flat line, suggesting limited evidence for skill decay. Inspection of the plots of growth curves, and of the results from the slope coefficient from the multilevel model, are in congruence.

To analyze participants' performance on procedural knowledge across time for OB in more detail, Figure 15 and Figure 16 break out performance for each participant and show LOESS (locally estimated scatterplot smoothing) curves averaged over all tasks. We filtered out participants that indicated in their end-of-study survey that they had received training or instruction on the topic. The x-axis represents the number of days since sd0. Individual participants are indicated by the color and the number on the curves and are the same across the plots. The *probe acceleration* metric is not shown because is perfectly correlated with *scan path length*. We can see individual differences. For example, between the first (sd0) and last (sd6) skill decay timepoints, Participant 11 shows no skill decay or increase—the curves have a small variance and are relatively flat. Participant 8 shows a high curve variance with an overall skill increase between sd0 and sd6. Participant 1 shows a high curve variance with an overall skill decrease between sd0 and sd6.

These findings could form an argument that there exist individual differences between performers. Additionally, there are still many other factors that will influence performance as mentioned in the introduction.

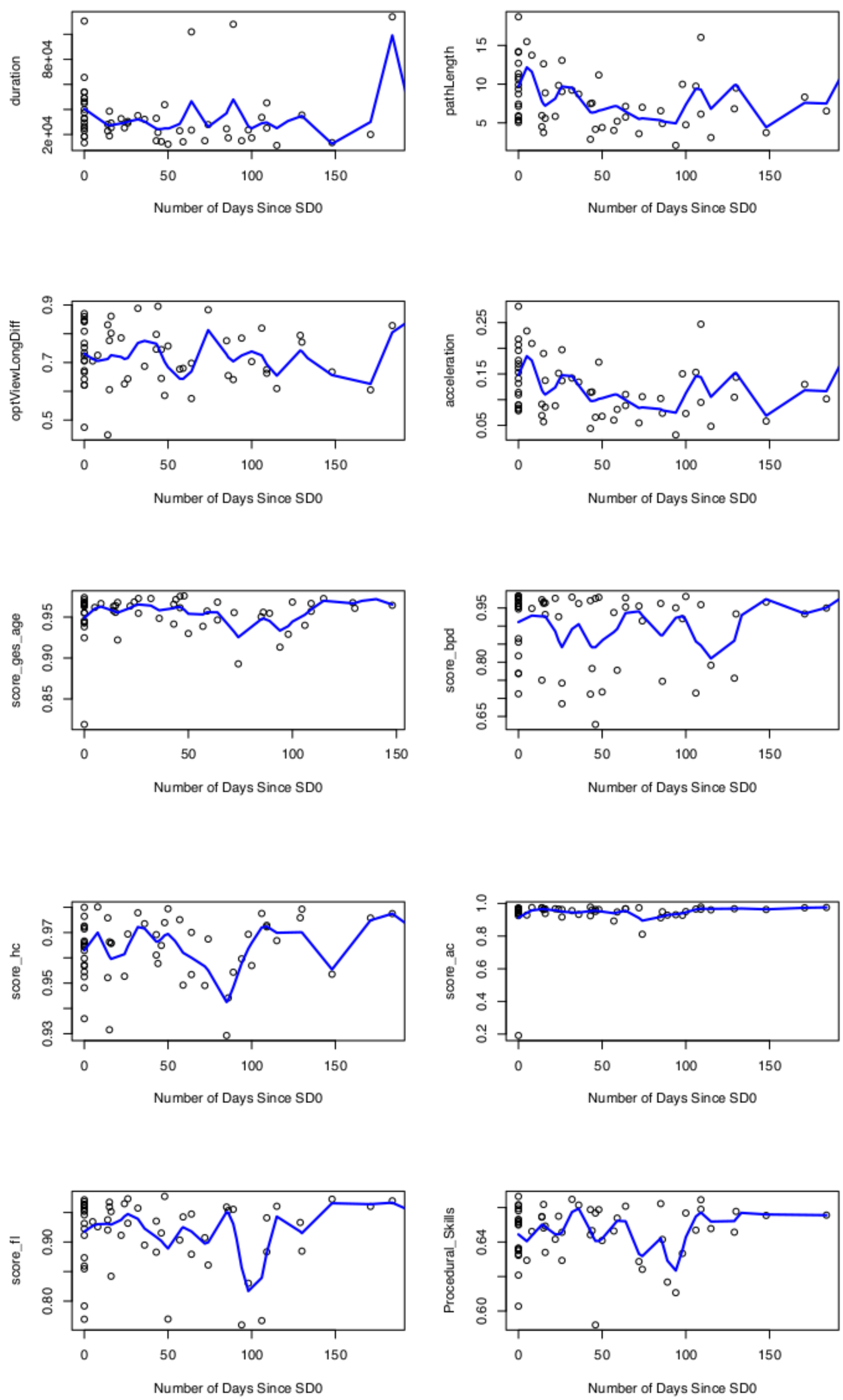


Figure 14. Skill decay plots for the OB tasks, averaged over all participants.

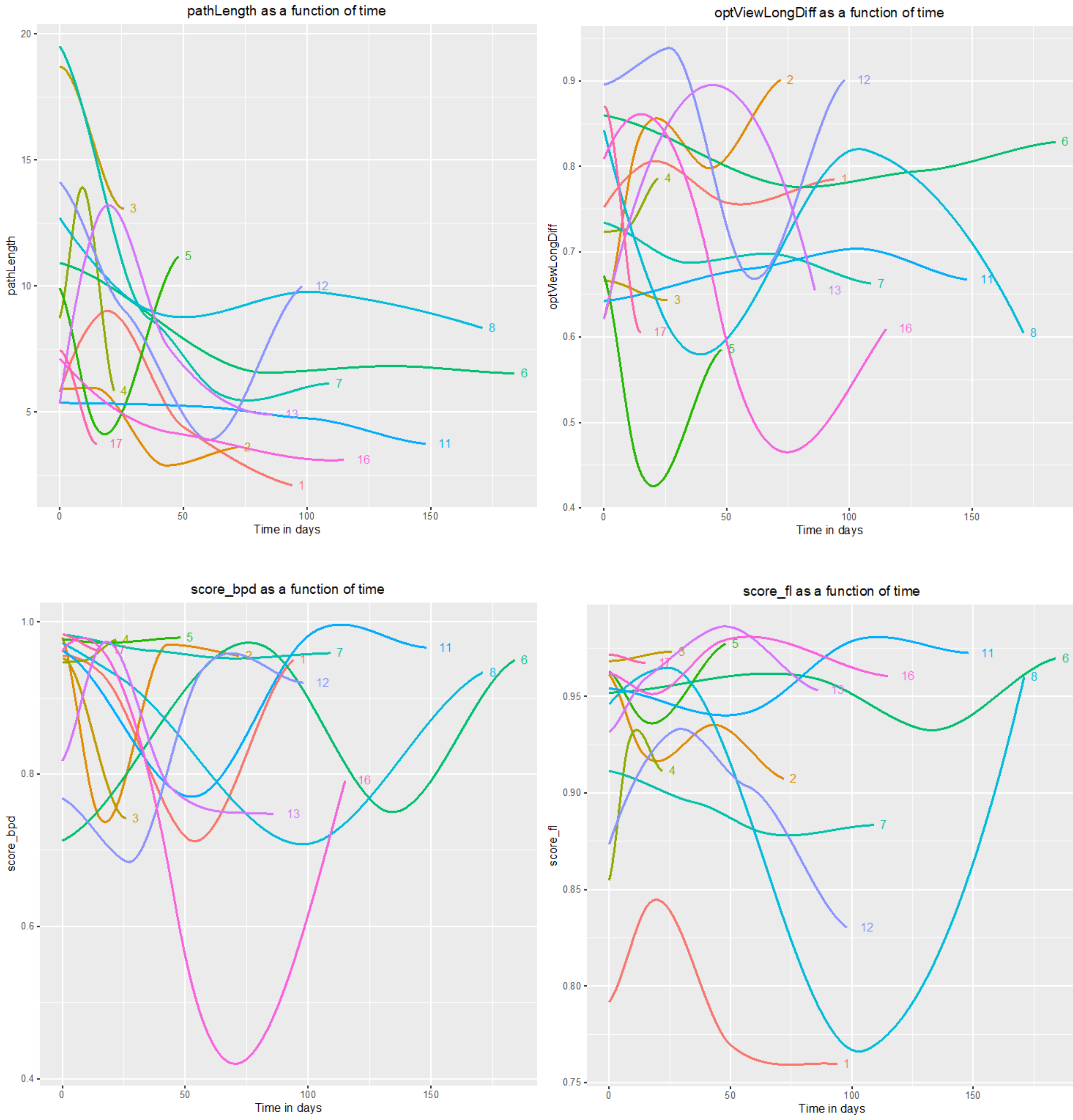


Figure 15. Skill decay plots for the OB tasks, by participant.

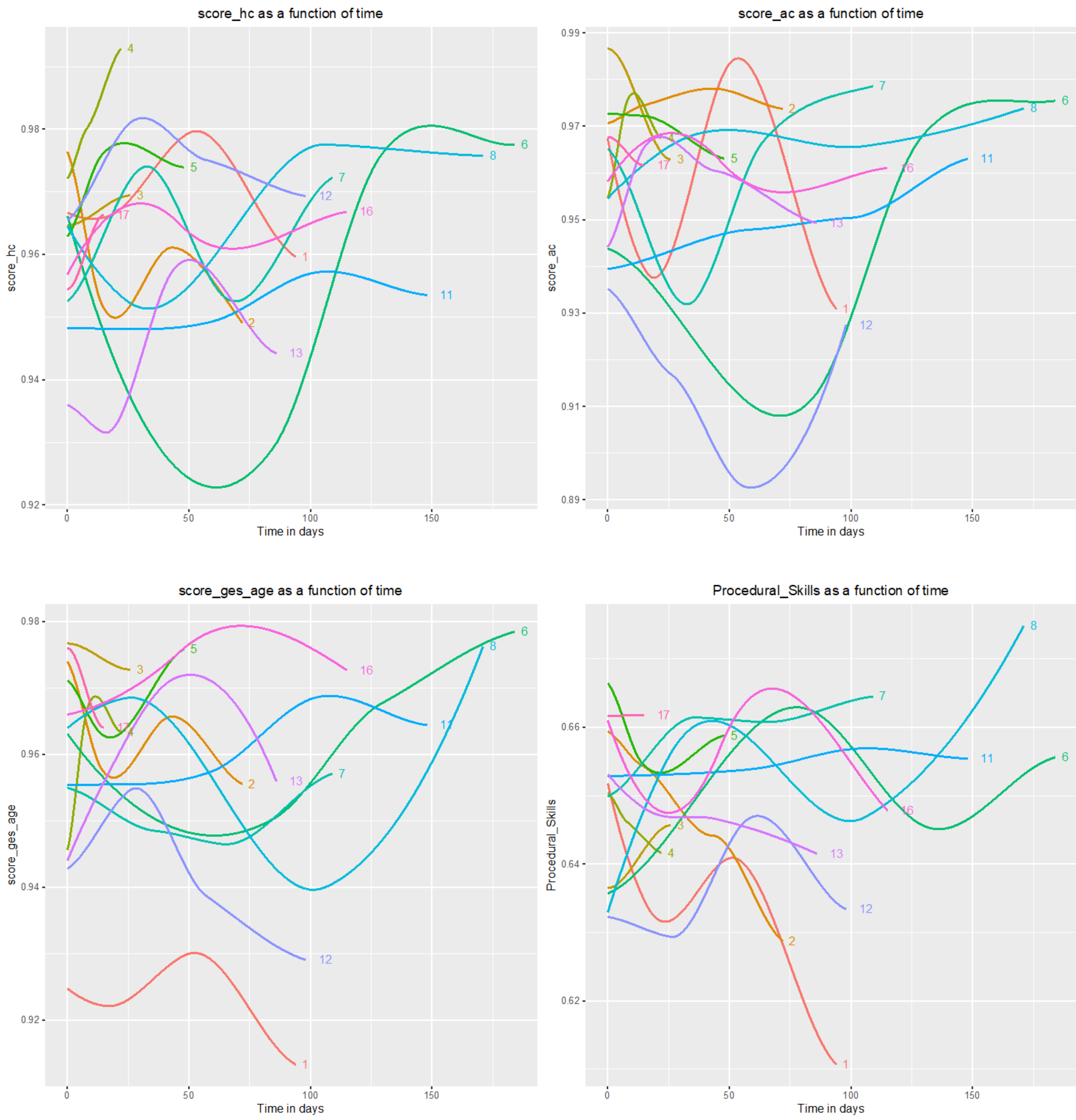


Figure 16. Skill decay plots for the OB tasks, by participant. (continued from Figure 15)

Skill Decay Results (SD: IJV)

Participants (SD: IJV). Nineteen individuals participated including two residents (1.5%) and 17 medical students (89.5%). There were just four participants that indicated a specialty, the most common being emergency medicine with two participants (1.5%). Two worked in clinical settings (1.5%), one in nonclinical settings (5.3%), and three worked in both (15.8%). Their ages ranged from 22 to 35, with the majority within the 23- to 26-year-old age range. Fifty-three percent of participants were female and 42% were male. The majority of participants had less than one hour of OB ultrasound experience and less than one hour of ultrasound-guided IJV experience. Overall, most participants (89.5%) had less than six hours of ultrasound experience when starting the study. In participants' self-assessment, 68.4% stated that their skills before the study for image acquisition were very poor or below average; for image interpretation, 68.4% stated that their skills were very poor or below average; and for ultrasound-guided procedures, 73.7% considered themselves to be very poor or below average.

Changes in procedural skills (SD: IJV). We ran random intercept multilevel models with time points nested within participant to understand the extent to which participants' slopes changed over time. In Table 10, the coefficient for time represents the slope parameter. The random effect variances are also depicted in the table. σ^2 is the within-participant residual variance, τ_{00} is the between-participant variance (the variation between the participant's intercept and the average intercept), and the ICC represents the intraclass correlation (the amount of variance in the outcome that the participant ID explains). Across some outcomes for IJ, it appears as though there is no statistically significant effect of time on participants' outcome suggesting that skill decay is not occurring. There is a statistically significant and meaningful difference in the effect of time on scan duration and path length suggesting that as participants have had more practice, the scan duration and path length of their probe gets smaller.

Table 10

Growth Curve Multilevel Model for IJV

	scan duration (msec)		scan path length (rad)		probe displacement (mm)		needle insertions		probe acceleration (rad/sec^2)		BN probability of IJV skills	
Predictor	Estimates	*SE*	Estimates	*SE*	Estimates	*SE*	Estimates	*SE*	Estimates	*SE*	Estimates	*SE*
(Intercept)	55101.58***	6402.65	1.06***	.75	.18***	.02	2.05***	.18	.19***	.02	.57***	.00
Time	-1267.95***	354.23	-.25**	.08	-.01**	.00	-.03	.02	-.01**	.00	.00	.00
Random Effects												
σ^2	135793644.39		6.68		.00		.49		.00		.00	
τ_{00}	$723367265.84_{part.}$		$6.50_{part.}$		$.00_{part.}$		$.29_{part.}$		$.00_{part.}$		$.00_{part.}$	
ICC	$.84_{part.}$		$.49_{part.}$		$.55_{part.}$		$.37_{part.}$		$.32_{part.}$		$.56_{part.}$	
Observations	59		59		59		59		59		59	
Marginal R^2 / Conditional R^2	.038 / .848		.087 / .538		.064 / .583		.018 / .384		.101 / .387		.029 / .574	

Note. The subscript "$_{part.}$" means relative to participant ID.
**p* < .05. ***p* < .01. ****p* < .001.

Figure 17 shows participants' performance on procedural knowledge across time for IJV. The x-axis represents the number of days since sd0. We averaged participants' scores across all tasks. Note that we omitted scores that were higher than the 95th percentile to account for outliers in the data. We then applied a smoothing spline to produce a mean trend line across all participants to get a better understanding of trends in procedural knowledge over time.

Overall, similar to OB, performance across time on many of the variables reveal a flat line, suggesting limited evidence for skill decay. The plots of growth curves and of the results from the slope coefficient from the multilevel model are in congruence.

In order to analyze participants' performance on procedural knowledge across time for OB in more detail, Figure 18 and Figure 19 break out performance for each participant and show LOESS (locally estimated scatterplot smoothing) curves averaged over all tasks. We filtered out participants that indicated in their end-of-study survey that they had received training or instruction on the topic. The x-axis represents the number of days since sd0. Individual participants are indicated by the color and the number on the curves and are the same across the plots.

We can see individual differences. For example, Participant 6 has a small skill curve variance and Participant 8 has higher variance; both show a skill increase between sd0 and sd6. Participants 12 and 13 show a steady, high success score overall. However, the number of needle insertions increase from between one and two insertions to between two and three insertions. This could indicate that the sensitivity of the needle is too high or that the needle insertion depth threshold in our software is too low. In addition, we see that the BN probability of IJV skills values only range between .54 and .6, which is a very small range, considering that the range between 0 and 1 is available.

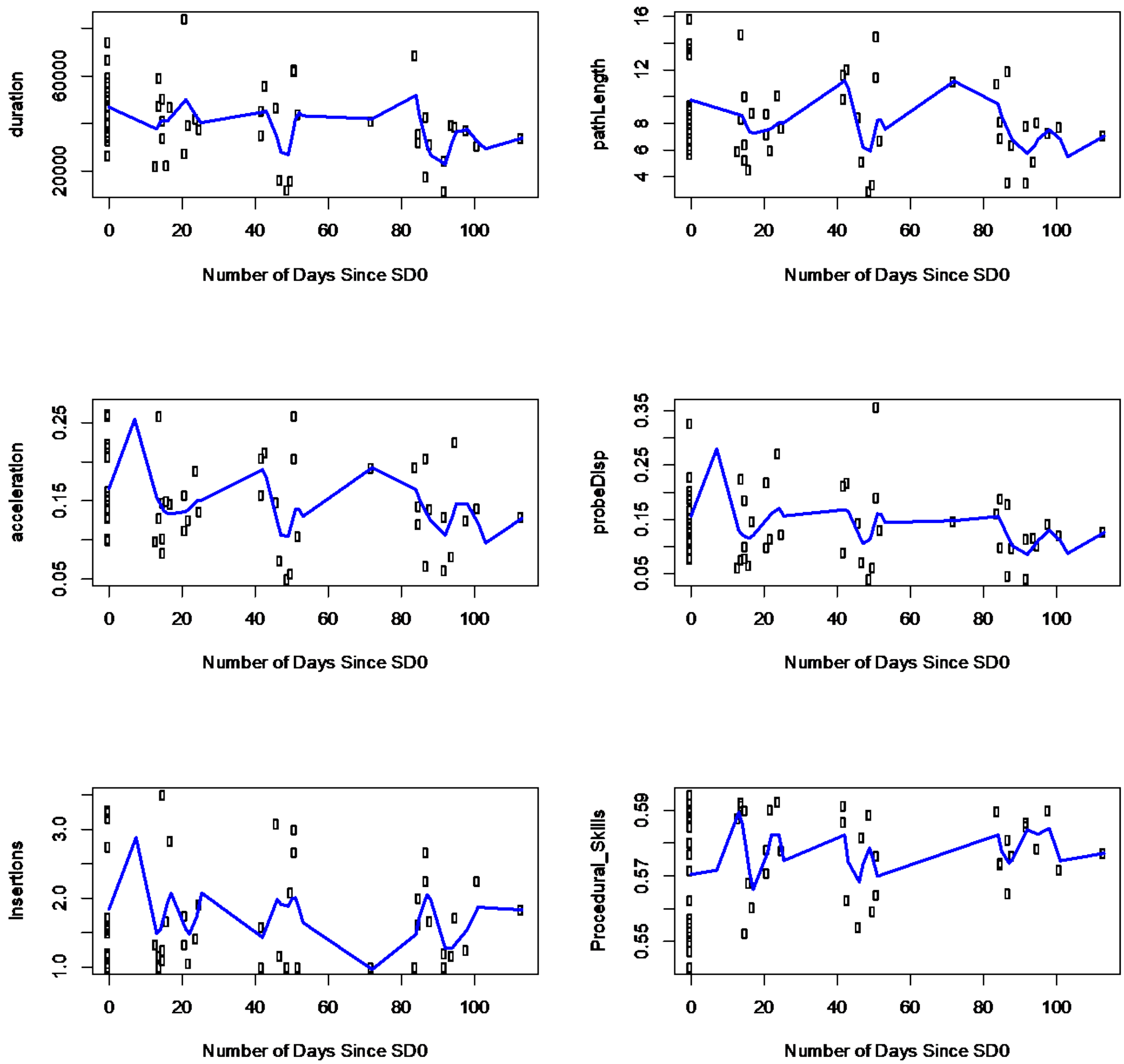


Figure 17. Skill decay plots for the IJV tasks, averaged over all participants.

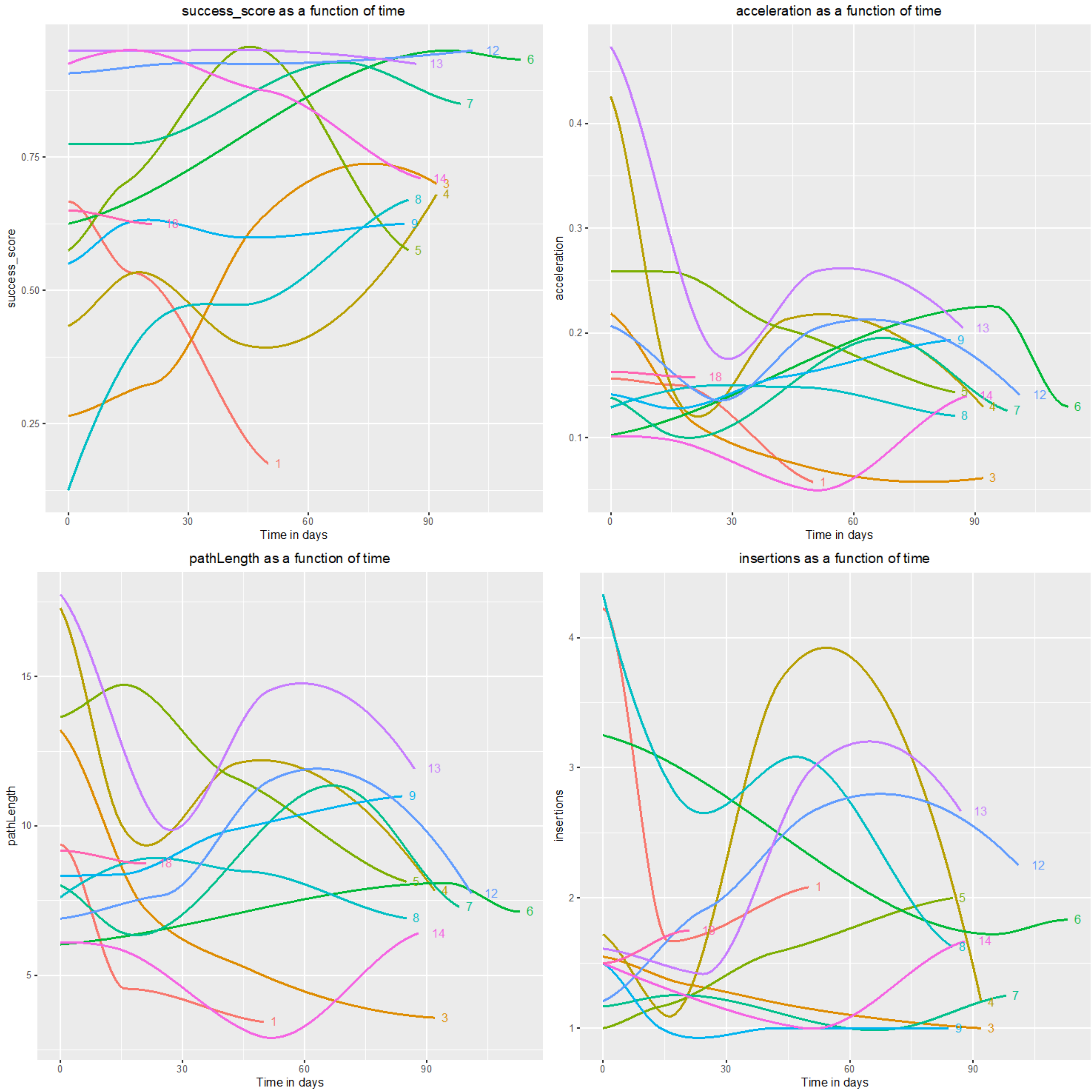


Figure 18. Skill decay plots for the IJV tasks, by participant.

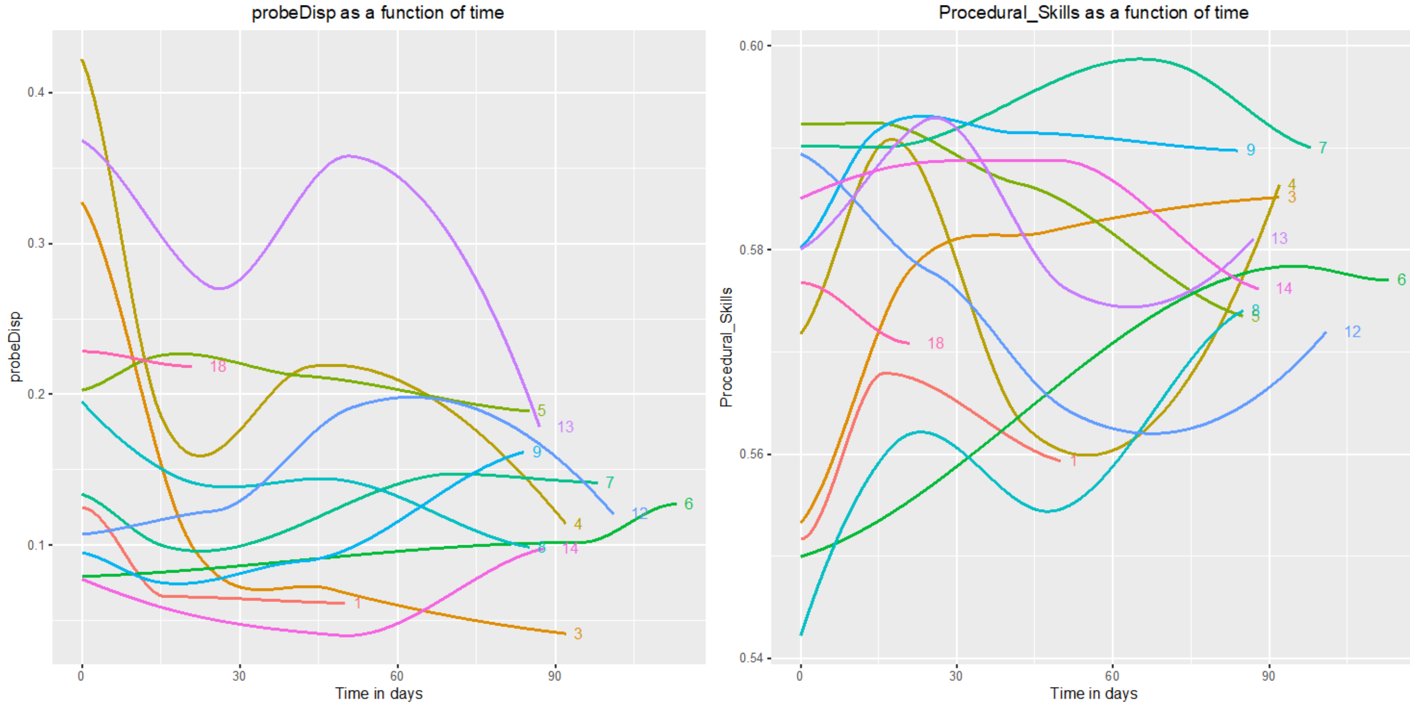


Figure 19. Skill decay plots for the IJV tasks, by participant. (continued from Figure 18)

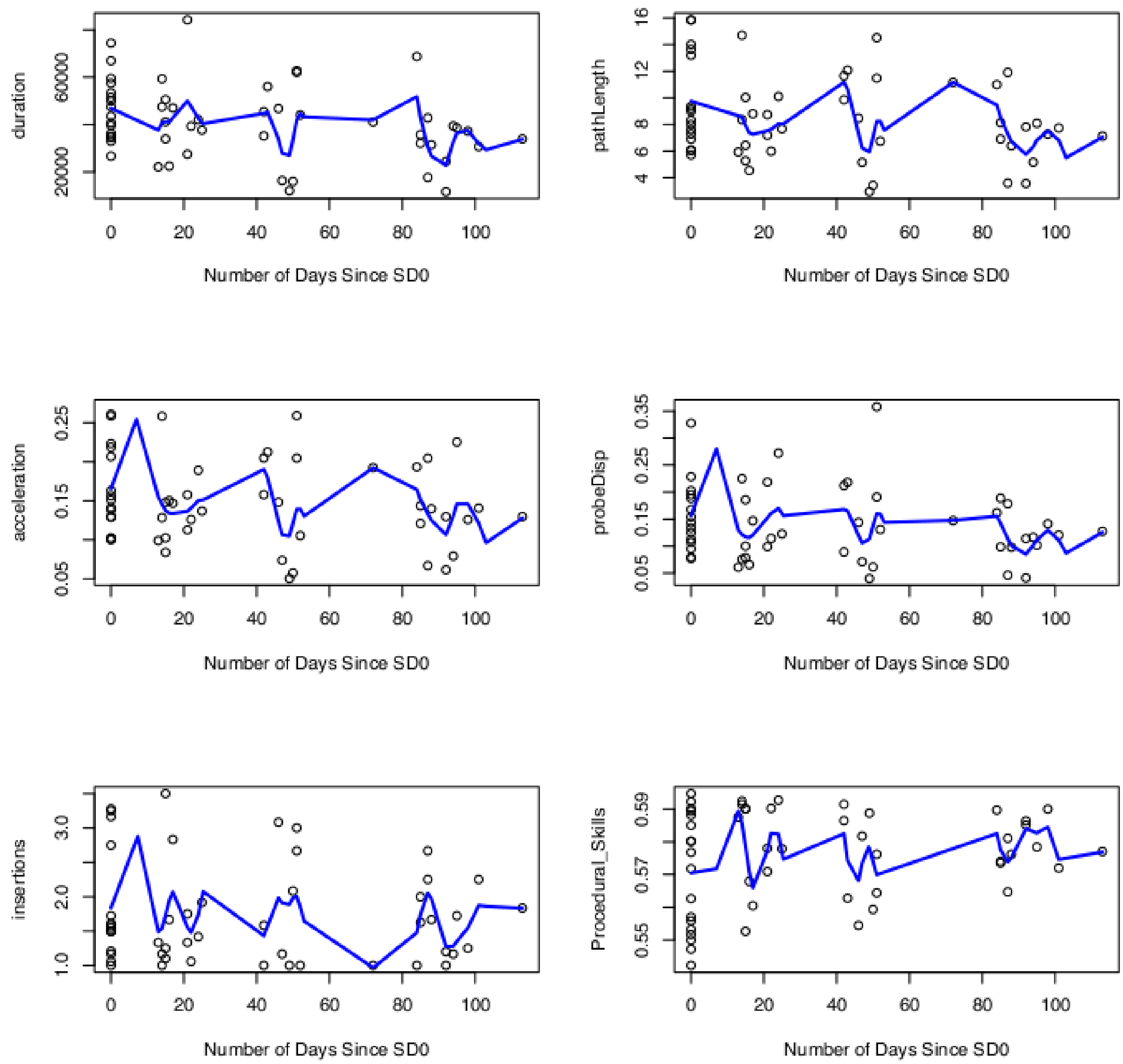


Figure 20. Skill decay plots for the IJV tasks.

LIMITATIONS

Limitations of this study were the small sample size for the skill decay part of the study and some technical issues (e.g., needle and probe calibration software, missing data, which were addressed and fixed, but still delayed and complicated data collection to some degree). For future skill decay studies, it would beneficial to be able to enroll whole classrooms or cohorts to achieve the needed sample size.

In addition to above issues, the cognitive load on proctors was high, since they had to be aware of many simultaneous components: is the video recording, did the participant complete the various online assessments and questionnaires, what is the next assessment module on the participant's timeline, are the needles charged, etc. For future assessments, an integrated software solution or learning management system that manages all the steps and prerequisites would be very helpful because it would increase the standardization of data collection.

SUMMARY AND OUTLOOK

We assessed a total of 86 participants, 58 on the OB topic and 37 on the IJV topic with nine participants taking part in both topics. We collected data related to the following metric types: background information (i.e., scanning experience), declarative knowledge (performance on SonoSim® course mastery questions), procedural skills (scanning performance in the SonoSimulator®), and model inferences (overall inferred procedural skills from Bayesian network).

In the model validation part of this study, we found that scanning experience, declarative knowledge, procedural skills, and model inferences strongly correlated with each other overall, which confirmed our hypotheses that more experience is correlated with better performance (declarative and procedural) and with higher performance scores predicted by our model.

Experienced participants performed better on all critical metrics and exhibited smaller variations in performance. After analyzing the results, we consider the following metrics to be critical:

- For the OB topic, gestational age and all biometric measurements, except head circumference.
- For the IJV topic, cannulation success and the number of needle insertions.

The following, noncritical metrics did not perform well as discriminators between experienced and inexperienced participants and had generally lower or nonsignificant correlations:

- scan path length
- probe acceleration (which is highly correlated with path length)

- probe displacement
- scan duration (not presented in this report, since highly correlated with path length).

We think that the criticality of metrics is task dependent and that for other tasks, where, for example, one needs to be quick, the scan duration metric will be a critical factor of performance.

In the skill decay part of this study, we found small but not statistically significant skill decay that seems to level off, which could be indicative of skill retention rather than decay. It is possible that there might be a testing effect (Karpicke & Aue, 2015) to some degree, where repeated testing actually has a practice effect of refreshing and retaining skills. The participants, by virtue of being physicians or in medical school, are likely knowledgeable and skilled in medical procedures and good at retaining skills over time. It is also possible that we found a manifestation of individual skill decay (Sackett, Lievens, Van Iddekinge, & Kuncel, 2017), where skill decay is dependent on the individual.

The Bayesian network model used in both model validation and skill decay portions of the study shows promise for use as a criterion measure for proficiency in the assessment of procedural skills for those topics. However, the restricted range of the current inferences points to the need for tuning of the conditional probability tables.

In summary, this study has shown that our assessment approach of OB and IJV skills using the SonoSimulator® can provide a viable criterion standard and method for proficiency assessment for use in military and civilian contexts, for example, point-of-care ultrasound.

Future studies could explore other models of assessment. Latent variable models such as the diagnostic classification model (DCMs) (Rupp, Templin, & Henson, 2010) or DBN-DCM combinations (Levy, 2014) would be potential candidates. To be able to train models with data and use more sophisticated analyses, a bigger sample size would be required, that is, to obtain an overall evaluation of the quality of the models (Williamson, Almond, & Mislevy, 2000).

REFERENCES

Ahmed, O. M. A., O'Donnell, B. D., Gallagher, A. G., Breslin, D. S., Nix, C. M., & Shorten, G. D. (2016). Construct validity of a novel assessment tool for ultrasound- guided axillary brachial plexus block. *Anaesthesia, 71*(11), 1324–1331. https://doi.org/10.1111/anae.13572

Almond, R. G., Mislevy, R. J., Steinberg, L., Yan, D., & Williamson, D. (2015). *Bayesian networks in educational assessment.* New York, NY: Springer.

American Institute of Ultrasound in Medicine. (2019). Curriculum for the performance of ultrasound-guided procedures. *Journal of Ultrasound in Medicine*, *38*(8), 1951–1969. https://doi.org/10.1002/jum.15089

Andersen, D. K. (2012). How can educators use simulation applications to teach and assess surgical judgment? *Academic Medicine*, *87*(7), 934-941.

Arthur, W. J., Bennett, W., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, *11*(1), 57–101. http://doi.org/10.1207/s15327043hup1101_3

Arthur, W. Jr., Day, E. A., Villado, A. J., Boatman, P. R., Kowollik, V., Bennett, W. Jr., & Bhupatkar, A. (2007). *Decay, transfer, and the reacquisition of a complex skill: An investigation of practice schedules, observational rehearsal, and individual differences* (Tech. Rep. No. AFRL-HE-AZ-TR-2008-0001). Mesa, AZ: Air Force Research Laboratory. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA482766

Bollen, K. A., Curran, P. J. (2006). *Latent curve models: A structural equation perspective.* Hoboken, NJ: Wiley-Interscience.

Chenkin, J., Lee, S., Huynh, T., & Bandiera, G. (2008). Procedures can be learned on the Web: A randomized study of ultrasound-guided vascular access training. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, *15*(10), 949–954. https://doi.org/10.1111/j.1553-2712.2008.00231.x

Chiem, A. T., Soucy, Z., Dinh, V. A., Chilstrom, M., Gharahbaghian, L., Shah, V., … Fox, J. C. (2016). Integration of ultrasound in undergraduate medical education at the California medical schools: a discussion of common challenges and strategies from the UMeCali experience. *Journal of Ultrasound in Medicine, 35*(2), 1–13. https://doi.org/10.7863/ultra.15.05006

Chung, G. K. W. K., Gyllenhammer, R. G., Baker, E. L., & Savitsky, E. (2013). The effects of practicing with a virtual ultrasound trainer on FAST window identification, acquisition, and diagnosis. *Military Medicine, 178*(10S), 87–97.

Culbertson, M. J. (2015). Bayesian networks in educational assessment: The state of the field. *Applied Psychological Measurement, 40*(1), 3–21. http://doi.org/10.1177/0146621615590401

Dinh, V. A., Fu, J. Y., Lu, S., Chiem, A., Fox, J. C., Blaivas, M., & Hoppmann, R. (2015). Integration of ultrasound in medical education in United States medical schools: A national survey of directors. *Ultrasound in Medicine & Biology*, *41*(4, Supplement), S19–S2. doi:https://doi.org/10.1016/j.ultrasmedbio.2014.12.123

Dong, B., Savitsky, E., & Osher, S. (2009, November). A novel method for enhanced needle localization using ultrasound-guidance. In *International Symposium on Visual Computing* (pp. 914-923). Berlin, Germany: Springer.

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77*(5), 615–622.

Iseli, M. R., & Cai, L. (2016). *Creation of a knowledge and skills ontology for fundamentals of ultrasonography* (Final deliverable, CRESST Pélagique project report: SBIR Phase II). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Iseli, M. R., & Jha, R. (2016). Computational issues in modeling user behavior in serious games. In H. F. O'Neil, E. L. Baker, & R. S. Perez (Eds.), *Using games and simulations for teaching and assessment: Key issues* (pp. 21–40). New York, NY: Routledge.

Iseli, M. R., & Savitsky, E. (2015, August). *Creation of an ontology for the assessment of fundamentals of medical ultrasound.* Presentation at the 2015 CRESST conference, Redondo Beach, CA.

Iseli, M. R., Savitsky, E., & Schenke, K. (2019). *Simulation-based assessment of psychomotor skills for ultrasound competency evaluation.* Manuscript submitted for publication.

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review, 27*(2), 317–326. doi:10.1007/s10648-015-9309-3

Levy, R. (2014). *Dynamic Bayesian network modeling of game based diagnostic assessments* (CRESST Report 837). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

MacGregor, S. N., & Sabbagha, R. E. (2009). Assessment of gestational age by ultrasound. *The Global Library of Women's Medicine*. https://doi.org/10.3843/GLOWM.10206

Mackay, F. D., Zhou, F., Lewis, D., Fraser, J., & Atkinson, P. R. (2018). Can you teach yourself point-of-care ultrasound to a level of clinical competency? Evaluation of a self-directed simulation-based training program. *Cureus, 10*(9). doi:10.7759/cureus.3320

Nelson, B. P., & Sanghvi, A. (2016). Out of hospital point of care ultrasound: Current use models and future directions. *European Journal of Trauma and Emergency Surgery*, *42*(2), 139–150. https://doi.org/10.1007/s00068-015-0494-z

Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology* (Knowledge Systems Laboratory Tech report KSL-01-05). Palo Alto, CA: Stanford University.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications.* New York, NY: Guilford Press.

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology, 102*(3), 254–273.

Savitsky, E. (2013). *U.S. Patent No. 8,480,404*. Washington, DC: U.S. Patent and Trademark Office.

Selig, D. J., Collins, J., Church, T. L., & Zeman, J. (2019). An editorial review of mobile health: Implications for the US military health system. *Military Medicine*, *184*(7–8), E253–E258. https://doi.org/10.1093/milmed/usz073

Spears, W. D. (1985). Measurement of learning and transfer through curve fitting. *Human Factors*, *27*, 251–266.

Todsen, T., Tolsgaard, M. G., Olsen, B. H., Henriksen, B. M., Hillingsø, J. G., Konge, L., … Ringsted, C. (2015). Reliable and valid assessment of point-of-care ultrasonography. *Annals of Surgery*, 261(2), 309–315. https://doi.org/10.1097/SLA.0000000000000552

Williamson, D. M., Almond, R. G., & Mislevy, R. J. (2000, June). Model criticism of Bayesian networks with latent variables. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence* (pp. 634-643). San Francisco, CA: Morgan Kaufmann.

Wisher, R., Sabol, M., Ellis, J., & Ellis, K. (1999). *Staying sharp: Retention of military knowledge and skills* (Special Report 39). Alexandria, VA: U.S. Army Research Institute.

APPENDIX A: BAYESIAN NETWORK NODE DESCRIPTIONS

Node Name	Node Type	Description
Procedural Skills	Latent	Overall procedural skills.
P.IND	Latent	Not Assessed.
P.PREP	Latent	Not Assessed.
P.ACQ	Latent	Image acquisition skills.
P.ACQ.MOT	Latent	Motor skills.
P.ACQ.MOT.3	Latent	Perform transducer movement fluently and smoothly. (fluency)
P.ACQ.MOT.4	Latent	Perform transducer movement efficiently and economically. (movement efficiency)
P.ACQ.MOT.9	Latent	Move transducer and another instrument bilaterally in a coordinated manner. (bilateral coordination)
P.ACQ.VMOT	Latent	Visuomotor skills.
P.ACQ.VMOT.1	Latent	Execute a transducer movement to match the orientation of an object as closely as possible, given visual feedback.
P.ACQ.VMOT.4	Latent	Take accurate biometric measurements. (accuracy)
P.ACQ.VMOT.5	Latent	Simultaneously track and modify the progress of an instrument towards a target structure.
P.ACQ.WINV	Latent	Acquisition of windows and views.
P.ACQ.WINV.1	Latent	Use appropriate anatomic planes.
P.ACQ.WINV.4	Latent	Acquire appropriate scanning windows.
P.ACQ.WINV.5	Latent	Obtain appropriate scanning views.
P.ACQ.VSPAT	Latent	Visuospatial skills.
P.ACQ.VSPAT.7	Latent	Acquire optimal (final) scanning orientation.
P.INT	Latent	Interpretation skills.
P.INT.DDF	Latent	Interpretation of directly demonstrated findings.
P.INT.DDF.5	Latent	Perform accurate pregnancy dating.
P.PRES	Latent	Not assessed.
P.INTG	Latent	Not assessed.
ProbStead	Observable	Probe steadiness/acceleration score.
PaLen	Observable	Path length score.
ScanDur	Observable	Scan duration score.
ProbDisp	Observable	Probe displacement length score.
ViewDiff	Observable	Distance to optimal view score.
CanSuccess	Observable	Cannulation success score.
NdIns	Observable	Number of needle insertions score.
BPDpctErr	Observable	Biparietal diameter accuracy score.
HCpctErr	Observable	Head circumference accuracy score.
ACpctErr	Observable	Abdominal circumference accuracy score.
FLpctErr	Observable	Femur length accuracy score.
GestAgeErr	Observable	Gestational age accuracy score.

APPENDIX B: MASTERY QUESTIONS EXAMPLES

1) Which of the following parameters are used for pregnancy dating during fetal biometry?
    A. Stomach diameter
    B. Femur length
    C. Biparietal diameter of fetal skull
    D. A and B
    E. B and C

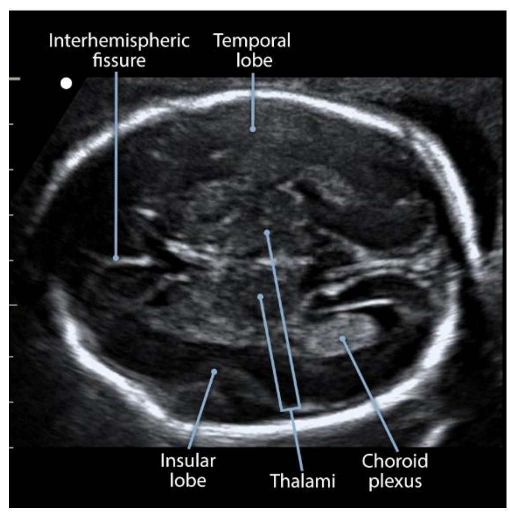
Answer: E

Already in Mastery test as Question #3:

2) Which of the following statements best characterizes sonographic assessment of basic fetal neuroanatomy?
    A. Cerebral tissue has a highly echogenic sonographic appearance.
    B. The fetal skull normally has an elliptical echogenic appearance.
    C. The choroid plexus and pia, arachnoid, and dura mater layers have a bright, hyperechoic appearance.
    D. A and C
    E. B and C

Answer: E

3) Which of the following structures does the indicator in this transverse-view ultrasound image of a fetal head identify?
    A. Cerebellum
    B. Third ventricle
    C. Thalami
    D. Cavum septi pellucidi
    E. None of the above



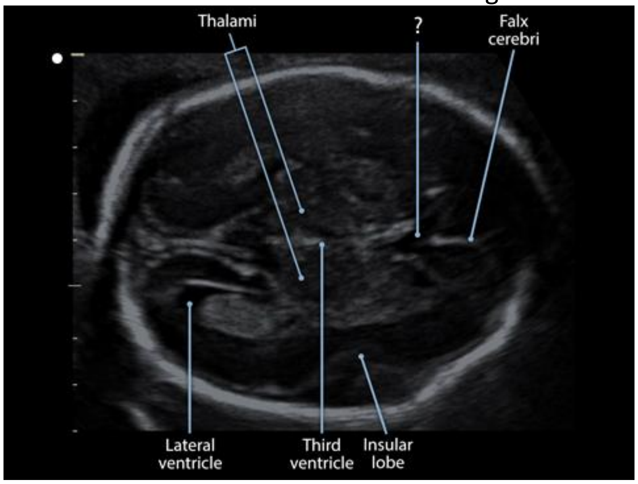
Answer: C
Image edits: Remove “Thalami” label and replace with a question mark

4) Which of the following statements best characterizes the ventricles of the brain?
    A. The brain contains three ventricles.
    B. Each lateral ventricle consists of a frontal horn, occipital horn, temporal horn, body, and atrium.
    C. The walls of the ventricles are lined with the echogenic leptomeninges.
    D. A and B
    E. B and C

Answer: E

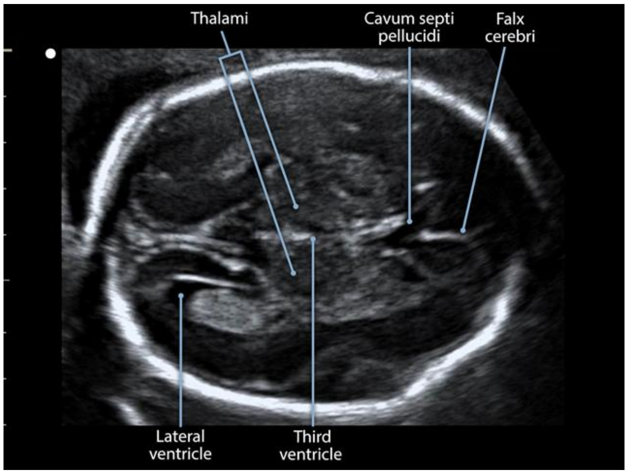
Already in Mastery test as Question #6:

5) Which of the following structures does the indicator in this transverse-view ultrasound image of a fetal head identify?
    A. Cavum septi pellucidi
    B. Third ventricle
    C. Foramina of Luschka
    D. Foramen of Magendie
    E. Fourth ventricle



Answer: A

6) Which of the following structures does the indicator in this transverse-view ultrasound image of the fetal head identify?
    A. Interhemispheric fissure
    B. Lateral ventricle
    C. Cerebellum
    D. Falx cerebri
    E. None of the above



Answer: D

Image edits: Remove “Falx cerebri” label and replace with a question mark

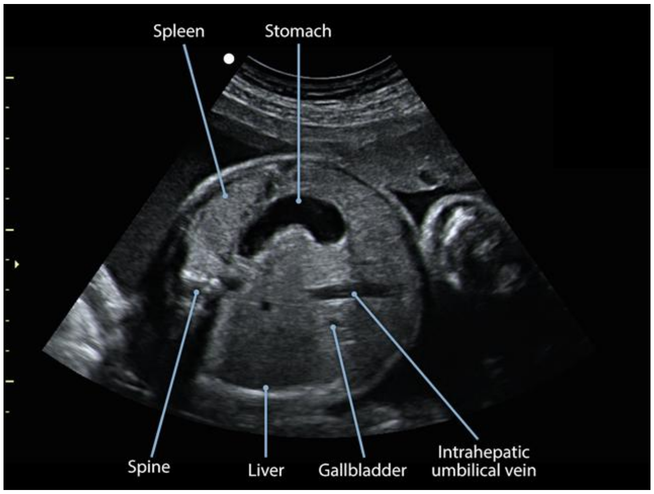
7) Which of the following sonographic landmarks can be seen when scanning along the imaging plane used to measure biparietal diameter?
    A. Falx cerebri
    B. Thalami
    C. Cerebellum
    D. All of the above
    E. A and B only

Answer: E

8) Which of the following statements best characterizes using biparietal diameter as a biometric measurement parameter for pregnancy dating?
    A. Biparietal diameter should be measured along a transcerebellar transverse-oblique plane.
    B. Relevant sonographic landmarks include the falx cerebri and cavum septi pellucidi.
    C. The cerebellum should typically be seen when in an optimal biparietal diameter imaging plane.
    D. All of the above
    E. A and C only

Answer: B

9) Which of the following structures does the indicator in this transverse-view ultrasound image of the fetal abdomen identify?
    A. Heart
    B. Spleen
    C. Stomach
    D. Liver
    E. Gallbladder



Answer: C
Image edits: Remove all labels except spine, intrahepatic umbilical vein, and indicator line for stomach. Replace "Stomach" label with question mark.

10) Which of the following statements best characterizes anatomy of the fetal abdomen?
    A. The fetal liver occupies most of the right abdomen and partly transcends into the left abdomen.
    B. The fetal spleen and fetal liver have a similar echotexture
    C. Fetal large intestine has a hypoechoic appearance relative to surrounding structures.
    D. All of the above
    E. None of the above

Answer: D

11) Which of the following variables determines the echogenicity of fetal bone on ultrasound?
    A. Degree of bone mineralization
    B. Distance of bone from transducer
    C. Angle of insonation
    D. All of the above
    E. A and B only

Answer: D

12) Which of the following statements best characterizes biparietal diameter measurement of the fetal skull?
    A. The measurement should be done from the outer edge of the proximal parietal bone to the outer edge of the opposing parietal bone.
    B. The line connecting the two caliper icons should run perpendicular to the axis of the midline falx cerebri.
    C. The calipers should align with the level of the thalami, since this is typically the widest skull diameter.
    D. A and B
    E. B and C

Answer: E

13) The thalamic plane can be used for both biparietal diameter and head circumference measurements.
    A. True
    B. False

Answer: A

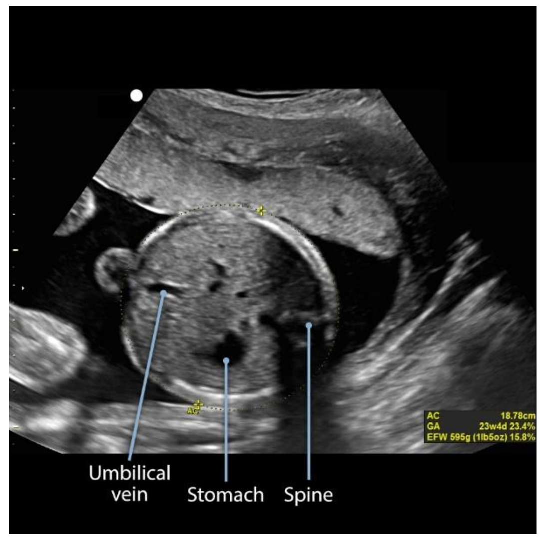
14) When measuring fetal head circumference, the ellipse should be positioned along the inner edge of the calvaria.
    A. True
    B. False

Answer: B

15) Which of the following sonographic landmarks are typically used to identify the correct axial plane for measuring fetal abdominal circumference?
    A. Spine
    B. Stomach bubble
    C. Intrahepatic segment of the umbilical vein at the level of the portal sinus
    D. Fetal ribs
    E. All of the above

Answer: E

16) Which of the following structures does the indicator in this transverse-view ultrasound image of the fetal abdomen identify?
    A. Heart
    B. Umbilical vein
    C. Inferior vena cava
    D. Aorta
    E. None of the above



Answer: B
Image edits: Replace "Umbilical vein" label with a question mark

17) Which of the following statements best characterizes umbilical and portal vein circulation?
    A. The umbilical vein delivers oxygenated blood from the placenta to the fetus.
    B. The umbilical vein has a direct anastomosis with the inferior vena cava.
    C. The umbilical vein courses cephalad along an oblique plane and joins the portal sinus, which leads to the portal veins.
    D. All of the above
    E. A and C only

Answer: E

18) When measuring the abdominal circumference using an ellipse, the ellipse should extend to the outer skin edge of the abdominal wall.
    A. True
    B. False

Answer: A

19) Which of the following structures is used for fetal femur length determination during fetal biometry?
    A. Femoral diaphysis
    B. Non-ossified distal femoral epiphysis
    C. Femoral metaphysis
    D. A and B
    E. A and C

Answer: E

20) Which of the following statements best characterizes femur-length biometric measurement with ultrasonography?
    A. Visualize the entire length of the femoral diaphysis and metaphysis prior to obtaining a measurement.
    B. Do not include the epiphysis in the biometric measurement.
    C. Do not include cartilage in the biometric measurement.
    D. All of the above
    E. A and C only

Answer: D

APPENDIX C: FETAL BIOMETRY PROCTOR SCRIPT

OB Ultrasound Simulation Validation Study Proctor Script

Model Validation

Before the participant arrives, check and see if they've completed their at-home portion of the study. If they have not, reschedule the appointment.

Overview of Participant Onsite Tasks

1. Simulator Training: watch video and practice on SonoSimulator®
2. Demonstrate use of 7 skills
3. Simulator Assessment
4. Survey

Simulator Training

When the participant arrives, get them settled at the workstation. The following script will introduce them to the segments.

Script **[[SAY]]**

Hello, my name is ___________________. I will be your proctor for the on-site segment of the study. This study is about the assessment of ultrasound knowledge and scanning skills using a simulator.

The study has two parts. The first part collects data that is used to train computational models which are used to infer and predict ultrasound knowledge and skills from measurements of performance, such as hand motion data and multiple-choice question answers. The second part of the study analyzes how knowledge and skills decay over time when not used or refreshed.

We're going to do a quick overview of the tools you will be working with.

The external monitor will display the necessary instructions and tutorials to guide you through this segment. The laptop screen will display the ultrasound simulation software called the SonoSimulator®, which will assess your scanning skills. On the external monitor, we will begin with a video tutorial that will walk you through all the features of the SonoSimulator®. The laptop screen and the monitor are connected, so you can use your mouse to navigate between the two.

[[SHOW participant how to navigate between screens]]

Please follow along with the tutorial video to practice using the tools and features of the SonoSimulator®. After each introduced feature, click pause and practice what you learned in the SonoSimulator®. Then, after you've practiced a few times and feel you've learned the skill or feature, resume playing the video. You can also scroll back in the video if you need to review a specific skill or feature.

Please place your probe on the scan pad while scanning in the Simulator.

Once you've completed the tutorial, please let me know and I will navigate you to the next segment of the study.

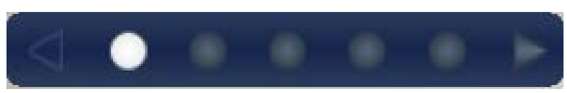
[[END]]

When a participant says they are finished with the hands-on tutorial segment **click the 'Next' button** at the bottom of the page on the external monitor, and continue reading the script.

[[SAY]] Before we move onto the assessment segment, let's see what you've learned from the tutorial. I am going to ask you to demonstrate each skill taught in the tutorial. The page on the right lists all the skills taught in the video. If you need to review a skill, you can watch its associated video as a refresher.

Let's begin. **[[END]]**

Navigate back to the first point and press the calibration button for participant.



[[SAY]]

1. Please demonstrate probe calibration. [probe in stand, press calibration button]

2. Please demonstrate how to recall the tasking instructions. [press Case history button]
3. Please demonstrate that you are comfortable scanning through the ultrasound image [moving probe]
4. Please demonstrate how to freeze the image. [Use the F key]
5. Please select the head circumference (HC) calipers. Please place the calipers. Now select the femur length (FL) calipers. Please place the calipers.
6. Please demonstrate how to save your image and measurement. [press save button]
7. Please demonstrate how to navigate to the next task. [press arrow on right of the dots]

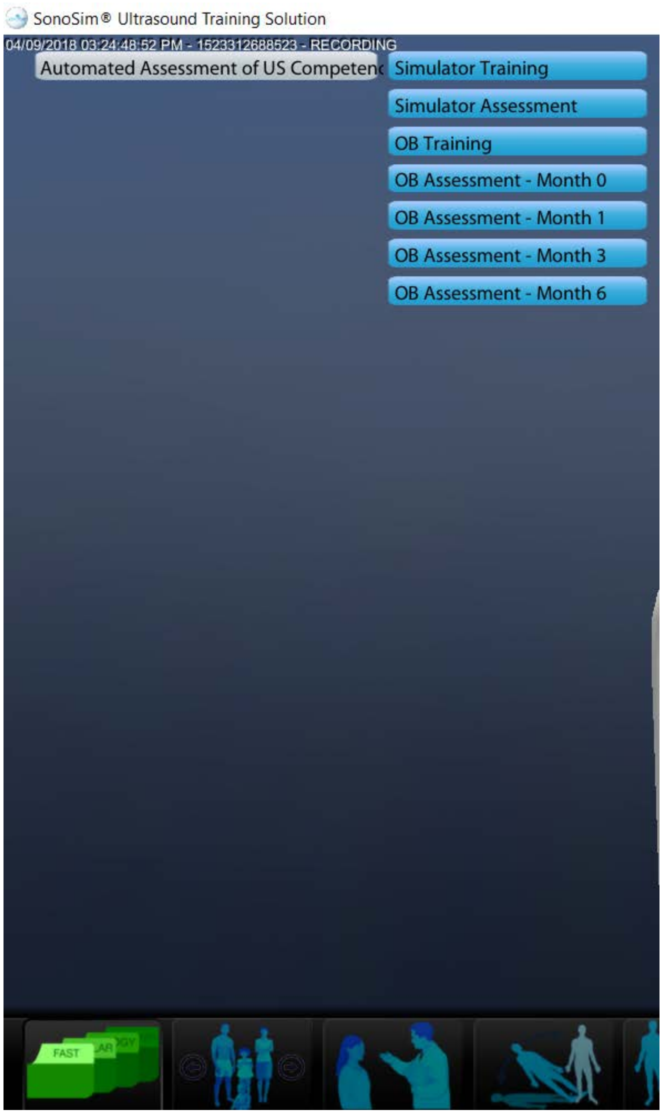
[[END]]

- **If the participant fails** to properly demonstrate any of the above skills, have them review the appropriate refresher videos and then confirm mastery.

- **If the participant has demonstrated mastery** of all the above skills, continue below.

Simulator Assessment

1) Navigate the participant to the assessment tasks by clicking Case List > Automated Assessment of US Competence > Simulator Assessment.



2) Confirm the Simulator Assessment cases are loaded. You should see fifteen dots (tasks) on the bottom above the icon toolbar and the Simulator Assessment Case 1 Task 1.

[[SAY]] We are now going to move onto the assessment segment of the study. During this next segment, you will perform the skills you have just demonstrated for 15 tasks. You will no longer be following along with a video. However, you can refer back to the refresher videos on the right, if needed. You also have a printout with the same information. **[[SHOW PRINTOUT]]**

For this segment, please be mindful to use economy of probe motion. Be as time- and movement-efficient as possible when acquiring the desired image for measurements.

Do you have any questions before we move forward? **[[END]]**

Now have participant start cases.

[SAY] There are 15 cases, you can start now. [cases will already be loaded]

User Completion Survey

Have participant complete user survey online: <link here>

Register for Skill Decay Part

[SAY]

Thank you for your participation. Now that you completed the first part of the study, you can continue to Part II of the study in which you will receive additional instruction and scanning opportunities so that you will be able to consolidate your knowledge and skills. Would you like to participate in Part II?

[END]

If answer is YES, try to set up an appointment right now on the Acuity website: <link here>

Skill Decay - Month 0

Before the participant arrives, check to see they completed their at-home portion of the study for Month 0. If they have not, reschedule the appointment.

Overview of Participant Onsite Tasks

1. Review 7 skills
2. OB Training: watch video and practice on SonoSimulator®
3. Month 0 assessment

Review 7 Skills

When the participant arrives, get them settled in the workstation. Refresh participant's memory on the seven skills:

[[SAY]]

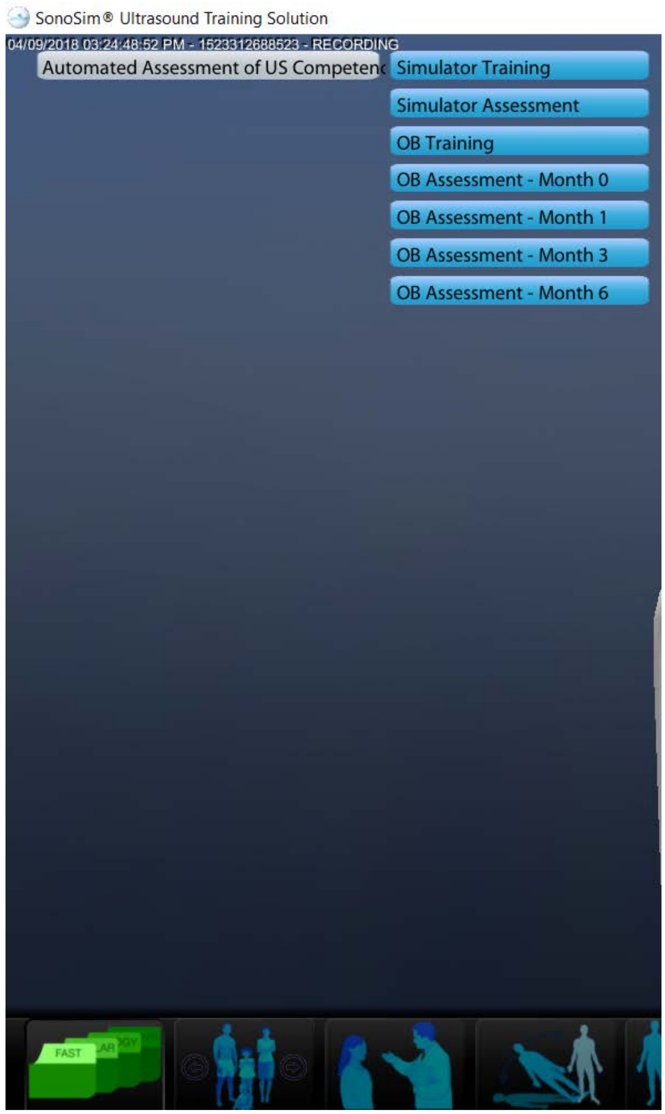
1. Please demonstrate probe calibration. [probe in stand, press calibration button]

2. Please demonstrate how to recall the tasking instructions. [press Case history button]
3. Please demonstrate that you are comfortable scanning through the ultrasound image [moving probe]
4. Please demonstrate how to freeze the image. [Use the F key]
5. Please select the head circumference (HC) calipers. Please place the calipers. Now select the femur length (FL) calipers. Please place the calipers.
6. Please demonstrate how to save your image and measurement. [press save button]
7. Please demonstrate how to navigate to the next task. [press arrow on right of the dots]

[[END]]

OB Training

1) Navigate the participant to the assessment tasks by clicking Case List > Automated Assessment of US Competence > OB Training.



2) Confirm the "**OB Training**" tasks are loaded.

3) Press the **"Next"** arrow in the online instructions [on external monitor].

[[SAY]] We are going to move onto the simulator training segment of the study. During this segment, you will perform the skills you learned during the model validation segment of the study. However, this time there are a few new features that are provided to complement your training. Please watch the tutorial video. You will also have a printout with the same information if needed.

Then, please go through the cases in the simulator. As you do, please watch the Findings Videos to learn more about the tasks. You can also use the Probe Guide, Layers tool, and case legend if they are helpful to you.

Do you have any questions before we move forward? **[[END]]**

OB Assessment - Month O

1) Once they are done, then go to Case List > Automated Assessment of US competency > OB Assessment - Month 0

2) Confirm the "Skill Decay: Month 0" tasks are loaded.

3) **Select the next button** at the bottom of the online instruction.

   [[SAY]] We are now going to move onto the Assessment segment of the study. During this segment, you will perform 15 tasks, using the skills you have just learned.

   You will not be following along with a video, and both the probe guide tool and the findings video tool will no longer be available to you. However, this page will remain on the screen so you can refer back to the refresher videos if needed. You will also have a printout with the same information if needed.

   For this segment, please be mindful to use economy of probe motion. Be as time- and movement-efficient as possible when acquiring the desired image for measurements.

   Do you have any questions before we move forward? **[[END]]**

2) Once they have finished the OB Assessment for Month 0, then they are done. Please let them know we will be sending a reminder email to schedule their Skill Decay--Month 1 appointment.

Skill Decay – Month 1, 3, or 6

Before the participant arrives, check and see if they have completed the at-home portion of the study for that session. If not, have them take the mastery test on site.

[[SAY]] "You will be retaking the mastery posttest for the 2nd and 3rd Trimester Pregnancy. Please let me know if you have any questions and when you are finished" **[[END]]**

You will need to have them click on the content first and then take the test link to start the test

Overview of Participant Onsite Tasks

1. Review of 7 skills

2. Month 1, 3, or 6 assessment
3. Survey [If month 6]

Skill OB Assessment - Month 1, 3, or 6

1) Press Case List > Automated Assessment of US Competence > OB Assessment – Month 1
2) Confirm the "Skill Decay: Month 1, 3 or 6" tasks are loaded.
3) Select the next button at the bottom of the online instruction.

[[SAY]] We are now going to move onto the Month 1/3/6 Assessment segment of the study. During this next segment, you will perform the skills you have just demonstrated for 15 tasks.

You will not be following along with a video, and both the probe guide tool and the findings video tool will no longer be available to you. However, this page will remain on the screen so you can refer back to the refresher videos if needed. You will also have a printout with the same information if needed.

For this segment, please be mindful to use economy of probe motion. Be as time- and movement-efficient as possible when acquiring the desired image for measurements.

Do you have any questions before we move forward? **[[END]]**

6) If Month 6, also have them fill out the post survey

[STOP THE VIDEO at the end of the session]

APPENDIX D: INTERNAL JUGULAR VEIN CANNULATION PROCTOR SCRIPT

IJV: Ultrasound Simulation Validation Proctor Script

Model Validation

Check STATUS to confirm did at-home portion.

If they did not do the at-home portion, ask them to reschedule.

Overview of Participant onsite tasks

1. Simulator Training: watch video and learn 7 skills; practice on SonoSimulator®
2. Demonstrate use of 7 skills
3. Simulator Assessment
4. Take survey

When the participant arrives

INTRODUCE yourself; confirm their name and

- Offer to let them put things down wherever they like.
- Check if they need bathroom/water before beginning
- Ask them if they have completed the at-home portion to confirm.

[SETTLE participants at workstation]

[SAY]

I'm _______________and I will be your proctor for the on-site part of the study. This study is about the assessment of ultrasound knowledge and procedural skills using a simulator.

The study has two parts. The first part is called Model Validation and collects data that is used to train computational models. These are used to infer and predict ultrasound knowledge and skills from measurements of performance. Examples of performance include such as hand motion data and multiple-choice answers.

The second part is called Skill Decay. That portion of the study analyzes how knowledge and skills decay over time when not used or refreshed.

[SHOW participant how to navigate between screens]]

[SAY]

Before we begin, there is a Study we're going to do a quick overview of the tools you will be working with.

[[SHOW the participant the Torso, TrackPad, syringe, probe, gel, towels. Make sure they know the *area of the torso to probe*. (not the black parts)]]

We'll begin with the tutorial. The laptop screen will display the ultrasound simulation software called the SonoSimulator®, which will assess your scanning skills. On the external monitor, we will begin with a video tutorial that will walk you through all the features of the SonoSimulator®. The laptop screen and the monitor are connected, so you can use your mouse to navigate between the two.

Please do the demonstrated skill along with the tutorial video to practice using the tools and features of the SonoSimulator®. After each introduced feature, click pause and practice what you learned in the SonoSimulator®.

Then, after you've practiced a few times and feel you've learned the skill or feature, resume playing the video. You can also scroll back in the video if you need to review a specific skill or feature.

Please place your probe on the TrackPad while scanning in the Simulator.

Once you've completed the tutorial, please let me know and I will navigate you to the next segment of the study.

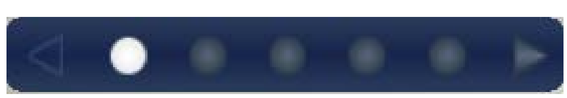
Review 7 Skills

When a participant says they are finished with the hands-on tutorial segment, **click the 'Next' button** at the bottom of the page on the external monitor, and continue reading the script.

[[SAY]] Before we move onto the assessment segment, let's see what you've learned from the tutorial. I am going to ask you to demonstrate each skill taught in the tutorial. The page on the right lists all the skills taught in the video. If you need to review a skill, you can watch its associated video as a refresher.

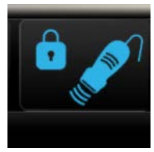
Let's begin.

Navigate back to the first point and press the calibration button for participant.



[[SAY]]

1. Please demonstrate probe and needle & syringe calibration. [probe and syringe in stand, press calibration button]

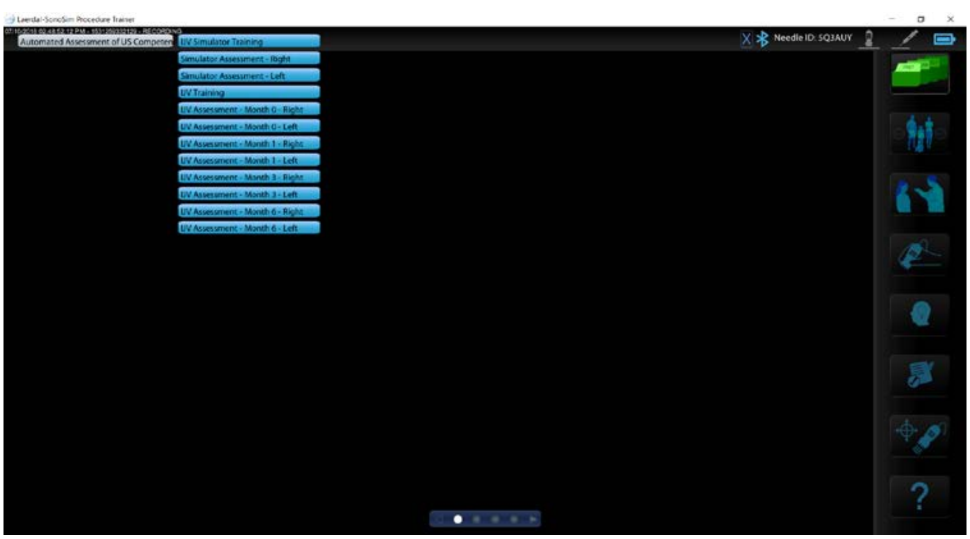


2. Please demonstrate that you are comfortable scanning through the ultrasound image [moving probe]
3. Please demonstrate how access Doppler [pressing Doppler button and selecting one of the options]
4. Please demonstrate how to insert the needle
5. Please demonstrate how to complete cannulation [Pull up on the syringe]

- **If the participant fails** to properly demonstrate any of the above skills, have them review the appropriate refresher videos and then confirm mastery.
- **If the participant has demonstrated mastery** of all the above skills, continue below.

Simulator Assessment - Right

Navigate the participant to the assessment tasks by clicking Case List > Automated Assessment of US Competence > **Simulator Assessment - Right**



Confirm the Simulator Assessment Right cases are loaded. You should see 6 dots (tasks) on the bottom above the icon toolbar and the Simulator Assessment Case 1 Task 1.

> **[[SAY]]** We are now going to move onto the assessment segment of the study. During this next segment, you will perform the skills you have just demonstrated for 6 tasks. You will no longer be following along with a video. However, you can refer back to the refresher videos on the right, if needed. You also have a printout with the same information. **[[SHOW PRINTOUT on WALL]]**
>
> For this segment, please be mindful to use economy of probe motion. Be as time- and movement-efficient as possible when acquiring the desired image for measurements.
>
> Do you have any questions before we move forward?
>
> There are 6 cases, you can start now.

Simulator Assessment: Left

After the participant has competed the 6 tasks on the right neck, Go to Simulator Assessment Left. You should see 12 dots (tasks) on the bottom above the icon toolbar and the Simulator Assessment Left Case 1 Task 1. Have the participant go through the 12 tasks.

Survey

Make sure user takes the survey which is at the end of the instructions.

[[STOP THE VIDEO if it is the end of the session]]

Skill Decay Portion of the Study

SD Month 0

Check STATUS to confirm participant did at-home portion.

At-home portion for Month 0 includes:

1. Introduction to Ultrasound-Guided Procedures
2. Ultrasound-Guided Internal Jugular Vein Cannulation

Overview of Participant onsite tasks

1. Review 7 skills
2. IJV training: watch video and practice on SonoSimulator®
3. Take Month 0 assessment - right neck
4. Take Month 0 assessment - left neck

If the participant has <u>not</u> completed the at-home portion ask them to reschedule their appointment and complete the at-home portion before their appointment.

If the participant <u>has</u> completed the at-home portion:

Review 7 Skills

Refresh participant's memory on the seven skills:

[[SAY]]

1. Please demonstrate probe calibration. [probe and needle and syringe in stand, press calibration button]

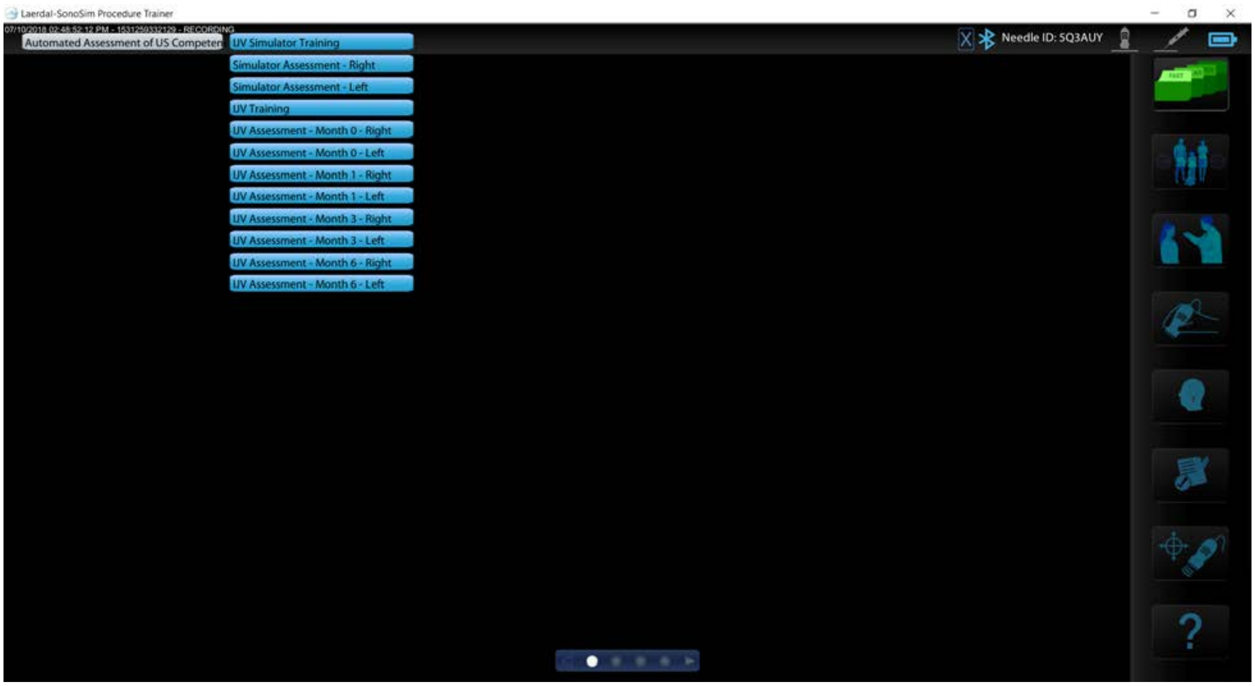
2. Please demonstrate how to recall the tasking instructions. [press Case history button]
3. Please demonstrate that you are comfortable scanning through the ultrasound image [moving probe]
4. Please demonstrate compression
5. Please demonstrate how access Doppler [pressing Doppler button and selecting one of the options]

6. Please demonstrate how to insert the needle
7. Please demonstrate how to complete cannulation [Pull up on the syringe]

[[END]]

IJV Training

1) Navigate the participant to the assessment tasks by clicking Case List > Automated Assessment of US Competence > IJV Training.



2) Confirm the "**IJV Training**" tasks are loaded.

3) Press the **"Next"** arrow in the online instructions [on external monitor].

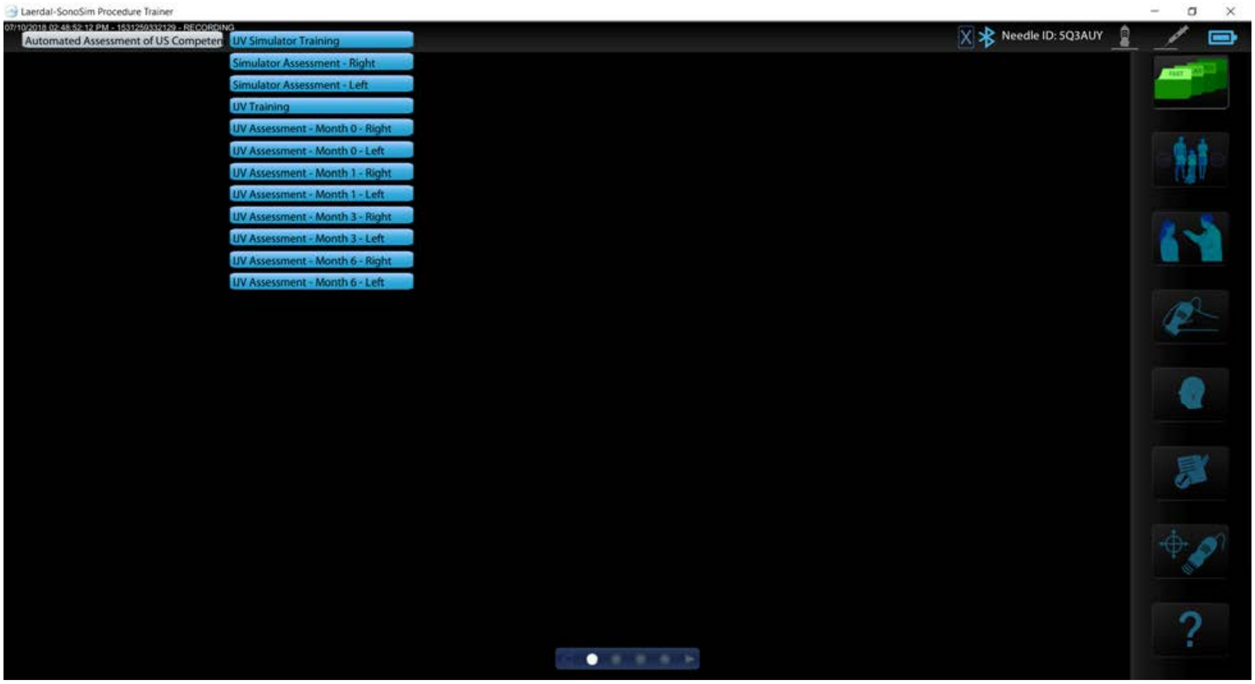
[[SAY]] We are going to move onto the simulator training segment of the study. During this segment, you will perform the skills you learned during the earlier model validation segment of the study. However, this time there are a few new features that are provided to complement your training. Please watch the tutorial video.

Then, please go through the cases in the simulator. As you do, please watch the Tutorial Videos to learn more about the tasks. You can also use the Probe Guide, Layers tool, and case legend if they are helpful to you.

Do you have any questions before we move forward? **[[END]]**

IJV Assessment

4) Press Case List > Automated Assessment of US Competence > IJV Assessment – Month 0 Right.



5) Confirm the "IJV Assessment Month 0 Right" tasks are loaded.

6) Select the next button at the bottom of the online instruction.

[[SAY]] We are now going to move onto the Month 0 Assessment segment of the study. During this segment, you will perform 6 tasks, using the skills you have just learned.

You will not be following along with a video, and both the probe guide tool and the findings video tool will no longer be available to you. However, this page will remain on the screen so you can refer back to the refresher videos if needed. You will also have a printout with the same information if needed.

For this segment, please be mindful to use economy of probe motion. Be as time- and movement-efficient as possible when acquiring the desired image for measurements.

Do you have any questions before we move forward? **[[END]]**

7) Once they have finished the 6 tasks, load IJV Assessment Month 0 Left and have the participant complete the 12 tasks.

SD Month 1, 3, or 6

At-home portion for Month 1, 3, or 6 includes:

1. Month 1, 3, or 6 Post-Test: Introduction to Ultrasound-Guided Procedures - Study Version
2. Month 1, 3, or 6 Post-Test: Ultrasound-Guided Internal Jugular Vein Cannulation - Study Version

Overview of Participant onsite tasks

1. Review 7 skills
2. Month 1, 3, or 6 IJV assessment - right neck
3. Month 1, 3, or 6 IJV assessment - left neck
4. [Month 6 only] Survey

Before the participant arrives, check and see if they completed the at-home portion of the study for their session.

If they haven't done the Mastery Test at home, log them into TTLMS with appropriate login sbir_X_y (e.g. sbir_01_4 for Month 1)

a. Have them take the Mastery Test

[[SAY]] "You will be retaking the mastery posttest for the Ultrasound Guided Procedures. Please let me know if you have any questions and when you are finished" **[[END]]**

You will need to have them click on the content first and then take the test link to start the test

If they have done the Mastery Test at home, or after completed onsite, continue as follows:

Review 7 skills

1. Refresh participant's memory on the seven skills:

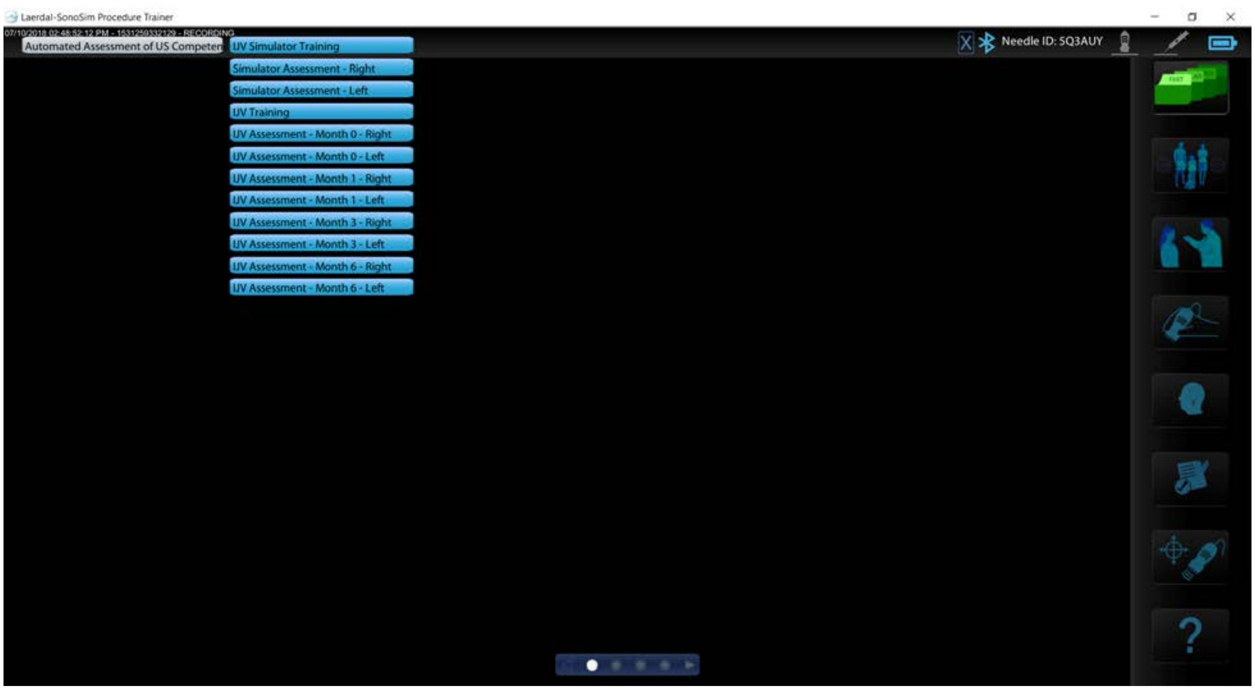
[[SAY]]

1. Please demonstrate probe calibration. [probe and needle and syringe in stand, press calibration button]

2. Please demonstrate how to recall the tasking instructions. [press Case history button]
3. Please demonstrate that you are comfortable scanning through the ultrasound image [moving probe]
4. Please demonstration compression [pressing on trackpad with probe]
5. Please demonstrate how access Doppler [pressing Doppler button and selecting one of the options]
6. Please demonstrate how to insert the needle
7. Please demonstrate how to complete cannulation [Pull up on the syringe]

IJV Assessment

3) Press Case List > Automated Assessment of US Competence > IJV Assessment – Month 1, 3, or 6 Right.



4) Confirm the "IJV Assessment Month 1, 3 or 6 Right" tasks are loaded.
5) Select the next button at the bottom of the online instruction.

[[SAY]] We are now going to move onto the Month 0/1/3/6 Assessment segment of the study. During this next segment, you will perform the skills you have just learned for 7 tasks.

You will not be following along with a video, and both the probe guide tool and the findings video tool will no longer be available to you. However, this page will remain on the screen so you can refer back to the refresher videos if needed. You will also have a printout with the same information if needed.

For this segment, please be mindful to use economy of probe motion. Be as time- and movement-efficient as possible when acquiring the desired image for measurements.

Do you have any questions before we move forward? **[[END]]**

6) Once they have finished the 6 tasks, **load IJV Assessment Month 0, 1, 3 or 6 Left** and have the participant complete the 12 tasks.

7) If Month 6, also have them fill out the post survey. **Click Next** on instructions to locate.

[STOP THE VIDEO at the end of the session]

APPENDIX E: SAMPLE POST SURVEY

Part 1 - Post Survey

University of California, Los Angeles - Automated Ultrasound Assessment Validation Study

16. Self- Assessment of your Skills After Study: *

	Very Poor	Below Average	Average	Above Average	Excellent
Image acquisitio...	○	○	○	○	○
Image interpretat...	○	○	○	○	○

17. What did you like about this study? *

Long answer text

18. What could be improved in this study? *

Long answer text

19. Are you interested in participating in any future studies? *

○ Yes

○ No

20. How long have you been playing video games? (On gaming systems or computer, not phone) *

- ○ Never
- ○ Less than a year
- ○ 1-4 years
- ○ More than 4 years

21. Which game types have you been playing (select all that apply)? *

- ☐ First-person perspective or shooter
- ☐ Top-down perspective or God view
- ☐ Not applicable

22. Which game genres have you been playing (select all that apply)? *

- ☐ Strategy (emphasizes strategic, tactical, and sometimes logistical challenges)
- ☐ Action (emphasizes physical challenges, including hand–eye coordination and reaction-time)
- ☐ Role playing (player controls the actions of a character immersed in some well-defined world)
- ☐ Sports (simulates the practice of sports)
- ☐ Adventure (player assumes the role of a protagonist in an interactive story driven by exploration and puzzle-s...
- ☐ Simulation (copies various activities from real life in the form of a game)
- ☐ Not applicable
- ☐ Other...

APPENDIX F: PARTICIPANT DATA COLLECTION TASKS

OB Ultrasound Study - Participant Data Collection Tasks

On site participant tasks	**At home participant tasks** Proctor check STATUS of at-home portion to see if this needs to also be done on site for months 1-3. Reschedule if at-home section not completed for MV and month .
OB Model Validation 1. Simulator Training: watch video and practice on SonoSimulator® 2. Demonstrate use of 7 skills 3. Simulator Assessment 4. Survey	OB Model Validation 1. Introductory Survey 2. Pre-Test: Fundamentals of Ultrasound 3. Fundamentals of Ultrasound - Study Version 4. Pre-Test: Second- & Third-Trimester Pregnancy: Part I - Study Version
OB Month 0 (0 weeks) 1. Demonstrate use of 7 skills 2. OB Training: watch video and practice on SonoSimulator® 3. Month 0 assessment	OB Month 0 (0 weeks) 1. Second- & Third-Trimester Pregnancy: Part I - Study Version
OB Month 1 (actually 2 weeks) 1. At home portion if not completed 2. Demonstrate use of 7 skills 3. Month 1 assessment	OB Month 1 (actually 2 weeks) 1. Month 1 Post-Test: Second- & Third-Trimester Pregnancy: Part I - Study Version
OB Month 3 (actually 6 weeks) 1. At home portion if not completed 2. Demonstrate use of 7 skills 3. Month 3 assessment	OB Month 3 (actually 6 weeks) 1. Month 3 Post-Test: Second- & Third-Trimester Pregnancy: Part I - Study Version
OB Month 6 (actually 12 weeks) 1. At home portion if not completed 2. Demonstrate use of 7 skills 3. Month 6 assessment 4. Survey	OB Month 6 (actually 2 weeks) 1. Month 6 Post-Test: Second- & Third-Trimester Pregnancy: Part I - Study Version

IJV Ultrasound Study - Participant Data Collection Tasks

On site participant tasks	**At home participant tasks** Proctor check STATUS of at-home portion to see if this needs to also be done on site for months 1-3. Reschedule if at-home section not completed for MV and month .
IJV Model Validation 1. Simulator Training: watch video and practice on SonoSimulator® 2. Demonstrate use of 7 skills 3. Simulator Assessment for right neck 4. Simulator Assessment for left neck 5. Survey	IJV Model Validation 1. Introductory Survey 2. Pre-Test: Fundamentals of Ultrasound 3. Fundamentals of Ultrasound - Study Version 4. Pre-Test: Introduction to Ultrasound-Guided Procedures - Study Version 5. Pre-Test: Ultrasound-Guided Internal Jugular Vein Cannulation - Study Version
IJV Month 0 (0 weeks) 1. Demonstrate use of 7 skills 2. IJV training: watch video and practice on SonoSimulator® 3. Month 0 assessment - right neck 4. Month 0 assessment - left neck	IJV Month 0 (0 weeks) 1. Introduction to Ultrasound-Guided Procedures 2. Ultrasound-Guided Internal Jugular Vein Cannulation
IJV Month 1 (actually 2 weeks) 1. At home portion if not completed 2. Demonstrate use of 7 skills 3. Month 1 assessment - right neck 4. Month 1 assessment - left neck	IJV Month 1 (actually 2 weeks) 1. Month 1 Post-Test: Introduction to Ultrasound-Guided Procedures - Study Version 2. Month 1 Post-Test: Ultrasound-Guided Internal Jugular Vein Cannulation - Study Version
IJV Month 3 (actually 6 weeks) 4. At home portion if not completed 5. Demonstrate use of 7 skills 6. Month 3 assessment - right neck 7. Month 3 assessment - left neck	IJV Month 3 (actually 6 weeks) 1. Month 3 Post-Test: Introduction to Ultrasound-Guided Procedures - Study Version 2. Month 3 Post-Test: Ultrasound-Guided Internal Jugular Vein Cannulation - Study Version
IJV Month 6 (actually 12 weeks) 1. At home portion if not completed 2. Demonstrate use of 7 skills 3. Month 6 assessment - right neck 4. Month 6 assessment - left neck 5. Survey	IJV Month 6 (actually 12 weeks) 1. Month 6 Post-Test: Introduction to Ultrasound-Guided Procedures - Study Version 2. Month 6 Post-Test: Ultrasound-Guided Internal Jugular Vein Cannulation - Study Version

UCLA

CRESST

NATIONAL CENTER FOR RESEARCH ON EVALUATION, STANDARDS, AND STUDENT TESTING

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)

Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522

(310) 206-1532
www.cresst.org