Computer Assisted Learning in Obstetric Ultrasound.

Brian Patrick Dromey

A thesis submitted for the degree of

Doctor of Philosophy

of

UCL

Research Department and Centre:

[Research Department of Maternal Fetal Medicine]

Principal Supervisor: [**Professor Donald M. Peebles**]

Secondary Supervisor: [**Professor Danail V. Stoyanov**]

Start Date: [**01.08.2018**]

Table of Contents

Declaration

I, Brian Patrick Dromey, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

I wish to thank my supervisors Professor Donald Peebles and Professor Danail Stoyanov. I will always be indebted for their constant encouragement, endless support and boundless patience. Without their motivation, wisdom and immense knowledge this thesis would simply not have been possible.

Besides my supervisors I would like to thank Professor Anna David. Her open-door policy and advice was very much appreciated throughout my PhD, but especially in the early months. Mr Raffaele Napolitano shaped this thesis and provided invaluable feedback during the upgrade viva, for which I am very grateful.

I would also like to thank Francisco, Eddie, Sophia, Lydia, Chiara Manuel and all my collaborators at WEISS. The breadth and depth of knowledge and enthusiasm for learning is inspiring and something I will try to embody throughout my career.

I would like to extend a special thanks to the clinical staff at UCLH who welcomed me warmly during uncertain and challenging times. It was a privilege and a pleasure to come to work alongside you all each day and I feel tremendously lucky to have worked with wonderful people and brilliant clinicians.

Lastly, I must thank my family. My mom Gertie for her boundless selflessness throughout my education and my Dad John for his proof reading. To my Fiancé Russell, without the confidence, self-belief and bottomless coffee you supply me none of the following 200 pages would have happened. Thank you.

Abstract

Ultrasound is a dynamic, real-time imaging modality that is widely used in clinical obstetrics. Simulation has been proposed as a training method, but how learners performance translates from the simulator to the clinic is poorly understood. Widely accepted, validated and objective measures of ultrasound competency have not been established for clinical practice. These are important because previous works have noted that some individuals do not achieve expert-like performance despite daily usage of obstetric ultrasound. Underlying foundation training in ultrasound was thought to be sub-optimal in these cases. Given the widespread use of ultrasound and the importance of accurately estimating the fetal weight for the management of high-risk pregnancies and the potential morbidity associated with iatrogenic prematurity or unrecognised growth restriction, reproducible skill minimising variability is of great importance.

In this thesis, I will investigate two methods with the aim of improving training in obstetric ultrasound. The initial work will focus on quantifying operational performance. I collect data in the simulated and clinical environment to compare operator performance between novice and expert performance. In the later work I developed a mixed reality trainer to enhance trainee's visualisation of how the ultrasound beam interacts with the anatomy being scanned. Mixed reality devices offer potential for trainees because they combine real-world items with items in the virtual world. In the training environment this allows for instructions, 3-dimensional visualisations or workflow instructions to be overlaid on physical models.

The work is important because the techniques developed for the qualification of operator skill could be combined in future work with a training programme designed around educational theory to give trainee sonographers consistent feedback and instruction throughout their training.

Impact statement

Obstetric ultrasound is essential to the safe and effective practice of modern obstetrics. Ultrasound can be used with minimal training to visualise the fetal heartbeat or the fetal presentation in labour. Assessments of fetal growth and wellbeing require training, rigorous assessment and ongoing quality control. Multi-centre, randomised-control studies, such as intergrowth-21 have established standards, tools and quality assessment methodology for the use of ultrasound in obstetrics. International organisations such as International Society of Ultrasound in Obstetrics and Gynaecology (ISUOG) and The American Institute of Ultrasound in Medicine (AIUM) have proposed training outcomes. Much less research interest has been spent understanding how trainees progress through their training and if there is a baseline assessment that represents competency.

This thesis sets out a metric which could be used to quantify a sonographer's performance over time. A series of experiments were undertaken in simulated and clinical environments to observe operator performance when performing fetal biometry. Clinically, these three standard planes are most commonly used for estimating fetal weight. Sonographers hand movements and images were recorded and analysed while they attempted to image these standard planes in the fetal head, abdomen and leg. Dimensionless Squared Jerk, a derivative of position with respect to time correctly classified expert and novice sonographers.

The second part of the thesis developed a mixed reality training system which combines a hitherto low fidelity Ultrasound training phantom with probe motion guidance superimposed into the users field of view. The use of mixed reality allows for real-time instruction. The Lightweight and unobtrusive Hololens2 is worn by the trainee. Hololens2 uses on-board cameras to optically track the position of the ultrasound probe and the known position of the standard planes in the phantom. The sonographer is presented with directional guidance based on the position of their hand/probe relative to the standard planes for the estimation of fetal weight. They are also given explanation as to why they are moving the probe in that direction and the expected effect it will have on their on-screen image. Mixed reality trainers offer trainee sonographers consistency of instruction and aims to given the sonographer an understanding of the relative position of the ultrasound beam to the fetal anatomy, which is fixed in nether time or space.

The methods developed in this thesis could combine to offer trainees and trainers a relatively affordable high-fidelity simulator which is used in conjunction with clinical ultrasound systems, unlike many of the commercial simulators which are available today. Objective measurements of competence such as DSJ offer objective assessment of progress in training for trainees and their trainers. These technologies increase training opportunities where time or resource are constrained, particularly in the developing world where access to trainers is even more constrained. The addition of machine-learning-based image analysis would allow for automated quality control of real clinical scans and even training opportunities without the need for a phantom at-all. In the future mixed reality would allow sonographers to receive guidance and feedback, in real-time, as they scan a clinical case.

Contributions; published and presented work from this thesis.

Quantifying Expert Performance in Obstetric Ultrasound and comparing it to Novice Performance : Data to Improve Training. Brian Dromey, Shahnaz Ahmed, Fransisco Vasconcelos, Evangelos Mazomenos, Danail Stoyanov, Anna David, Donald Peebles. BMFMS, Edinbugh, 2019.

1st Prize – Oral Presentation in Fetal Medicine Category. British Maternal and Fetal Medicine Society (BMFMS), Edinburgh, March 2019.

A Comparison of Expert and Novice Performance in Obstetric Ultrasound Using Probe Tracking Systems: Data to Improve Training. Brian P Dromey, Shahanaz Ahmed, Francisco Vasconcelos, Evangelos Mazomenos, Anna L David, Danail Stoyanov, Donald M Peebles. CRAS, Genoa, Italy 2019.

A Systematic Review and Meta-analysis of the Use of High-Fidelity Simulation in Obstetric Ultrasound. Dromey BP, Peebles DM, Stoyanov DV. *Simul Healthc*. 2021;16(1):52-59. doi:10.1097/SIH.0000000000000485

Simulation and Beyond–Principles of Effective Obstetric Training. Jaufuraully, Shireen, Brian Dromey, and Danail Stoyanov. *Best Practice & Research Clinical Obstetrics & Gynaecology*(2021).

AutoFB: Automating Fetal Biometry Estimation from Standard Ultrasound Planes. Bano, S., Dromey, B., Vasconcelos, F., Napolitano, R., David, A. L., Peebles, D. M., & Stoyanov, D. (2021, September). In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 228-238). Springer, Cham.

Dimensionless squared jerk: An objective differential to assess experienced and novice probe movement in obstetric ultrasound. Dromey, B. P., Ahmed, S., Vasconcelos, F., Mazomenos, E., Kunpalin, Y., Ourselin, S., ... & Peebles, D. M. (2021). *Prenatal Diagnosis*, *41*(2), 271-277.

Dimensionless Jerk - An Objective Differential to Assess Experienced and Novice Ultrasound Operators in a Clinical Setting.Brian P Dromey, Francisco Vasconcelos, Sebastien Ourselin, Anna L David, Danail Stoyanov, Donald M Peebles. ISUOG 2020

CAL-Tutor: Investigating the utility of a mixed reality platform for training in obstetric sonography. Brian Dromey, Manuel Birlo, Eddie Edwards, Anna L. David, Donald Peebles and Danail Stoyanov. MICCAI 2021

CAL-Obs, A Prospective Study Investigating The Use Of Dimensionless Square Jerk For The Assessment Of Expertise In Obstetric Ultrasound. Brian P Dromey, Francisco Vasconcelos, Lydia Neary-Zajiczek, Anna L David, Danail Stoyanov, Donald M Peebles. CLINICCAI 2021

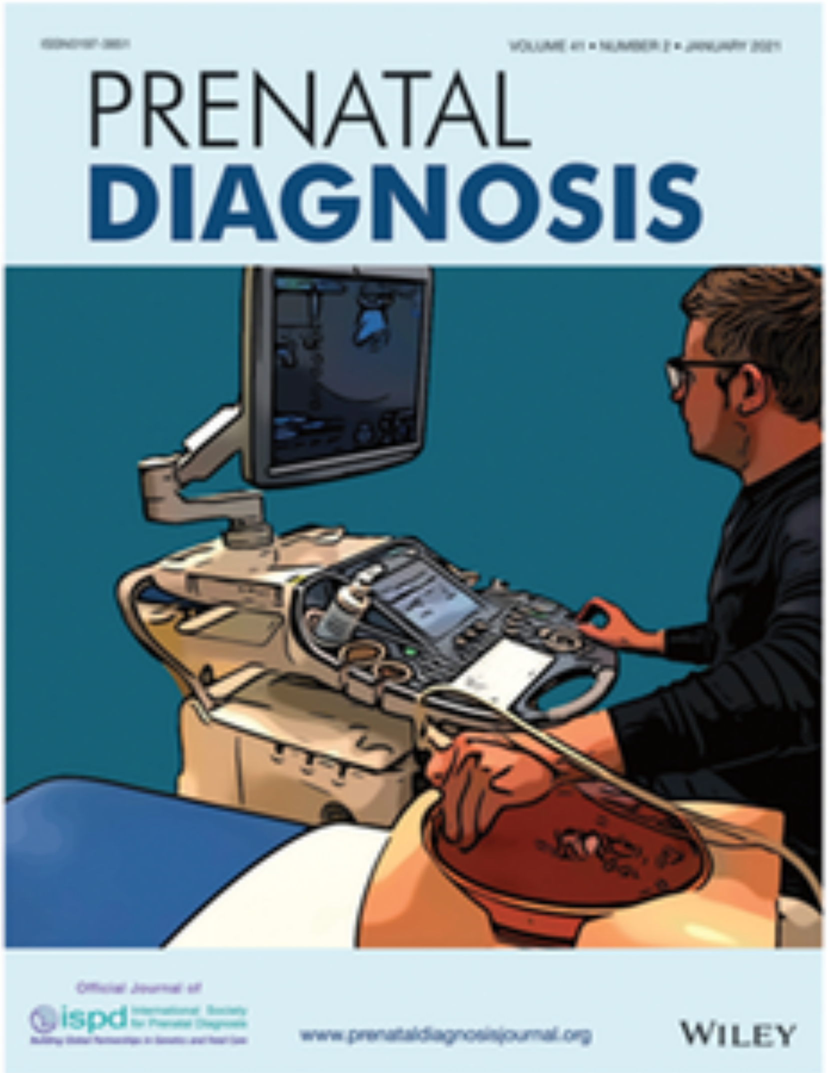
Deep learning-based plane pose regression in obstetric ultrasound. Di Vece, C., Dromey, B., Vasconcelos, F., David, A. L., Peebles, D., & Stoyanov, D. (2022). International Journal of Computer Assisted Radiology and Surgery, 1-7.

Automating Fetal Biometry Estimation from Standard Ultrasound Planes. Bano S, Dromey B, Vasconcelos F, Napolitano R, David AL, Peebles DM, Stoyanov D. AutoFB: In International Conference on Medical Image Computing and Computer-Assisted Intervention 2021 Sep 27 (pp. 228-238). Springer, Cham.

CAL-Obs : The use of mixed reality headsets for training in obstetric sonography Brian Dromey, Manuel Birlo, Eddie Edwards, Anna L.David, Donald Peebles and Danail Stoyanov. Oral Presentation, RCOG World Congress, London, June 2022.

CAL-Tutor: A mixed reality platform for training in obstetric sonography. Brian Dromey, Manuel Birlo; Philip J Edwards; Anna L. David; Donald Peebles; Danail Stoyanov. Oral Presentation at IPCAI 2022, Tokyo, Japan.
1st Prize – Digital surgery – Long Abstract

Journal covers from this thesis.



The cover image is based on the Original Article *Dimensionless squared jerk: An objective differential to assess experienced and novice probe movement in obstetric ultrasound* by Brian P Dromey et al., https://doi.org/10.1002/pd.5855.

Glossary

2D	Two Dimensions
3D	Three Dimensions
AC	Abdominal Circumference
AIUM	American Institute of Ultrasound in Medicine
ANOVA	analysis of variance
AR	Augmented Reality
ATSM	Advanced Training and Speciality Module
CNGOF	French National College of Gynaecologists and Obstetricians
CRL	Crown Rump Length
DSJ	Dimensionless Squared Jerk
EBCOG	European Board and College of Obstetrics and Gynaecology
FL	Femur Length
HC	Head Circumference
HEE	Health Education England
HMD	Head Mounted Display
IUGR	Intra Uterine Growth Restriction
ISUOG	The International Society of Ultrasound in Obstetrics and Gynaecology
MERSQI	Medical Education Research Study Quality Instrument
MIS	Minimally Invasive surgery
MR	Mixed Reality

NHS	National Health Service
NICU	Neonatal Intensive Care Unit
PACS	Picture Archiving and Communication System
PICO	Population, Interventions, Comparisons & Outcomes
PPIAG	Patient and Public Advisory Group for Guided Instrumentation for Fetal Therapy and Surgery (GIFT-Surg) Study
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PROSPERO	International Prospective Register of Systematic Reviews
RCOG	The Royal college Of Obstitricians & Gyanecologists
REC	Research Ethics Committee
UCLH	University College London Hospital
US	Ultrasound
USU	Obstetric Ultrasound Screening Unit
VR	Virtual Reality

1 Introduction

1.1 Background - the use of ultrasound in obstetrics.

A recent consensus statement considered ultrasound (US) essential to the safe, timely and effective practice of Obstetrics and Gynaecology[1], but acknowledged that training is challenging. In obstetrics, ultrasound can be used in acute clinical care to perform basic tasks such as confirmation of the fetal heartbeat or assessment of fetal presentation. Away from the delivery suite, intermediate level skills include monitoring fetal growth and wellbeing. Advanced applications include the diagnosis of major congenital abnormality, generally performed by doctors with a specialist interest in fetal medicine. These uses have a higher training demand and require ongoing assessment of competency and quality assurance[2]. Advanced uses are beyond the requirements of most sonographers and are best concentrated in specialised units where a relatively high caseload can be maintained. For context congenital abnormalities are found in approximately 1% of livebirths. They represent 10% of the total admissions to the neonatal unit (NICU). In one series the authors reported that congenital abnormalities accounted for 13% of the total NICU patient-days, 26% of the total NICU mortality and 32% of all deaths within the first year. In contrast, the authors reported that 35% of all NICU infants had intrauterine growth retardation (IUGR)[3] in their cohort. Growth restricted and small for gestational age fetus are over-represented in terms of NICU admissions and contribute more patient-stays than congenital abnormalities. The overwhelming majority of livebirths, 99% will not require advanced obstetric ultrasound skill for the diagnosis and treatment of a structural abnormality. Nevertheless, robust training and robust referral pathways for expert opinion when abnormalities are suspected is required. Besides structural abnormalities, a very large burden of NICU workload, morbidity and mortality is due to fetal growth restriction. The focus of this thesis will be on training in the context of normal anatomy, allowing trainees to gain an appreciation of the appearance of normal fetal anatomy and the ability to accurately measure fetal weight. Therefore, I will focus on training outside the context of specialised centres and considering how performance can be quantified, training and skills certified. The ultimate aim is that this data would contribute to trainee assessment. This approach is inkeeping with broader medical skills training, as doctors are expected to recognise the limitations of their training and competence. Medical professionals are required to practice in this way and escalate cases when the limits of their abilities are reached. Conversely without education in basic ultrasound skill and consistent use of ultrasound in clinical practice the trainee will not have the confidence to use ultrasound and their skills will wane rather than develop. In the context of the UK system this may only mean a delay of a small number of hours or days while the woman waits for an appointment with a sonographer. In the global context this could mean that the woman is denied any kind of ultrasound at all or be sent to a referral unit which may be distant from her home.

1.2 Clinical Challenges for trainees in Obstetric Ultrasound

Training, certification and quality control are of upmost importance, but are poorly standardised across the ultrasound community. Training is challenged by a number of issues. These are discussed in the context of the UK system where most ultrasound examinations are performed by sonographers, rather than doctors. While in many European countries and in North America ultrasound is performed by the woman's obstetrician or midwife the delivery of care within the NHS is delivered by a team of healthcare professionals. Women who are classified as "low risk" at the time they reported their pregnancy to their local midwifery team may not see a doctor for the duration of their pregnancy. When triaged by their midwife to see an Obstetrician the woman will generally be referred to a hospital, or care team rather than a named individual. Thus, they may see a number of doctors within the hospital, rather than seeing an individual doctor. Nevertheless, there is an increasing expectation that obstetricians would be able to perform basic and intermediate ultrasound in the acute setting. Indeed, it is a requirement for the progression through specialist training in the UK[4]. The training in obstetric ultrasound has been identified as problematic, not only in the context of the RCOG curriculum for trainees, but across the speciality globally. More specifically the issues are considered in the context of assessment of fetal wellbeing. These skills are considered "intermediate" by the RCOG training curriculum for Obstetricians & Gynaecologists[4]. There are several factors that trainees find difficult. These relate to the training but also maintenance of US skill.

- Time available to trainees away from their "day job". This is a problem across medical specialities.
- Time for trainers to train the trainees. Many consultant posts include time for personal development and the supervision of trainees, but training lists or lists with reduced capacity to allow for instruction and discussion of cases are much less common. These have been used widely in training undergraduate medical students and GP training.
- Steep learning curve and diagnostic nature of US means that trainees provide limited clinical service to the training organisation until late in their training.
- Rotational nature of UK training means that the plateau is rarely reached in a single unit. Initiatives such as lead employers could allow trainees to return to a designated training site to complete their training. This model could be adapted from the ATSM programme for UK trainees in the last two years of their clinical training.
- The majority of UK obstetric sonography is provided by sonographers rather than doctors. These are not rotational posts and the training investment is retained within the hospital.
- The skill set of doctors in training allows them to assist in a number of locations. This might include in-patient wards, surgical theatre, labour ward, ante-natal clinic or gynaecology out-patient clinics. This breath of

skills is combined with 88% of units reporting gaps in their middle grade rota in 2017[5]. Cross covering and being 'pulled' from a training or educational session to cover service provision is unsurprising and often unavoidable. Middle grade doctors tend to be the most flexible to redeploy at short notice and many can work without direct supervision in a number of the clinical settings already mentioned.

- Lack of standardisation and certification standards make it difficult to retain skills with UK trainees when they move hospitals as trainees are not expected to be able to contribute to service provision.
- UK trainees have competing, mandatory training demands which are required to progress through training waypoints.
- In the UK the demand for private ante-natal sonography services is limited as comprehensive screening for fetal abnormality is carried out free of charge at the point of care. This is different to many other countries where the majority of sonography is provided by private healthcare providers. There is not the same incentive for trainees to scan if the health service or patients do not reward or recognise this skill through career progression or financial reward.

1.3 How historical training patterns contributes to issues in contemporary training.

When considering the training issues as listed above it is clear that many of the issues around training, competence and certification are shared across surgical, or practical medical specialities. Some of these have been discussed in the context of surgical training. Obstetric ultrasound shares some parallels with surgical specialities, in particular laparoscopic surgery. Ultrasound examinations, much like minimally-invasive surgery require the operator to interpret a dynamic image produced by the three-dimensional (3D) position and motion of the ultrasound probe by means of a two-dimensional (2D) visual display. The analogy continues. The operator must form a diagnosis based on the information presented to them on-screen. Views of the anatomy may be sub-optimal or the procedure technically challenging. The trainee may be affected by fatigue or stress. The barriers to safe, reproducible practice have been considered extensively in minimally-invasive surgery[6,7]. Training challenges have also considered the impact of reduced training hours[8]. The effect of which has led to concerns that current specialists-in-training will have had fewer training hours than their predecessors[9].

Historically, an apprenticeship style of training was commonplace, particularly in surgical specialities. Trainees learned from watching and listening to the master, a more senior surgeon. Working time directives and industrial agreements have made consistent team-based structures impractical. These structures are no longer common in contemporary UK practice. Thus, training must evolve, recognising that the continuity for trainers and trainees once provided by the firm structure has been lost. The loss of a firm-based structure is not inevitably negative for trainees, who may find themselves exposed to a breath of approaches and techniques during their training. During training, the assessment of technical proficiency is based upon the subjective opinion of senior colleagues, which is subject to bias and inconsistency. Exposure to a wider pool of trainers could be beneficial in this scenario. The maintenance of a procedure logbook by the trainee will not often consider the quality of the trainees performance or fully detail the specific involvement of the trainee. Both methods therefore lack validity and are unable to robustly assess a surgical trainee's capability[10]. Structured curricula with specific competencies have been developed in medical and surgical specialities, including obstetrics and gynaecology.

Training Body	RCOG	ISUOG
Module Name	*Basic Fetal Assessment.*	*ISUOG Education Committee recommendations for basic training in obstetric and gynecological ultrasound*
Tasks	Be able to perform and interpret standard fetal measurements: • HC (BPD) • AC • FL	Obtain standardized planes for anatomical and biometric evaluation • BPD, HC • AC • FL • Cervical length
Assessment	3x OSAT completed by senior clinician, at least one of which should be a consultant.	A minimum of 100 hours of supervised scanning, to include: A minimum of 100 obstetric scans covering a wide spectrum of obstetric conditions

Table 1.1 – Adapted table of ISUOG[11] and RCOG[12] US curricula in basic ultrasound. The skills expected to be performed at the basic level and their assessment criteria are included.

Such curricula are an opportunity to capture the breadth and depth of a trainee's experience over their career. Several have been developed specifically for obstetric ultrasound. Two are compared in Table 1.1. The details serve to highlight the similarities of the tasks the trainee is required to learn, but differences in the assessment of the overall clinical competency between the organisations.

1.4 Using simulation as a training tool

The currently available curricula have been developed within the constraints of contemporary delivery of medical care in the UK. Because these curricula face many competing challenges, the majority have been agreed by consensus. A pragmatic approach to the difficulties faced by trainees gaining adequate time in the ultrasound department, rather than derived from educational theory was required. Consequently, most remain to be validated and correlate poorly with trainee confidence and clinical competence for independent practice. This is particularly true for obstetric ultrasound, taken in the context of UK post graduate doctor training, which is the focus of this thesis and the environment in which the studies were carried out.

In the 2019 curriculum the RCOG has made efforts to introduce pragmatic training requirements and reduce overly prescriptive and "tick-box" annual appraisal requirements. Despite this, there remains a great dissatisfaction expressed by trainees[6] in respect of their training and confidence in obstetric ultrasound. Simulation has been proposed as a solution to improve training and to reduce training time. For this approach to be effective training needs and certification need to be identified in an objective way. This basic certification ought to be transferrable between units. Although considerable investment in simulation has been made by UK training bodies, there is little validation of the efficacy of the purchased systems. Use by trainees is limited and falls short of the projected demand for simulator facilities[13]. The lack of construct or face validity implies that the current generation of simulators cannot be used to identify areas of strength or weakness in trainee performance[14]. This may contribute to trainees' expression of enthusiasm for simulation facilities, but there is limited evidence that trainees utilise them, find them useful and therefore benefit from the availability of simulators for learning.

1.5 Objective metrics & learning curves to define progress in training

When considering training the concept of "practice makes perfect" is widely understood. Unfortunately, perfection is difficult to define and impossible to achieve. Educational theory proposes that all learned skills reach a plateau, eventually. The point where further investment of time does not improve performance[15,16]. The rate of change of skill can be graphed against time. This is called a learning curve. Most novices should achieve the same skills performance plateau of experienced surgeons with continued practice [15,16].

Defining the learning curve for any surgical skill proves challenging. It seems intuitive to plot time on the X axis, but the metric for the Y axis is less clear. The metric should be objective and reproducible. Within the literature, the most commonly reported metrics are path length of the laparoscopic instrument or ultrasound probe, time to completion of task and "complication" rates. Some studies have also considered trainee-reported feedback. The rate at which the convergence towards expert performance occurs and the metrics that most reliably reflect the trainee competence from conscious incompetence to unconscious competence are the initial work in this thesis. The following work explores how a clearer understanding of learning curves could be used to improve training in obstetric ultrasound.

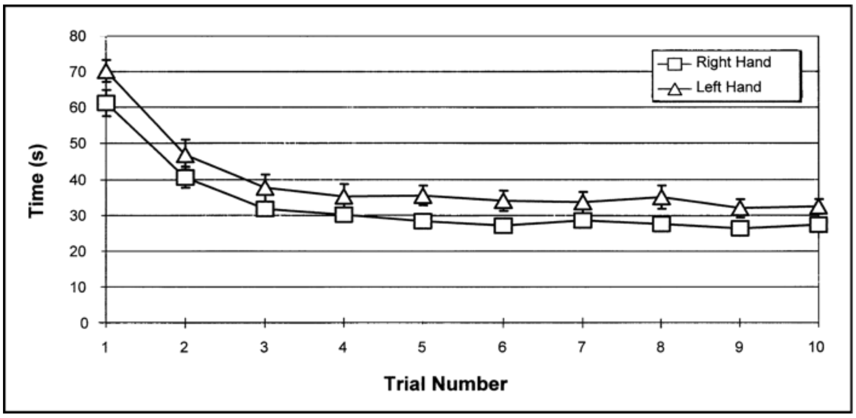


Figure 1.1 - Trial Vs Time to completion of task, adapted from 'Sensitivity of Smoothness Measures to Movement Duration, Amplitude and Arrests' Hogan and Sternad, 2009[17].

Path length has been discussed at length in the literature. While time can discriminate between expert and novice operators, alone it does not offer any comment on the quality of the image obtained or the overall competency of the operator [18],[19]. Time to task completion demonstrates that the operator has performed the task quickly, but no comment is offered on the competence or the confidence with which the task was completed. *Figures 1.1* & *1.2* show the concept of a learning curve, while performance improves over time, the progress is not necessarily linear. The authors also raise the possibility that performance might be positively, or negatively, impacted by a number of confounders. van Epmell et at considered the path length travelled by a laparoscopic instrument[18]. They recorded how it was affected by increased familiarity with the task, but also confounded by requiring it to be performed by both the left and right hand. The experiments were not performed while the operator was distracted, disturbed or tired, factors which could also affect performance and may begin to explain why learning curves aren't linear and performance appears to vary slightly between attempts.

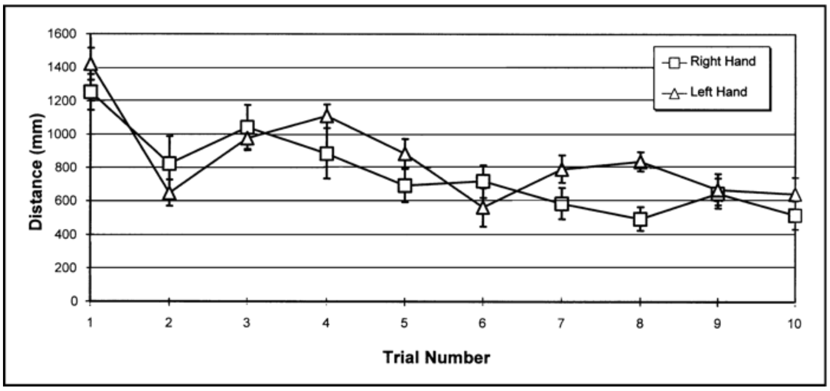


Figure 1.2 – Trainees were asked to repeatedly perform laparoscopic tasks, in this series they were asked to perform the tasks with both their left and right hand. Over time their distance travelled by the laparoscopic instrument decreased. This is not a linear process and not achieved at an equal rate by all trainees. Adapted from 'Sensitivity of Smoothness Measures to Movement Duration, Amplitude and Arrests' Hogan and Sternad, 2009[17]

Various calculations and methodologies have been proposed to quantify motion metrics. All of these metrics seek to link the intentionality of hand movements to expertise. Many of these consider smoothness of the operator movement, rate of acceleration, number of movements made during a task and jerk[17]. The overriding theme being that expert operators intuitively understand the consequences of each movement their hand makes. This understanding means that each movement has a reasoning. The effect is that operation time and path length are shortened. I hypothesise that if the intentionality of the operators movement can be measured, I can investigate how it relates to the expertise of the operator.

1.6 Proposing a novel objective metric for assessment of training

I want to establish a definition of expert versus novice performance in a simulated setting when performing obstetric ultrasound. Beyond defining expert versus novice performance in a simulated setting I aim to correlate this to clinical performance. Computer-based systems are ideal to record and display data relating to an individual's training over time. This data can display a personal learning curve for the training task. This can be achieved with multiple data streams because computer-based systems can simultaneously record large amounts of data for individual users. This data can also be pooled between groups of trainees to allow for benchmarking performance against colleagues or the wider training community. Such a dataset could provide a basis for the establishment of defined thresholds for expert-like behaviours when performing obstetric ultrasound. Over time a cohort of novices could be followed through their training. This would allow me to define a learning curve for a number of parameters for the acquisition of obstetric ultrasound skill. The PhD aims to define bookends around progression through training and establishing learning curves. Beyond this I want to understand how it might be possible to benchmark trainee progression through training against a validated, objective metric learning curve. A novel approach, aiming to teach trainees expert-like behaviours and integrating training into clinical workflow is somehow required. I aim to impart an understanding of how the ultrasound beam interacts with the fetal anatomy, the effect of probe motion on the on-screen image and the optimisation of the image once the ideal plane has been obtained. I aim to use computer learning and vision techniques to accelerate the training process and thereby reduce barriers to training. This approach requires an understanding of the objective measures of operator performance which can be used to assess baseline performance, monitor the trainee learning curve and benchmark them against expert operators.

1.7 Dimensionless Squared Jerk as a candidate metric.

My first research objective is to identify objective hand motion metrics, that could be used to demonstrate differences between novice and expert sonographers. To achieve this a series of metrics ought to be proposed, tested and confirmed or refuted as useful in the assessment of obstetric ultrasound skill. As discussed earlier, when a new skill is being learned, a learning curve can be plotted. I have explained the concept that skill eventually reaches a plateau, assuming continual practice of that skill[15,16].

Furthermore, novice performance should converge towards expert performance as time goes on[16]. A number of metrics have been explored in the literature in respect of surgical skill and parallels were established between surgical skill and obstetric ultrasound skill.

A number of objective measurements have been proposed. The most commonly reported are; path length of the laparoscopic instrument or ultrasound probe, time to completion of task and “complication” rates. Authors have proposed various calculations and methodologies to quantify motion and experience in an objective way. Many of these relate to smoothness of the operator movement, rate of acceleration, and the number of movements made during a task and jerk[17].

Simpler measurements such as the number movements required to complete a task seem attractive as they are easily understandable. Reproducibly counting and recording them has proved challenging and poorly discriminatory for experience. ‘Intentional movement’ is an evolution of this idea. This requires a threshold under which movement is accidental, inconsequential and can be disregarded from datasets. It has been described as the number of deliberate movements above a threshold acceleration value. One study considered an arbitrary value of 10 m/s^2 as the threshold value to detect deliberate hand movements of a surgeon[20]. Another study considered the minimum acceleration value from each participant as the threshold value to detect the deliberate hand movement[21]. These selections were arbitrary, a key limitation of use of this parameter for skill assessment. Furthermore, these parameters have not been demonstrated to be reproducible for defining expert and novice operators.

Movement smoothness is a quality related to the continuality or non-intermittency of a movement. Movement intermittency is directly related to the movement’s temporal organization or coordination. Thus, a valid smoothness measure must change monotonically to changes in movement intermittency. Intermittency in this context refers to movements that alternately decelerate and accelerate, and more intermittency corresponds to un-smooth movements. It is, therefore, a reflection of the intentionality of operators hand movement[22]. Smoothness reflects a lack of hesitancy, implying intention and thus experience. Movement intermittency is typically observed as dips in a movement's speed profile or finite non-zero periods of zero speed (i.e. movement arrest) during an ongoing movement. A dip in a speed profile is a point where the second derivative of position goes to zero, at this point there is no acceleration. It highlights a period of deceleration followed by acceleration, which is a mark of movement intermittency. All of the measurements of smoothness explored in this thesis are independent of movement amplitude and duration, i.e. *dimensionless*.

Jerk, the time derivative of acceleration has been investigated to quantify smoothness and coordination and proposed as a surrogate for experience. In physics, jerk is defined as a rate of change for acceleration. Therefore, it is a derivative of acceleration with respect to time and distance and as such is the second derivative of velocity or the third derivative of position. Hogan and Sternad noted that jerk could quantify the smoothness of motion related to hand coordination, with superior thoroughness[17]. The sensitivity needed to be dimensionless, so that there would be no natural dependency of movement duration, extent, and spurious peak[23]. This has been shown in endoscopic surgery[16] and has been shown to reliably differentiate novice from experienced operators[20].

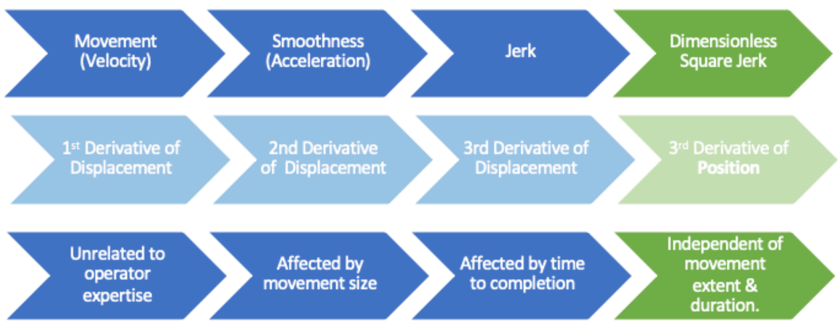


Figure 1.3 – Comparison of metrics proposed in the literature for the assessment of operator hand movements.

I propose to investigate Dimensionless Square Jerk, DSJ, as a suitable metric to evaluate operator experience. DSJ has been selected because it is independent of movement duration and amplitude. It is independent of time. It is described by Hogan and Sternad[17] and compared to four other measurements of jerk described in the literature [24]. DSJ was designed to be less dependent on time and to place more emphasis on movement. In physics, jerk is defined as a rate of change for acceleration. Therefore, it is a derivative of acceleration with respect to time and distance. It is the second derivative of velocity or the third derivative of position. Hogan and Sternad demonstrate how several squared-jerk measures of smoothness vary with movement duration *D* while movement amplitude *A* and the number of "hesitations" *n* remain constant. The figure is reproduced here as *Figure 1.4*. The main point to note is that, while the several jerk measures with units (integrated squared jerk, mean squared jerk, mean squared jerk normalized by peak speed, mean squared jerk normalized by mean speed) differ in their precise details, all exhibit an essentially similar variation with movement duration: as movement duration increases, all these measures become dramatically smaller. In contrast, DSJ is independent of movement duration. This is crucial, as discussed earlier that speed is not demonstrably related to competence, quality or confidence. In the context of my research, it would be expected that familiarity with performing ultrasound and the US equipment may allow the operator to complete the examination more quickly. DSJ untethers the intentionality of the movement from the speed of completion of a task.

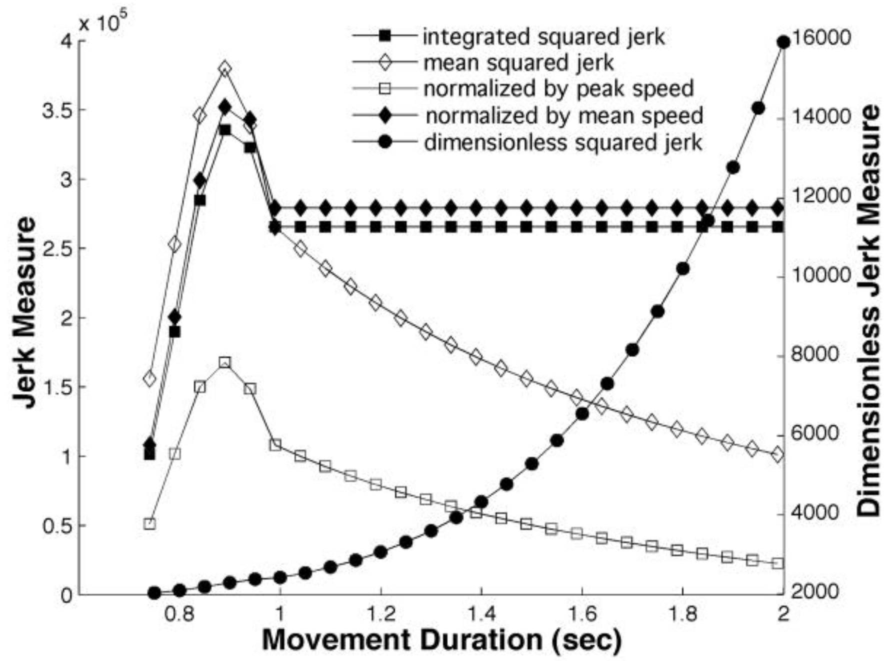


Figure 1.4 - A graphical depiction of how several squared-jerk measures vary with movement duration. DSJ increases monotonically with the temporal separation of movements. Demonstrating that as the time between the movement grows larger the motion profile becomes less smooth.

Hogan and Sternad further examined the effect of small, unintentional movements on measurements of jerk. *Figure 1.5* shows how several squared-jerk measures vary with movement duration. Each movement lasted 500ms and was separated by variable amounts of time. The time interval varied from shorter than the intentional movements to greater than the movement itself. When the two submovements overlap in time, all four of the measures with units (integrated squared jerk, mean squared jerk, mean squared jerk normalized by peak speed, mean squared jerk normalized by mean speed) exhibit a non-monotonic variation with duration. When the two submovements are separated in time, integrated squared jerk remains constant—the rest periods contribute nothing to the integral—while mean squared jerk declines with total movement duration. Similarly, mean squared jerk normalized by mean speed does not vary as separation increases while mean squared jerk normalized by peak speed declines with total movement duration. In contrast, DSJ increases monotonically with the temporal separation of the two submovements. Unlike the four measures with units, this measure increases monotonically even when the two submovements overlap. It also continues to increase when the two submovements are separated in time. This properly reflects the change of movement shape with duration. Demonstrating that, as the separation of submovements grows longer, the movement profile becomes progressively less smooth. In practical terms this equates to the hesitancy moving the US probe or laparoscopic instrument. The less smooth the movement by the operator, the higher the DSJ measurement will be.

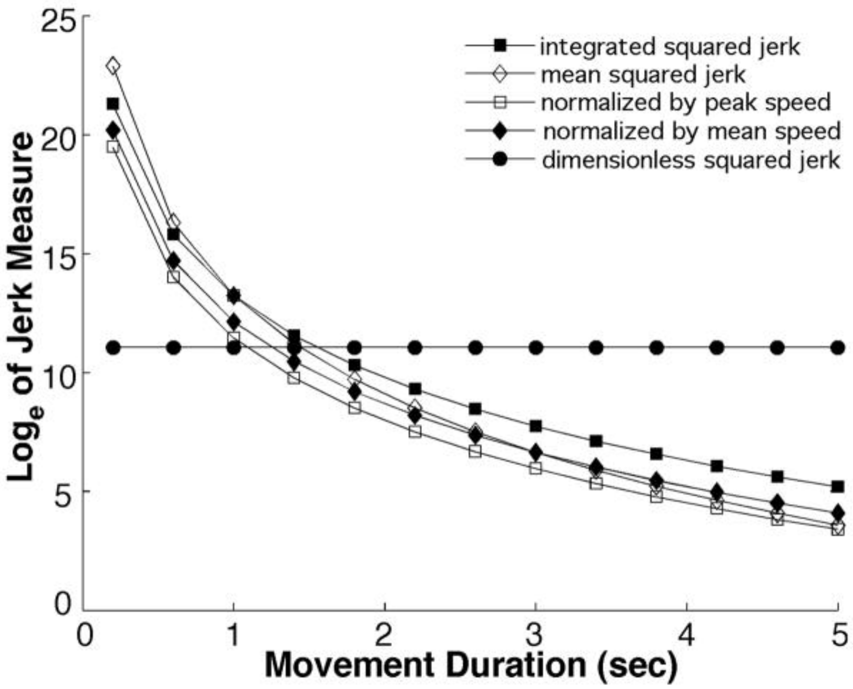


Figure 1.5 - Variation of several measures of smoothness based on squared jerk with movement duration while movement amplitude and shape (number of speed peaks, deviation of speed profile from a cycloidal movement) remain constant. For graphical presentation the ordinate is plotted on a logarithmic scale.

1.8 Research Hypothesis & Objectives

I hypothesise that computer assisted methods can be used to quantify competence in obstetric ultrasound. Further, that this quantification can be used to enhance the effectiveness of simulation-based training.

The objectives for this thesis are:

To identify objective hand motion metrics, that could be used to demonstrate differences between novice and expert sonographers.

To quantify novice and expert performance levels in simulated and clinical environments.

To identify if probe-handling skills acquired in simulation can be transferred effectively into the clinical environment.

To quantify the effect of training on image quality, based on a validated QC scoring system.

To examine the use of contemporary Artificial-Intelligence and Augmented-Reality methods in training and competence assessment.

To explore how such contemporary methods can be developed to deliver enhanced training in the clinical environment.

1.9 Conclusions

There are training issues which are likely to be unique to training in obstetric ultrasound. In the first instance these must be explored. The literature review discusses these further and explores simulation as a potential tool for the training of obstetric sonographers. The subsequent hypothesis is based on the assumption that training in obstetric ultrasound and surgery have parallels. To test this, I investigated DSJ in the context of obstetric ultrasound. Some of these data sets have already been published in peer-reviewed international journals. Having established that this relationship appears to be valid, later chapters will discuss the experiments undertaken to establish DSJ and other metrics, which illustrate the learning curve in obstetric ultrasound. Once these metrics have been established, further understanding is required to manipulate these learning curves with the aim of improving training for clinicians. This lead to the development of a Mixed-Reality training system, CAL-Tutor.

Finally, I explain how the last phase of my PhD uses this knowledge of learning curves to assess the training impact of an Augmented-Reality ultrasound trainer.

2 Literature review - "A Systematic Review and Meta-analysis of the Use of High-Fidelity Simulation in Obstetric Ultrasound"

2.1 Introduction

In the introductory chapter I discussed barriers to training in obstetric ultrasound. Traditional teaching of ultrasound, like surgery, has taken the form of "see one, do one, teach one"[25], initially under the supervision of a more experienced operator. The outcomes of curricula, where present, are often based on the number of clinical cases completed, rather than objective outcomes of competence[26]. The current assumption is that competence is directly related to clinical experience, or numbers of cases completed. However, this is not necessarily the case. Tolsgaard et al[27], remarked that some experienced clinicians did not display expert-like behaviours despite daily use of obstetric ultrasound in their clinical practice. The authors hypothesized that poor basic training may be a root cause of this, suggesting that the operators did not have the correct foundation to benefit from later clinical training. The authors further hypothesized that the expected improvement in performance was not seen because sustained, deliberate practice rarely occurs in clinical practice. The findings undermine the basic premise of training.

Ultrasound examinations, much like laparoscopic surgery require the operator to interpret a dynamic image produced by the three-dimensional (3D) position and motion of the ultrasound probe by means of a two-dimensional (2D) visual display. It has been reported that laparoscopic skill and performance metrics improve with training and experience[9]. Similarly, it might be expected that an ultrasonographers' performance would improve with training and practice. It is hypothesized that as a novice gains experience and familiarity with a technique that their performance evolves[28]. This is often referred to as a learning curve. I have discussed learning curves in Chapter one. Establishing a learning curve is difficult and specific to a single task. A learner's progress along the curve is not always linear, the reasons for this are complex, related to familiarity with the task at hand, limitation of the surgical equipment used and an appreciation of normal anatomy.

Simulation has been proposed as a strategy to shorten skill acquisition time and to allow clinicians learn in a safe, blame-free environment. Given the complexities inherent in clinical practice the acquisition of basic skills without exposing trainees or patients to unnecessary risk seems a sensible approach. Simulation suites are commonplace in UK teaching hospitals[13]. Despite the large investment in simulation facilities by HEE, uptake has been disappointing [13]. This might be because little attention has been focused on how to effectively integrate simulation into modern training curricula. A recent survey of UK trainees in Obstetrics & Gynaecology reported that 79% considered simulation essential for training in ultrasound and that 90% would participate in a formal simulation-based training program. When provided, 76% of trainees found the simulator useful for improving clinical skills. 54% never, or rarely, used the ultrasound simulation facilities available to them, citing a lack of formal guidance; unawareness of facilities; inconvenient access times, clinical workload and time pressures as barriers to participation[13]. In short simulation facilities are nice to have but often an imperfect and impractical solution for trainees with high clinical workloads and limited protected learning time.

The aim of this review chapter is to investigate the use of high-fidelity simulation in obstetric ultrasound, to identify its usability for learners and to establish if the skills obtained in a simulated environment can be translated to improved clinical performance.

The central question in this review is: *Do training tools enhance training and prepare trainees for clinical practice?*

The secondary questions are if skills can be transferred to the clinical setting and if transferred skills are robust and sustained in the medium and long term?

2.2 Methods

Protocol & Registration

A systematic review was conducted of the currently available literature. The review was completed in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards for quality of reporting systematic reviews[29]. The protocol was registered on the International prospective register of systematic reviews (PROSPERO)[30] database in February 2019 as, "High-fidelity ultrasound simulation in obstetric ultrasound. Serious training tools or gaming toys? A review of the current literature", reference number CRD42019122974. The registered protocol is available on the Prospero database at https://www.crd.york.ac.uk/prospero/.

2.2.1 Eligibility Criteria

Studies considering the use of simulators in the training or assessment of ultrasound operators were eligible for inclusion. The PICO (Population, Interventions, Comparisons and Outcomes) model was considered when designing the search strategy[31]. The population was any trainee in ultrasound, these individuals may be doctors or allied health professionals. Interventions considered suitable were any use of a simulator, either before commencing clinical training or concurrent with clinical training. Suitable comparators included cohorts not trained on simulators, either in a parallel or crossover design. Outcomes showing a positive, negative or no correlation on performance after the use of ultrasound simulators were considered suitable for inclusion.

2.2.2 Information Sources

The search was completed on 30th of October 2018. The search strategy used four database search tools, PubMed, EMBASE, Scopus and Web of Science. Publications for inclusion were identified using the search terms "Simulat*" & "Training" & "Obstetric*", either as keywords or contained within the manuscript tittle. The “obstetric*” wildcard was used to capture variations including “obstetrician”, “obstetrics” and “obstetric”. “Simulat*” wildcard was used to capture variations such as simulated, simulation and simulator. The search terms were combined using the Boolean operator “OR”. The search was limited to articles in English and duplicates were removed by the author (BD) as part of the screening procedure to assess full-text articles for inclusion. No further papers were identified by examining the bibliography of the papers read in full.

2.2.3 Search

The literature search process is represented in *Figure 2.1*. 2,581 records were identified. 2,470 were excluded by screening the titles of the abstract. The reasons for exclusions were Non-English, Different Topic, Non-Obstetric Ultrasound, Conference/Congress Abstract (full text not available) and Communication to Editor. From a pool of 2,581 results 111 results were retrieved from the search engine results for screening. Once duplicates were excluded and abstracts were examined for relevance 39 papers were deemed suitable for inclusion. Three full-text articles were excluded as the content was not relevant to simulation in ultrasound.

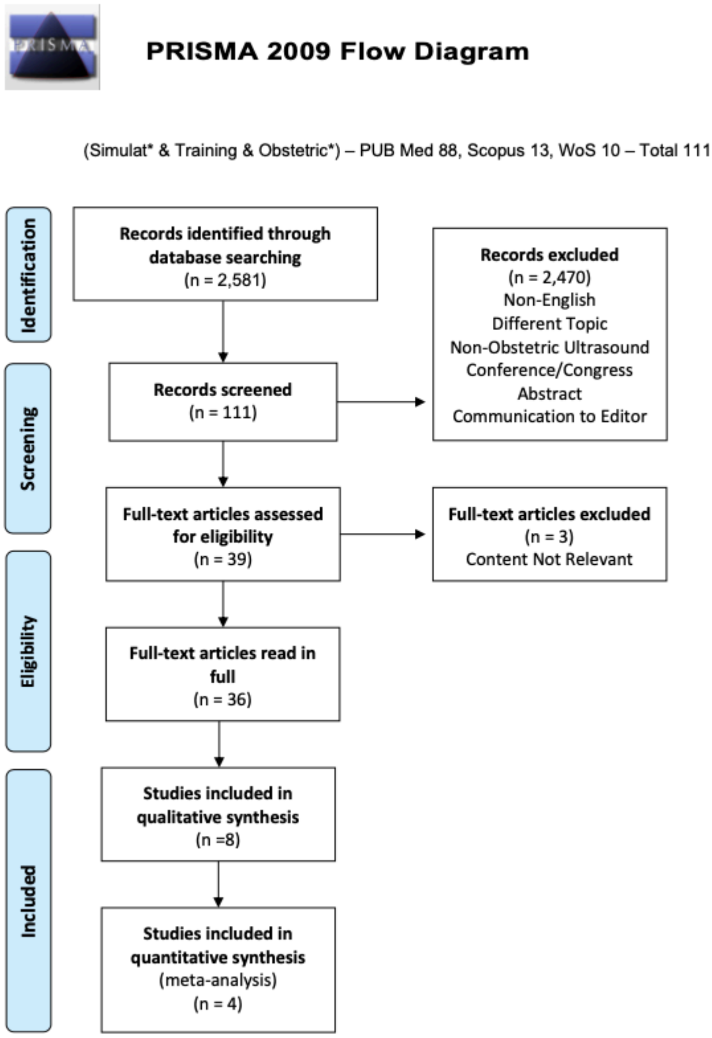


Figure 2.1 - PRISMA Flowchart detailing the search process used for the literature search.

2.2.4 Study Selection

The remaining 36 articles were read in full. The motivation for this review was, as stated earlier, to determine if the literature has reported behaviours which could be used to establish the utility of simulators in obstetric ultrasound training. Studies which considered the use of high-fidelity simulators in ultrasound were considered for inclusion. The concept of 'fidelity' refers to the realism of a particular simulator, or how closely the simulator replicates the task being learned. All simulators replicate one, or more, parts of a clinical task for the purposes of education. High fidelity simulators generally have some degree of computer control, interactivity or trainee feedback. High fidelity simulators are thought to increase realism and to have greater educational value because of this. Although there is wide variation in the design of ultrasound simulators all are, by their nature, high fidelity simulators. No studies were excluded based on the type of simulator used.

Studies examining the use of simulators in obstetric ultrasound or systematic reviews on the topic were eligible for inclusion. All the studies included for analysis included novice operators in their study population. Study design was varied. Authors chose to compare novice and expert performance when using a simulator, while others chose to observe novice behaviour before and after using a simulator. Studies were not excluded based on the type of medical professional selected to form the novice/inexperienced group as we recognize that obstetric ultrasound is performed by clinicians from a variety of backgrounds, including radiology, obstetrics, midwifery and by sonographers.

No studies were excluded based on their date of publication, as commercially available, high fidelity ultrasound simulators are relatively new to the market. All studies were published between 2002 and 2018.

Studies were excluded if their primary outcomes were not in obstetric ultrasound. Studies were also excluded if the study did not include an educational intervention using a simulator. Although ultrasound validation studies were included in the qualitative analysis, these were excluded from the quantitative analysis as the primary outcome measured simulator performance rather than the learners change of performance.

2.2.5 Data Collection Process

Two researchers independently reviewed the 36 full-text articles. Discrepancies were resolved by discussion of the validity of the methods and quality of the content within the manuscript. After discussion, eight studies were included in the qualitative analysis (Burden[32], Todsen[33], Chalouhi[14] Pittini[34], Jensen[35], Madsen[36], Monsky[37], Maul[38]), four studies were included in the quantitative analysis as four studies did not report findings in a format suitable for inclusion in the meta-analysis.

2.2.5.1 Data Items

A database of the 36 included papers was created using Microsoft Excel. For each full-text article read, the following data were recorded; Title, Author, Article Title, Journal Title, Keywords, Problem Statement, Research Method, Statistical Methods Used, Number of included participants, Author Conclusions, Findings in relation to past research, reviewer summary and reviewer notes.

2.2.5.2 Statistical Analysis - Risk of Bias

As part of the data collection and meta-analysis analysis process included studies were scored using the Medical Education Research Study Quality Instrument (MERSQI) tool[39]. MERSQI is an instrument developed for measuring the quality of education research studies. The maximum score is 18, made up from the flowing domains, Study design (3), Number of institutions sampled (1.5), Follow-up (1.5), Outcome assessment (3), Validity evidence (3), Data analysis (3) and Outcome type (3). A score of ≥12 is considered an indication of high study quality. The MERSQI authors describe their assessment of 210 medical education research studies published in 13 peer-reviewed journals. Over a fifteen-month period the mean MERSQI score was 9.95 (SD, 2.34; range, 5-16). We calculated the mean MERSQI score for included manuscripts of 11.88 (SD, 1.81; range, 9.5-15). In this context the articles included are, at least, reflective of study quality seen in broader medical education.

2.2.5.3 Statistical Analysis - Summary Measures & Synthesis of Results

Review Manager 5.3[40](The Cochrane Collaboration, 2014.) was used to produce forest plots of the included studies. Meta-Essentials [41] running on Excel (Microsoft Excel for Mac Version 16.32) was used to perform the meta-analysis and to calculate the sensitivity and specificity of each included study. The results are shown in *Figure 2.2,* finding favourable effect for improved accuracy of biometry in obstetric ultrasound following simulation training.

All the included studies had similar methodology and all included novice participants. In all studies a group of novice operators was asked to complete a specified training package. Their performance was compared before and after completion of the training package. No study compared novice with expert performance, either before or after the training. No study compared objective clinical performance before and after training. All studies were completed in a training centre, or simulation suite, none were undertaken in a clinical area. Measures of heterogeneity indicated moderate heterogeneity. Cochrane's Q value was calculated at 6.73.

Eight studies were included in the qualitative analysis, all eight studies recruited doctors. None of the included studies recruited nurses, sonographers, midwives or students. Five studies recruited doctors from Obstetrics & Gynaecology[32,33,14,36,38] with the remaining studies recruiting trainees from Emergency Medicine[33], and Radiology[37]. One study recruited any post graduate year 0-5 doctor[34]. The calculated I^2 value of 40% indicates moderate heterogeneity between the studies, despite difference in design, methodology and reporting. In total six models of simulator were used, UltraSim, VimedixTM US simulator, Canadian Amnio Model, Scantrainer, UltraSim and SonoTrainer. A summary of the findings of the qualitative analysis is presented in *Table 2.1*.

2.3 Results

Table 2.1 – Summary of the qualitative analysis of the included manuscripts. The table includes the stated purpose, design and findings of each study.

Study Author	**Burden**[32]	**Lous**[26]	**Chalouhi**[14]	**Pittini**[34]	**Jensen**[35]	**Madsen**[36]	**Monsky**[37]	**Maul**[38]
Year	2013	2017	2016	2002	2018	2014	2002	2004
No of participants	26	30	29	30	25	28	16	45
No of Experts	8			15	0	12	0	
No of Novice	18	30	29	12	25	16	16	21
Purpose of the study	To assess the usability of virtual-reality (VR) simulation for obstetric ultrasound trainees.	assess progress made in the ultrasound (US) measurement of femur length (FL) by students after one hour of training on US obstetric simulators.	To test the validity of an obstetrical ultrasound simulator as a tool for evaluating trainees following structured training	How to teach procedural skills without compromising patients health	To investigate the learning curves for novices training the FAST protocol on a virtual-reality simulator.	To assess the validity and reliability of performance measures	The purpose of our study was to evaluate the effectiveness of a sonographic simulator in evaluating residents before their taking overnight call.	To evaluate the effectiveness of the SonoTrainer ultra-sound simulator as a training method for first-trimester screening
Participant group	Obstetrics & Gynaecology Trainees	Obstetric Ultrasound Trainees	Obstetric Ultrasound Trainees	Doctors – PGY 0-5	Emergency Medicine Trainees	Obstetrics & Gynaecology Trainees	Radiology Trainees	Certified Obstetricians
Study design and control	Comparative Study, no control	Prospective Single Centre Study	Comparative Study, no control	Prospective Single Centre Study	Case Control	Comparative Study, No control.	Case-control observational	Case-control with intervention applied. One control group

MERSIQ Score	11.5	10.5	11	11	13.5	15	9.5	13
Type of simulated training program (intervention)	Fetal Biometry Early Pregnancy	FL at 20 weeks	LF, HC, AC. 4 Chamber, cardiac apex, heart crux, pulmonary vein, descending aorta, ROI.	Amniocentesis	TA USS (FAST)	TV USS	TV and TA USS	TA USS
Type of simulator	UltraSim	VimedixTM US simulator	Vimedix	Canadian Amnio Model	Scantrainer	Scantrainer	UltraSim	SonoTrainer
Trimester of pregnancy	1, 3	2	2	2	1	1	1	1
Time interval between training and testing	5 attempts. First and last measurements taken	Immediately before and after 1 hour training session.	Nil. Training completed prior to attending for exam.	Immediately before and after training session.	Immediately before and after training session.	2 months	10 hours of self directed study	Unclear - ? Education modules completed.
Alternate training program as comparator to simulated training	Nil	Nil	Nil	Nil	Nil	Nil	Conventional clinical training	Theoretical Training Package
Outcome Measures	Difference in time and accuracy of CRL, BPD, OFD, FL and Time between first and last scans	Time to obtain femur length before and after training. Femur Length Measurement obtained.	Score-based evaluation of 6 morphological planes in fetal cardiac imaging	Score-based evaluation of Ultrasound guided Amniocentesis	Time to achieve mastery learning level, corresponding to the performance level of a group of ultrasound experts.	Which metrics distinguish reliably between expert and novice	Mean Test Scores Comparing Two Consecutive Classes of Eight Residents	Differences in Nuchal Translucency (NT), Crown Rump Length (CRL) and Mean duration per examination.

The results of the meta-analysis find that superior performance has been achieved after training using high-fidelity ultrasound simulation. All the evaluated results considered performance before and after a training event using an ultrasound simulator, as shown in *Figure 2.2*.

As detailed in the methodology, eight studies Burden[32], Todsen[33], Chalouhi[14], Pittini[34], Jensen[35], Madsen[36], Monsky[37]and Maul[38] were included in the qualitative analysis. Five outcome measures from four studies were included in the quantitative analysis[32,33,37,38]. In total 214 participants were recruited to the four studies, 129 were novice participants (56%). All four studies reported positive effect on operator performance. Specifically, the performance improvements were noted in the measurement of Crown Rump Length (reported in three studies) and in Femur Length (reported in two studies). These improvements were seen, regardless of the model of simulator used. Across the eight studies six models of simulator were used.

All studies had similar aims, but the subsequent training or instruction differed. All studies established baseline performance for each user and all studies did this using a simulator. All studies used a single model of simulator. The participants undertook assessment and training on the same model of simulator. Studies by Burden[32] et al, Todsen et al[33], Chalouhi et al[14], Pitttini et al[34] and Jensen et al[35] required participants to attend a single simulator session, these studies did not compare simulator-based training to other training methods.

Study name	Heges'g	CI Lower limit	CI Upper limit	Weight
Burden (2013) CRL	-0.428301	-1.300339	0.4157542	15.61%
Monsky et al (2002) CRL	-1.544364	-2.33813	-0.820886	17.91%
Maul (2004) CRL	-0.755759	-1.302792	-0.226732	25.39%
Lous (2017) FL	-0.421251	-0.942697	0.0891016	26.40%
Burden (2013) FL	-1.020013	-1.951358	-0.155312	14.68%
Total (95% CI) n = 214				

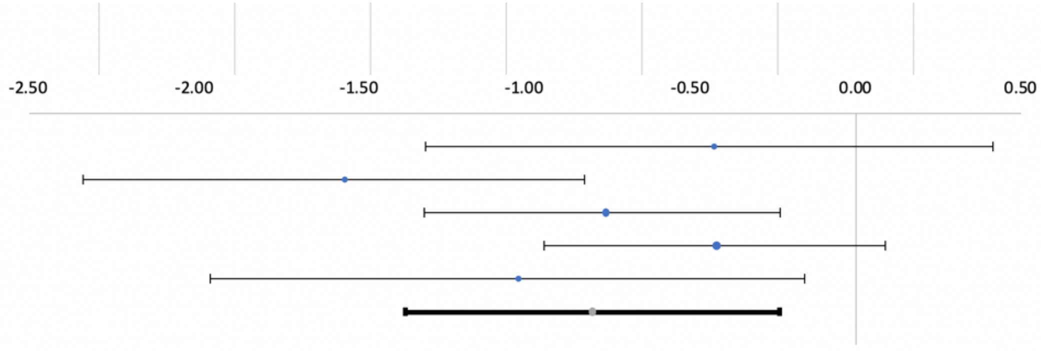


Figure 2.2 – Forest plot diagram of Meta-Analysis. Four studies reported outcomes of fetal biometry which were suitable for inclusion in the analysis.

Madsen et al[36] repeatedly assessed participants over two months while Monsky et al[37] required participants to compete ten hours of self-directed learning using the simulator and compared final performance to doctors of similar grade who had not participated.

Three studies examined operator performance in the first trimester of pregnancy measuring the Crown Rump Length (CRL). The remaining two studies examined performance in fetal biometry in the second trimester. The meta-analysis is detailed in *Figure 2.2.* One study specifically reported Femur Length but other measures of fetal biometry were not reported. Some studies used expert operators as a control group. One study compared the use of a high-fidelity ultrasound simulator to a theoretical training package, one study compared 10 hours of self-direct learning using the UltraSim to conventional clinical training.

As stated earlier, the aims of this review were to investigate the use of high-fidelity simulation in obstetric ultrasound, to identify its usability for learners and to establish if the skills obtained in a simulated environment can be translated to improved clinical performance, which is sustained over time. The papers included in the qualitative review have been scored against these aims in *Table 2.2.* The study design used by authors predominantly focused on the functionality and usability of ultrasound simulators. Most studies have not focused on how skills are translated from the simulation suite into the clinical environment, how the acquired skills translate to clinical practice and if the skills are maintained over time.

Study Author	Burden[32]	Lous[26]	Chalouhi[14]	Pittini[34]	Jensen[35]	Madsen[36]	Monsky[37]	Maul[38]
Year	2013	2017	2016	2002	2018	2014	2002	2004
1. The use of high-fidelity simulation in obstetric ultrasound,	Paper specifically used obstetric ultrasound.	Fetal biometry using obstetric ultrasound.	Measurements of biparietal diameter, abdominal circumference, and femur length as well as reference planes for cardiac 4-chamber and outflow tracts, kidneys, stomach/diaphragm, spine, and face.	Amniocentesis, not specifically examining ultrasound skills.	Simulator used, but for the FAST protocol. FAST protocol can be used to identify free fluid in the abdomen associated with Ectopic pregnancy.	Transvaginal US – mainly used in early pregnancy or Gynaecology.	General Radiology.	Paper specifically used obstetric ultrasound.
2. To identify usability for learners	The user questionnaire identified timetabling issues as a barrier to training.	All participants were preparing for national exam.	All participants were preparing for national exam.	A curriculum was constructed to allow trainees gain skill.	90% of participants completed the learning program.	Mapped performance metrics shipped with the system to expert performance.	Completion was mandatory at departmental level.	Comparison of a didactic training curriculum with an integrated alternative.
3. To establish if the skills obtained in a simulated environment can be translated to improved clinical performance.	Improved performance on the simulator, but no clinical correlation undertaken.	Shorter image acquisition time, higher skill and all trainees used zoom function after training in assessment.	Similar dexterity scores were achieved, higher scores for image quality in assessment.	Checklist and training improved performance in simulated setting.	The 'mastery' level was benchmarked against experts asked to perform the FAST protocol on that particular simulator.	With time and repeated practice improved performance on the simulator was shown.	The participants had to be assessed as competent before being allowed to perform independent ultrasound in clinic.	The expert level was benchmarked against experts asked to perform ultrasound on that particular simulator.

4. Can skills be transferred to the clinical setting?	No clinical correlation undertaken.	No clinical correlation undertaken. Technique and image optimization better after training.	Not directly asked or answered.	Unclear	Mastery level achieved by the trainees is based on expert performance.	Unclear.	Better performance was seen in the group exposed to simulator training than the group not given US simulator training.	Unclear - assumed
5. if transferred skills are robust and sustained in the medium and long term? "	No follow-up cohort	No follow-up cohort.	No follow-up.	Assumed, but no follow-up of the trainees.	No follow-up.	No follow-up.	No statistical significance between two groups, but subjective reported assessment and trainee confidence improved.	No follow-up.

Table 2.2 – Tabulation of the qualitative analysis of each of the included papers against the aims of the review. The use of simulators by learners and the motivations for learners to use the simulators have been considered by all authors. Some consideration has also been given to how the learner can be assessed in the simulated environment. Only Monsky et al considered how the skills acquired in the simulated setting compared with those acquired by learners who had not been exposed to simulation.

2.4 Discussion

All the included studies look to validate the concept of using simulation for training or assessment in obstetric ultrasound. The meta-analysis shows that skills can be acquired, improved and assessed by means of a high-fidelity simulator. In particular, the findings suggest that simulation can be best be used for acquisition of technical skills[32] and image optimisation[42]. Superior technical ability may accelerate a learner's time to competence[36]. The review of the literature finds that simulation training can be used to equip novice ultrasound practitioners with sufficient skills to perform basic obstetric ultrasound in a clinical environment under direct supervision.

Our findings suggest that consideration ought to be given to integrating simulation training into the clinical curriculum. Even in research settings trainees reported clinical commitments as barriers to engaging with simulation training[13]. The highest levels of engagement, 90%, were seen when participation was mandated by the faculty by Monsky et al[37]. The authors undertook simulator-based assessment of Radiology Residents before taking overnight call. The authors were surprised to find that their findings challenged established beliefs within the radiology department that Residents were suitably and adequately trained prior to taking up semi-autonomous clinical practice. The participant survey also highlighted Residents' concerns about their own preparedness for overnight calls. As a result, the authors modified the Residency training program at their hospital. The redesigned curriculum addressed these concerns. An additional 8 weeks of targeted, clinical training, focusing specifically on transvaginal ultrasound was provided. Twelve months later, the experiment was repeated. The authors found that residents performed significantly better on the simulator and reported higher confidence in performing ultrasound. Senior clinicians also reported higher subjective performance scores for Residents when being assessed.

Studies by Bernardi et al[42] and Maul et al[38], showed that even novice operators could achieve competent performance in obstetric ultrasound when being trained by means of simulation alone. The authors compared their simulation-based curriculum to conventional didactic teaching of ultrasound theory and practice.

The example of simulator use in pilot training is often used as justification for the use of simulation in medical education. It is true that high fidelity simulators are universally used for training airline pilots. When considering the use of simulation in medicine it is important to understand that full-motion flight simulators are integrated into pilot training, assessment and licensing. Initial pilot training and recurrent assessment in a simulator take place every six months for commercial pilots. Mandatory emergency simulator sessions allow trainers to create an entirely immersive experience, recreating the systems and motion of the aircraft and the human factors which have been recurrent contributors to accidents and near-misses. None of the simulators described to date have addressed the clinical context in which the trainee will eventually work. The current devices focus on technical skills proficiency, while ignoring communication with patients and colleagues, distractions and clinical management which contribute to overall clinical performance. Our review finds that that trainees in obstetric ultrasound can benefit from the use of a high-fidelity simulator but that these tools are not formally integrated into medical education curricula. It is preferable that training programs be based on objective outcomes, rather than trainer reports and arbitrary numbers of cases recorded in a logbook.

We suggest that high-fidelity ultrasound simulation can be used to train users more quickly. However, this review is limited by the heterogeneity of the evidence base. The wide disparity in maternal-fetal medicine training curricula globally is reflected in the heterogeneity of the studies and reported outcomes. These limit the generalizability of the results. We were able to include four studies and a total of 214 participants in the meta-analysis. Even with these limited numbers we were able to show a positive effect for simulation training. The positive result may reflect that by using a simulator the participants were gaining tuition and experience that they would not otherwise have been exposed to. The effects seen might be attributable to additional intentional practice, rather than the simulator itself. Because all studies carried out baseline assessment, training and subsequent assessment on the same model of simulator, it is possible that the results reflect user familiarity with the simulator, rather than a true improvement in clinical skill. The limitations of the study highlight the need for future research to consider how skills acquired in the simulation setting translate to a clinical setting.

2.5 Conclusion & Relationship to the current work

The results of the review finds evidence of benefit for high-fidelity ultrasound simulation. The evidence for deployment in training is limited, but some authors have found their own training curricula challenged by the introduction of simulation-based training and assessment. In these instances, simulation has been used to augment traditional learning, with a strong focus on specific, objective and measurable clinical outcomes, audit and revision of the curriculum based on learner feedback.

Further investigation of ultrasound simulation in training should follow models closer to pilot training, were training and ongoing assessment are routine, mandatory and completed by all grades. The challenges of inertia to change, suspicion of simulation as a valid means of learning can be challenged by considered design of further studies now that the utility and validation of this equipment is established.

It is likely that simulation is best considered as a process to allow the learner to transition to semi-autonomous practice in a supervised, clinical setting. By integrating ultrasound simulation into training curricula and promoting self-directed learning trainees could contribute to the clinical service while learning a complex skill. Integrating ultrasound training into clinical workflow would allow us to establish if skills acquired in the simulated environment correlate with clinical performance and if skills are maintained in the longer term, which has been poorly considered by the literature to date.

The literature review demonstrates that simulation can be an effective training tool. The lack of validation and limited understanding of how best to deploy investment in simulation are two areas which require further research. The following chapters aim to improve this understanding, defining metrics which could be used as metrics for training. These must be defined in the simulated environment and in the clinical scenario, with a broader understanding gained of how performance in a simulator reflects performance in a clinical scenario.

3 Dimensionless squared jerk: An objective differential to assess experienced and novice probe movement in obstetric ultrasound.

3.1 Introduction

Ultrasound is a dynamic, real-time imaging modality that is widely used in clinical obstetrics for screening of congenital anomalies, to monitor fetal growth and to assess fetal well-being. The quality of an ultrasound examination is known to be operator dependent and to have high inter-operator variability[43]. This directly affects the information available to the clinician. Consequently, training and competence assessment are of great importance to ensure effective, reproducible and safe clinical practice.

The earlier chapters have presented how the acquisition of complex skills and psychomotor tasks follows a learning curve. Ultrasound examinations, much like in minimally invasive surgery (MIS), require the operator to interpret a dynamic image produced by the three-dimensional (3D) position and motion of the ultrasound probe by means of a two-dimensional (2D) visual display. The challenge of delivering high quality training leading to safe, reproducible practice has been considered extensively in minimally-invasive surgery[6,7]. Obstetric ultrasound requires a detailed understanding of the capabilities and limitations of ultrasound equipment, as well as knowledge of normal maternal and fetal anatomy. This must be combined with an appreciation of abnormal findings and their implication on fetal or maternal outcomes.

The literature review highlighted parallels between performance of ultrasound a laparoscopic surgery. Namely, that US examinations, much like MIS approaches require the operator to interpret a dynamic image produced by the three-dimensional (3D) position and motion of the ultrasound probe or camera by means of a two-dimensional (2D) visual display. The experienced operator will be able to combine this knowledge with the technical ability to manipulate transducer and display settings to optimise the image, leading to higher objective image quality. It is accepted that performance improves with training and experience[9,44,8], but this observation has not been objectively measured or classified. This chapter aims to set out how expert performance differs from the novice. Metrics including time taken to complete the study, the distance the probe travelled during the examination, quality of the images obtained and smoothness of probe movement were considered.

Inspired by comparable training challenges in endoscopic surgery and the approaches used in that speciality, an experiment was designed to observe operator behaviour during the performance of obstetric ultrasound, in a simulated environment.

Dimensionless squared jerk (DSJ) has been proposed as an objective parameter to discriminate between expert and novice operators in MIS[45]. DSJ is a measure of deliberate hand movements and is a derivative of acceleration with respect to time and distance while remaining independent of spurious peaks and dimension within the movement[17]. DSJ is normalised with respect to time duration, so that it is only the shape of the trajectory, not the dimension or extent that contributes to the metric. DSJ is calculated as the squared third derivative of tool or probe position with respect to time, or the squared derivative of tool acceleration, multiplied by a ratio of total movement amplitude and time duration such that the result is dimensionless and independent of total distance travelled and elapsed time DSJ quantifies common deviations from smooth, coordinated movement and it has been accepted as an objective parameter to quantify hand motion in different disciplines, such as parkinsonism, kinetics, and optometry[23,46]. In healthcare it has been used to differentiate between expert and novice endoscopic surgeons[17]. To my knowledge, quantification of US operator performance using DSJ has not previously been published in the literature. I hypothesised that dimensionless squared jerk could be used to differentiate experienced from novice operators. I have compared the results between experienced and novice ultrasound operators with the aim of establishing specific, measurable and reproducible performance differences between them.

3.2 Methods

We undertook a prospective, observational study of medical practitioners who were accredited specialists or specialty trainees in Obstetrics & Gynaecology at University College London Hospital NHS Foundation Trust, London, UK (UCLH). Participants were empirically divided between either experienced in fetal ultrasound (n=10, >200 fetal ultrasound examinations, 'experienced') or novice operators (n=10, <25 ultrasound examinations, 'novice'). These numbers were selected as The European Board and College of Obstetrics and Gynaecology (EBCOG) guideline on obstetric training recommends that trainees complete a logbook as part of their training which contains 200 obstetric ultrasound examinations[47]. The study was undertaken at the Obstetric Ultrasound Unit at UCLH. Clinical experience varied from Foundation Year 2 Doctors to Consultants in Fetal Medicine. The study was exempt from review by the NHS Research Ethics Committee as it was not performed on patients. Informed consent was obtained from each participant.

Each participant was asked to obtain standard 2D fetal measurements using a clinical GE Voluson E8 ultrasound machine (GE Healthcare, Chicago, Illinois, United States) on a commercially available second trimester phantom (SPACE-FAN ST, Kyoto Kagaku Co., Ltd, Kyoto, Japan). The phantom was chosen as it simulates the fetal skeleton and the key structures required to obtain technically adequate cross-sectional images. The simulator consists of a fetal skeleton fixed within an oval shaped abdomen. Prior to commencing the scan each participant was given written instructions, including images of the required planes, these are shown in *Figure 3.1,* and up to five minutes to familiarise themselves with the operation of the ultrasound scanner. Both novice and expert operators were allowed unlimited time to complete the task. Participants were permitted to refer to the instructions at any time during the task, if they wished to.

Figure 3.1 - Images of the fetal anatomy in the second trimester phantom. Image A represents the Transventricular Plane, used for measurement of the Biparietal Diameter (BPD), Image B represents the Transabdominal Plane, used for measurement of the Abdominal Circumference (AC), Image C represents the required view of the Femur used to measure the length of the bone (FL).

The required views were:

- The trans-ventricular plane, with bi-parietal diameter (BPD) measurement *(Figure 3.1 A)*.
- The trans-abdominal plane, on which the Anterior Posterior Abdominal Diameter (APAD), the Transverse Abdominal Diameter (TAD) and the Abdominal Circumference (AC) can be measured. *(Figure 3.1 B)*
- A view of the femur and a measurement of its length (FL). *(Figure 3.1 C)*

These views were chosen as these are the views necessary to calculate fetal weight, in accordance with Hadlock B formula and are in accordance with guidance on the estimation of fetal weight from ISUOG[2]. The Hadlock B formula was used to calculate the Estimated Fetal Weight as it is widely used to estimate fetal weight and it has a lower error rate than other formulae[48].

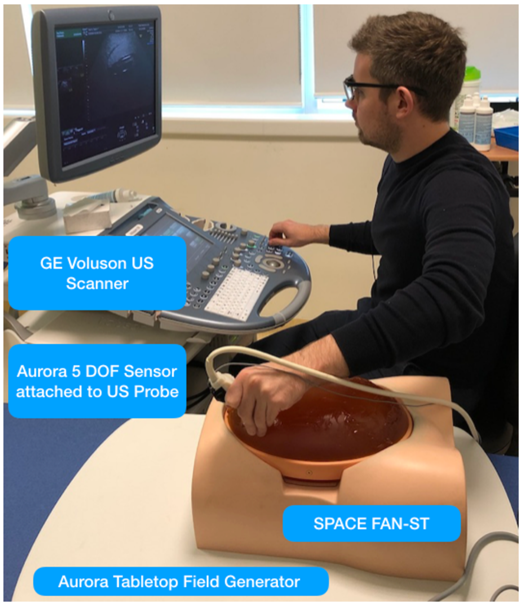


Figure 3.2 (a) The experimental set-up, showing the training phantom, ultrasound scanner and the electromagnetic probe tracking system.

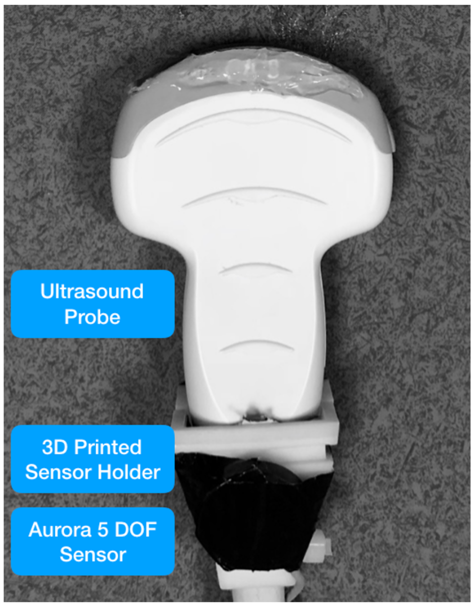


Figure 3.2 (b) A close-up of the ultrasound probe with an Aurora 6-DOF attached using a custom 3-D printed holder.

The position of the probe was tracked using the Aurora electromagnetic tracking system (NDI Inc, Ontario, Canada). The experimental set-up is shown in *Figure 3.2(a).* A close-up of the probe and the custom 3-D printed sensor mount is shown in *Figure 3.2(b)*. To track the probe a 6 degree-of-freedom (DOF) sensor was attached to the ultrasound probe by means of a custom 3-D printed holder, as shown in *Figure 3.2(b).* The sensor used was 6DOF sensor, (Aurora Part Number: 610066), in practice, only translation (3 DOF) was used to compute dimensionless jerk and path length. The phantom was placed on an Aurora planar field generator on the movable couch used clinically to simulate the exact position of a patient, this is shown in *Figure 3.2(a)*. This allowed the position of the probe to be recorded using NDI's "Aurora ToolBox" software. Each image was saved to the memory of the ultrasound scanner and later all captured images were transferred to the research database. The measurements (BPD, AC and FL) obtained by each participant were compared to the specifications provided by the manufacturer of the phantom, assumed to represent the ground truth. Two clinicians (BD and YK) were asked to score the images individually for quality control in fetal biometry, using the scoring system described by Salomon et al[49].

Transventricular Plane (HC): One point was assigned for each of Symmetrical Plane, Visible Thalamus, Cavum Septi Pellucidum Visible, Absence of Cerebellum, Head occupying >30% of the Image (Zoom).

Abdominal Circumference (AC): Symmetrical Plane, Stomach Bubble Visible, No Kidney Visible, Abdomen Occupying >30% of the image (Zoom), Callipers Placed Correctly.

Femur Length FL: Both ends of the bone visible, angle <45 degrees, Femur Occupying >30% of the Image, Callipers placed correctly.

Both reviewers were blinded to the participants clinical experience or whether the images being scored were from the novice or expert groups. The maximum possible score was 15.

Dimensionless squared jerk (DSJ) [17,50] was calculated for each novice and expert participant, using the formula described by Hogan and Sternad.

$$Jerk(i) = \frac{\left(\int_{t1}^{t2} \dddot{x}(t)^2 dt\right) D^5}{A^2}$$

Where: *A* is movement amplitude or extent, t1 = initial time; t2 = final time; t2 - t1 = completion time; D = path length, x is position in 3D and d= $t_2 - t_1$ is duration

3.3 Results

	Expert ±SD	**Novice ± SD**	***P***
BPD (mm)			
58 (Manufacturer Spec)	60.21±1.54	62.59±3.21	0.03
AC (mm)			
177 (Manufacturer Spec)	181.52±4.05	183.12±5.95	0.79
FL (mm)			
37 (Manufacturer Spec)	44.32±1.87	44.04±2.50	0.56
Estimated Fetal Weight			
515g (Hadlock)*	633.3±33.94	630.7±78.18	0.93
Image Score	11.8±1.87	10.2±1.46	0.04
Time (Sec)	176.46±47.31	666.94±490.36	0.0004
Probe Path Length (mm)	521.23±27.41	2234.82±188.50	0.007
Dimensionless Jerk	19.26 ± 3.02	22.08 ± 1.05	0.01

Table 3.1 - Mean biometry results, image results, estimated fetal weight, time to completion and path length for Expert and Novice Operators.

Experienced operator performance was associated with a shorter time to task completion and probe travel. On average a novice operator took almost four times longer to perform the task. The mean experienced operator time to completion was 176.46 seconds (SD 47.31) while in novices it was 666.94 seconds (SD 490.36, p=0.0004). In the novice group the ultrasound probe travelled a further distance on the phantom, on average 3.3 times as far 2234.82mm SD 188.50, versus 521.23mm SD 27.41, p=0.007 for experienced operators.

When evaluating the images used to estimate fetal weight in accordance with the scoring system detailed in the methodology, we found that experienced operators achieved a higher quality image score (mean score 11.8, SD 1.87) than novice operators (10.2 SD 1.46, p=0.04) (*Table 3.1*). We found that more experienced operators achieved a higher score for the submitted images, but this did not correlate with the accuracy of assessment of the Abdominal Circumference (AC) or Femur Length (FL). The estimated fetal weight did not differ significantly between the groups 633g SD 33.94 and 630g SD 78.18, p = 0.93 for the respective groups.

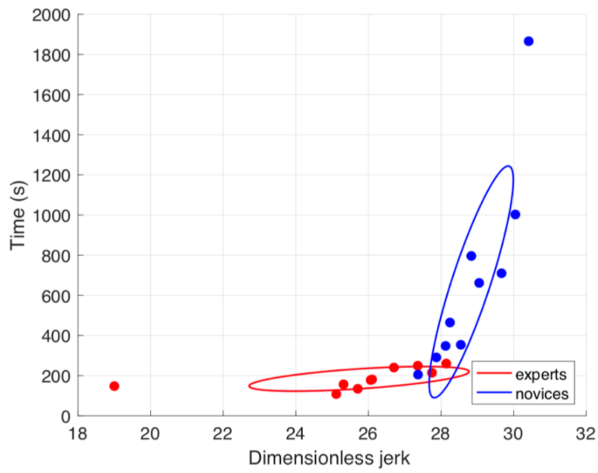


Figure 3.3 - Dimensionless Jerk Plotted Against Time.

The mean DSJ for experienced operators was 19.26 SD 3.02 and 22.08 SD 1.05 for the novice group respectively (p = 0.01). The plotted DSJ against time to task completion is shown in *Figure 3*.3. The DSJ value and time taken are smaller for the experienced group and larger for the novice group. The covariance of the data is represented by the overlaid ellipses. Each ellipse contains results within an 95% confidence interval. The experienced and novice groups show little overlap and a distinct pattern of distribution, as seen in *Figure 3.3*.

3.4 Discussion

Our results indicate that metrics such as time to completion, probe path distance travelled and DSJ show significant difference between novice and experienced ultrasonographers. When time and dimensionless squared jerk are considered together, the experienced group had lower dimensionless squared jerk, indicating that experienced operators had less unwanted and more purposeful movements than novices. Our results allow us to conclude that performance changes with experience can be quantified using objective measures.

The outliers in both the novice and expert groups are likely to represent the extremes of their respective groups. The expert outlier represents performance of obstetric biometry quickly and efficiently while the novice outlier represents an individual relatively inexperienced in performing these tasks. On the contrary it is the area where the two ellipses overlap that represent interesting areas for future research. The DSJ offers an objective measurement which could be combined with more traditional assessments such as trainee logbooks, patient feedback, audit of clinical images and trainer assessments such as OSATs to determine progression through training. As we reported in our literature review[51], it is worth noting that some clinicians will never achieve expert levels of performance despite daily use of ultrasound. We remind the reader that each ellipse contains results within an 95% confidence interval and, as such, outliers are to be expected. Our aim at this stage is not to try to move the outliers towards the mean but rather identify the likely extremes we might encounter if this methodology were to be used in a clinical setting and where a threshold between expert and novice performance might lie.

It is reassuring that our findings reflect the findings of similar studies describing the use of DSJ in minimally invasive surgery, where the path length of the arthroscope, the time taken to complete the task and the dimensionless squared jerk values were significantly different between experienced and novice groups. The authors expressed a view that the assessment of technical skills alone is unlikely to give an assessor a complete picture of clinical performance[50]. In obstetric ultrasound, particularly when used as a point of care examination this is likely to be more pertinent. Further management of the pregnancy is likely to be influenced by the ultrasound findings. Technical skills which cannot be assessed by motion metrics, such as the interpretation of the obtained images and accurate placement of measurement callipers are combined with non-technical skills such as communication with patients and colleagues, report writing and treatment planning to form the overall clinical competence. These cannot be assessed by motion metrics alone[52,53]. The current RCOG curriculum examines these using Workplace Based Assessments (WPBA). An objective measure of skill is likely to contribute to this assessment, rather than replace it.

3.5 Strengths and Limitations

This study used clinical ultrasound systems, a commercially available probe tracking system and the participants were reflective of the skills mix in many ultrasound departments. The use of clinical ultrasound machines and commercially available equipment makes the study reproduceable by other groups. The study is limited by small numbers in both groups and although completed in a clinical area, the ultrasound phantom was not able to simulate fetal movements, variations in Body Mass Index, or other challenges such as patient anxiety and distraction which may be associated with patient communication. Despite this, we have proposed a metric which is a first step in demonstrating performance differences between novice and experienced ultrasound operators. Repeating this experiment in patients rather than using a phantom may provide insight into how skills assessed in a simulated environment translate to clinical practice. Performing a scan in a clinical environment may induce stress or distractions which also impact on operator performance. Additional variables, such as gestational age, fetal position, maternal body habitus and fetal movements could affect operator performance.

3.6 Conclusion

In this chapter I describe how metrics such as time to completion, probe path distance travelled and DSJ show significant difference between novice and experienced ultrasonographers in a simulated setting. When time and DSJ are considered together, the experienced group had lower DSJ, indicating that experienced operators had less unwanted and more purposeful movements than novices. The results suggest that in a simulated setting operator performance changes with experience. Using DSJ we can quantify the difference.

During the collection of data using the phantom some users noted that the ultrasound phantom was not able to simulate fetal movements. The phantom was also unable to simulate variations in body mass index, or other challenges such as patient anxiety which may be associated with operator distraction, cognitive overload and impaired performance. As such, this was a significant limitation of the study. I proposed that repeating the methodology using patients rather than a phantom could confirm or refute the hypothesis that performance observed in a simulated environment fully translate from the clinical and phantom study.

4 Assessing objective metrics in simulated and clinical settings when performing obstetric ultrasound.

4.1 Introduction

In the previous chapter I discussed a methodology and experiments to investigate the reproducibility of objective metrics for the assessment of operator experience. Although clinical US systems were used, the experiment was limited as it was undertaken on a phantom, rather than a live patient. Some of the shortcomings identified in the literature review could not be avoided when using a phantom. The ultrasound phantom was not able to simulate fetal movements. The phantom was also unable to simulate variations in body mass index, or other challenges such as patient anxiety or operator distraction. These are unavoidable in clinical practice and have been associated with cognitive overload and impaired performance[54]. To understand how, or if the metrics described in the previous chapter could be reproduced in a clinical setting a clinical study was designed. Prior to undertaking the experiments approval was obtained from the UCL and UCLH Research Ethics Committee. The study was approved under the title "Computer Assisted Quantification of Learning Curves in Obstetric Ultrasound Scanning.", "CAL-Obs" for short.

The study was an observational study comparing clinicians of different levels of expertise performing obstetric ultrasound in a clinical environment. Around 8,000 women book their pregnancy each year at UCLH. All women who book at UCLH are offered ultrasound at approximately 12 and 20 weeks' gestational age as part of the NHS national fetal anomaly screening programme. I recruited ultrasonographers performing fetal ultrasound at UCLH to the study. The study was undertaken when the woman attended for her 20-week scan. The scan was undertaken with standard clinical US equipment currently used at the Fetal Ultrasound Screening Unit (USU) at University College London Hospital by staff who are familiar with the department and the departmental protocols in obstetric scanning. The ultrasound scanner was not modified. The methodology in the clinical setting translated the equipment used in the phantom version of the study to a clinical scan room setting. As the phantom had been brought to UCLH the same room layouts and US machines were used in both clinical and phantom experiments, minimising confounding factors.

4.2 Study schedule

The study has been reviewed and approved by The Joint Research Office at UCLH and the regional Research Ethics Committee, reference R&D 120750 (IRAS 253474). Both the patient and the ultrasound operator were given information leaflets and written consent forms to sign prior to enrolment into the study. An informed consent form was created for patients and for operators. A specific participant information leaflet was provided to each group. This ensured that each group of participants got the tailored information they needed to make an informed decision on their participation. The patient information leaflet and consent forms were reviewed by the members of the Patient and Public Advisory Group (PPIAG) for Guided Instrumentation for Fetal Therapy and Surgery (GIFT-Surg). Further information on the GIFT-Surg project can be found at https://www.gift-surg.ac.uk.

Ultrasound operators were recruited from the Obstetric Ultrasound Screening Unit and the Fetal Medicine Department at The Elizabeth Garrett Anderson Wing, University College London Hospital. Operators were asked to enrol on a voluntary basis. The operator was provided with a copy of both the participant and the patient information sheets and invited to take time to consider their decision. The operators were informed that they could withdraw from the study at any time. Operators were not financially compensated in any way for their participation. Participants were informed that they could withdraw their consent at any stage and neither their training or employment at UCLH would be effected in any way.

Ordinarily, the ultrasound operator will manipulate the transducer and machine settings until they are satisfied that the area of interest is optimally displayed. This image is captured and measurements can be made using calliper tools which are built-into the ultrasound machine. The images used to make the measurements and the measurements obtained are stored in the hospitals Picture Archiving and Communication System (PACS). In our study, this procedure did not change. Additionally, the entire ultrasound was captured as a video file. The still images and measurements captured by the operator were transferred to the research database and were stored for scoring. The time taken to complete the procedure was recorded as the time from the beginning of the ultrasound video to the end. The process is based on a framework developed in earlier work by the GIFT-Surg project, classed GIFT-Cloud. The process used for the collection, pseudonymisation, transfer and storage of images is being used by other research groups within the EGA wing at UCLH.

Women attending UCLH for a routine pregnancy scan may be scanned by a Junior Fellow (a doctor with less than 12 months experience) or a Senior Fellow (a doctor with sufficient ultrasound experience to complete the scan independently). Junior Fellows are supervised by a Senior Fellow on an individual basis. In keeping with normal departmental procedure, if patients are assigned to be scanned by a member of the 'novice' group they will be supervised by an appropriate clinician. Outside of the research setting women are not aware if they will be scanned by a Junior or a Senior Fellow.

On arrival in the department patients received a consent form. In the event of wishing to withdraw from the study the scan was performed as planned by the assigned operator. In this scenario no element of the scan, or personal data was recorded.

Participants were pregnant with a single baby and attending for a routine pregnancy scan. This is sometimes known as the "20 week" or "anatomy" Scan. Participants had an Estimated Due Date (EDD) calculated by ultrasound in the first trimester of the current pregnancy. Every pregnant woman is offered a detailed scan of her pregnancy between 18 and 21 weeks. The scan screens for abnormalities in the fetal anatomy:

- Head & brain
- Spine & back
- Lips
- Stomach & Abdominal Wall
- Heart
- Kidneys

This scan was chosen as it is offered widely around the world, has specific views of the fetal anatomy which are required and by limiting the gestational age the study population was more heterogeneous. The use of ultrasound to monitor fetal wellbeing in the third trimester varies globally and a wide spread of gestational ages from 28 to 36 weeks would have introduced variability and reduced generalisability at this stage.

When the patient entered the scan room, I or the operator explained what will happen during the scan and took verbal consent to perform the scan, as would be done for a scan outside the study setting. The scan did not differ from the scan performed on women who did not take part in the CAL-Obs study. Once the operator has completed the scan the patient will be free to leave the department. A report will be generated by the operator and assigned to the woman’s electronic patient record, as per standard departmental protocol.

4.3 Methods

Competent performance of ultrasound requires a detailed understanding of the capabilities and limitations of ultrasound equipment, as well as knowledge of normal maternal anatomy and an awareness of potential abnormal findings. The experienced operator will be able to combine this knowledge with the technical ability to manipulate the ultrasound transducer and adjust relevant display settings to optimise the image, leading to higher objective image quality. It is accepted that performance improves with training and experience[9,44,8], but this observation has not been objectively measured or classified. Dimensionless squared jerk (DSJ) has been proposed as an objective parameter to discriminate between expert and novice operators in endoscopic surgery[45] and functions as a quantitative kinematic measure of deliberate hand movements. DSJ is calculated as the squared third derivative of tool or probe position with respect to time, or the squared derivative of tool acceleration, multiplied by a ratio of total movement amplitude and time duration such that the result is dimensionless and independent of total distance travelled and elapsed time[17]. DSJ quantifies deviations from smooth, coordinated movement and has been accepted as an objective parameter to quantify hand motion in different disciplines, such as endoscopic surgery, observing the effects of parkinsonism and optometry[23,46]. A prospective, observational study of ultrasound practitioners performing fetal biometry in a simulated setting was detailed in Chapter 3[55], finding that metrics such as time to completion, probe path distance travelled and DSJ are measurably different for novice and experienced ultrasonographers. We now translate that methodology to the clinical setting to ascertain if DSJ can reliably differentiate between expert and novice operators outside of a simulated environment.

In line with the approved study schedule, I recruited both experienced and less experienced ultrasound operators. Participants were empirically divided between either experienced in fetal ultrasound (n=41, completed >200 fetal ultrasound examinations) or novice operators (n=25, independently completed <25 ultrasound examinations). The threshold for experience was selected based on the European Board and College of Obstetrics and Gynaecology (EBCOG) guideline on obstetric training, which recommends that trainees complete 200 obstetric ultrasound examinations[47]. As outlined in the introduction, the majority of routine ultrasound scans in the UK are not performed by doctors, therefore recruiting consultant obstetricians would not, necessarily, capture individuals experienced in ultrasound. Simply recruiting Obstetricians or even Doctors would also have excluded a significant part of the workforce as it would exclude non-medical sonographers with significant experience and considerable skill. Thus, 200 completed ultrasounds could be considered the minimum number of examinations performed prior to independent practice to be classified as an experienced operator in this methodology.

Principal inclusion criteria

1) Singleton pregnancy

2) Referral for anomaly scan as per the fetal anomaly screening programme, otherwise known as "20 week" or "anatomy" scan

Principal exclusion criteria

1) Women under 16 years of age

2) Not pregnant

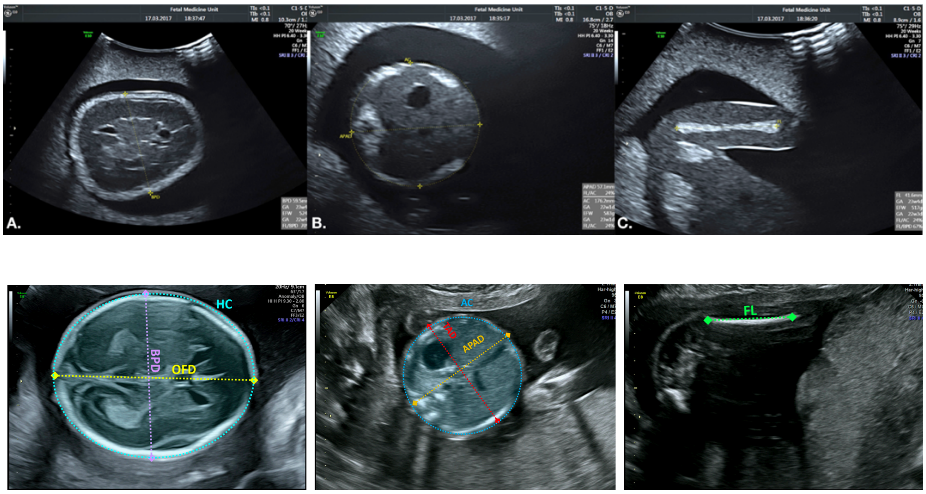
3) Known fetal structural abnormality which would normally require assessment by a specialist in fetal medicine.

4) Multiple pregnancy

5) Withdrawal of patient consent to participate at any time during the study

6) Women with a disability meaning they are not capable of providing informed consent

7) Women who do not understand English to a high enough level to fully understand the research and provide informed consent, unless there is a clinical translator present to explain (and co-sign the consent)



(a) Head (b) Abdomen (c) Femur

Figure 4.1. Images of a phantom model of fetal anatomy in the second trimester.
(A) Transventricular plane, used for measurement of the biparietal diameter (BPD) (B) Transabdominal plane, used for measurement of the abdominal circumference (AC), (C) Required view of the femur used to measure the length of the bone (FL). The second row depicts the clinical images of the same structures with required measurements overlaid.

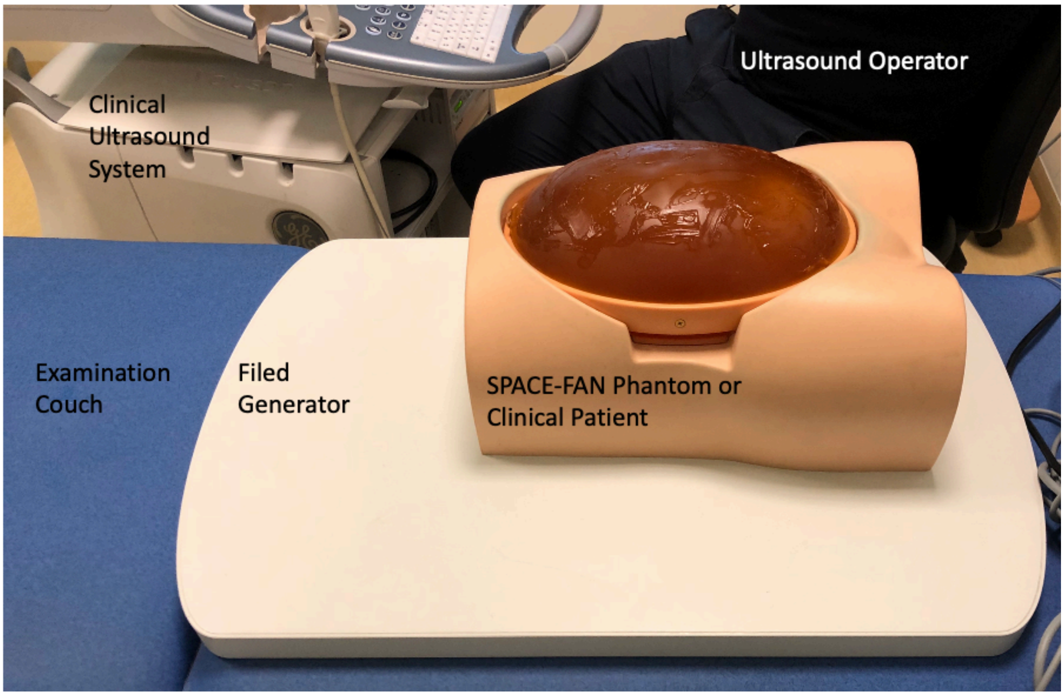


Figure 4.1(d) A close-up view oft he clinical ultrasound system and the movable examination couch. On the couch lies the Aurora field generator, upon which the Phantom can be placed, or the clinical patient can lie fort eh duration oft he ultrasound examination.

Each clinical participant was asked to obtain standard 2D fetal measurements using a clinical GE Voluson E8 ultrasound machine (GE Healthcare, Chicago, Illinois, United States). The position of the ultrasound probe was tracked using the Aurora electromagnetic tracking system (NDI Inc, Ontario, Canada). To track the probe's movement, a 6 degree-of-freedom (DOF) sensor was attached to the ultrasound probe by means of a custom 3-D printed holder, as shown in *Figure 3.2(b).* The patient was asked to lay on an Aurora planar field generator placed on top of the movable examination couch, this is shown in *Figure 4.2(d)*. This allowed the position of the probe to be recorded while a video stream of the scan was captured simultaneously, generating a synchronized dataset of ultrasound images and real-time kinematic information of the probe.

Datasets were captured using a custom software program written in LabVIEW (National Instruments, Austin, Texas, United States). The ultrasound video stream was captured using an Epiphan AV.io HD frame grabber (Epiphan Video, Palo Alto, California, United States). For the Aurora tracking system, a C++ interface program was written that read the probe's position over USB and broadcast it over a virtual serial port to be collected and recorded by the main LabVIEW control program. Ultrasound images were saved at a frame rate of 5 frames per second (fps), while tracking data was recorded at 40 fps. Full data capture was carried out using a Windows 10 laptop with 16 GB of RAM and two USB-C ports. The E8 allows for recording of ultrasound images through its DVI port, which was connected directly to the Epiphan frame grabber. The frame grabber outputted HD video via an HDMI port and was connected to the laptop via an HDMI to USB-C converter. The Aurora tracking system was connected to the laptop via a USB-B to USB-C cable.

A typical experimental set-up is shown in *Figure 3.2(a)* and a detailed view of the sensor attached to the ultrasound probe is shown in *Figure 3.2(b).* The experimental set-up was identical for phantom or patient examinations. During the scan each participant was required to obtain a series of ultrasound planes, as detailed in *Figure 4.1*. The required views were:

- The trans-ventricular plane for the bi-parietal diameter (BPD) measurement,

- The trans-abdominal plane for the anterior posterior abdominal diameter (APAD), the transverse abdominal diameter (TAD), and the abdominal circumference (AC) measurements,
- A view of the femur and a measurement of its length (FL).

These views were chosen as the International Society of Ultrasound in Obstetrics & Gynaecology (ISUOG) recommend these views for the assessment of fetal growth. The views are necessary to calculate fetal weight, following guidance from ISUOG[2]. It was shown that experienced operators can be differentiated from novice operators in a simulated setting using objective metrics, specifically the quantity of dimensionless squared jerk (DSJ) as a discriminator when performing obstetric ultrasound on a phantom. DSJ[50] was calculated for each novice and expert participant using the methodology described in section 3.2

4.4 Results

When performing fetal biometry on the phantom, the mean DSJ for experienced operators was 19.26 ± 9.16 and 22.15 ± 1.01 for the novice group. The results are detailed in *Table 4.1.* DSJ values were found to be statistically different at a significance level of p = 0.01. Typical DSJ values for both groups were lower when performing the same tasks in a clinical setting; the mean DSJ for experienced operators was 18.34 ± 3.42 compared with 19.06 ± 2.69 for the novice operators. The DSJ values for the two groups were again statistically different at a significance level of p=0.006.

	Clinical Expert	*Clinical Novice*	*Phantom Expert*	*Phantom Novice*
Mean	18.34	19.60	19.26	22.15
Variance	3.43	2.69	9.16	1.02
Observations	31	31	10	7

Hypothesized Mean Difference	0	0
df	59	12
t Stat	2.83	-2.80
P(T<=t) two-tail	0.006	0.016
t Critical two-tail	2.01	2.19

Table 4.1 – Distribution of DSJ values for different scenarios with corresponding statistical metrics (t-test: two-sample assuming unequal variances)

DSJ was not found to differ significantly between the clinical or simulated setting for experienced operators. A mean DSJ in the expert group of 19.26 ± 9.15 in the phantom setting compared to 18.34 ± 3.43 for experienced operators in the clinical setting. This did not meet the threshold for statistical significance (p=0.3). When considering the expert group, the outcome of Cochran's Q-test, did not support statistical difference in performance in the clinical or phantom arm (see *Table 4.3*).

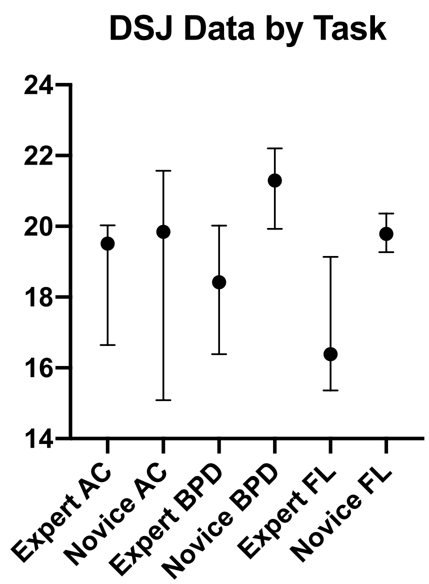


Figure 4.2 - Dimensionless Squared Jerk by task.

4.5 ANOVA

ANOVA						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
DSJ Experts	31	607.64665	19.601505	2.6945123		
DSJ Novice	31	568.65782	18.343801	3.4259191		
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	24.52	1	24.52	8.01	0.006	4.00
Within Groups	183.61	60	3.06			
Total	208.13	61				

Table 4.2 – Analysis of variance between expert and novice operators when performing obstetric ultrasound in a clinical setting.

There was a statistically significant difference between groups as determined by one-way ANOVA (*F*(8.01) = 2.75296975, *p* = .0006), *Table 4.2*. A Tukey post-hoc test revealed that the DSJ values were statistically different when comparing the expert group with the novice group in both clinical and simulated scenarios. (18.34 ± 0.347, 20.11 ± 0.46, *p* = .015) (19.26 ± 0.611, 22.15 ± 0.73, p = 0.018). Difference in performance did not reach statistical significance when expert performance in a clinical scenario was compared to phantom performance (18.34 ± 0.35, 19.26 ± 0.61 p = 0.56). Similarly, no difference was seen when novice performance was compared across clinical and simulated settings (20.11 ± 0.45, 22.15±0.73 p= 0.09).

	Image QC Score	Path Length	Time	DSJ
Abdominal Circumference	No Difference	No Difference	No Difference	No Difference
p	0.4	0.3	0.2	0.3
Transventricular Plane (BPD)	Difference	Difference	Difference	Difference
p	0.01	0.05	0.001	0.0002
Femur (FL)	No Difference	No Difference	Difference	Difference
p	0.5	0.3	0.000003	0.00003
			Novice	**Expert**
Participants			27	20
Images Obtained			43	60

Table 4.3 - Comparison of DSJ to conventional metrics in the assessment of Estimated Fetal Weight (EFW).

When considering individual tasks which the clinicians are required to perform to estimate the fetal weight, DSJ differentiates between novice and experienced operators when assessing the femur and Transventricular planes. *Table 4.3* demonstrates DSJ against more conventional measures of operator performance, such as image quality, path length and time to completion of task.

4.6 Discussion

Using the same clinical methodology in both a phantom and a clinical environment, I found that DSJ differentiates between experienced and novice operators. Statistical difference was maintained when performing biometry during routine obstetric ultrasound at 20 weeks' gestation.

When I reported the findings using this methodology in a simulated setting I concluded that objective measures of operator performance correlate with experience[55]. I noted that the ultrasound phantom was not able to simulate fetal movements, variations in body mass index, or other challenges such as patient anxiety which may be associated with operator distraction, cognitive overload and impaired performance. As such, this was a significant limitation of the previous study. I proposed that repeating the methodology using patients rather than a phantom could test the hypothesis that performance observed in a simulated environment fully translates to clinical practice. These results confirm my hypothesis that DSJ as a performance metric can differentiate between experienced and inexperienced operators in the clinical setting.

When considering novice performance, I found significant differences between DSJ values obtained in simulated and clinical environments. This is likely to reflect the overall experience of the operator. I previously hypothesised that performing a scan in a clinical environment may induce stress or distractions which also impact on operator performance. Additional variables such as gestational age, fetal position, maternal body habitus and fetal movements could also affect operator performance. This mirrors findings in the aviation industry, where pilot age and experience were the significant predictors of pilot performance in a simulator. The groups were assessed in a simulator against younger and/or less experienced pilots[56], [57]. The performance differences were especially marked in domains concerned with spatial orientation and communication skills, domains which are also important in performance, interpretation and decision-making in obstetric ultrasound. Non-technical skills such as communication with patients and colleagues, report writing and treatment planning all contribute to the assessment of clinical competence, but these cannot be assessed by motion metrics. It is likely that an objective measure such as DSJ would be most appropriately used as part of a larger global skills assessment, considering both technical and non-technical competencies [52,53].

In the experienced operator group DSJ does not significantly differ between clinical and simulated settings. This is in contrast to novice operators where DSJ is higher in the clinical setting, i.e. there is less intentional and more random movements. I suggest that this represents a degradation of novice operator performance in the clinical setting compared to the simulator. Experienced operators' skills appear to be more robust across different scenarios. This implies that with time and experience, an experienced operator has developed strategies and techniques to improve their chance of obtaining an acceptable image. These are derived from repeated exposure to a large variety of fetal positions and maternal body characteristics. Such strategies would be conventionally described as "experience". For experienced operators changing the environment in which they are performing the ultrasound is of little consequence. Experienced operators understand the task they wish to perform, the anatomy they are seeing at any given time and how to move the ultrasound probe to obtain the view of interest. While most of the participants will achieve this level of performance, it has been seen in similar experiments (but not in this cohort) that some operators never achieve expert-like objective performance. These participants would present as outliers. In terms of patient safety, however, their performance is likely to be of little consequence, as the inter-operator variability between the images remains small and difference in QC score was statistically insignificant. All participants in our experienced operator group acquired images of the planes (BPD, AC, FL) required for the assessment of fetal biometry and assessment of fetal weight, even under unfavourable clinical circumstances.

In the clinical setting we found that DSJ differentiates between expert and novice operators during specific tasks such as assessing the Transventricular plane to determine the BPD and measuring femur length (FL). Traditional methods of assessing competence in obstetric ultrasound use a combination of subjective and objective observations. My results suggest that single observations, considered in isolation, are poorly predictive of expert-like behaviour, as shown in *Table 4.3*. From my results it appears that the process of obtaining an image is more discriminatory than the ultimate image obtained. This reflects the complexity of the task being performed. Ultrasound examinations, much like laparoscopic surgery, require the operator to interpret a dynamic image produced by the three-dimensional position and motion of the ultrasound probe by means of a two-dimensional visual display.

4.7 Conclusion

In this chapter I have described a series of experiments and recorded performance data while trainees are performing obstetric ultrasound. This is important because, if novice operators can be shown to imitate expert performance during their training, they should be performing to the level approaching that of an expert operator. If such performance can be reliably quantified, it would represent a point where trainees can further consolidate their performance in the clinical environment. Once a basic level of competence has been achieved they could receive further training. Examples would include assessment of the fetal brain, limbs or heart. It is also worth noting that participants had access to "ideal" images and adequate time to compete the assessment of fetal weight. Taking a global view, high quality images taken in a reasonable time with high reproducibility would represent a novel assessment of performance.

I conclude that DSJ is a metric which can be used to differentiate between expert and novice operators in a simulated or clinical setting. This makes it a candidate for the assessment of learning outcomes. This would be in the context of a training programme or training intervention. My statistical analysis of measured DSJ values supports the hypothesis that the process of performing an ultrasound scan is highly variable between the novice and experienced operator. I evolve the paradigm and demonstrate that performance in a simulated environment is a reliable surrogate for performance in a clinical environment. DSJ is likely to be most discriminatory when considered as part of learning outcomes which attempt to assess trainees' knowledge of the anatomy, an understanding of the limitations of ultrasound, image optimisation and confidence in performing fetal biometry.

In this chapter I have compared performance of expert and novice operators in clinical and phantom settings. I had established that expert performance does not differ significantly in either setting. This understanding opens up avenues for further investigation. The first is to gain an understanding of how novice transitions to expert like behaviours. How long is this transition and is it linear? In other words can a learning curve be defined? The second question follows from establishing a learning curve. Once a learning curve has been established subsequent investigations could establish how such a curve can be manipulated. Can the learning curve be made steeper? That is to say can the time in training be made shorter? Alternatively, can the skills plateau be made higher? Can sonographers be assessed against a clearly defined standard? Further work in this thesis investigates how a learning curve can be manipulated or altered to improve trainees experience prior to performing ultrasound in the clinical setting.

5 Proposing DSJ for the longitudinal assessment of training in Obstetric Ultrasound.

5.1 Introduction

For reasons discussed at length in the literature review chapter the challenges associated with training and maintenance of competency in obstetric ultrasound are multifaceted. The standard and frequency of assessment and revalidation are inconsistent globally. A validated, objective and reproducible approach to competency assessment could form the basis of a move towards standardisation. Standardised assessment could lead to increased transferability of skills when doctors in training move hospitals. Research might be easier to conduct as standardised training protocols would be readily available.

Developing skills in ultrasound is difficult because ultrasound examinations, much like minimally invasive surgery (MIS) require the operator to interpret a dynamic image produced by the three-dimensional (3D) position and motion of the ultrasound probe by means of a two-dimensional (2D) visual display. Successful interpretation of clinical obstetric ultrasound requires an understanding that the fetus, a 3D object, fixed in neither time nor space is being represented on a fixed 2D grey-scale monitor. Operator experience, combined with the effects of probe motion and homogeneity of ultrasound images contributes to high inter- and intra-operator variability. Although the concept of standard planes for the assessment of fetal weight has been well established and validated in multi-centre trials, to date there has been little standardisation in training and assessment of operators beyond didactic lectures[47,58]. Data presented earlier in this thesis suggests that image quality alone is a poor differentiator of experience or expertise[59]. This should be of little surprise as quality control systems are retrospective and assess the quality of the output, in this case a standard plane. Assessing sonographer skill solely by image quality is analogous to death rates for surgeons. These outcomes are very important but their macro nature does little to explain the factors which contribute to variations in inter-operator performance.

A longitudinal means of assessment is required to allow for objective and reproducible assessment of progress in a training programme. An objective methodology which can differentiate between expert and novice operators must be the starting point. Further, this metric would evolve over time and with operator experience. It would allow trainees and their trainers to assess their progress against standardised learning outcomes and benchmark them against their peers. Today, the assessment process typically considers surrogates of competency, such as economy of motion or time to task completion. Alternatively, patient outcomes such as discomfort, rates of complications or length of stay in hospital are often reported[60]. In earlier work[55] I have shown that Dimensionless Square Jerk (DSJ) can differentiate between expert and novice operators when performing biometry on a phantom. I have also shown this methodology can be successfully used in a clinical setting. I now investigate DSJ as an objective measurement of operator performance as they advance through training.

5.2 Learning Curves in Obstetric Ultrasound

The data from historical attempts at defining learning curves in ultrasound have indicated that expert-like performance on a simulator could be achieved with as little as 5 hours of training[61]. This was not obstetric ultrasound, however. A longitudinal study recruited experienced sonographers and re-trained them to perform US assessment of abdominal hernias suggested that their performance converged with expert-like performance between 40 and 70 independent ultrasound examinations. The study was conducted over a 4 year period. In this time the sonographers continued to perform ultrasound on other areas of the body. This study recruited sonographers with a range of experience varying between 11 and 17 years of clinical US practice. While this experience is not directly comparable to the novice ultrasound operator, or a doctor needing to develop ultrasound skill as part of a wider learning portfolio, this study did illustrate the difficulty in setting time aside for training and assessment. In contrast, the CAL-Obs study protocol did not require any experience in ultrasound and none of our participants had undertaken formal training prior to recruitment in the study.

In summary, the data currently available does not allow for estimation of how long a typical trainee might take to achieve competency in the practice of obstetric ultrasound.

5.3 Methods

To investigate learning curves, we repeated the previously reported methodology in Chapter 4. The experiments were repeated at 3 monthly intervals in the clinical setting. In line with the approved study schedule, we undertook a prospective, observational study of medical practitioners who work in the Ultrasound Screening Unit (USU) at University College London Hospital NHS Foundation Trust, London, UK (UCLH). We recruited both experienced and less experienced ultrasound operators who were employed as ultrasound training fellows. During the study recruitment window five new scan fellows joined the department and consented to participate in the longitudinal study. Five trainees in obstetric ultrasound had three assessments in clinical practice over a six month period following their induction on a phantom. All the sonographers in this study were medically qualified and had commenced specialist training in Obstetrics & Gynaecology, all worked at UCLH. None had yet completed their training, or were eligible for inclusion on the UK specialist register in Obstetrics & Gynaecology or Radiology. We compared the progress of training fellows receiving didactic, practical and theoretical training in a full-time post. In this post training fellows spent 80% of the week in the ultrasound department and 20% of the week in the Obstetrics & Gynaecology service.

Independent performance of obstetric ultrasound was not required for acceptance on the fellowship program at UCLH. None of the participants had previously had formal ultrasound training prior to commencing their fellowship. During the period of fellowship the post-holder is supervised by a team of Consultants and experienced doctors in Fetal Medicine. Following 12 months of didactic, practical training and assessment it is expected that the Fellow would progress to independent practice of obstetric ultrasound. All the participants in this cohort had completed formal assessment, graded against the RCOG OSAT for the performance of fetal biometry by the end of the 6 month period[62]. I previously described a methodology for calculating DSJ values for novice and experienced operators in Chapter 3 and 4 respectively.

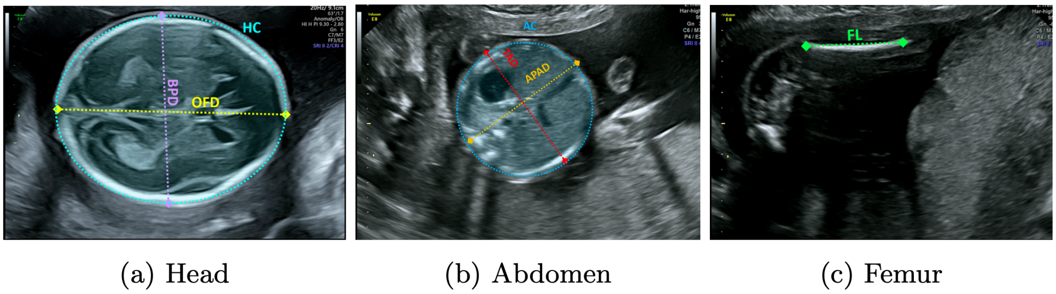


Figure 5.1.

Images of fetal anatomy in the second trimester.
(a) Transventricular plane, used for measurement of the biparietal diameter (BPD)
(b) Transabdominal plane, used for measurement of the abdominal circumference (AC),
(c) Required view of the femur used to measure the length of the bone (FL). The second row depicts the clinical images with measurements overlaid.

Each participant was asked to obtain standard 2D fetal measurements using a clinical GE Voluson E8 ultrasound machine (GE Healthcare, Chicago, Illinois, United States). Example images and the measurements the participants were required to obtain are detailed in *Figure 5.1 (a), (b) & (c).* A more detailed example of the experimental set-up is shown in *Figure 5*.2. The position of the ultrasound probe was tracked using the Aurora electromagnetic tracking system (NDI Inc, Ontario, Canada). To track the probe's movement, a 6 degree-of-freedom (DOF) sensor was attached to the ultrasound probe by means of a custom 3-D printed holder, as shown in *Figure 3.2(b).* The patient was asked to lie on an Aurora planar field generator placed on top of the movable examination couch. This allowed the position of the probe to be recorded while a video stream of the scan was captured simultaneously, generating a synchronized dataset of ultrasound images and real-time kinematic information of the probe.

Datasets were captured using a custom software program written in LabVIEW (National Instruments, Austin, Texas, United States). The ultrasound video stream was captured using an Epiphan AV.io HD frame grabber (Epiphan Video, Palo Alto, California, United States). For the Aurora tracking system, a C++ interface program was written that read the probe's position over USB and broadcast it over a virtual serial port to be collected and recorded by the main LabVIEW control program. Ultrasound images were saved at a frame rate of 5 frames per second (fps), while tracking data was recorded at 40 fps. Full data capture was carried out using a Windows 10 laptop with 16 GB of RAM and two USB-C ports. The GE Voluson E8 allows for recording of ultrasound images through its DVI port, which was connected directly to the Epiphan frame grabber. The frame grabber outputted HD video via an HDMI port and was connected to the laptop via an HDMI to USB-C converter. The Aurora tracking system was connected to the laptop via a USB-B to USB-C cable.

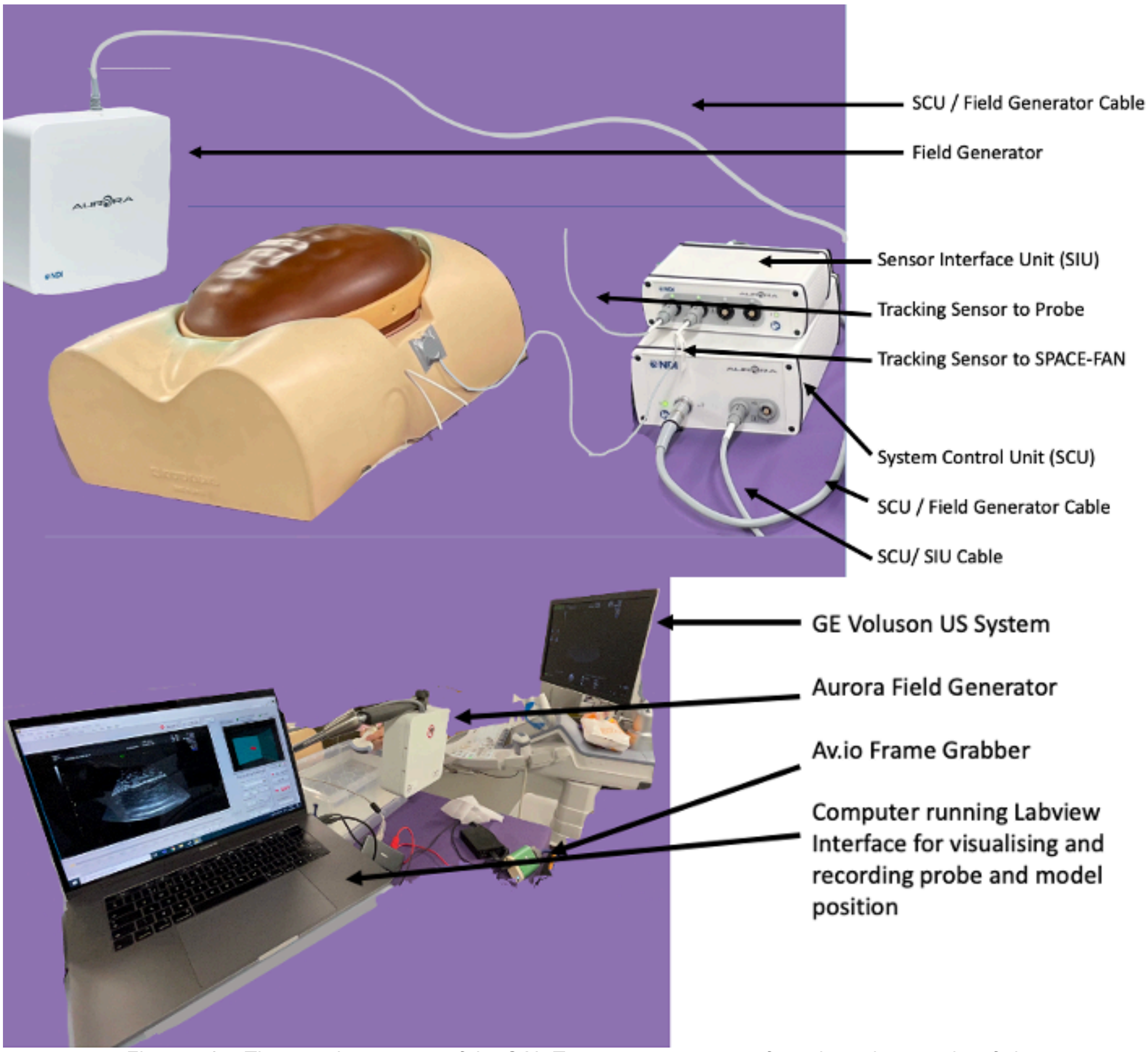


Figure 5.2 – The constituent parts of the CAL-Tutor system, as seen from the trainees point of view.

.

A typical experimental set-up is shown in *Figure 3.2(a)* and a detailed view of the sensor attached to the ultrasound probe is shown in *Figure 3.2(b)* in chapter 3. Further detail of the hardware set-up and collection equipment is shown in *Figure 5.2.* During the course of the scan each participant was required to obtain a series of ultrasound planes, as detailed in *Figure 5.1*. The required views were:

- The trans-ventricular plane for the bi-parietal diameter (BPD) measurement,
- The trans-abdominal plane for the anterior posterior abdominal diameter (APAD), the transverse abdominal diameter (TAD), and the abdominal circumference (AC) measurements,
- A view of the femur and a measurement of its length (FL).

These views were chosen as the International Society of Ultrasound in Obstetrics & Gynaecology (ISUOG) recommend these views for the assessment of fetal growth. The views are necessary to calculate fetal weight (EFW). Our previous work has shown that experienced operators can be differentiated from novice operators, specifically the quantity of dimensionless squared jerk (DSJ) as a discriminator when performing fetal biometry. DSJ[50] was calculated for each novice and expert participant using the formula described in section 3.2.

5.4 Results

When performing fetal biometry on the phantom, the mean DSJ for experienced operators was 19.26 ± 9.16 and 22.15 ± 1.01 for the novice group. These were found to be statistically different at a significance level of p = 0.01. Typical DSJ values for both groups were lower when performing the same tasks in a clinical setting; the mean DSJ for experienced operators was 18.34 ± 3.42 compared with 19.06 ± 2.69 for the novice operators. The DSJ values for the two groups were again statistically different at a significance level of p=0.006.

Figures 5.3 (A), (B) and(C), suggest that novice ultrasound performance trends towards expert performance by six months of daily performance of obstetric ultrasound. In this post training fellows spent 80% of the working week in the ultrasound department and 20% of the working week in the Obstetrics & Gynaecology service. Based on the clinical workload at the hospital, this approximates to 20 growth scans per week, or 480 singleton growth scans over 6 months.

By repeatedly assessing trainees as they progress through the initial six months of a training fellowship, we show that DSJ values trends towards that of the expert operator. By six months there is no statistical difference between trainees and expert operators when obtaining views of the Head Circumference and femur. The analysis of variance (ANOVA) results were significant at the 5% level, indicating that the expected values for DSJ differed significantly between the groups when assessing the Biparietal Diameter (BPD) and Femur Length (FL) at 20 weeks gestation. As seen in the phantom and earlier clinical studies assessment of the abdominal plane (AC) does not differentiate between novice and expert sonographers at $P<0.05$.

When comparing DSJ values for the trainees, following 6 months of training, there was no statistical difference at the 0.05 level when considering BPD ($F(1,11)= 0.01$, $p=0.93$) and AC ($F(1,11) = 0.02$, $p = 0.90$) between them and the experienced operator group. The results are detailed in *Table 5.1*. Results when obtaining the FL remained statistically different.

	ANOVA for Biparietal diameter (BPD) at 6 months after training commenced against Expert operators.						ANOVA for abdominal circumference (AC) at 6 months after training commenced against Expert operators.						ANOVA for Femur Length (FL) at 6 months after training commenced against Expert operators.					
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.02	1.00	0.02	0.01	**0.93**	4.84	0.05	1.00	0.05	0.02	**0.90**	4.84	4.78	1.00	4.78	2.19	**0.17**	5.12
Within Groups	27.05	11.00	2.46				30.87	11.00	2.81				19.65	9.00	2.18			
Total	27.07	12.00					30.92	12.00					24.43	10.00				

Table 5.1 – DSJ ANOVA for biparietal diameter (BPD), abdominal circumference (AC) and femur length (FL) at 6 months after training commenced compared against Expert operators.

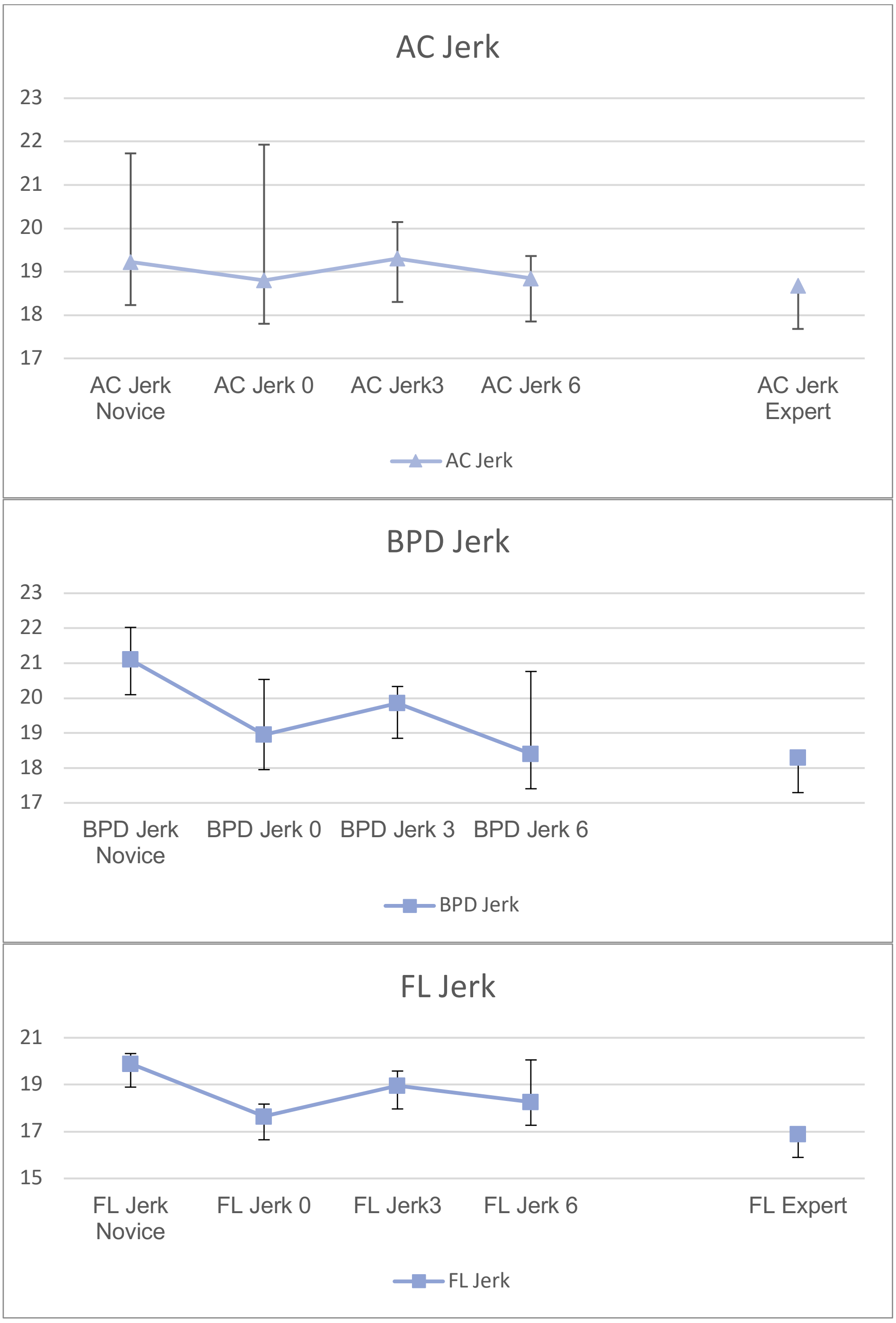


Figure 5.3 (A), (B) and (C) - DSJ values collected at 3 month intervals as ultrasound fellows commenced the initial six months of an obstetric ultrasound training fellowship. Each trainee was asked to obtain the biparietal diameter (BPD), abdominal circumference (AC) and femur length(FL).

5.5 Discussion

Using the same experimental methodology in a phantom and a clinical environment, we demonstrate that DSJ differentiates between experienced and novice operators. We previously reported our findings using this methodology in a simulated setting and concluded that objective measures of operator performance correlate with experience[8]. We proposed that repeating the methodology using patients rather than a phantom could test the hypothesis that performance observed in a simulated environment translates to clinical practice. Our results confirm our previous hypothesis that DSJ as a performance metric can differentiate between experienced and inexperienced operators in the clinical setting.

I already discussed in Chapter 4 that DSJ does not significantly differ between clinical settings and simulated environments for experienced operators. In contrast, DSJ values in the novice group was different. In this group, DSJ was higher in the clinical setting than in the phantom setting, i.e. there is less intentional movement in the clinical compared to the phantom environment. I suggested that the difference in DSJ represented a degradation of novice operator performance in the clinical setting compared to the simulator. Had the difference in DSJ been seen in both experienced and novice operators I might have concluded that the simulated environment was somehow 'easier' or poorly representative of the clinical environment that it was purporting to simulate. On the contrary the simulated environment seems to be representative of the clinical reality. Experienced operators' skills appear to be more robust across different scenarios. This implies that with time and experience, an experienced operator has developed strategies and techniques to improve their chance of obtaining an acceptable image. It is likely that these are derived from repeated exposure to a large variety of fetal positions and maternal body characteristics. For experienced operators changing the environment in which they are performing the ultrasound is of little consequence. Experienced operators understand the task they wish to perform, the anatomy they are seeing at any given time and how to move the ultrasound probe to obtain the view of interest. Such strategies develop over time and these are the skills I am attempting to capture and train the inexperienced operators to mimic.

The results suggest that single observations, considered in isolation, are poorly predictive of expert-like behaviour, as shown in *Table 4.3, Chapter 4*. In the clinical setting we found that DSJ differentiates between expert and novice operators during specific tasks such as assessing the transventricular plane to determine the BPD and measuring femur length (FL). Traditional methods of assessing competence in obstetric ultrasound use a combination of subjective and objective observations. From our results it appears that the process of obtaining an image is more discriminatory than the ultimate image obtained. This reflects the complexity of the task being performed.

As noted, the DSJ values recorded showed significant difference between experienced and novice at the start of the training period. Over a period of a six month training period, the data demonstrates a convergence of novice toward the experienced operators. The data, the parallels to laparoscopic surgery and previously published trainee guidelines suggest that time and training resource dedicated to US training are underestimated by training bodies, trainers and trainees. Independent performance of fetal biometry is a stated aim of the UK training curriculum, yet 4.3% of the trainee population had competed the intermediate ultrasound module in 2018[15]. In common with other complex skills, such as laparoscopic surgery, consistent and regular training is required for safe and independent practice. In the absence of a dedicated period out of training, our data suggests that at least half a day per week is required over a period of 2 years to achieve independent practice of fetal biometry, equivalent to the RCOG intermediate module.

In this chapter, I have presented a series of experiments to study how skills develop over time in the case of obstetric ultrasound. The strengths of this study are that we used a standard clinical ultrasound system with a commercially available electromagnetic tracking system to track probe movement. The study demonstrated that these can be safely and practically used in a busy clinical department in a large NHS teaching hospital without interrupting patient capacity of the department. The acceptability of this type of study has been demonstrated to patient representatives and pregnant patients, as seen by both groups being willing to consent to participation. The use of clinical ultrasound machines and commercially available equipment makes the study reproducible by other research groups. This study is limited by relatively small sample sizes. Recruitment to the study and all other research studies was suspended by the hospital due to the COVID-19 pandemic, and that I was redeployed to assist the clinical service. The limited numbers of trainees recruited limited the generalisability of the model we have proposed.

In this chapter I performed a longitudinal variant of our previous reported methodology to develop learning curves for the performance of fetal biometry using ultrasound. I have shown that operator performance can be quantified at the beginning and the end of the training period and that it evolves in a predictable way. This data decouples learning outcomes from assessments of image quality, logbooks and assessment by senior clinicians, with their associated limitations. The data demonstrates that trainee performance improves on the tasks assessed over time. This is an essential exercise toward the development of both training curricula and metrics that are discriminative of operational skill. This data proposes an objective learning outcome based on experienced operator performance and would allow individual trainees to assess their individual training progress against these benchmarks.

In the context of currently published educational curricula and given the operator-dependant nature of US, effective training is vital to engender consistent, effective performance[43]. Learning curves and the expected rate of progression through training have not been previously published when using an objective metric such as DSJ. Previous studies have attempted to quantify the numbers of US cases required to achieve a predetermined level of competence in emergency examination[11], or the use of transvaginal US[13] but none of these studies describe objective metrics as a learning outcome.

Previously published works have examined image quality as a proxy for operator experience. In previous works we examined the utility of QC scores to differentiate between expert and novice operators. We found that QC score was not a reliable differentiator for operator experience[55]. This is unsurprising, as these studies have been in the context of quality control, rather than training[63]. Operators quickly gain an understanding of what a clinically diagnostic image should look like. If they are unable to obtain this, they can refer to a more senior or more experienced colleague, for example. For certification and quality control operators are often required to submit several images which they obtained. These images are not chosen at random and are usually selected by the operator themselves. In this methodology I assessed the trainee on a longitudinal basis and no allowances are made for the trainee to select the case beforehand, avoiding issues with operator bias. Perhaps secondary to this, our results show an unexpected deterioration in performance between the initial assessment and the second. This phenomenon has been described by authors in other disciplines, such as in laparoscopic surgery. Bingener-Casey et al described a temporary performance deterioration after technical competence has been achieved. The authors reasoned that undertaking more difficult cases and over confidence may have contributed to lapses in technique or judgement[60].

The data in this chapter can clearly be applied to clinical scenarios, as noted. DSJ values showed significant difference between experts and nonexperts at the start of the training period. We demonstrated a convergence of novice toward the experienced operators as the training period progressed. Our data, the parallels to laparoscopic surgery and previously published trainee guidelines, suggest that time and training resource dedicated to US training are underestimated by training bodies, trainees and their trainers. Independent performance of fetal biometry is a stated aim of the UK training curriculum, yet only 4.3% of the trainee population had completed the intermediate ultrasound module in 2018[64]. In common with other complex skills, such as laparoscopic surgery, consistent and regular training is required for safe and independent practice. In the absence of a dedicated period out of training, our data suggests that at least one half day a week is required over a period of 2 years to achieve independent practice of fetal biometry, equivalent to the RCOG intermediate module.

5.6 Conclusion

The data presented in this chapter supports the conclusion that DSJ is a robust metric which can differentiate between expert and novice operators in a simulated and clinical setting. This makes it a candidate metric for assessment of learning outcomes for a training programme or intervention.

The study demonstrates that trainee performance improves with training time and plots this change on a learning curve. The description of learning curves is an essential exercise toward the development of training curricula and metrics that are discriminative of a operators skill. As noted, the measures of DSJ showed large differences between experts and novice operators at the start of the training period. We demonstrated a convergence of novice toward the experienced operators as the training progressed.

DSJ is likely to be most discriminatory when considered as part of learning outcomes. Collectively, these attempt to assess trainees' knowledge of the anatomy, an understanding of the limitations of ultrasound, image optimisation and confidence in performing fetal biometry. Future work in this area will be to use DSJ to assess training performance in a series of inexperienced volunteers. These trainees will have been trained exclusively in a simulated environment. I will compare the DSJ during training with subjective self-assessment from trainees and Objective, Structured Assessments of Training (OSAT) currently used by the RCOG curriculum.

In this chapter, I have presented a method to compute learning curves from recordings of probe tracking data during obstetric ultrasound. The study demonstrates that trainee performance improves with training time. This description of learning curves is an essential exercise toward the development of both training curricula and metrics that are discriminative of operational skill. As noted, the measures of DSJ showed large differences between experts and nonexperts at the start of the training period. We demonstrated a convergence of novice toward the experienced operators as the training progressed.

The final chapters will aim to build on the longitudinal data presented in the initial chapters. Now that a timeframe and a slope of a learning curve has been estimated future experiments will be designed with the intention to manipulate the trajectory of the learning curve to make training more effective. DSJ is a robust metric which can differentiate between expert and novice operators in a simulated or clinical setting. We have shown that objective measurements can be used over time to monitor trainee progress in a structured training environment. However, DSJ is likely to be most discriminatory when considered as part of a training package which attempt to assess trainees' knowledge of the anatomy, an understanding of the limitations of ultrasound, techniques for image optimisation and confidence in performing fetal biometry. The subsequent chapters will examine the effect that a mixed reality training device has on the learning curves. Of particular interest is the possibility that the early training phase, up to the point where the trainee gains familiarity and proficiency with US could be replaced by simulator or classroom-based training. This would be signified by the temporary performance deterioration mapped onto the expected learning curve, for example.

6 Proposing, Developing and Exploring Mixed Reality As A Training Tool In Obstetric Ultrasound.

6.1 Introduction

The work described thus far has demonstrated that it is possible to develop objective metrics which correlate with operator experience. In addition, I have performed repeated assessments of these metrics over 6 months of a fellowship training programme in obstetric ultrasound. The intention was to understand how these metrics change as the operator gains experience.

The results of the literature review suggest that simulators can be a useful tool for training. A confounding factor for measuring the effectiveness of training interventions is that most trainees achieve technical proficiency regardless of the training methodology. The varied approaches to training in obstetric ultrasound further support this. This is further supported by comparable rates of poor obstetric outcome in developed countries, which ought not to be the case if one training methodology was superior. Simply considering macro-level patient outcomes may falsely reassure, however. I have described high levels of dissatisfaction with training and high levels of trainee dropout in earlier chapters. There have been reports of senior clinicians that never achieve mastery levels of technical skill, purportedly due to poor basic training[35]. In the earlier chapters, I have shown the use of DSJ to differentiate between experienced and novice operators in both clinical and simulation settings. Although this work developed a methodology that could record trainee progress over time, it was not clear that being able to record progress would, improve trainees' skill or competency.

6.2 Identifying a training need

The literature review highlighted barriers to training beyond the provision of simulation equipment and how it is used by trainees and their trainers. The most relevant barriers include:

- Trainees lack consistent and iterative feedback on their performance in relation to their progress in training.
- Trainees are unable to benchmark their own performance against their peers or trainers in an objective way.
- Trainee engagement with training can vary. Sometimes this is secondary to clinical workload. Some trainees may not identify ultrasound skills as valuable for their long-term career goals.
- Access to training equipment, which is close to the clinical environment, serviceable and readily available.
- Simulators rarely represent a close approximation of the equipment they would use in clinical practice.

With these barriers in mind a questionnaire was circulated to trainees in obstetrics & gynaecology to understand how these factors directly impacted their training in obstetric ultrasound and to give a clearer understanding on areas to be targeted by a training intervention.

6.2.1 Initial survey of trainee learning experience

The initial survey sought to understand how general trainees in obstetrics and gynaecology at UCLH reported their US abilities against the RCOG curriculum for the intermediate obstetric ultrasound module. The survey was hosted on the SurveyMonkey platform. It was open for one week between the 1st and 6th of March 2021. In total there were 5 responses. Trainees were made aware of the survey by e-mail sent to all junior doctors in the department. This totalled 43 individuals. The response rate was 11.6%. This is typical for this kind of survey[65], where the response rate varies between 5 and 30%. Of the replies submitted 60% of trainees reported that they had formal training in obstetric ultrasound. Two respondents reported satisfaction with their ultrasound training. Four of respondents felt that formal, timetabled scanning sessions would be the optimal method for ultrasound learning. The responses are detailed below in *Tables 6.1 (a), (b) & (c)*. A minority of post graduate doctors were dissatisfied with their training. *Figure 6.1* details responses by post graduate doctors on their ability to perform specific ultrasound tasks. While trainees reported that the could perform US functions such as measuring abdominal circumference, femur length and performing doppler, trainees self-reported that they were unable to optimise the image, such as adjusting depth, gain and focal depth. These contradictory answers are of note because they suggest that trainees are working in the "unconscious incompetence", or "conscious incompetence" which are the early stages of training.

	Yes	No
Are you a doctor in training	60%	40%
Have you had formal ultrasound training/fellowship	60%	40%

Table 6.1(a) – User responses to the initial questionnaire.

	Very Satisfied	Satisfied	Neutral	Dissatisfied	Very Dissatisfied
Thinking about your US Training. Are you…		40%	40%		20%

Table 6.1(b) - User responses to the initial questionnaire.

	Informal bedside teaching with specified learning outcomes	Timetabled scanning session in the US department	Other (free text)
Preferred teaching method	20%	80%	-

Table 6.1(c) - User responses to the initial questionnaire.

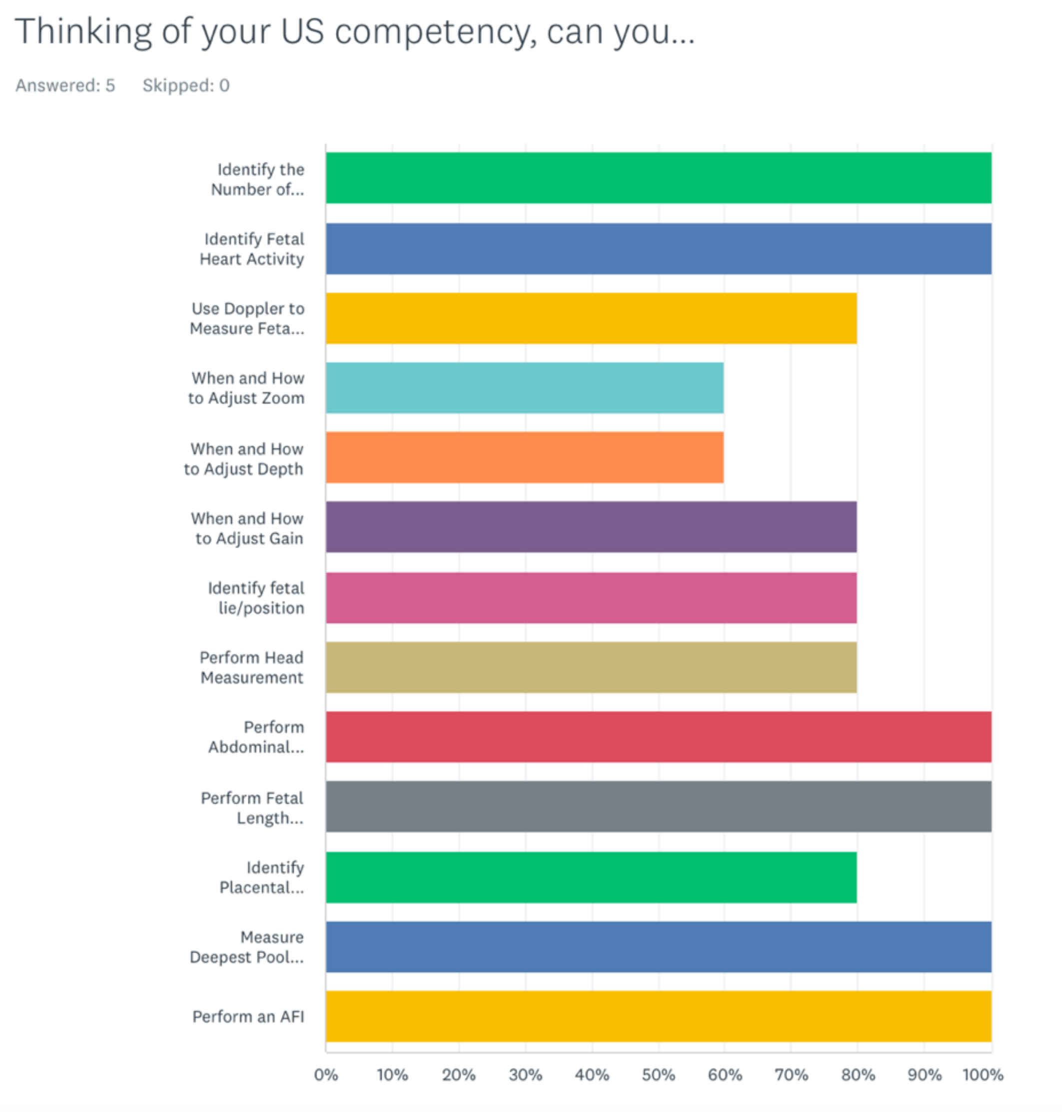


Figure 6.1- trainee responses of common US tasks which they can perform independently.

Trainees reported that they would prefer timetabled training sessions in the ultrasound department, rather than ad-hoc, or informal training sessions, even when defined learning outcomes were provided. Work investigating training in surgical specialties suggests that the most effective way to maximise training opportunities is to allocate a block of training to trainees. The benefits of a dedicated training block include the ability to complete theoretical learning, simulation and clinical practice in a short, focussed time-frame. A specific block of time dedicated to ultrasound training would allow trainees to gain an understanding of the anatomical knowledge to understand the ultrasonic appearance of normal fetal anatomy. Once this is appreciated the trainees would still have enough time in a training block to practice under supervision of a more experienced clinician, gaining in confidence, before their return to general training, or rotating to another sub-speciality. This approach would be supported by trainees, based on the scoping exercise.

6.2.2 Considering trainee experiences with available research

Given the limited training time that trainees have for education, it must be utilised as efficiently and effectively as possible so that trainees can achieve their desired learning outcomes. To maximise training opportunities in a clinical setting, the aim ought to be to train operators such that they are competent, gaining in confidence and moving towards independent practice before a clinical attachment begins. What effect would standardised instruction and explicit training outcomes have on trainee confidence and trainee capability? Would this impact assessor opinions of the trainee? By achieving certain metric values in simulation would the trainee be better equipped and ready to contribute to the clinical service before ever performing an ultrasound in the clinical environment?

Toolgard et al found that clinical experience and length of stay in ultrasound units were predictors of trainees' ultrasound confidence ($p<0.001$)[85]. For trainees to feel 'confident' in performing transvaginal and transabdominal ultrasound scans approximately 12–24 days of training in ultrasound units over 24 months of clinical experience was required. They identified three factors related to ultrasound confidence:

Technical aspects,

Image perception,

Integration of scan into patient care.

The idea of requiring trainees to demonstrate their performance in a simulated setting prior to commencing clinical practice is not as far-fetched as might be expected. This approach has been used in pilot training for decades. Pilots, for example, may have competed their training in a simulator and never have operated a particular aircraft type in commercial operations before their first flight carrying fare-paying passengers. This approach gives trainees, their trainers and senior clinicians who may not be directly involved in their training confidence that each sonographers' skills, image interpretation and resultant clinical decisions meet a consistent, minimum standard for their stage in training. Importing an ideology from established training methods in the aviation industry is attractive but requires a defined learning outcome and minimum performance levels. As introduced earlier aberrant fetal growth is seen in up to one third of pregnancies, millions of annual cases globally and growth restriction is attributed to a significant proportion of neonatal intensive care unit workloads. Given the morbidity and mortality associated with fetal growth restriction and the standardised approach to the US assessment of growth restriction, utilising standard planes (SP) to estimate fetal weight, was the learning outcome selected.

6.2.3 How current training deficiencies can be improved by Mixed Reality

In earlier chapters I have described the practical difficulties in training clinicians and high levels of trainee dissatisfaction[66]. While training remains challenging, there is a growing body of evidence for the utility and use of ultrasound at the bedside[67]. In the emergency medicine setting, for example point of care (POC) US findings influence clinical treatment plans in approximately half of cases[68]. Poor, or non-existent training, will lead to poor decision making and impact patient safety. The competing demands of limited training time and large clinical workload is obvious. Any proposed solution must recognise that training time and resources are finite. To offer a real improvement over current training CAL-Tutor ought to balance these resources to better meet the trainees learning need. Trainees ought to be thought anatomical knowledge, technical skill and image interpretation skills in a setting which closely mirrors the clinical environment in which they will be working. The skills taught should be relevant for the setting in which they will be used. The training content should closely reflect the cases a trainee might be expected to encounter in clinical experience. Consideration must also be given to the equipment they will have access to in the clinic. The ideal training platform would be mobile, be at the patient bedside and give real-time instruction to the trainee in a consistent way. To deliver this the training material should be validated for learning and assessment of progress.

NHS FASP provides a suite of pictures which are considered “ideal” for the purposes of estimating fetal weight, these are shown in *Table 6.8*. Commonly, the trainee is given these images and is expected to scan the fetus until they obtain the ideal image, or a close approximation of that image. Early in their training the operator will usually be directly supervised and offered instruction by their trainer. Instruction can be highly variable between trainers. Individual trainers may use inconsistent terminology and instruction. Sometimes guidance is limited by clinical workload and the trainee is not adequately debriefed by their trainer. The trainee may misinterpret or become confused by such inconsistency. Surprisingly, contemporary high fidelity ultrasound simulators, such as “ScanTrainer”, have the same problem. The focus remains on submitting a close approximation of the ideal image, rather than an understanding of how to make that image. Whither training using a simulator, or in clinic the problem remains the same. The trainee is presented with the ideal image, they may have an appreciation for its appearance, but limited dexterous skill to obtain it. The trainee is not offered direction in how to move the probe, or a logic-based explanation to achieve a better image. This leads to inevitable frustration, negative reinforcement of poor performance and even the abandonment of training. As training progresses, the trainee must develop strategies and techniques to intentionally achieve the ideal image, rather than find it by chance. Simulators offer no benefit to the trainee because they do not know where the US probe is in relation to the anatomy of interest. They cannot, therefore, offer any guidance in how to navigate. As their understanding of how the ultrasound beam “cuts” through the anatomy and the on-screen appearance of anatomy develops, sonographers demonstrate increased intentional movement of the probe. This is the behaviour that CAL-Tutor aims to emulate by presenting the US beam “cutting” through the fetal anatomy using an Augmented Reality (AR) device to enhance the fidelity of the current generation of simulator.

The earlier chapters focussed on assessing performance difference between expert and novice operators and understanding training needs, as perceived by doctors in Obstetrics & Gynaecology. I proposed a plausible learning curve for the attainment of skills in obstetric ultrasound using an objective metric, DSJ. That timeline forms the basis of the learning program. The competing priorities of trainees, their trainers and what can be realistically developed, tested and deployed in a single PhD research project have shaped the outcome. This work attempts to avoid the limitations of previous devices such as their poor representation of the clinical environment, current generation ultrasound equipment and unvalidated assessments of training. The vision is to standardise training in obstetric ultrasound and increase trainee confidence prior to commencing supervised clinical practice. I aim to train inexperienced individuals to the point where their confidence allows them to perform US under indirect supervision. The is reflected by the inflection point in skills assessment discussed in Section 5.5. I aim to use a mixed reality head mounted display to deliver instruction to trainees while they performed ultrasound. The system, known as CAL-Tutor, was developed in the ultimate part of this PhD. Mixed Reality is an interesting research area because high quality US requires the trainee to appreciate the appearance of fetal anatomy on an ultrasound display. The trainee needs to appreciate the difference in appearance between the anatomy they are visualising in the present compared to the anatomy they want to see. They must appreciate how to move the probe to make this happen.

6.2.4 Using standard planes as a training target.

As discussed previously, there are three key planes used for the estimation of fetal weight by ultrasound. These are the Bi-Parietal Diameter (BPD), the Abdominal Circumference (AC) and Femur Length (FL). These planes are well described by many authors and have been extensively used in research and quality control studies. These views have also been used in the earlier experiments detailed in this thesis. For these reasons BPD, AC and FL will be the planes of interest for the training experiments. The RCOG curriculum encompasses a broader suite of images that the intermediate-level trainee should be able to achieve independently. Some aspects, such as adjusting gain are subjective. Other US findings, such as fetal lie and position are implicit if fetal biometry can be achieved. For this reason, training efforts in this thesis are focussed on these three standard planes.

In previous chapters I have described how I combined a clinical ultrasound system with a commercially available obstetric ultrasound phantom. The results allowed me to better understand how clinical experience affects probe motion and to a lesser extent US image quality. I aim to train the user to mimic expert behaviour, rather than try to match the image to an “ideal” or “perfect” image. I aimed to build a system that seeks to give the trainee consistent, real-time feedback on the position of the ultrasound probe relative to the anatomy being scanned. I hypothesised that this would be a viable experiment because expert ultrasound operators have been shown to move the ultrasound probe more frugally. This implies that the experienced operator has a greater understanding of what would happen to an image given a certain probe movement. By giving the trainee consistent instruction and asking them to move the probe in a single direction at a time, the goal is to allow the trainee to understand the consequence of probe motion on the anatomy being scanned. This system would act as a tutor for the trainee. The CAL-Tutor system replicates the apprentice and master training relationships discussed earlier, which have been eroded by changes in the way doctors work and are trained. This approach allows for consistent trainee instruction rather than being delivered by a single person at a single training facility.

6.3 Proposing Mixed Reality as a training tool

I have described obstetric ultrasound as a difficult skill to gain competence in and how experienced US operators move the probe more efficiently, probably ebcaue they have built a virtual mind map of what major anatomical structures normal fetus would look like on an ultrasound. This allows them to orientate themselves within the fetus quickly. Trainees would benefit from visual queues emulating this skill. Some US simulators, such as ‘ScanTrainer’ present this visualisation, but it is not offered in the users direct field of view. Nor is any instruction offered on where to move the probe to achieve the desired standard plane. Virtual, Augmented and Mixed reality offer the potential to present information on probe position to the trainee. In the following discussion I discuss the technologies and the implementation in CAL-Tutor.

6.3.1 What is Mixed Reality (MR)?

Augmented reality (AR) is described as an enhanced version of the real world. AR adds digital elements to a live view of the real world. This is often achieved by using the camera and screen of a smartphone or tablet. Mobile gaming such as Pokémon Go is a popular consumer-grade examples of AR. The user’s world is augmented through the lens and screen of a smartphone, tablet or similar device. This concept can be applied outside gaming, or visualising objects such as furniture or paint colours in the home. When used in medical education Virtual Reality (VR) and AR are able to give the impression that organs or tissue layers have been removed or made transparent. For example, AR can be used to overlay the appearance of an object inside an otherwise solid or opaque object. AR differs from VR because performed task lies within in the real world. VR is an entirely immersive experience. VR removes the individual from their physical reality and all interaction happens in a digital, or virtual, environment. With the removal of external cues VR allows for the use of sham probes or tools. The real world cannot be seen when wearing a VR headset. In contrast, AR presents the user with an interactive overlay onto their real environment. This approach allows the physical model and an enhanced visual perception to be blended into a single training environment.

Mixed Reality (MR) goes a step further and combines the ability to allow physical objects and digital objects to interact with each other. In the last 10 years a number of MR devices have been developed by Google, Facebook and Microsoft. The current state of the art is the MS Hololens2, a development of the original Hololens. Hololens2 is considered an untethered device. It can be considered a computer in its own right, unlike some earlier devices which required a wired, or wireless connection to an external computer. Such tethered devices can be considered as a monitor worn by the user. Hololens2 does not require another device to function. The standalone headset is controlled by the wearer through a holographic interface. Microphones and cameras allow for voice and gesture control rather than conventional keyboard and mouse interfaces. Recently described uses in medical training have been in laparoscopic surgery, endotracheal intubation, joint injections, and assistance in placing local anesthesia[69].

Hololens2, detailed in *Figure 6.2* has capabilities which make it suitable for training. Hololens2 can present information to the user while tracking the position of their hands and other physical objects by way of optical tracking cameras built-into the device. As Hololens2 can record and track eye gaze, hand movements and speech a rich dataset can be recorded without using additional recording or tracking equipment. A final feature is the HoloLens2 depth-sensing ability, which can be used to support remote pointing and enhance procedural guidance.

The Hololens2 has some disadvantages to be considered. These include the weight of HoloLens2, discomfort when wearing the headset due to heat and even pain are recurrently seen in user feedback[70, 71]. Additionally, the battery life of HoloLens2 is limited to approximately 100 minutes of use, or less. HoloLens2 will sometimes close down an application due to the limited memory size. Given these resource constraints the performance of the device, as perceived by the user, can be disappointing, even frustrating. Careful consideration has been given to resource allocation and the data I collected to ensure that the augmented reality environment which is presented to the learner is stable and runs smoothly. A balance between what is technically possible and practical had to be struck to ensure that the CAL-Tutor user experience augmented, rather than impeded the trainees learning. These issues notwithstanding a superior, cheaper and more powerful device is not currently available and the MS Windows-based architecture is familiar to many software developers. The relatively large and growing research community which is using Hololens makes development easier and troubleshooting more efficient. Much of the software already developed in research settings is open-source and available in online repositories, again easing the development pathway.

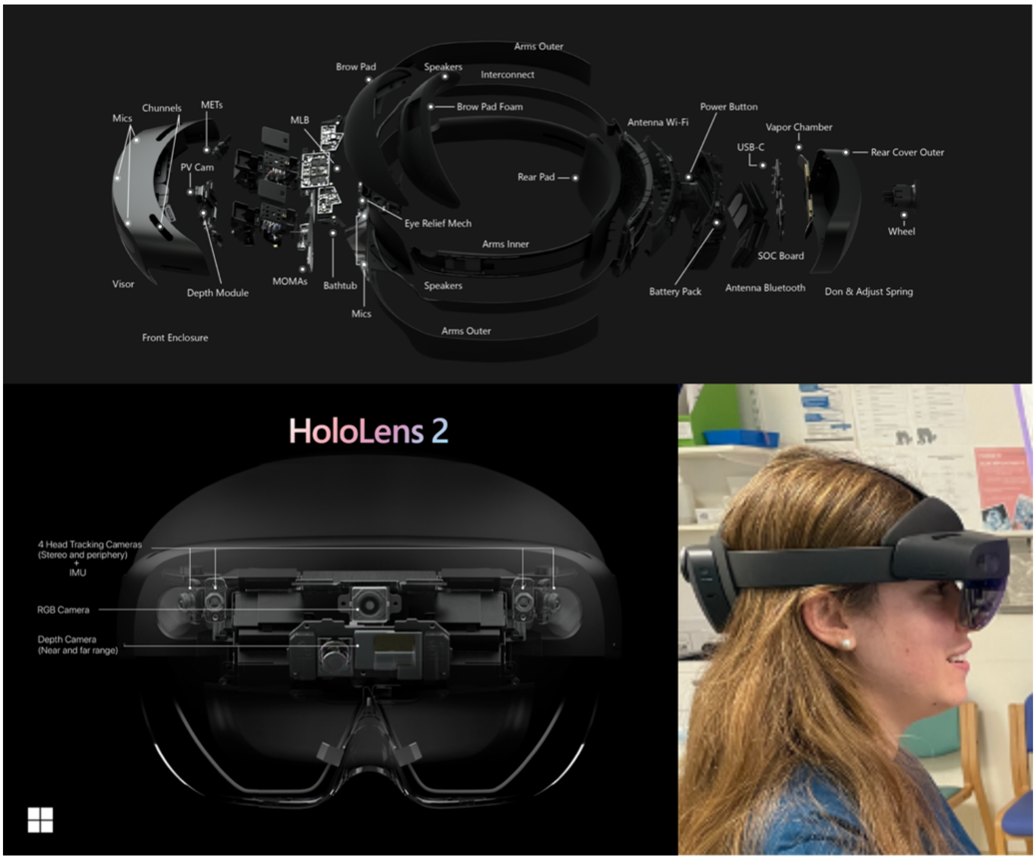


Figure 6.2 – Exploded view of the components of Hololens 2 and the device being worn by a user. Adapted from Hololens2 support documents from Microsoft. https://docs.microsoft.com/en-us/hololens/hololens2-hardware

The fundamental concept of the MR device was to allow trainees to understand the spatial relationships between fetal anatomy and the ultrasound beam. By allowing trainees to appreciate the relationship between the US beam and the fetal anatomy it is hypothesised that the location of the standard plane can be more easily appreciated by the trainee and then navigated to. MR technology can deliver such a learning experience. It offers a generational improvement over currently available simulators and allows trainees to interact with ultrasound devices and a phantom in ways not previously possible. CAL-Tutor helps to create authentic simulated experiences, using clinical US machines and presenting probe guidance within the trainee's gaze. This is an approach not used by previous simulators or implementations of MR technology in ultrasound education.

6.3.2 The evidence for AR in medicine

As previously noted, workplace conditions for learning are limited by clinical workload. While simulation is predominantly used to train professionals away from the patient environment, AR has primarily been used to train or educate medical professionals in clinical environments[72]. Most authors describe uses as a navigation tool during surgical procedures[73] or to enhance visualization at the operating theatre[74]. AR has seen limited use in training sonographers, but it is a relatively new technology in educational methodology[75]. The contribution of this work is the novel combination of a mixed reality device with a physical US simulator. By using Hololens2 as the basis for a training device, the advantages of a physical mannequin, clinical-grade equipment and virtual reality simulation can be combined. The resultant guidance is brought into the direct view of trainee sonographers. The physical presence of the phantom is used for haptic feedback and to allow the clinical US system to perform as closely as possible to how it would when scanning *in vivo* while 3D visualization is provided through Hololens2 head mounted display.

Several authors have shown the potential of AR to bridge the gap between achieving the actual competence needed in the real working environment and training them in a virtual context. In 2016 a systematic literature review reported on seven AR applications that have been developed for medical training[74]. While the methodologies relied heavily on subjective reports of trainee satisfaction, no follow-up studies on retention of skills were identified. The authors also failed to find evidence of subsequent clinical improvement of trainees from the studies. The review concluded that *“To date, it is unclear if the use of ARAs in training medical professionals is likely to contribute to patient safety. However, as training methods become more engaging and reliable, learning curves may be expected to become steeper and patients will ultimately benefit.”* More recently, a similar review was conducted in 2020[76] which found similar results. The authors identified 100,807 articles and of these 36 were deemed of sufficient quality for inclusion in the review. None of the studies included in the review considered all five assessments of validity for simulated devices. As in the 2016 review, my literature review the authors reported that the volume of publications is growing at a rapid rate but that methodology is poor and this reduces the generalisability of findings for trainees.

6.3.3 The evidence for AR in US.

The evidence for the use of AR in ultrasound is limited by the relative infancy of the technology. The available evidence follows much the same trajectory as wider medical education. There seems to be wide interest in the technology but an unclear implementation, poor validation of training interventions and limited follow-up or comparison to conventional training. A recent study examined how an AR application could improve performance in a group of 66 medical students[77]. The newly developed augmented reality ultrasound simulator mobile app provides a useful add-on for ultrasound education and training. The results indicated that medical students' use of the mobile app for training purposes improved the quality of kidney measurements. The importance of validating new tools within the field of medical education is noted by authors. The importance of validation is illustrated by the fact that validity assessments of the device have indeed been undertaken, especially since 2011. However, no follow-up studies on retention of skills was identified. Consequently, no evidence of subsequent clinical improvement or sustained skill of trainees be retrieved from the available literature.

6.3.4 How does the initial work support a mixed reality US training system?

I believe that the ambition to allow trainees to achieve a level of skill that would allow them to perform obstetric ultrasound using high-fidelity, interactive simulation is a realistic prospect. This belief is not without supporting evidence. Studies by Chalouhi et al[42] and Maul et al[38], showed that even novice operators could achieve competent performance in obstetric ultrasound when being trained by means of simulation alone. The authors compared their simulation-based curriculum to conventional didactic teaching of ultrasound theory and practice. Currently, there is no clearly defined pathway for trainees to further develop the skills acquired in a simulated environment to a clinical setting. CAL-Tutor attempts to resolve this by combining theoretical practical and assessments into a single programme.

The concept of CAL-Tutor is to enable consistent feedback combined with visualisation of the ultrasound probe, beam and fetus. This will allow the learner to better understand the complex relationship between the 2-D ultrasound image presented to them on-screen and the 3-D anatomy of the fetus. The combination of a phantom and instructions presented through an HMD allows the learner to scan and receive guidance simultaneously. The feedback is given in the context of the task the trainee is trying to perform. Psychologists and those interested in adult education have extensively

studied language, gestures and how closely the learning environment reflects the ultimate task[78], [79]. The development of CAL-Tutor follows directly from these conclusions by integrating the training environment into daily clinical workflow.

The CAL-Tutor system, as discussed earlier has the potential to deliver improved learning outcomes for trainees by allowing trainees to visualise how the US beam produced by the probe intersects through the fetal anatomy. A deeper understanding of this spatial relationship would allow the trainees to visualise where the probe is now in comparison to where it needs to be. Using consistent instruction the system guides the trainee to this location. Over time this instruction can be removed as the trainee will have learned the required motion for themselves. Finally, an objective measure of their progression through the simulated part of their training will allow trainees and trainers to assess their progress. This progress can be plotted against the learning curves described earlier. These curves are based on DSJ data to assess the effect of a training package.

CAL-Tutor is a computer system that guides an ultrasound user through the steps of obtaining a diagnostic quality ultrasound image of a target anatomical structure, providing real-time feedback on whether each of these steps is being performed correctly. The system will contain a set of user-selectable target planes that are standardised in clinical practice and that are routinely used for biometry measurements. The instructions will be provided on a screen or HMD that also displays a 3D representation of fetal anatomy and the ultrasound probe trajectory. Guidance arrows change in size and colour to guide the trainee to the ideal plane. Deviations from the desired probe trajectory will be detected and the user guided how to move the probe so that they can obtain the plane of interest. The guidance instructions given by CAL-Tutor uses consistent terminology to describe probe movement. Consistency of instruction is important because inconsistent instruction to trainees can lead to confusion for the trainee and slow the acquisition of skill.

Given the other parameters recorded by Hololens2, including eye gaze, time to completion of tasks, number of attempts to acquire a satisfactory image and overall QC score could be included in a multi-variate analysis. This could be correlated with user reported scores and trainer OSATS to assess success of achieving my aim of delivering high-quality, reproducible training in a clinical environment.

Ultimately, the novice sonographer should more consistently achieve the desired image, regardless of fetal position or patient habitus. As a team, my collaborators and I have developed a simulation environment that;

- provides real-time feedback on the US probe position relative to a fetal model
- establishes a hand/probe motion plan towards aligning the scan with the desired target image
- assesses the quality of the images saved by the sonographer

- provides a training summary and additional guidance, as required. This guidance will be based on NHS FASP guidelines, or other high-quality peer-reviewed literature.

By combining the on-screen instructions given to trainees with a real-time classification architecture I believe we can train novice operators to imitate expert-like behaviours more quickly than a conventional apprentice model. While highly effective, it is resource intensive and requires a large department with stable staffing and high motivation to training. These conditions are not routinely available for doctors in training in the UK.

6.4 Assessing the validity of Mixed Reality as a training system.

In the literature review in Chapter 2 I discussed that the face, construct and content validity of many of the commercially available ultrasound simulators has not been evaluated or established. To avoid this issue and to ensure that a valid training environment was being proposed a number of validity assessments were considered for the CAL-Tutor system. The assessments were carried out during development of the hardware and software for the CAL-Tutor HMD. Given the weaknesses in previously described literature, I felt that it was important to undertake the assessments against established criteria before self-directed learning material and the assessment modules associated with the CAL-Tutor system were finalised.

The results of the validity assessments directly led to the creation of an online training portal where the trainees could find guidance and information on theoretical and technical aspects of learning obstetric ultrasound. The design of the modules was conceived to be iterative. Initially the trainee would learn fundamental theory, fetal anatomy and functionality of an ultrasound system. They would then have the opportunity to use the MR system to train in a simulated environment. A clinical assessment of their ability to perform fetal biometry would be undertaken at the end of the course. The intention was twofold. The clinical assessment would give trainees a learning outcome to achieve and the assessment would allow them to benchmark their progress against that outcome.

Despite the development of a technologically novel platform for training in obstetric sonography, there was a risk that a standalone device would face many of the same issues as simulators which are already available. Simulators tend to assume a certain level of knowledge and skill. This is not always the case, my initial survey asked trainees what ultrasound abilities they had. The results suggested that some trainees did not feel confident that they could operate the basic functions of an US machine. Some of the answers given by trainees to questions were contradictory, further suggesting a poor understanding of the functions available to optimise clinical US images. To adequately fulfil the "Learning & Knowledge" aspects of the validity requirements the pre-existing knowledge and ability of trainees could not be assumed. The responses to the initial survey of trainee experience suggested that anatomical knowledge and appreciation for the appearance of images on screen formed only a single part of the trainees challenge. Trainees did not understand how to use the basic functions available for image optimisation, zoom and callipers for measurement of fetal structures.

Beyond an understanding of the US machine functionality, trainees expressed interest in performing more advanced tasks beyond biometry. These included doppler assessment of the middle cerebral artery or umbilical artery of the fetus and transvaginal cervical length. These are examinations which can be used acutely at the patient bedside, can be used to manage the timing of delivery or to diagnose pre-term labour. The interest in performing these kinds of ultrasound examinations suggests that trainees desire the ability to perform assessments which are of immediate use to patients in acute settings. These might include the assessment of fetal wellbeing and measuring cervical length to predict the risk of pre-term labour. An example would be the QUIPP App which uses the length of the cervix as one of the factors for calculating the risk of pre-term labour in the subsequent 24 hours. Despite clinical tools being available which alter patient care and which have been demonstrated to improve outcomes for babies[80], contemporary training fails to deliver even the basic understanding of ultrasound physics and machine capabilities, much less the competencies required to assess fetal weight, cervical length or obtain doppler values. This is despite trainee recognition of the importance of US in acute settings and enthusiasm for its use.

The CAL-Tutor system was benchmarked against four aspects of validity reported in the literature, these were content, criterion, construct and face validity. Their application in relation to CAL-Tutor is detailed in *Table 6.2*. *Table 6.3* details the how CAL-Tutor will address the learning needs from *Table 6.2* and knowledge gaps from the trainee survey in *Figure 6.1*. This level of prospective needs assessment and benchmarking has not been reported to date. This work itself identifies the need for a systematic and robust methodology to be used during the design of simulators and training material in medical contexts.

Validity	Definition	Application to CAL-Tutor
Content	Test items and format constitute a relevant and representative sample of the domain of tasks	Use of SPACE-FAN phantom with a clinical ultrasound system. The operator is manipulating the probe and machine settings. Previous investigations have shown that for novice operators there is no statistical difference in DSJ between clinical or simulated settings for novice operators, reflecting its suitability as a training tool in early training.
Criterion	Correlation between actual test scores and the "true" (criterion) score	There are no defined test outcomes from CAL-Tutor (these may be added after subsequent analysis of the collected user data). Data to be collected by the CAL-Tutor/Hololens • Attendance (number of attendances) • Pre-Course Quiz Assessment Score • Time spent using CAL-Tutor • Images Saved to the US drive • User Eye Tracking • Probe Tracking Metrics Trainee performance between the 'entry' and 'exit' scan will be compared by means of • Assessor OSAT • Trainee reported confidence • Images obtained & Scored • Probe Tracking metrics (DSJ)
Construct	Scores vary as expected based on an underlying psychological construct	Assessed by • user Questionnaire • trainer Questionnaire
Face	extent to which a test is subjectively viewed as covering the concept it purports to measure.	Assessed by • change in OSAT scores • change in DSJ metrics

Table 6.2 - Validity assessment, definition and knowledge gap assessment for the CAL-Tutor learning system.

	Construct and proposed interpretation	Make explicit the intended decisions	Define the interpretation-use argument, and prioritize needed validity evidence	Identify candidate instruments and/or create/adapt a new instrument	Appraise existing evidence and collect new evidence as needed	Keep track of practical issues including	Formulate/synthesize the validity argument in relation to the interpretation-use argument	Make a judgment: does the evidence support the intended use?
	Learners' knowledge of US indications and risks	*Performance Feedback for the learner*	"pass" allows the attempt of supervised US on a real patient.	*Course Structure*		*Location*	scoring and generalization evidence	
Training/Learning Aspect	Pre Course Material (online)	*Cal-Tutor GPS Arrows & Consistent Direction*	Pre Course Material (online)	Pre Course Material (online)		Charles Bell House / EGA Wing	Outcomes being scored against RCOG OSAT, being recorded in trainee portfolio.	
Assessment	Pre-Course MCQ	Post Course OSAT	Obtain 3x planes (Head, Abdomen and Femur)	pre-course work attended, pre-course score	RCOG OSAT for Intermediate Module	Engineering Support		
	Learners' ability to perform US		*Appreciation of the required standard of obtained images*	CAL-Tutor & Hololens			extrapolation evidence	
Training/Learning Aspect	*Cal-Tutor GPS Arrows & Consistent Direction*		Pre Course Material (online)	*Cal-Tutor GPS Arrows & Consistent Direction*			Direct comparisons between simulation and real-world performance based on metrics recorded for each individual trainee in both environments.	
Assessment	*Post Course OSAT*		Score the images saved to the US memory based on INTERGROW criteria.	collect data such as images, image score, DSJ, probe path length, time to completion, time spent on simulator, number of attempts at training, number		Can this be self directed?		

Table 6.3 – Proposed aspects of training to be addressed by the CAL-Tutor system, as identified by the validity assessments and knowledge gaps..

6.5 Technical Development of the CAL-Tutor System

The CAL-Tutor system was developed in an iterative fashion, over time. The final system represents a collaboration. The collaboration developed CAL-Tutor from my original concept of a device which would track learners' performance over time and offer feedback on their performance and progress. The technical development was carried out at the WEISS centre at UCL. This thesis describes the process of the initial concept and mock-up using research standard equipment and the subsequent transition to using clinical ultrasound systems and a mixed reality headset. The discussion of the technical challenges are brief as they are the work of my collaborators. Principally Manuel Birlo and Chiara Di Vece who are PhD students under the supervision of Professor Stoyanov. Their research and innovations will be detailed in their theses. In the interim, details of the technical challenges and innovations are included in the submission to the MARRS competition and to IPCAI 2022, where the long abstract won the best presentation award in the digital surgery category.

6.5.1 Initial Concept & Early Mock-up.

The CAL-Tutor training system could have been developed on a conventional monitor adjacent to the clinical US machine rather than on a head mounted display (HMD), such as the Hololens2. In either the conventional monitor or HMD version the user would be able to see how the motion of the ultrasound beam 'cuts' through the fetus and how that 'cut' is affected by their movement of the probe. With a conventional monitor placed alongside the US system there are several disadvantages. Locating the monitor such that it is in the gaze of the trainee, adjacent to the US scanner but not impacting on the patient positioning and therefore the ergonomics of the machine is difficult. A display monitor would require an external computer to run the guidance and tracking software. Tracking hardware, either electromagnetic or optical would have to be installed in the scan room. The additional hardware required and the impact on the ergonomics of the US system make a conventional display less mobile, more cumbersome to install in a clinical setting. This directly impacts the design goals of a training system which was easy to set-up, use and portable.

The conceptual mock-up in *Figure 6.3* represents the information that trainees would require, including information on the standard plane to be obtained, spatial information on where the probe is relative to the fetal anatomy and the next probe motion required for the trainee to navigate toward the standard plane.

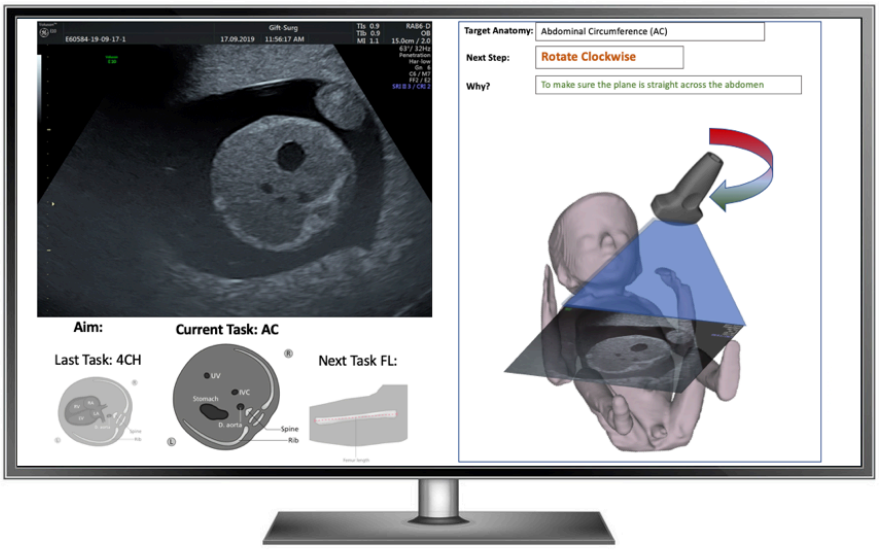


Figure 6.3 - Representation of the User Interface of "CAL-Tutor" on a conventional monitor. The data displayed in this mock-up was preserved in the HMD version of the system.

By combining the graphical instructions given to trainees with a real-time image classification and scoring system I believe that novice operators can be trained to imitate expert-like behaviours. The ultimate iteration of the concept offers trainees feedback both on the position of their probe relative to the fetus and the quality of the image that they have produced. This approach gives trainees the certainty of consistent direction and allows them to track their progress as their training advances.

When worn by a trainee, the HMD offers the trainee the ability to see the fetal anatomy and navigation instructions overlaid on the maternal abdomen. I have already introduced the idea that instruction would best be given in the trainees' direct line of sight. There is some evidence in medical[81] and aviation[82] settings that information presented in a head-up-display can improve clinician and pilot situational awareness. Because I wished to explore the effect of the trainees understanding of the fetal anatomy in relation to the US probe, a HMD-based system offered a viable device to deliver the concept in *Figure 6.4(a & b)*. Hololens2 was selected over conventional display options.

An initial conceptual diagram is shown in *Figure 6.4(a).* The aim of this diagram was to illustrate how the ultrasound emitted from the US probe (A) interacts with the structure of the fetus. Three standard planes (B) are shown to illustrate their relative positions in the fetal anatomy. The novelty arises from the directional arrows (C) which guide the sonographer towards the standard plane of interest.

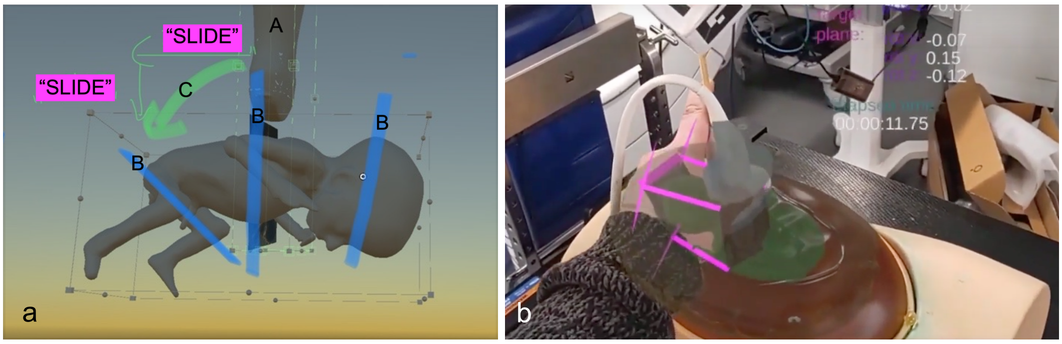


Figure 6.4 – An early conceptual view on the 3-D objects placed in a unity scene on the Left. On the right is the final version of the system as tested. Legend: A is the Ultrasound Probe, B are the three planes for fetal biometry, C represents the direction the probe should be moved to acquire a plane (in this case the Femur Length FL).

Figure 6.3 (a) on the left, does not depict the view the end user might see Rather it represents an early concept of the overall instruction to be given to the trainee sonographer. *Figure 6.3 (b),* on the right, represented the final evolution, as presented to participants in this study.

The main component of this system is a software package that runs on a dedicated HMD and is connected to an unmodified ultrasound machine. The HMD includes an optical tracking system with a tracking device that can be attached to the ultrasound probes used in the Fetal Medicine Unit at UCLH (GE C1-5). The tracking device is attached to the probe by means of a non-permanent 3-D printed holder, as the electromagnetic trackers were attached to the US probe in the DSJ studies.

6.5.2 The Proof Of Concept With A Conventional Display.

The initial, conventional display version of CAL-Tutor is described below and is shown in *Figures 6.5 a & b.* Hardware readily available within WEISS was used and a demonstration system was created. At this stage a conventional display was used. The hardware used is detailed below

- SPACE-Fan obstetric ultrasound training phantom
- Ultrasonix Sonix MDP US machine
- 3-D Slicer model of the fetus, images from the Ultrasonix US system were streamed to an external laptop. Within slicer, a virtual US probe controlled the location of the US steam on-screen.
- The on-screen virtual probe was manipulated by using the real-world US probe. This was achieved by tracking the position of the probe using, an optical tracking system.

The hardware combined with software gave learners an idea of how the US beam and the baby interact. It acted as a proof of concept that the visualisation of the US as it moved in relation to the fetus was viable.

Figure 6.5– Initial mock-up using '3D Slicer', optical tracking and an ultrasonix US system.

The initial version of CAL-Tutor used an Ultrasonix US system, which is a research system and not suitable for use in clinical practice. The machine functions and button layout differed from the GE systems in use at the hospital site. The initial models of the US probe and fetus were of low resolution and limited fidelity to clinical systems. This led to limitations when the virtual and real models were being calibrated. The alignment of the virtual fetus and the real maternal abdomen was done manually. The only visualisation offered to the trainee was of the US beam as it moved around the surface of the SPACE-Fan. There were no instructions or direction on how to move the probe.

As research software, the nature of 3-D slicer made it difficult to reload quickly in the event of an error or unexpected system shutdown.

Unity, a game engine was suggested by the engineering team as a more stable, adaptable and contemporary platform to develop further versions of CAL-Tutor. It has advantages including being able to display 3-D renders which had already been created of the various US probes, the fetus and the maternal abdomen model. The engineering team were able to import these items into a single Unity scene. *Figure 6.6* shows the initial view through Hololens2 of the 3-D model of the fetus placed in relation to the physical phantom. This was the first view we obtained in which CAL-Tutor made use of mixed reality technology. In *Figure 6.4(b)* the virtual and world phantom can be seen together. The CAL-Tutor concept is built on the potential for the real and virtual worlds to interact, the remainder of this thesis focusses on that potential.

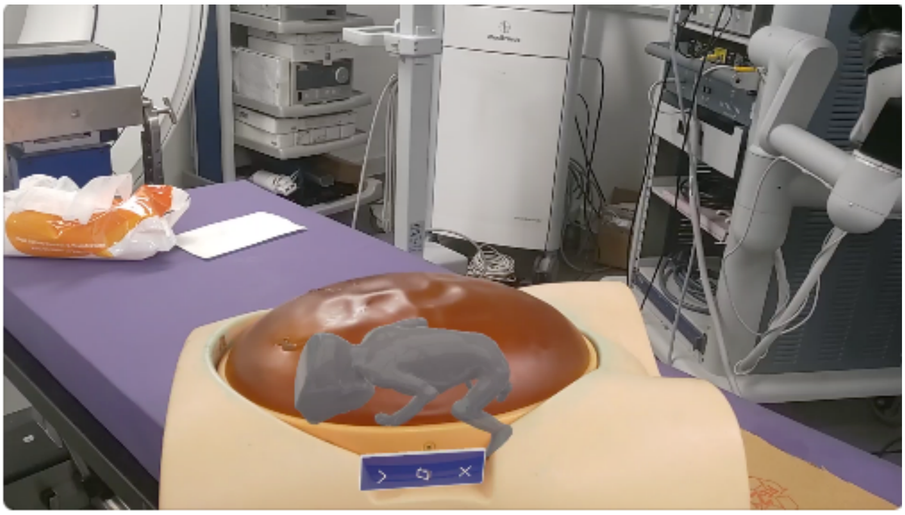


Figure 6.6 - First view through the Hololens of the fetus approximately aligned with the SPACE-Fan phantom.

6.5.3 Transition to Hololens2-based system

This thesis aims to explore if a mixed reality training environment would be superior to a purely physical or virtual training approach. I hypothesise that allowing a trainee to visualise how the US beam emitted from the probe interacts with anatomical structures would allow the trainee to build a` 3-D spatial model of the fetus. Whither training using a simulator, or in clinic the training issues remain the same. The trainee is provided with an ideal image, as shown in *Figure 6.7 (A)-(C)*. While they may have an appreciation for the appearance, they have limited dexterous skill to obtain it. The trainee is not offered direction in how to move the probe, or a logic-based explanation of how to achieve a better image. This leads to inevitable frustration, negative reinforcement of poor performance and even the abandonment of training[38]. I propose a Unity game engine-based HoloLens application using imported 3D models and external tracking software. Similar technologies have previously been described in ultrasound simulation, surgical workflow planning, US guided procedures and transthoracic ultrasound[75,83,84,85]. The proposed solution creates an environment where holograms interact with physical US systems and US training phantoms. Real-time tracking of the ultrasound probe allows a visual association of the ultrasound probe in relation to the fetus and the ultrasound image this generates on-screen, see *Figure 6.7* (D). As currently described CAL-Tutor is made up of three physical components; a low fidelity US phantom, a Holographic mixed reality headset and a commercially available US machine. In this case the GE Voluson E10.

HoloLens 2, announced in 2019, brings a number of improvements with respect to the first-generation device. The HoloLens 2 also includes a larger field of view, a custom Deep Neural Network core, fully articulated hand tracking and eye gaze tracking. The DNN core was introduced to avoid latency and lag in machine learning and vision applications. A detailed schematic of the device is seen in *Figure 6.*2.

The device features a second-generation custom-built Holographic Processing Unit (HPU 2.0), which enables low-power, real-time computer vision. The HPU runs all the computer vision algorithms on device (head tracking, hand tracking, eye gaze tracking, spatial mapping etc.) and hosts the DNN core. It is located on the front part of the device, near the sensors (*Figure 6.2, top*). The CPU on the SoC (a Qualcomm SnapDragon 850) remains fully available for applications. The SoC is located on the rear (*Figure. 6.2, bottom*).

The device is equipped with a depth and an RGB camera, four grayscale cameras, and an Inertial Measurement Unit (IMU), as shown in Fig. 6.6. Audio is captured with a microphone array (5 channels).

- Inertial Measurement Unit (IMU):
- Accelerometer, used by the system to determine the linear acceleration along the x, y and z axes as well as gravity.
- Gyroscope, used by the system to determine rotations.
- Magnetometer, used by the system for absolute orientation estimation.

6.5.4 Visualising fetal anatomy using mixed reality

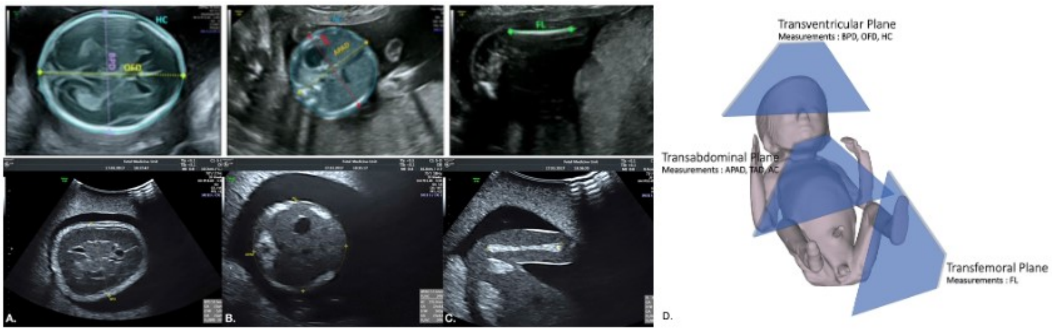


Figure 6.7 - 2d: Images of fetal anatomy in the second trimester.(A)Transventricular plane, used for measurement of the biparietal diameter(BPD) (B)Transabdominal plane, used for measurement of the abdominal circumference(AC), (C) Required view of the femur used to measure the length of the bone (FL). The top row shows the planes as achieved on the SPACE-Fan training phantom used in this study. (D) An illustration of the shape of the US beam and how it 'cuts through' the fetal anatomy to achieve the US images A, B and C.

As I have previously explored, US is widely used in Obstetrics[86] but training is challenging[33]. This is because US examinations require the operator to appreciate the US appearance of normal anatomy, understand how to optimise an image using the various machine settings and be able to build a mental three-dimensional (3D) model of the fetal anatomy to navigate the US probe effectively. To do this the operator must understand that US images depict slices through the anatomy. US penetrates the skin and so US does not produce a picture of the surface of the anatomy as a conventional camera does. Rather, the probe in the sonographers hand produces a thin, fan-shaped beam of sound, as illustrated in *Figure 6.7 (D)*. US images are produced based on the reflection of the waves from the interface of tissues of differing density. The resulting US image is a cross section 'through' the anatomy, rather than a reflective image of the surface. The perspective of ultrasound images shares more with CT and MRI images than camera systems. Only a single cross-section can be visualised on-screen at any given time. Furthermore, the user is free to move the probe in any of the 6 degrees of motion. The relationship between the fetal anatomy and the US probe is demonstrated in *Figure 6.7 (D)*.

In summary each US image represents a very thin slice through the fetal anatomy, roughly 1mm in depth. Because neither the fetus nor the probe are fixed with reference to maternal anatomical landmarks the operator must build a 3-D map of the fetus in their mind to allow them to navigate effectively. This map must be sufficiently flexible to allow the trainee to alter course when the position of the baby changes during the scan. This is especially important because certain anatomical structures, including certain views of the heart can best be visualised from the anterior aspect. This concept is known as an ‘acoustic window’ and further limits the options available to the sonographer, in addition to the other limitations.

6.6 Design & Development of the mixed reality environment.

The mixed reality environment combines a clinical US system, a physical phantom and virtual models. Therefore, CAL-Tutor has multiple constituent parts. These are

- The Physical Simulator & Clinical Ultrasound Machine
- The Virtual Environment, Hololens
- The CAL-Tutor Mixed Reality Trainer

These are displayed to the user through an HMD. The mixed reality environment combined a low-fidelity ultrasound simulator (SPACE-Fan) and an unmodified clinical US system. This was achieved using a Hololens2 mixed reality headset (HMD) running a Unity game engine-based application. ARToolkit tracking allowed the US probe to be tracked in real time. The position and motion of the US probe relative to the 3-D model of a 23-week fetus could then be displayed to the trainee in real time through the HMD.

6.6.1 Physical Simulator & Clinical Ultrasound Machine

A SPACE Fan-ST, Kyoto Kagaku, Kyoto, Japan, was used for this study. The SPACE Fan-ST allows the foetus to be placed in one of four fixed positions. These are demonstrated in *Figure 6.4.* The SPACE-Fan phantom contains skeletal structure,of a 23 week fetus. This allows for trainees to visualise the anatomical structures required by fetal anomaly screening programs, which take place around this gestation. The model can be used to image brain, 4-chamber view of the heart, lungs, spleen, kidneys, aorta, UV, UA, and external genitalia, as shown in *Table 6.5.* We combined the low-fidelity ultrasound training simulator with an unmodified clinical US system. The US system used was the GE Voluson E10. Trainee instructions were shown on the mixed reality HMD, a Microsoft Hololens2. The phantom has been designed to simulate the ultrasonic appearance of a 23-week pregnancy. I have previously validated its use for the purpose of training in obstetric ultrasound, showing that it can be used by both expert and novice sonographers. Performance differences can be shown between groups of expert and novice sonographers, when considering DSJ, time to task completion and probe path length. The GE Voluson system was chosen because they are the systems used in clinical practice at our institution. This creates a training environment that closely approximates the clinical reality. In the review chapter several contemporary ultrasound simulators were introduced. Unlike these systems which use laptop, or other consumer-grade computer hardware with haptic devices to simulate a training environment CAL-Tutor can be set-up in a clinical scan room. The trainee uses a clinical ultrasound system and places the SPACE-Fan phantom on the scan couch.

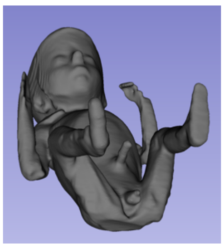
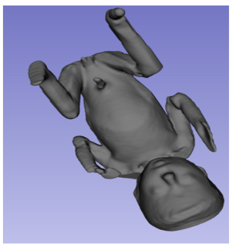
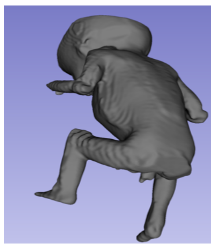
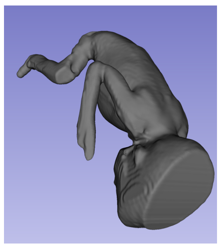
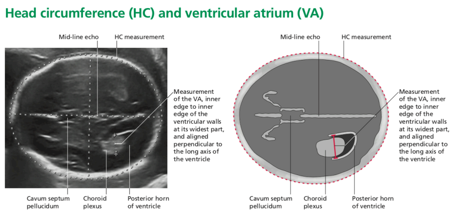
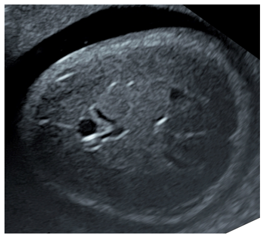
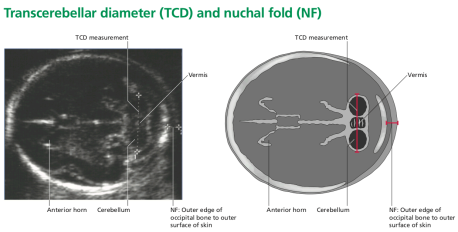
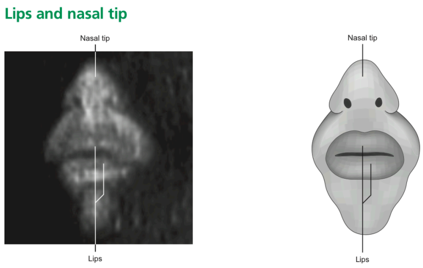
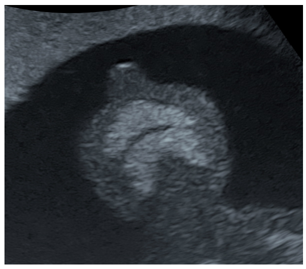
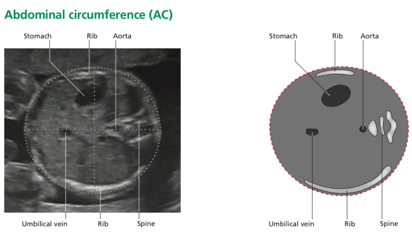
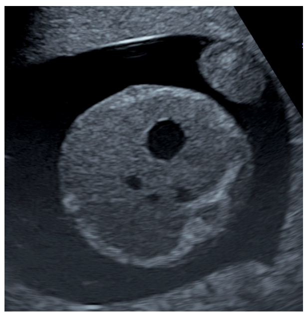
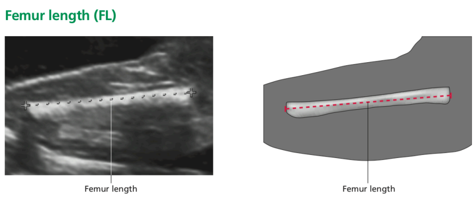
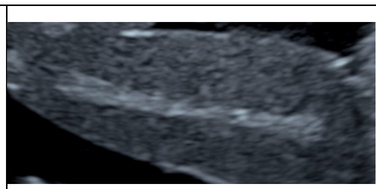
			
Breech, Sacroposterior	Cephalic, Sacroposterior	Breech, Sacroanterior	Cephalic, Sacroanterior

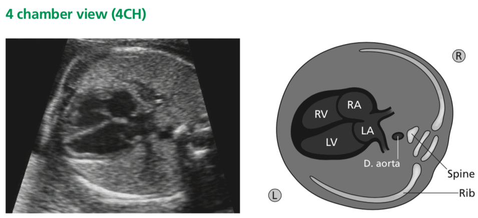
Table 6.4 - Fetal Positions which can be simulated using the SPACE-FAN ultrasound phantom.

In this study all participants were given relevant excerpts from the NHS Fetal Anomaly Screening Programme (FASP) handbook, as shown in in *Table 6.5* This handbook details the required planes for accurate and reproducible fetal biometry for use in all fetal screening provided by. NHS service providers. There are also posters depicting these planes in each scan room at the study site.

<table>
<tr><th>NHS FASP Reference Image</th><th>RANDO FAN Image</th></tr>
<tr><td>
</td><td></td></tr>
<tr><td>
</td><td></td></tr>
<tr><td>
</td><td></td></tr>
<tr><td>
</td><td></td></tr>
</table>







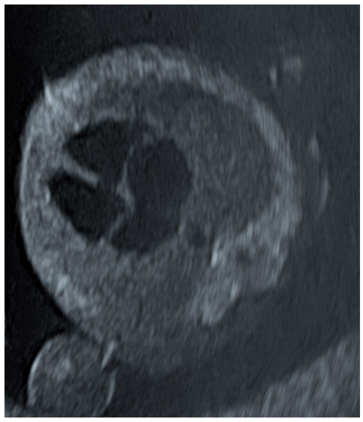


Table 6.5 – Views to be achieved by the trainees using "CAL-Tutor". The left columns shows excerpts from the NHS Fetal Anomaly Screening menu. The left hand side shows comparable views achieved on the SPACE Fan-SP phantom. Images on the left are adapted from the NHS FASP Handbook[87].

The experiments aimed to establish how a trainee's learning curve is affected by training based on imitating expert behaviours, rather than image matching. To do this, I invited novice operators to enrol in the CAL-Tutor training course. After enrolment, they were required to undertake a scan of a normal pregnancy at twenty weeks gestation as a baseline assessment of their ability to obtain the views shown in *Table 6.5*. I chose this scan because it is offered to all pregnant women and has a standardised set of images to be acquired. Ethical approval for the enrolment of pregnant women and sonographers was sought from the research ethics committee (REC) and had been granted before recruitment commenced, as detailed in section 3.2.

6.6.2 Virtual Simulator, Hololens.

The following software was used to implement the mixed reality software that is displayed in the HoloLens:

- Unity game engine v2020.3.14[88]
- Mixed Reality Toolkit (MRTK) v2.7.2.0 [89]
- HoloLensARToolKit: A Unity based marker tracking software for the HoloLens2 that uses the front-facing cameras[90].

Unity is a cross-platform game engine that allows users to create 3D scenes with 3D objects and user interaction logic. The MRTK is an asset containing mixed reality components and features that can be imported into Unity. It contains for example components that allow users to manipulate (move, rotate, scale) objects. The HoloLensARToolkit is a HoloLens adaptation of the ARToolkit open-source object tracking library that can be imported into a Unity scene as an additional asset, like the MRTK and contains features that allow 3D objects to be tracked using printed markers. A user-eye view of the physical and virtual environments are shown in *Figure 6.8*

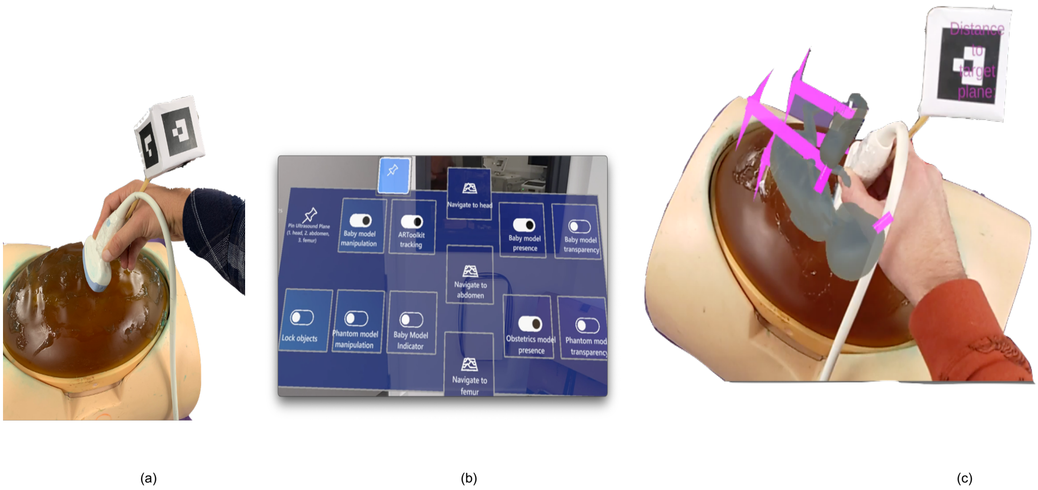


Figure 6.8 – Views of physical and virtual components as seen by the user through the HMD.
(6) Voluson ultrasound device with cube marker attached to the probe and an obstetrics phantom with a baby model inside. (b) CAL-Tutor Menu, (c) Voluson probe with attached marker cube in use with active ARToolkit tracking.

The virtual model seen by users in the Hololens2 does not contain anatomical detail. This is contained within the physical simulator. The anatomy can only be visualised by using an ultrasound scanner. This is in contrast to other training systems which use a simulated US machine or a sham probe. This approach allows the trainee to train using a clinical US system, rather than a simulator device with synthetic images, as seen in many high-fidelity simulators. The advantage is that none of the rendering or image processing is done by the simulator, or on the Hololens. This frees the limited computational resources of the Hololens for displaying navigation.

6.6.3 The CAL-Tutor Mixed Reality US Trainer

As explained in the previous section the Hololens tracks the US probe and displays guidance to the user. The US machine, buttons and controls behave exactly as it would when scanning a real patient. This is because the clinical US system and SPACE-Fan phantom are not modified. Once the guidance system has been activated the user is presented with arrows which are intended to guide them towards the standard planes for the estimation of fetal weight. The user views are shown in *Figure 6.8(c).* The instructions the user is shown are shown in *Table 6.6*. The nomenclature used has been based on AIUM terminology. This nomenclature has been reached by consensus of expert sonographers and was selected as it offered a recognisable standard which could provide consistent instruction to trainees when they were using the guidance feature of CAL-Tutor.

On-Screen Instruction: Slide (Up/Superiorly/Cranially) Slide (Down/Inferiorly/Caudally) **What to do:** Move the probe towards the head or feet *of the mother* without changing the angle you are holding the probe at. **How to move:** Movement from the arm, keep the wrist fixed.	

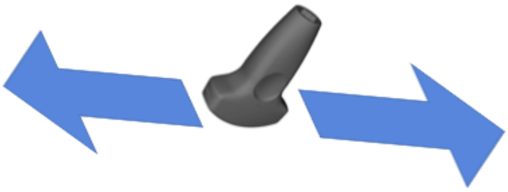
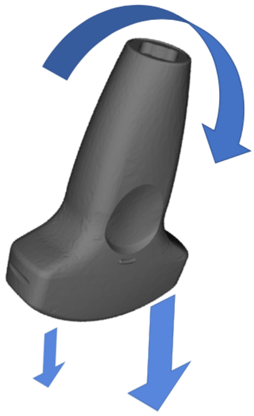
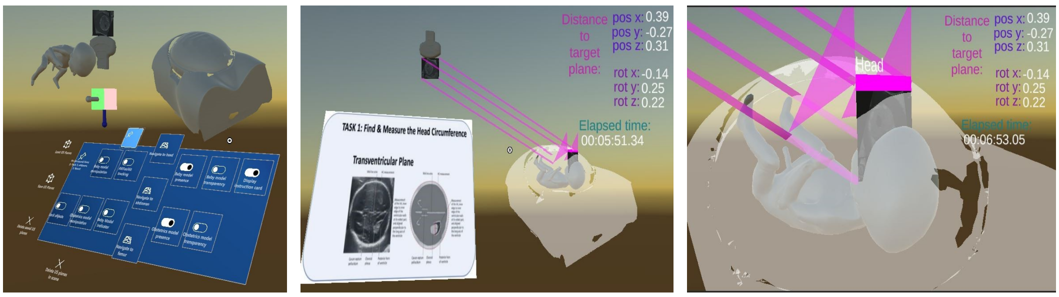
On-Screen Instruction: Slide Laterally **What to do:** Move the probe away from the midline *of the mother* without changing the angle you are holding the probe at. **How to move:** Movement from the arm, keep the wrist fixed.	
On-Screen Instruction: Rocking (Heel-Toe-Tilting) **What to do:** Without sliding the probe move the probe so that one of the short sides is lower than the other. You can rock clockwise (lower the left side of the probe) or anticlockwise (raise the left side of the probe. **How to move:** Movement from the wrist, keeping the arm fixed.	
On-Screen Instruction: Sweep (Fan) **What to do:** Without sliding, move the probe so that one long is lower than the other. You can sweep up, or down. **How to move:** Movement from the wrist, keeping the arm fixed.	
On-Screen Instruction: Rotate (clockwise or anti-clockwise)	

Table 6.6 – Standardised probe movement instructions provided to the trainees.

6.7 User Experience within the Mixed Reality Environment

The central design aspect of the creation of a mixed reality ultrasound training approach was an easy-to-use workflow that doesn't require advanced computer science knowledge. Therefore, all user interaction options are gathered on one holographic menu that shows all available options (without hidden sub menus) and follows the user's eye gaze but can also be pinned to remain at a fixed 3D position. *Figure 6.9* shows the basic Unity components as displayed in the Unity game view. A close-up of the user panel is shown in *Figure 6.10.* After 3D reconstructive post-processing of CT scans, 3D objects of a baby, obstetric phantom and a Voluson Ultrasound probe were imported into the Unity scene as .obj files. The objects were scaled to match the size of their physical counterparts. The US probe has an associated plane that represents the US beam produced by the probe. A holographic cube with coordinate axes is rigidly attached to the probe model in a fixed distance, which is used to facilitate a user's visual confirmation of virtual to real world alignment when tracking of the probe is enabled.



(a) (b) (c)

Figure 6.9 Unity scene showing the various components of the holographic setup: (a) baby and obstetrics phantom, Voluson probe with attached ultrasound plane and reference cube with coordinate axes, menu. (b) Navigation to target ultrasound plane incl. navigation instruction card, navigation arrows, position and rotation offset between probe and target plane and elapsed time for the navigation task. (c) Closeup of the pined head ultrasound plane.

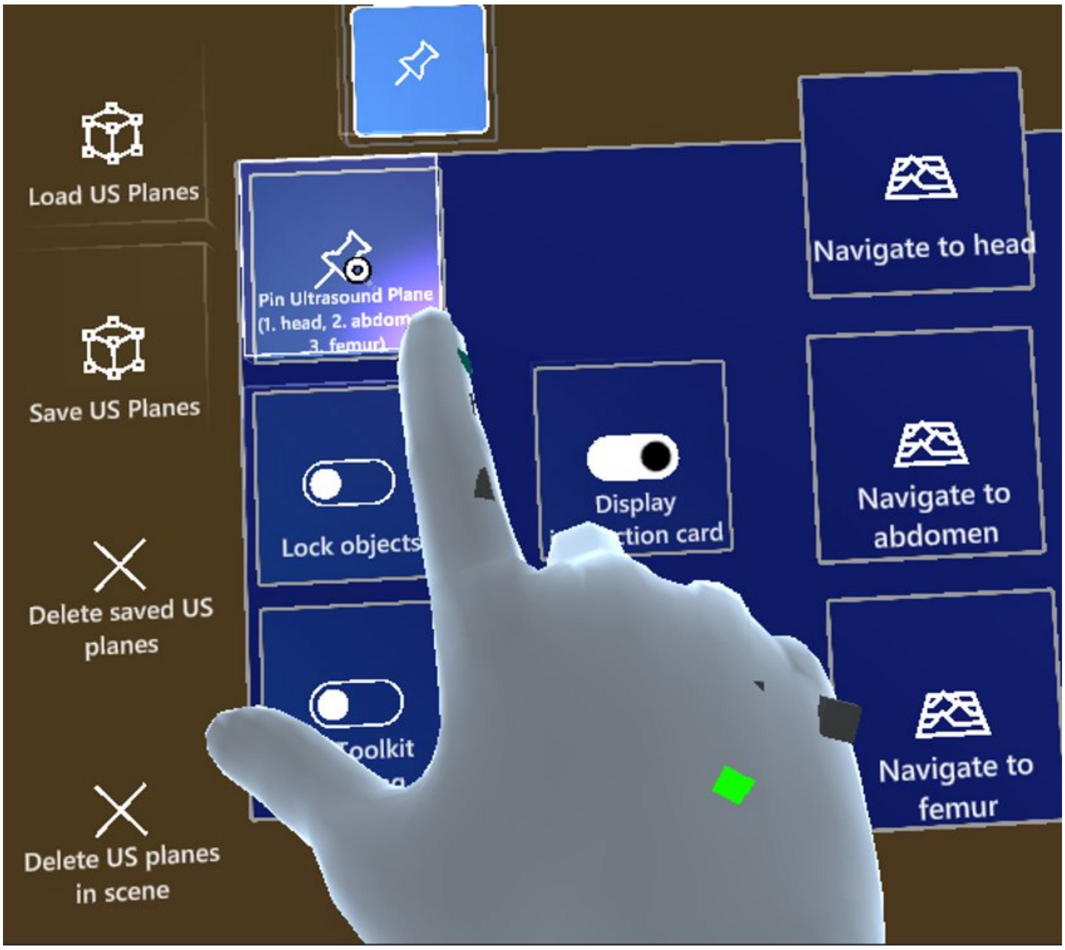


Figure 6.10 – A close up of the user interface panel. CAL-Tutor functions can be controlled using this menu. The trainee would primarily use the buttons "Navigate to head", "Navigate to Abdomen" and "Navigate to Femur".

The menu has multiple buttons allowing an ultrasound trainer or technician to do to the following actions:

- Turn on/off object manipulation of the baby and obstetrics phantom model (translation and rotation). This is needed to ensure correct alignment of the models in the real and virtual environment. This is shown as the "Baby Model Maipulation" function in *Figure 6.11*.
- Activate/deactivate the baby or phantom model presence. In case the presence of these models is perceived as an obstruction by the trainee. This is the function "Obstetrics model presence" in *Figure 6.11.*

- Pin target ultrasound planes. Currently, only 3 target planes are supported: 1. head (BPD), 2. abdomen (AC), 3. femur (FL). These are shown in *Figure 6.10.*
- Enable/disable baby and phantom model transparency. Useful if the user wants to see the physical phantom counterparts underneath the holographic overlay. Currently transparency doesn't work as expected on the HoloLens and is therefore not being used for the presented system.

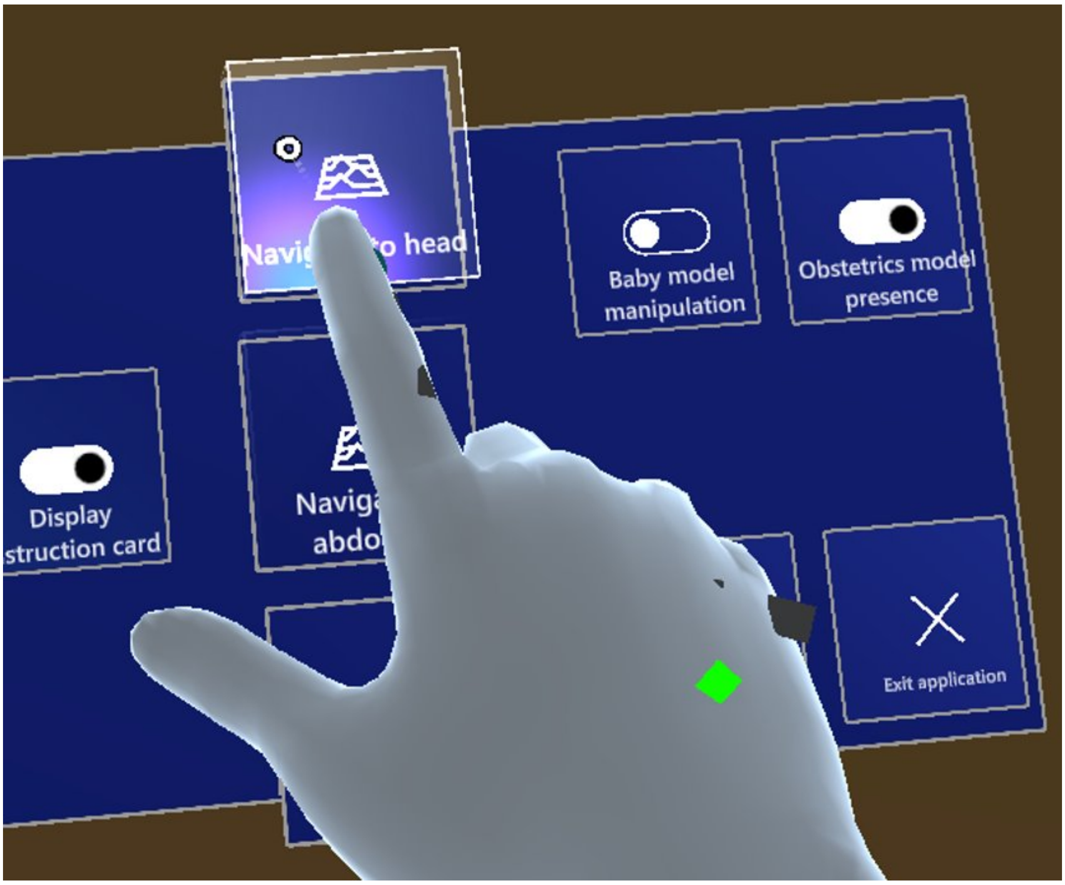


Figure 6.11 – A close up of the user interface panel. CAL-Tutor functions can be controlled using this menu.

- Turn on/off a directional indicator pointing to the baby model. This can be useful for trainers or technicians if the baby model is out of sight prior to correct alignment with the phantom. This is shown as "Display Instruction Card" in *Figure 6.10*

- Fix the relative position of the baby model and the pinned ultrasound planes. This is useful when the virtual overlay consisting of correctly placed baby model and target ultrasound planes becomes misaligned and needs to be re-aligned.
- Load and save the pinned ultrasound planes. In case the pinned planes are not at the desired location, they can also be deleted. Load US Plane in *Figure 6.10.*
- Deletion of the already saved ultrasound planes. This allows the saved planes to be replaced by new ones when the position of the fetus is changed, eg. from cephalic to breech. This is shown as "Delete US Plane" in *Figure 6.10.*

The holographic menu also offers a trainee interaction options during training:

- Start navigation to the three target ultrasound planes head, abdomen and femur. When navigation starts a new toggle button appears on top of the menu that allows the user to confirm when the target plane has been reached. This is shown as "Reached first plane" in *Figure 6.12.*

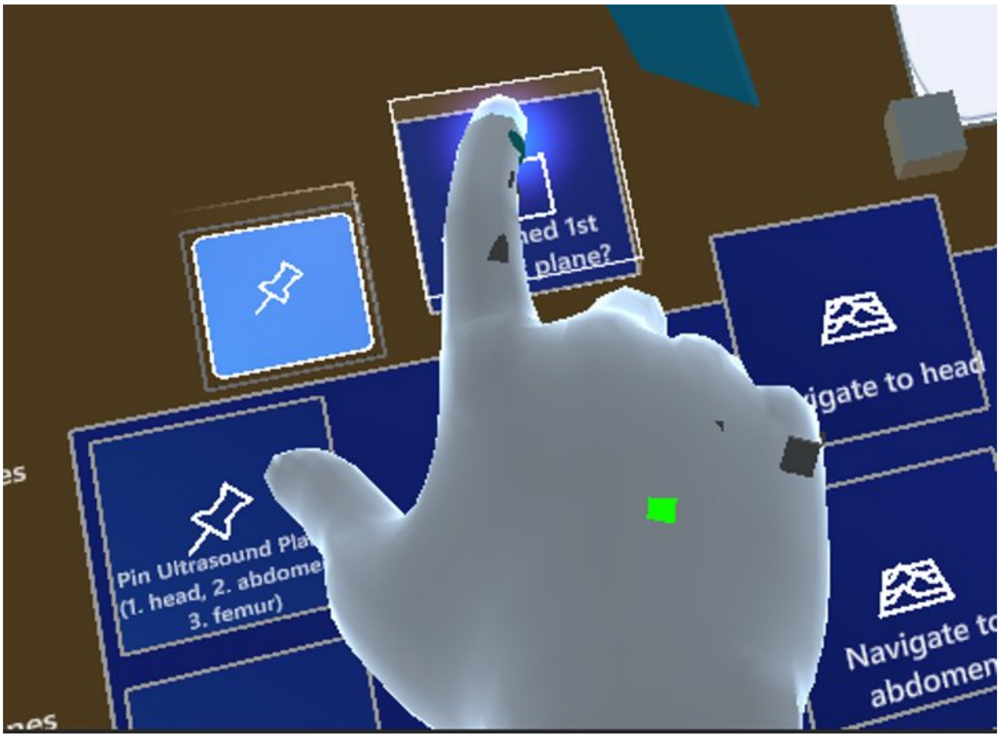


Figure 6.12 - – A close up of the user interface panel. CAL-Tutor functions can be controlled using this menu

-

- Enable/disable display of navigation instruction cards that appear above the menu and can be placed to another desired location and resized.

Figure 6.9 (a) shows the basic components of the Unity scene including holographic models of the baby, the obstetric phantom and the Voluson ultrasound probe. *Figure 6.9 (b)* shows navigation to the target ultrasound plane as would be seen by the user through the HMD. In this implementation guidance arrows connect all four edges of the probe to the target plane. Position and rotation offset data help the user to navigate quickly to the target plane. The hypothesis is that guidance information increases the user's spatial awareness of hand and arm motion in relation to the physical baby location. The elapsed time is an additional piece of information intended to alert the user to the time each navigation takes. The user can also display an instruction card (lower left corner of *Figure 6.9 (b)*) that contains relevant guidance information. In *Figure 6.9 (c)* a closeup of the pinned head ultrasound plane including navigation information is shown.

6.7.1 Unity Implementation Details

Game objects are the basic components of a Unity scene and represent 3D objects. Objects in a unity scene can also be used as a container for object interaction logic. Not all unity scene objects have a corresponding real world object. This is the case with the "MixedRealityToolkit" object. It is part of the MRTK and acts as a container for several mixed reality related settings allowing the Unity scene designer to configure mixed reality user input settings.

Other game objects in the scene represent actual 3D objects such as the baby the obstetrics phantom and the Voluson ultrasound probe which have all been scaled by the factor 0.001 in order to match the size of their physical counterparts. The baby and the obstetrics models can be moved and rotated to the desired location via user hand interaction. To this end both game objects have the components ObjectManipulator and BoundsControl attached that enables hand manipulation by the user when using the Hololens. fObject scaling is also possible but has been deactivated in order to maintain a correct scaling of 0.001. This preserves alignment of the unity scene objects with their real world counterparts when the end user is experiencing the scene using the Hololens HMD.

6.7.2 Probe Tracking

To to create a useful learning environment in which the user learns to place the ultrasound probe at a specified target location in an effective manner tracking of the probe is essential. To this end a marker based tracking software was used: the HoloLensARToolKit, a marker-based tracking approach that uses the HoloLens' front-facing camera. A cardboard cube marker was rigidly attached to the Voluson probe using a wooden stick, as shown in *Figure 6.8 (a) & (c)*. Due to the positioning of the cube at a fixed distance from the probe a permanent line of sight between the HMD front-facing camera and the cube is ensured which enables smooth uninterrupted tracking, throughout the expected range of motion and a probe grip styles.

6.7.3 Target Ultrasound Plane Navigation

To establish the position of the standard planes the trainer must scan the SPACE-Fan phantom prior to the training session, using the tracked Voluson probe. The trainer must 'pin' the location of each of the three target ultrasound planes. As discussed earlier, these depict the key locations of the baby's anatomy: the transventricular plane in the fetal head, a plane across the fetal abdomen and a plane in the fetal femur. This is seen in *Figure 6.10,* as the trainee would view them through Hololens2. The trainer presses the "Pin Ultrasound Plane" toggle button on the holographic menu, *Figure 6.11,* which creates a clone of the probe's ultrasound plane with the current image of the plane's video stream. A virtual line is also added to the pinned plane. This highlights the correct orientation of the probe with the target plane. Each plane is labelled "head", “abdomen” or “femur”. The trainee navigates to each of the three target planes in turn. Individual target plane navigation is activated by the trainee using toggle buttons labelled “Navigate to head”, “Navigate to abdomen” and “Navigate to femur”. Once selected, holographic guidance aids are displayed:

1.) guidance arrows connecting the edges of the probe and the target plane,

2.) position and rotation offset x, y, z values between the probe and the target plane and

3.) an optional plane with an instruction card explaining how to find the respective target plane.

In addition, the elapsed time is also recorded and displayed to the trainee. As soon as the trainee has correctly aligned the tracked probe to the target plane he/she can press a toggle button that confirms that the target has been reached.

6.7.4 Navigation to target ultrasound planes

Since the clinical trainer must be able to place target ultrasound planes at the correct anatomical locations of the holographic baby model, a Unity concept had to be developed that allows users to mark the location of the standard planes. This was done by allowing the trainer to create copies of the ultrasound plane that is attached to the probe model. Given that the probe's ultrasound plane displays the live ultrasound stream, a snapshot of the ultrasound video serves as the texture of the copied plane. To this end a Unity C# script named “CloneGameObjects.cs'' was written that combines the following steps:

- Creation of a new 2D plane with the same 3D location, rotation and scale of the probe plane

- Setup of event handlers for the probe plane's video player components that handle frame ready events: as soon as the next video frame becomes available a frame extraction event is triggered.
- Frame extraction coroutine: the current snapshot of the video stream is extracted from the video and fed into the new 2D texture of the newly created 2D plane.
- Creation of a navigation pink line that marks the probe sided edge of the copied ultrasound plane. This is demonstrated in Figure 6-10.

Figure 6.13 shows a Unity scene in which the three target planes "Head", "Abdomen" and "Femur" have been placed at the anatomical locations of the baby model. Each target plane is labelled with its target anatomy name and has the pink line at the probe sided edge of the plane. The uppermost part of the target plane has been highlighted in pink. The intention is to help the trainee to identify the correct orientation of the ultrasound plane. As can be seen from *Figure 6.13* the planes are neither parallel with each other nor in the same probe orientation. A highlight was added along one edge of the standard plane marker. This was done to avoid the trainee approaching the plane from an angle 90 degrees from where they should be. This would result in the image being rotated 90 degrees from the ideal orientation. Further, they might find that there is significant shadowing in the abdominal plane from the fetal spine or limbs. To further avoid confusion, the planes have been titled with names appropriate to their anatomical location, such as “head”, “abdomen” and “femur”. This acts as a further learning and visual cue for the learner. More importantly it allows the trainer to confirm with the trainee which target plane they are attempting to achieve. This is important from a training perspective as the trainer does not have a way to see what the trainee is visualising once they have put the Hololens2 on and started the training session. Hololens does have functionality which allows trainers to see what the Hololens2 wearer is seeing on a connected laptop. This iteration of CAL-Tutor does not include this functionality. We removed this functionality from CAL-Tutor due to the impact on frame rate and lag in instructions. Although the ability to see what the Hololens wearer was useful, the performance of the system degraded to the point where trainees would not have been able to use the system. Microsoft advertise an alternative functionality called "Spectator View" which offers similar functionality. Implementing this functionality requires additional coding cameras and hardware rigs which were not available during the study period.

Figure 6.13 - User view of the clinical training environment when they are wearing Hololens with CAL-Tutor software running. To the left is the clinical US system manufactured by GE Voluson and to the right is the training phantom with the fetus overlaid and the standard planes highlighted.

After the trainer has placed the three target planes trainees can navigate to these planes by pressing one of the navigation buttons on the holographic menu. If a user is supposed to navigate to the "Head" ultrasound plane, he/she has to press the ``Navigate to head'' button. *Figure 6.11* illustrates the navigation process: Four guide arrows, which start from the four corners of the plane of the ultrasound probe and point to their respective counterparts on the target ultrasound plane, help the trainee to navigate to the target plane. In addition, the distance to the target plane is visualized via position and rotation offset x, y, z values. As soon as the navigation button is pressed, the elapsed time is also displayed and serves as an additional aid for the trainee to develop a perception of the time it takes to reach a target plane. *Figure*

6.9(b) shows the optional visualization of an instruction card with written instructions that help the user to identify the target planes. The holographic instruction card replaces the same instruction cards that would otherwise be displayed on a separate physical screen and offers the advantage that the user can place them anywhere and scale them according to his needs.

In *Figure 6.12,* the navigation completion toggle button is shown, it is labelled as "Reached 1st Plane" . Trainees can complete each navigation sequence by pressing this button and continue with the next target plane navigation. After the button is pressed, user related navigation data is stored and guidance information (arrows, numbers, instruction card) disappears.

6.7.5 Saving, Reloading and Deletion of Target Ultrasound Planes

Since a trainer wants to be able to save the placed target ultrasound planes and reload them during another HoloLens session, two holographic buttons named "Save US Planes" and "Load US Planes" have been implemented (*Figure 6.14*, highlighted in red). The save button click event triggers extraction of the planes 3D position, orientation, scale and text label. This information is saved in a xml file. The ultrasound image is stored separately. The load event then creates a new plane game object and uses the saved image as a texture and pulls the rest of the information from the XML file.

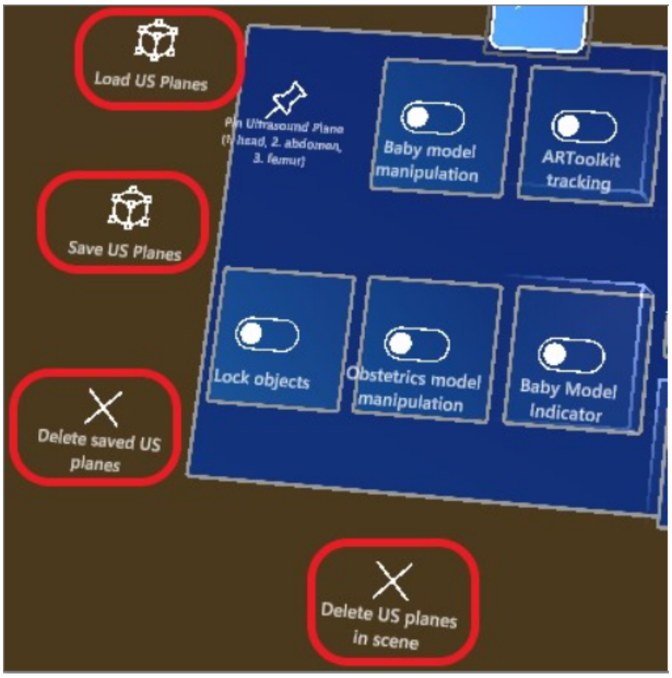


Figure 6.14 A close-up of the user menu. These functions allow the trainer to save the relative locations of the marked ultrasound planes and to reload them, as desired. It is also possible to delete the planes, if this is desired.

The user has also the option to delete all saved ultrasound planes from memory using the "Delete saved US planes" button, which is very useful for the trainer if he/she has placed one or more target ultrasound planes at an incorrect location and would like to delete these planes first before placing new ones.

We found that the position of objects when quitting and restarting the application were not always consistent. The Hololens2 has a reported issued with retaining spatial awareness and maintaining 3D objects at the same location. To build a 3-D environment the Hololens scans the room, unless the position of the Hololens2 is exactly replicated the scan generated by the Hololens2 is different on each occasion. Given the differences in height and possible head positions for each user an error is inevitably generated. Objects do not align correctly in the real and virtual environments. Therefore, the trainer has to be able to adjust the position of the baby model while keeping the relative position of the already placed target ultrasound planes intact. To this end a button named "Lock objects" has been added that makes all placed target ultrasound planes children of the parent baby model game object. This way, when moving the baby model to another location, all ultrasound planes keep their relative position.

6.7.6 User Data Recording

As soon as the user presses the “start navigation" toggle button the following user data is recorded:

- Time stamp
- The position and rotation x,y,z of the Voluson probe
- The eye gaze hit position (x,y, z) with respect to the target ultrasound plane

If the user presses the target confirmation toggle button data recording stops and a .csv file is created. The .csv file is stored on the HMD file system and can be downloaded via the HMD device portal. As the data was not recorded for all users it is not reported in this study. We intend to use this data to gain a better understanding of the user's motion profile and attention to the target plane. The parameters recorded should allow for comparison metrics such as probe path length, time to task completion and DSJ to be accurately calculated without the need for additional tracking equipment.

The idea behind collecting user data is to be able to carry out further analysis in order to understand the behaviour of the trainees. A potential application could be the classification of the trainees level of expertise based on this hands motion and eye gaze patterns. Further research is needed to evaluate this theory but it is anticipated that metrics described in the earlier chapters of this thesis could be used to evaluate trainee performance in future studies, using data captured directly by Hololens.

6.7.7 Optional Transparency of baby and obstetrics phantom models

We implemented optional alpha blending to provide transparent views, but this does not appear to work in the HoloLens 2 at present. Work remains ongoing.

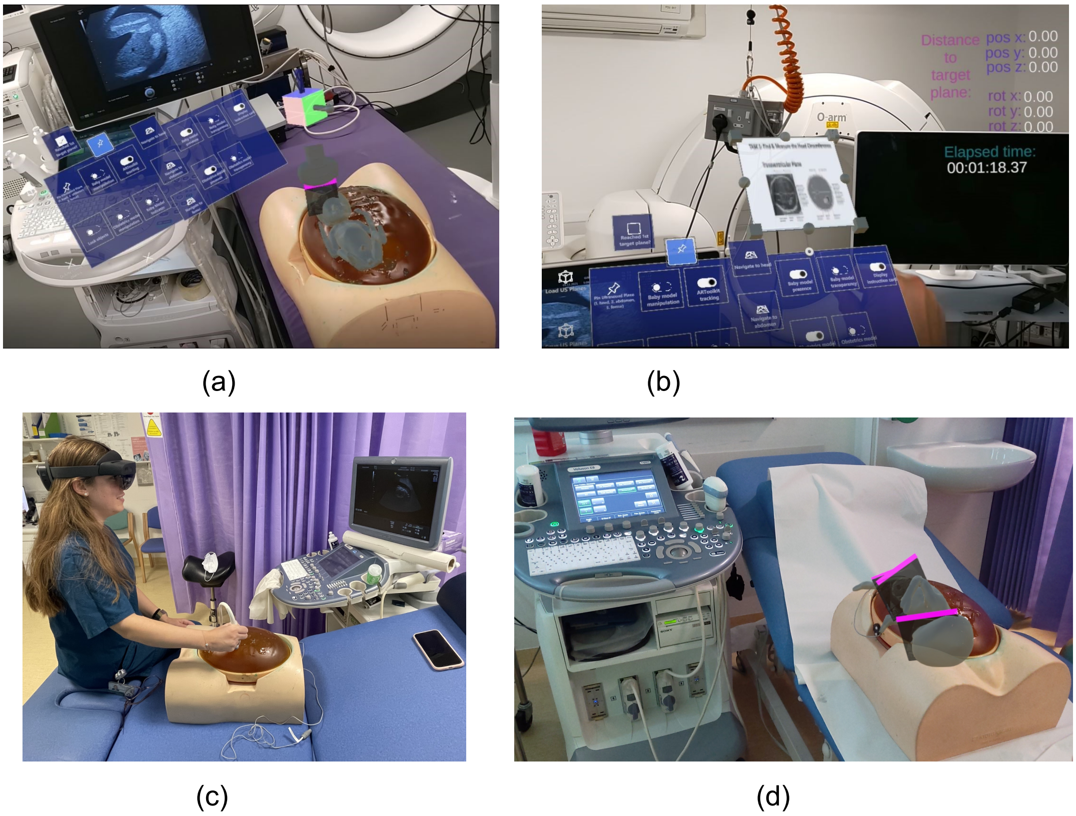


Figure 6.15 Views of the CAL-Tutor in use, as seen by the Hololens user: (a) An example target ultrasound plane has been set through the fetal head. The user control panel is also visible (b) An optional instruction card with guidance information and visual reference for the ideal plane. (c) CAL-Tutor in use in clinical environment. (d) View of the clinical set-up including views of the target planes, from the perspective of the HMD user.

6.8 Study Design for the validation of CAL-Tutor

In the previous section, 6.4, I detail the limitations in the literature in assessing the validity of mixed reality systems. To address this, I applied previously used tools for the assessment of validity to the initial scoping data and CAL-Tutor hardware. The development of the hardware and software have been detailed in the previous section. Following from this, a study assessing the simulator's realism and content based on user feedback was designed. That is detailed here. The study was conducted between September 2021 and October 2021. The planned study would be a non-randomised interventional cohort study. The primary outcome will be clinical performance and confidence having used the "CAL-Tutor" programme for training in obstetric ultrasound in comparison to conventional clinical training.

I designed the study to compare the "CAL-Tutor" programme to the current training of Ultrasound Fellows. The observation period was planned to be the initial 5 weeks from induction into the ultrasound department. Ten newly recruited ultrasound fellows will be invited to participate in the study. A similar number of ST1/2 trainees will also be recruited, these Doctors will not be working in the ultrasound department on a daily basis. Each participant will be assigned an anonymised user identifier. This identifier will be placed on the ultrasound images by the trainee, using the annotation tool built into the GE Voluson system. There will be two cohorts, one taken from core trainees in O&G, the other the ultrasound fellows, as described. Both groups will undertake a baseline growth scan, the captured images will be stored for scoring and the scan will be saved as a video file. Images will be scored and measurements compared with "expert" scanners on the same baby.

The baseline and exit scans will be undertaken with clinical equipment currently in use at the Ultrasound Screening Unit at University College London Hospital. Supervising staff are experienced in obstetric ultrasound, themselves undergo regular quality control assessments, are familiar with the department and the departmental protocols in obstetric scanning. The ultrasound scanner itself will not be modified. Any unexpected findings will be escalated by the Senior Ultrasound Fellow, as per normal departmental protocol. The standard of care will not be affected.

After the baseline scan has been completed the group using the "CAL-Tutor" system will undergo 4 training sessions, these are not anticipated to last more than one hour. The fetus will be in a different position each week, to allow the trainee to practice obtaining views of the fetus while it is in one of four positions. The possible positions are demonstrated graphically in *Table 6.4*.

In this study all participants will be given relevant excerpts from the NHS Fetal Anomaly Screening Programme (FASP) handbook, as shown in *Table 6.5*. This handbook details the required planes for accurate and reproducible fetal biometry.

Both groups will have access to the excerpts throughout. There are also posters depicting these planes in each scan room at UCLH. A flow-chart of the study is included as *Figure 6.16.*

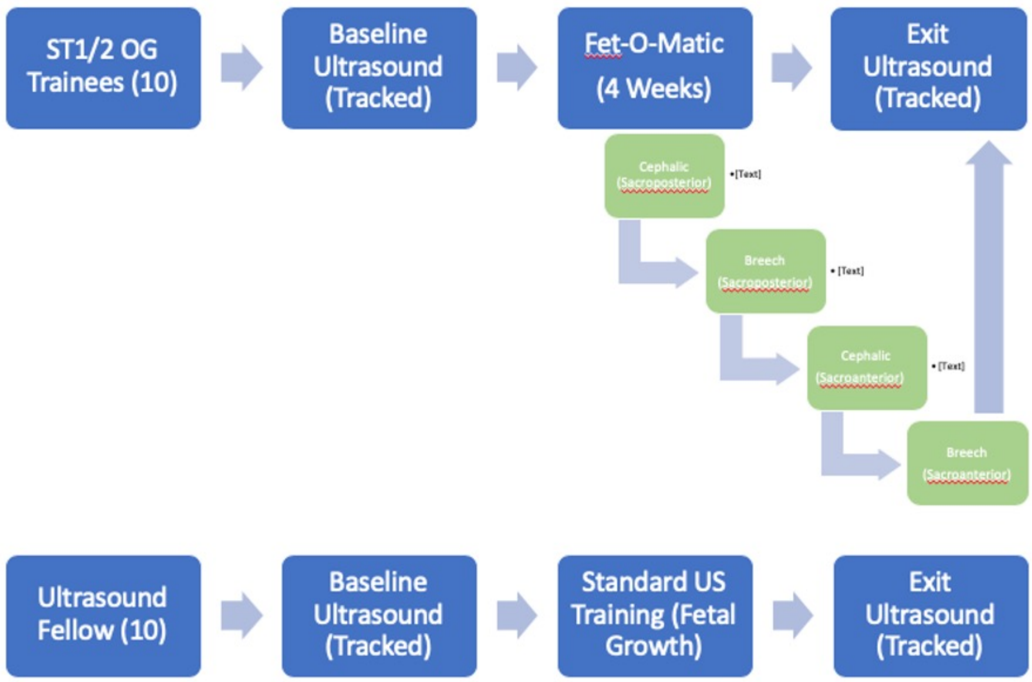


Figure 6.16 – Graphical representation of the "CAL-Tutor" study representing both the conventional training pathway taken by the Ultrasound Fellows and the proposed "CAL-Tutor" intervention.

6.9 CAL-Tutor learning website

From the initial scoping questionnaire and the validity assessments carried out it became clear that a single, standalone piece of hardware would be unlikely to solve the issues with ultrasound training. A comprehensive suite of learning materials would be required in addition to the hardware development. It needed to be easy to use and specifically aimed at the UK training without significant prior knowledge. To assist the delivery of learning material a portal was created using a software platform designed for the delivery of educational material over the internet. A pre-existing learning platform was attractive as no previous experience of building websites was required. Extensive coding experience was not required either. The Thinkific.com platform is a website designed to allow for the development, marketing and enrolling of online courses. I had previously used this platform as part of faculty delivering teaching remotely in Tanzania. I found the platform to be reliable, easy to navigate and the students enrolled on the course were able to access the website easily. Thinkific.com allows for each trainee to have their own account. Once registered each trainee can contact and be contacted using a private messenger function. Alternatively wider messages can be ‘broadcast’ to the entire learning group. Trainers can upload training material which is not limited to presentation slide decks. Trainers can upload videos and quizzes and links to external multimedia such as YouTube videos. Live tutorials can be hosted on platforms such as Zoom. Once tutorials have been delivered, they can be uploaded to Thinkific if they have been recorded by the tutor. The ability to upload material, videos and tutorials allows the trainees to consume material at their own pace. This is particularly attractive for this iteration of CAL-Tutor as none of the trainees enrolled have protected training time to participate. Shift-based working patterns make it difficult to mandate attendance at a specific time. However, progress and engagement can be monitored by the faculty by making completion of certain quizzes, tutorials or reading materials mandatory. Trainees can be contacted directly if there are issues with attendance or persistently low scores in the quizzes.

6.9.1 CAL-Tutor online course content

Following best practice in medical education, the content of the online course was constructed to be iterative[91]. Like a conventional textbook the course was arranged into chapters. Many of the topics introduced in the initial lecture are revisited in subsequent chapters. The course was structured based on the recommendations in the medical education distance learning guide[92]. Dependence on lecture-slides was minimal. These are relatively low in instructional value as viewing slides without access to the spoken dimension of a presentation often makes little sense. References to textbooks were also kept to as few texts as possible, which have been recently updated. Rather reference was made to best practice papers recently published and the NHS FASP screening materials which form the basis of care in the UK. These materials are updated periodically. I carefully curated links from external websites, for real-world scanning examples. Suggesting general search engine terms is best avoided, as search algorithms will typically identify the most viewed or linked-to content rather than the best quality. Search results can vary by the user's location, IP address, browsing history and other variables.

The learning chapters followed the structure below.

6.9.1.1 Introduction

I designed introductory pages to introduce the trainees to the CAL-Tutor learning platform, the CAL-Tutor hardware, the trainee learning outcomes and the assessments that they will be expected to complete. The trainees are invited to explore the chapters available to them and undertake an initial baseline knowledge quiz.

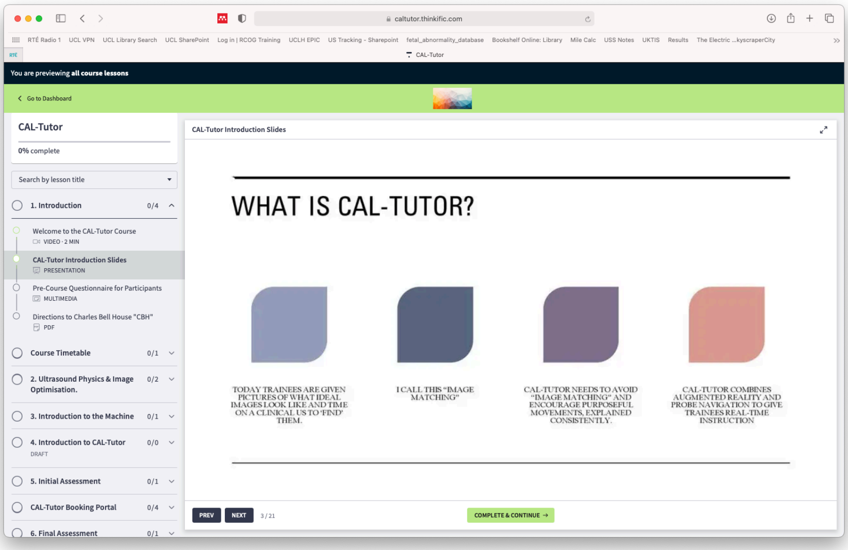


Figure 6-17 – Landing page for trainees when they log-onto the CAL-Tutor learning portal.

a. Welcome Video

b. Introduction Slides

These slides cover the purpose and learning outcomes trainees can expect from the course.

c. Pre-Course Questionnaire

6.9.1.2 Course Timetable

Here the trainees can see scheduled sessions for the entirety of the course. This is updated frequently and there are hyperlinks to live tutorials or recorded versions of the live tutorial once it has been delivered.

6.9.1.3 Initial OSAT

The assessment is based on the RCOG Intermediate US Assessment of the fetus OSAT. This is completed in the ultrasound department as part of the trainees education. Completion of the *basic ultrasound assessment of fetal size, liquor and the placenta* is required for all trainees to progress to intermediate training.

6.9.1.4 Ultrasound Physics & Image Optimisation

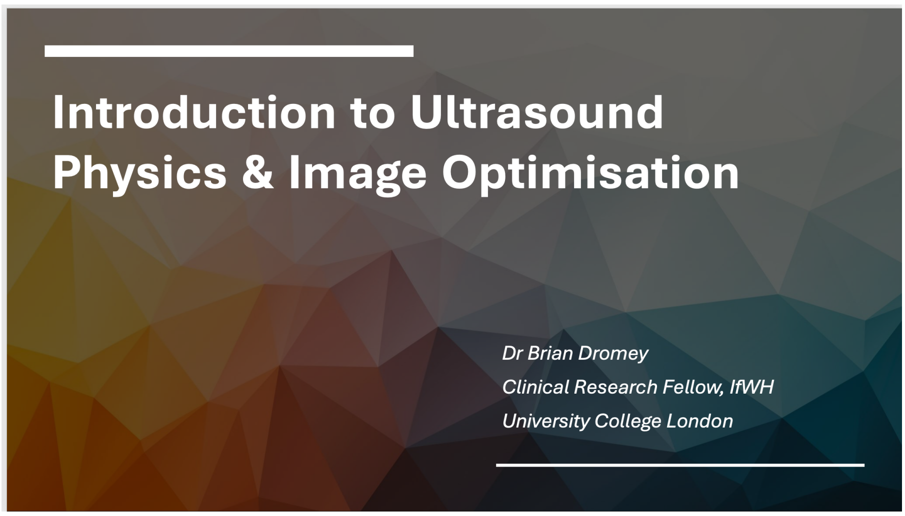


Figure 6.18 – Introductory lecture slide example.

d. This takes the form of a tutorial/lecture which is delivered by video conferencing. In this case a link to a Zoom meeting was embedded.

The meeting was recorded and uploaded for trainees who could not attend.

Topics covered included:

- Uses for Ultrasound in Obstetrics
- Brief History of US in Obstetrics
- Introduction to terms commonly used in US
- US Buttons & Controls
- Strategies to optimize images

6.9.1.5 Introduction to the GE Voluson

This chapter takes the form of a pre-recorded tutorial and introduces the trainees to the functions available on the Voluson machines. The tutorial is pre-recorded because of the challenges associated with multiple cameras and points of view required to clearly demonstrate the controls of the machine, the US probe orientation and on-screen US images. A still of the video is shown in *Figure 6.19.*

Figure 6.2 – Screenshot of the pre-recorded tutorial showing picture-in picture of the US examination overlaid on video of the machine controls.

6.9.1.6 Live Scan Tutorial

In this section the trainees are asked to attend at the given time. The trainees are shown how to assess for fetal position, placental location, abdominal circumference, head circumference and femur length. This is completed as a live, interactive session. It was initially planned that this session would happen in-person, but restrictions associated with the COVID-19 pandemic required teaching to be delivered on-line. Trainees were therefore not able to experience the "hands-on" aspect of the tutorial, as initially envisaged.

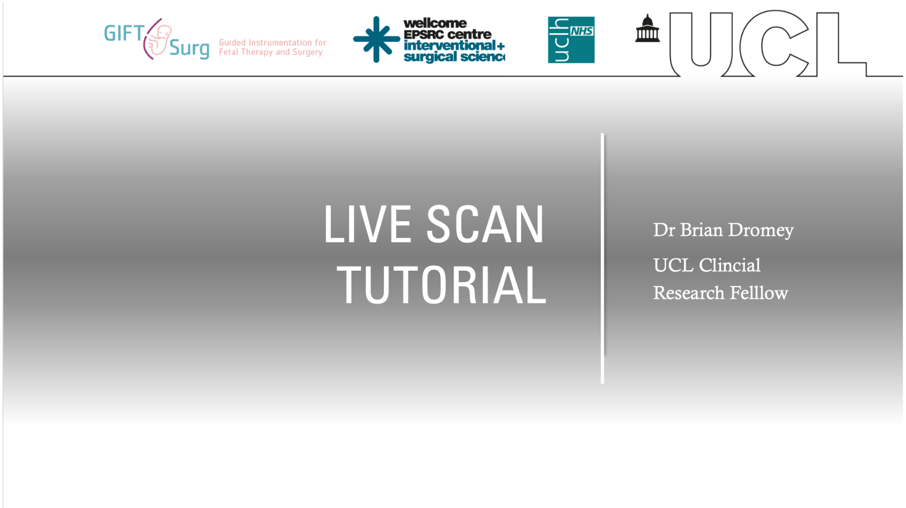


Figure 6.20 – Live Scanning Videos

6.9.1.7 CAL-Tutor Booking Portal

This is a live Outlook calendar that allows trainees to schedule 30 to 60-minute slots with the CAL-Tutor platform. The intention is to provide trainees with certainty that CAL-Tutor will be available for them when they arrive at the training centre. It is updated in real-time when there is a trainee cancellation or booking.

6.9.1.8 Final OSCE

The assessment is based on the RCOG Intermediate US Assessment of the fetus OSAT, as detailed earlier.

6.9.1.9 Course Feedback

This is a Microsoft office form which gives the trainees opportunity to reflect on their initial training goals and whether they have been met. Additionally, trainees are invited to report areas of the course which they were satisfied with and areas which have room for improvement.

6.10 Use of CAL-Tutor simulator

Clinicians working in Obstetrics at a single hospital site were asked to express their interest in learning how to perform fetal US, or improve the skills they had already acquired. They were informed of the research by an e-mail sent by one of the team (BD) and a brief presentation given at a departmental meeting. After enrolment trainees were asked to perform a 1 hour session using the HMD device and SPACE-Fan trainer. The fetus was in a cephalic presentation. Three planes associated with the estimation of fetal weight were already marked in the HMD environment as seen in Figure 6.1(d).

Trainees were asked to:

1. Identify the lie of the baby (Cephalic)
2. Fetal Biometry (AC, HC, FL),
3. Obtain a four-chamber view of the heart.
4. Locate the leading edge of the placenta.

The participants completed the exercise according to detailed in *Table 6.7* which are the standardised instructions given to trainees and procedure for training. This document was created to ensure that trainees were given consistent information and the tasks were reproduced by all trainees.

The trainees will be will be required to commence the scan at the point representing the pubic bone of the mother, with the probe in a longitudinal position. The training session will not commence until the system has detected that the probe is in the correct position.
To familiarise the user with the on-screen instructions, they will be instructed to move the probe according to instructions given by the on-screen guidance system. The instructions will be generated at random. They will not be scored.
The user will then be asked to find the head and to then perform the scan in a systematic way. The user will move from the head, to the face, the heart, abdomen, abdominal wall, bladder and femur. The trainee will be required to perform head, abdominal and femur length measurements as the scan proceeds.
While it will be possible to skip a task, the trainee will not be able to come back to this later. The aim of “CAL-Tutor” is to decrease operator variability and to equip operators with the skills necessary to obtain diagnostic images and accurate measurements by demonstrating the relevance of one view to another. This strategy would be undermined by allowing the trainee to skip between tasks.
A SPACE-FAN obstetric ultrasound trainer will be scanned. The trainee will use a GE Voluson Ultrasound system.
This simulator has been previously validated as a training tool. The fetus can be in one of four positions. Cephalic, Breech, Sacro Posterior, or Sacro-Anterior.
A 3-D representation of the fetus will be displayed on a secondary monitor, the position of the fetus will be selected to reflect the position of the baby in the SPCE-FAN model.
A 3-D representation of the ultrasound probe will also be displayed on-screen. This will move dynamically based on the motion of the probe.
Real-time feed of the ultrasound image will be shown on the screen using the video output port of the ultrasound system.
The user will be able to see how the motion of the ultrasound beam ‘cuts’ through the fetus and how that ‘cut’ is effected by their movement of the probe.
Dynamic Guidance arrows which change in size and colour will guide the trainee to the ideal plane.
They will not be given any feedback or scores at the end of each session.
Feedback will be collated at the end of the training period.

Table 6.7 - Standardised navigation instructions given to the trainee.

Participant demographics, clinical experience and previous ultrasound exposure were collected at baseline, using an online survey questionnaire hosted on the university Office 365 partition (forms.office.com). Upon completion of the study, feedback questionnaires were completed before the trainee left the scan room. The aim of the questionnaire was to allow participants to give their feedback and assess the usability of the pro- posed visualisation platform. Questions about the system were completed using Likert scales and free-text fields. Participants were asked to grade their agreement with several statements on a scale from 1 to 5 (1 representing Strongly Disagree, 5 Strongly Agree).

6.11 Results

I wanted to ensure that that a system was developed which would address trainees concerns about their training and adequately tackle the research objective of investigating trainees understanding of spatial orientation and machine functionality in the early part of their training. I was also mindful of the validity studies that I had undertaken earlier in the development process. I undertook a scoping questionnaire to understand the attitudes of trainees towards training in obstetric ultrasound. I wanted to understand the skills they had already attained during their training and how trainees would like to further develop their ultrasound skills. Both bedside teaching and dedicated skills sessions have strength. Bedside teaching offers training in clinical context, but the ad hoc nature makes reliable attainment of curriculum skills challenging. The ad hoc nature also means that dedicated and protected teaching time is difficult to achieve. Conversely dedicated skills sessions can be structured to allow time to be used efficiently. However, staffing shortages often mean that trainees find it difficult to attend some or all of the sessions with a resultant impact on their learning.

6.11.1 CAL-Tutor Trainee Demographics Questionnaire

Seventeen sonographers agreed to take part in the evaluation of CAL-Tutor. The enrolled clinicians averaged 4 years of clinical experience, as detailed in *Table*

6.8. Eleven of the 17 responses indicated prior experience with Voluson ultrasound machines. The group rated their confidence using this machine at 3.4 out of 5, where 5 represented "Very Confident". The majority of trainees self-reported as being only somewhat confident. 50% of trainees reported that they had completed over 100 ultrasound examinations of the fetus. This reflects the lack of formalised training in US techniques, image optimisation and machine-specific training. 57% had enrolled in an ultrasound course previously and 78% found the course useful.

6.11.2 Post-Usage Questionnaire for Clinicians

85% of respondents felt that CAL-Tutor helped them to understand the relationship between the US probe and fetal anatomy. While 78% felt that using the system would increase their confidence when performing fetal biometry. All participants agreed that CAL-Tutor represented a useful training tool and that mixed reality training was a viable training tool in clinical practice. The free text responses are detailed in *Table 6.9* and the Likert scale responses are shown in *Figure 6.21..*

What is your Training Grade (ST1/ST2)	Number of years of clinical experience (all specialities):	Do you wear glasses?	Have you used the GE Voluson before?	How confident do you feel using this machine? (1 = not at all, 5 = very)
Clinical fellow	7.00	No	Yes	4
St5	10.00	Yes	Yes	2
ST4	6.00	No	Yes	2
ST4	7.00	No	Yes	2
Sho	4.00	Yes	Yes	4
3rd year medical student	0.00	Yes	No	4
St5	5.50	Yes	Yes	4
Senior sonographer	9.00	No	Yes	4
ST4 OG	6.00	No	Yes	5
Medical student	0.00	No	No	1
Fellow	9.00	Yes	Yes	5
St5	7.00	Yes	Yes	2
Clinical Fellow in Ultrasound	5.00	No	Yes	4
ST6	6.00	Yes	Yes	5
Trainee in obstetric ultrasound	12.00	Yes	No	3
ST5	7.00	Yes	Yes	2
ST4	5.00	Yes	Yes	4
	6.205882353			**3.352941176**

Table 6.8 – Results from the pre-usage questionnaire detailing each of the participants clinical experience.

How was your interaction with CAL-Tutor? Considering your ability to orient the probe, physical aspects of the Hololens and US probes or the user interface - could these be improved in any way?	**Did you have any general thoughts or questions about mixed reality devices in medical training having used CAL-Tutor?**
Not seeing your hand image will help seeing the lines better	It's an amazing way to teach. Very impressed
I couldn't see my probe through my fake hand	It's great
I couldn't see the planes very clearly when the ultrasound probe was in the way.	Very innovative idea.
I found the planes a bit difficult to see and would have found it useful to have a picture or reference for the plane that I was trying to achieve.	Great potential for training.
At first difficult but then got better. To become more friendly with the user	No
Increased my understanding of ultrasound scanning and fetal anatomy. Great tool for those at a very beginner stage with minimal experience (including myself). Learnt very quickly how to orient the probe.	Really great tool, would be invaluable to my education and training.
Really slick I like it	I love it, excellent device
I think it's fairly good	No

Corresponding images on ultrasound scanning machine	N/A
Very interesting and helps to orient myself, especially as someone with no experience	It’s definitely a very helpful training tool to increase confidence without having to use patients
Very good	No
Yes	It's a very good device for teaching practical procedures.
Good	Think every little bit helps
Very easy	No
Overall good	No
The instructions were clear and the baby was also very clear. I've never seen the probe move around the baby like that. Its very helpful.	Could this be used with a real patient?
I feel More confident finding views	Very useful tool for learning
Difficult to get used to the virtual panel control nitially	Useful for training especially with no experience with us before doing it with real patients

Table 6.9 – Results from the post usage questionnaire.

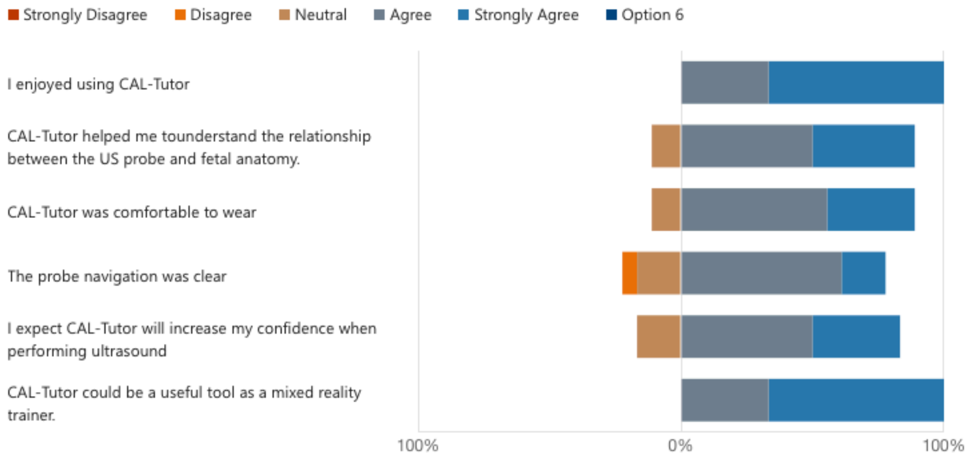


Figure 6-21 - Post-usage questionnaire results as completed by 18 clinical trainees. Each question was graded on a Likert scale where 0 represented “Strongly Disagree” and 5 represented “Strongly Agree”.

Once the prescribed tasks had been achieved the trainees were allowed free practice without direct instruction. They could remove aspects of the visual guidance if they wished. Two trainees (intermediate experience) removed the standard planes. One requested that the fetal model and the planes be removed. None of the novice trainees asked for guidance or other visual cues to be removed. It appears that more experienced trainees are less reliant on visual cues to obtain the desired standard plane. This supports our hypotheses that experts have developed a representation of the fetus in their mind and can recall this to navigate the US probe from the current location to a location where they will be able to obtain the desired view of the anatomy.

6.11.3 Post-Usage Questionnaire for Engineering Students

In parallel with the clinical study, undergraduate students in Computer Science were invited to a demonstration of the CAL-Tutor system as part of a WEISS centre open day. For students with an engineering background, 100% (n=6) responded that CAL-Tutor could be usefully deployed as a mixed reality trainer, see *Table 6.10*. In contrast to the clinicians, they spent more time reviewing the interface and technical novelty of the device while the clinicians feedback was more generalised and given in the context of a specific clinical application for training. The engineering students took more time to consider other applications and give a critical appraisal of the system hardware and software. They were also more forthcoming with alterative use cases for the MR headset and guidance, in addition to suggesting the inclusion of computer vision techniques to for automatic scoring of the ultrasound image, or for guidance in real-time clinical US on a fetus unknown to the system. Although much smaller in number the feedback from students with a technical background was more challenging than the feedback from clinicians, which tended to be brief and somewhat perfunctory. When designing future feedback surveys, we will consider the question structure carefully to elicit as much detail as possible from users without introducing bias

The HoloLens 2 could be a useful tool as a mixed reality trainer.	How was your interaction with the HoloLens 2? could these be improved in any way?	Did you have any general thoughts or questions about mixed reality devices in medical training having used the HoloLens?
Agree	Occasionally the probe was somewhat hard to grab but I think that might become less of an issue with practice	Seems like a very eloquent method to demonstrate tricky 3D issues
Agree	We did not calibrate it for my eyes and hence, I would be better equipped to comment if that was done.	They are definitely a boon to the industry.
Agree	I really liked the user interface (I think it was not 'filled' with too much much information which made the interface 'clear' in a way). The only thing I struggled a little bit with was the probe, there was a specific movement of hand (and fingers) that needed to be done in order for the user to be able to seize the probe so maybe, if possible, make the probe easier to hold. Nonetheless, once we are told how to put our hands to hold it, it works very well. in the mixed reality devices that help doctors train for a certain surgery.	
Agree	Maybe instructions can be shown near the the user virtually.	I think if it is operated with the real ultrasound probe will better.
Agree	I personally found the Hololens2 very comfortable to wear Yes, I was already interested in computer assisted surgery	
Strongly Agree	Good	Good
Strongly Agree	The UI is very intuitive and I was capable of performing basic functions within seconds. That said it still appears to have some issues and would not always recognize my hand movements correctly - this maybe due to the fact that I did not have time to calibrate the device.	I think the HoloLens are a great tool for conducting medical training where in areas where a good spacial awareness i necessary.

Table 6.10 - Free text feedback from post survey questionnaire completed by undergraduate engineering students.

6.12 Integrating automatic classification of fetal images & probe position estimation

6.12.1 Automatic Classification of fetal biometry.

CAL-Tutor system cannot currently classify the trainee-acquired images in real time from video feed. In this iteration of the system, it is necessary for the trainee to save the images to the image archive of the US machine. A member of the training team later feeds them to a stand-alone image classification module which has been developed alongside CAL-Tutor. This module is referred to as *AutoFB*: Automating Fetal Biometry Estimation from Standard Ultrasound Planes. The technical code behind AutoFB was developed by Dr Sophia Bano at WEISS. It was first presented at MICCAI 2021[93]. Briefly, AutoFB involves processing the three standard planes (Transventricular, transabdominal, and femur) to segment the head, abdomen and femur, and the extraction of the biparietal diameter (BPD), occipito-frontal diameter (OFD), head circumference (HC), transverse abdominal diameter (TAD), anterior-posterior abdominal diameter (APAD), abdominal circumference (AC), and femur length (FL). The classification is achieved by training a multi-class segmentation neural network that automatically identifies and segments the relevant anatomy structures within any of the three standard planes. The corresponding biometry is then extracted by applying scale recovery, using ellipse fitting (head or abdomen) and bounding box fitting (femur). In itself, AutoFB represents novelty, as it is the first framework to automate fetal biometry estimation from all three standard planes. The framework involves training state-of-the-art segmentation models for identifying the head, abdomen and femur anatomies. This is followed by shape fitting on the segmented regions, automated image scale retrieval and biometry estimation in millimetres units. Through retrospective scale recovery and shape fitting, we obtained the fetal biometry estimates. Comparison of the predicted versus clinically measured fetal biometry showed that the error in HC (2.67mm), AC (3.77mm) and FL (2.10mm) were minimal and was better than the ±15% error that is typically acceptable in fetus ultrasound assessment.

6.12.2 Automatic Estimation of the location of Standard Planes.

Currently, training and guidance in obstetric US are very focused on SPs recognition. I have already described an AR Tracking Toolkit methodology for tracking the US probe. This solution is attractive because it uses cameras already built-into Hololens2 and does not require further hardware, dongles or other paraphernalia which could impact on the usability, portability and stability of the system. I have previously described some technical limitations of Hololens2, primarily the computational resource available. Optical tracking is fundamentally limited by the necessity to have a clear line of sight, given the shape of the maternal abdomen at late gestations and the necessity for the user to observe the US screen, rather than the maternal abdomen. Decoupling of the tracking from Hololens hardware might be desirable in a larger study. Ciara Di Vece at WEISS has developed a regression CNN-based methodology to predict the 6D location of standard planes in the fetal brain[94]. The essence of this system is that a the standard planes can be identified without the need for real ground truth data in real-time or 3D volume scans of the patient beforehand. This emulates the potential use in clinic. Estimating the pose solely relative to the anatomy ensures independence from the considered reference frame. The experimental results demonstrate that our estimations are reliable. The maximum error is 8.77°for phantom (1.1) and 2.77°for real data (1.2) for planes acquired at arbitrary coordinates (Test A) in phantom volume thanks to a continuous rotation representation. Besides, the maximum error is lower for planes acquired around the annotated TV SP (Test B) in all the considered cases. This suggests that a selective denser sampling of planes in areas of interest, automatically controllable in our Unity 3D environment, guarantees more reliable performance. The regression CNN can reliably localise US planes within the fetal brain in phantom data and successfully generalises pose regression to an unseen fetal brain with- out the need for real ground truth data in real-time or 3D volume scans of the patient beforehand.

Given the promising results obtained in the transventricular plane further development is ongoing into volumes of the whole fetus to assess its potential for vision-based, freehand US assisted navigation when acquiring fetal SPs. The ultimate development of this concept would allow for navigation of the US probe without tracking apparatus at all as methodology would extrapolate the position of the US probe based on the incoming video stream from the US system.

6.12.3 Combined Automatic Image Classification and Probe Localisation.

High quality US requires the trainee to appreciate the appearance of fetal anatomy on an ultrasound display. The trainee needs to appreciate the difference in appearance between the anatomy that they are visualising in the present compared to the anatomy they want to see. They must also appreciate how to move the probe to make this happen. Contemporary high fidelity ultrasound simulators, such as “ScanTrainer”, require the trainee to submit as close an approximation of the SP as they can, rather than an understanding of how to make that image. Combining mixed reality guidance, AutoFB and POSE would offer significant potential for training of sonographers. Such a combination would allow them to train using clinical equipment, on real patients in a clinical setting at an earlier stage in their clinical career. By enabling CAL-Tutor to work on real-time video streams, it would be possible to build a system which can identify the likely location of the SP of interest, navigate the user to that location, then automatically classify, measure and perform quality assurance on that image. The ability to navigate the probe and interpret clinical images would represent significant novelty beyond the commercially available training systems of today.

6.13 Discussion

Obstetric ultrasound is a valuable technique for antenatal, intrapartum and postnatal care. There are multiple barriers to gaining skill and experience in obstetric ultrasound during training. In this thesis a mixed reality training environment has been developed to allow trainees to visualise the US probe in relation to the fetus. I hypothesise that this would be valuable because certain views of the fetal anatomy can best be achieved with a specific US probe orientation relative to the fetus. This position is not necessarily related to external landmarks on the maternal abdomen. Regardless of previous clinical background, the training challenge remains. The fetus is not fixed in space relative to any external anatomical landmark. Each trainee must understand how the 2-D image presented by the US screen represents the 3-D anatomy of the fetus they are scanning. This work represents an initial feasibility study. I show that mixed reality can offer new ways of learning relevant clinical skills. There is potential to provide, consistent direction using single vector instructions. Training is delivered using consistent and constant logic for the instructions.

Beyond probe navigation I envisage the development of an intelligent mixed reality assistance system. Image optimisation and image quality are scored against a validated multi-centre scoring system. Progress is recorded in an individual digital logbook, accessible to the trainee and the trainer. This would be able to monitor the user's behaviour longitudinally and tutor accordingly. User motion and eye gaze attention data, collected from many different users and gathered during ultrasound plane navigation, can be used to create a database that serves as the foundation for machine learning algorithms. These would classify the user's training requirements.

For example, novice users can be distinguished from expert users and mixed reality assistance can be offered accordingly. We believe that machine learning concepts that are based on 3D user data in a mixed reality environment offer exciting new ways of understanding user behaviour and allow for tailored support in order to achieve unified and standardized learning outcomes.

6.14 Conclusion

In this chapter I have detailed the methodology associated with the development of the mixed reality training system. The initial literature review identified deficiencies with training engagement, inconsistent curricula and validity of simulator-based training. Attempting to avoid these issues, a scoping exercise was carried out to identify areas that trainees themselves perceive as important and areas in which they desired further training. A combined approach of on-line learning for knowledge and mixed-reality based learning for skills followed from this. Ultimately, CAL-Tutor aimed to improve trainee image optimisation as well as giving trainees a better understanding of the three-dimensional relationship between the US probe and the fetal anatomy. CAL-Tutor therefore utilises a combination of self-directed online learning, direct one-to-one assessment and feedback in addition to a mixed reality training environment to deliver this training package.

The presented results suggest that a mixed reality environment utilising a HoloLens2 Head Mounted Display is useful for training in fetal biometry. There was wide acceptance of the device amongst the users. 78% of our user group felt that using the system would increase their confidence when performing fetal biometry. 85% of respondents felt that CAL-Tutor helped them to understand the relationship between the US probe and fetal anatomy. In particular, this could allow trainees to understand the spatial relationship between the US probe and the fetal anatomy.

Future work in this area will focus combining the on-screen instructions given to trainees with a real-time image classification architecture. We believe that we can train novice operators to imitate expert-like behaviours more quickly than a conventional apprentice model

I present a simulation environment that provides real-time feedback on the US probe position relative to a fetal model and establishes a hand/probe motion plan towards aligning the scan with the desired target image.

7 Summary & Possibilities for future Work

In this thesis I have explored the current state of the art applications of simulation in obstetric ultrasound. I developed novel technology for training obstetric ultrasound at the patient bedside and showed how these might be developed for use in real-time ultrasound scanning. The thesis does this by presenting an initial literature review identifying gaps in research knowledge concerning training in obstetric ultrasound. The literature review has been peer reviewed and published in *Simulation in Healthcare*[51]. In this thesis, I propose an objective measure for the assessment of skill and develop a novel, mixed reality trainer which could be more easily integrated into the clinical workflow and commitments of trainee sonographers. I have identified that there is significant deviation in the training, the assessment and the accreditation of sonographers in the developed world. I have proposed a series of metrics by which operators performance could be assessed in an objective way, including DSJ which has not previously been used in the assessment of ultrasound performance. The results of these studies have been published in *Prenatal Diagnosis*[55] and presented at the *ISUOG* Annual world congress[95]. I developed a novel guidance and training package to assist trainee sonographers to understand the complex spatial relationship between the physical location of the US probe on the maternal abdomen and fetal anatomy. The initial utility and usability studies have been presented at the RCOG World Congress 2022[96] and IPCAI 2022[97]. The data provides promising data for the direction of future research.

This work is important because it allows researchers to understand that training in ultrasound often follows an apprentice-type model and this results in disjointed training structures and certification requirements globally. Data presented in this thesis confirms that operator skill can be quantified objectively. Assessment should move from poorly predictive outcomes like image quality or time to complete a scan, to focus on trainee behaviours as they progress through training. Further, this work has shown that the way in which a trainee sonographer moves the probe evolves over time, a learning curve for these skills can be produced. Future research should investigate how to modify the slope of these learning curves and to study the extent to which training can effect them.

7.1 Metrics for the assessment of ultrasound performance

The requirement for objective metrics and validated approaches to training arose from the paucity of evidence highlighted by the literature review. Validated metrics for the assessment of trainee progress were not previously investigated in either the simulated or clinical environment. In this thesis, progress has been made towards establishing objective measures of operator performance, the most promising of which is DSJ. I have conducted experiments in both simulated and in clinical environments with promising results. A manuscript based on the phantom data has been published in *Prenatal Diagnosis*[55]*.* Abstracts reporting the methodology and results in a clinical setting have been accepted for oral and poster presentations at international conferences both in computer science and medical disciplines. The data suggests that operator hand motion and the fluidity of the hand movement evolves with experience and familiarity of the task being performed, in this case obstetric ultrasound. The rate at which these behavioural changes are seen is likely to be variable between individuals. A larger sample of trainees would be required to validate this hypothesis. Ideally, a cohort would be followed through their initial training and clinical practice. A power calculation based on the data collected to date estimates that this sample size would be in the region of 50 individuals. The results of such a study might provide insight into the number of scans, or hours required for trainees to move from a training environment where they are under constant instruction and close supervision to a clinical environment where direct, 1:1 supervision is not required. Here, the trainee can build their clinical confidence and abilities and move towards independent practice.

Establishing this transition point would allow training curricula to be based on robust, reproducible and clinically relevant learning outcomes which are clear to the trainee prior to commencing training and attainable in a reasonable period of time.

From the work completed to date I have identified weaknesses in the current approach to ultrasound training and explored how these contribute to trainee frustration and dissatisfaction with training. Further, I have established that training may affect a clinicians' skills in the long term. To counteract these, I have established a hypothesis for objective assessment of operator performance. I have collaborated with colleagues in engineering and surgical sciences to progress the vision of a computer-based tutor system for obstetric ultrasound. I have developed an understanding of how objective metrics of performance evolve with training. Together, we investigated the effect of integrating these metrics with the state-of-the-art hardware to address the gaps in training equipment and curricula which were identified in early chapters.

7.2 Mixed Reality Training for Obstetric Ultrasound

To build a mixed reality ultrasound trainer that can be used in for training, there were three key areas of work. These were Hardware, Software and Clinical Evaluation. Significant effort was spent considering the relative strengths and weakness of currently available commercial US simulators and how they are validated. We considered both head mounted displays and conventional monitors for this project, ultimately selecting the Hololens2. Software was developed to utilise the ability of Hololens2 to track objects using the on-board cameras. This data was combined with the known position of the fetus to present guidance arrows to the user. The system provides trainees with information of the position of the fetus relative to the US probe. Instruction and guidance by means of suggested movements were given to allow the trainee to navigate to standard planes used for assessment of fetal weight. Once a standard plane had been reached it was possible for the trainee to use the clinical US system to zoom the image, alter the gain (exposure) of the image and use the in-built calliper functions to estimate the fetal weight, just as they would in clinical practice. An initial iteration of the system was introduced to clinicians together with the associated learning material. It is in the refinement of this system and by observing its use by trainees that the scale of potential benefit can be identified and future evaluations in clinical practice planned.

7.3 Future Work – Monitoring & Grading Trainee Progress

From the initial trainee questionnaire it was clear that doctors in obstetrics and gynaecology feel the training available to them in ultrasound does not meet their educational needs. By extension many trainees do not feel confident or that they achieve competence in the use of ultrasound. Trainees reported that basic tasks including how to use the functions of an US scanner, including Doppler ultrasound, optimising and saving images were not clear to them. How to handle the probe, low-level maintenance of the machine such as cleaning and archiving of images were rarely considered by trainees. As I have previously identified, poor basic training can lead to sub-optimal performance for the remainder of one's clinical career. For these reasons it is imperative to consider the broader training landscape in obstetrics & gynaecology and how a mixed reality system might be deployed and used by trainees outside research setting.

The CAL-Tutor system automatically saves probe tracking data and a video of the ultrasound performed by the trainee each time they use the system. The performance metrics previously described can be calculated from this data. Thus, a database of performance can be built up over time. The data would include probe path length, time to completion of task and DSJ. The trainee will also save what they perceive to be the best possible image they can obtain for each of the three planes required to calculate fetal weight. The CAL-Tutor system will score the images submitted by the trainee against the standardised scoring system developed by the INTERGROWTH-21 study, which has been validated in a multi-centre study. Conventional analysis of these metrics could be used as objective measure of the learning outcomes. However, machine learning methodologies may be better suited to this task and identify previously unconsidered methodologies for assessment, such as eye tracking.

CAL-Tutor therefore offers several innovations compared to standard US training. Aside from visualising the relative positions of the fetal anatomy and the ultrasound beam the trainee is provided with instruction on how to move their probe in the most efficient way and given feedback on elements of the captured image which could be improved, as compared to the 'ideal', or 'perfect' image.

7.4 Conclusion

In conclusion this thesis has developed a novel metric for the assessment of trainee performance and a mixed reality platform with associated training materials. Dimensionless squared jerk was shown to differentiate between expert and novice operators and shown to achieve this in simulated and clinical environments. Initial work to track the evolution of DSJ over the initial period of training was also promising and rudimentary learning curves have been established. The development of the mixed reality trainer, CAL-Tutor opens new ways to train sonographers and brings consistent training language, instructions and training goals. Subject to further development and integration of machine learning and computer vision techniques it is possible that the methods could be used to provide real-time guidance and direction to trainees while they are performing ultrasound in the clinical environment. The ultimate realisation of this thesis would couple DSJ with mixed reality hardware to assess trainee competence in an objective and standardised way. This could improve trainee experience and allow trainees to train remotely from a training centre, at the patient bedside. This may be of particular advantage in remote centres, possibly even in the developing world. This technology could be deployed anywhere access to training is limited by financial or human resources. These methodologies give educators the ability to roll-out standardised training programmes with practical, clinically relevant educational goals for the identification and assessment of aberrant fetal growth, which is a feature in up to approximately one third of pregnancies, approximating to 30 million annual cases[98].

8 Bibliography

1. Abuhamad A, Minton KK, Benson CB, et al. Obstetric and gynecologic ultrasound curriculum and competency assessment in residency training programs: consensus report. *Ultrasound Obs Gynecol*. 2018;51:150-155. doi:10.1002/uog.18967

2. Salomon LJ, Alfirevic Z, Da Silva Costa F, et al. ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. *Ultrasound Obstet Gynecol*. 2019;53(6):715-723. doi:10.1002/uog.20272

3. Ling EW, Sosuan LC, Hall JC. Congenital anomalies: an increasingly important cause of mortality and workload in a neonatal intensive care unit. *Am J Perinatol*. 1991;8(3):164-169. doi:10.1055/S-2007-999369

4. Core Curriculum for Obstetrics & Gynaecology.

5. College of Obstetricians R. Obstetrics and Gynaecology Workforce Report 2017. 2017. www.rcog.org.uk/workforce2017. Accessed December 3, 2021.

6. Harries RL, Rashid M, Smitham P, et al. What shape do UK trainees want their training to be? Results of a cross-sectional study. *BMJ Open*. 2016;6(10). doi:10.1136/bmjopen-2015-010461

7. Harries RL, Williams AP, Ferguson HJMM, Mohan HM, Beamish AJ, Gokani VJ. The future of surgical training in the context of the 'Shape of Training' Review: Consensus recommendations by the Association of Surgeons in Training. *Int J Surg*. 2016;36:S5-S9. doi:10.1016/j.ijsu.2016.08.238

8. Ali MR, Tichansky DS, Kothari SN, et al. Validation that a 1-year fellowship in minimally invasive and bariatric surgery can eliminate the learning curve for laparoscopic gastric bypass. *Surg Endosc*. 2010;24(1):138-144. doi:10.1007/s00464-009-0550-z

9. Royal College of Obstetrics & Gynaecology Working Party report. *Working Party Report Tomorrow's Specialist*.; 2012. www.cla.co.uk]. Accessed October 23, 2018.

10. Jaffer A, Bednarz B, Challacombe B, Sriprasad S. The assessment of surgical competency in the UK. *Int J Surg*. 2009;7(1):12-15. doi:10.1016/j.ijsu.2008.10.006

11. isuog. org RECOMMENDATIONS ISUOG Education Committee recommendations for basic training in obstetric and gynecological ultrasound. doi:10.1002/uog.13208

12. RCOG. *Module 1: Basic: Early pregnancy (8-12 weeks) Ultrasound*.; 2013. https://www.rcog.org.uk/globalassets/documents/careers-and-training/ultrasound/ultrasound-modules/us_module1_curriculum.pdf. Accessed September 26, 2019.

13. D. C, H. P, E. M, N. W, K. J, A. G. Simulation training-Trainees want it but don't use it: A study by Midlands Research Collaborative in Obstetrics & Gynaecology. *BJOG An Int J Obstet Gynaecol*. 2016. doi:http://dx.doi.org/10.1111/1471-0528.14110

14. Chalouhi GE, Bernardi V, Gueneuc A, Houssin I, Stirnemann JJ, Ville Y. Evaluation of trainees' ability to perform obstetrical ultrasound using simulation: challenges and opportunities. 2016. doi:10.1016/j.ajog.2015.10.932

15. Howells NR, Auplish S, Hand GC, Gill HS, Carr AJ, Rees JL. Retention of arthroscopic shoulder skills learned with use of a simulator: Demonstration of a learning curve and loss of performance level after a time delay. *J Bone Jt Surg - Ser A*. 2009;91(5):1207-1213. doi:10.2106/JBJS.H.00509

16. Jackson WFM, Khan T, Alvand A, et al. Learning and retaining simulated arthroscopic meniscal repair skills. *J Bone Jt Surg - Ser A*. 2012;94(17):e132(1). doi:10.2106/JBJS.K.01438

17. Hogan N, Sternad D. Sensitivity of Smoothness Measures to Movement Duration, Amplitude, and Arrests. *J Mot Behav*. 2009;41(6):529-534. doi:10.3200/35-09-004-RC

18. van Empel PJ, van Rijssen LB, Commandeur JP, et al. Objective versus Subjective Assessment of Laparoscopic Skill. *ISRN Minim Invasive Surg*. 2013;2013:1-5. doi:10.1155/2013/686494

19. van Rij AM, McDonald JR, Pettigrew RA, Putterill MJ, Reddy CK, Wright JJ. Cusum as an aid to early assessment of the surgical trainee. *Br J Surg*. 1995;82(11):1500-1503. doi:10.1002/bjs.1800821117

20. Howells NR, Brinsden MD, Gill RS, Carr AJ, Rees JL. Motion Analysis: A Validated Method for Showing Skill Levels in Arthroscopy. *Arthrosc - J Arthrosc Relat Surg*. 2008;24(3):335-342. doi:10.1016/j.arthro.2007.08.033

21. Jung K, Kang DJ, Kekatpure AL, Adikrishna A, Hong J, Jeon IH. A new wide-angle arthroscopic system: a comparative study with a conventional 30{${}^\circ$} arthroscopic system. *Knee Surgery, Sport Traumatol Arthrosc*. 2016;24(5):1722-1729. doi:10.1007/s00167-015-3967-z

22. Balasubramanian S, Melendez-Calderon A, Roby-Brami A, Burdet E. On the analysis of movement smoothness. *J Neuroeng Rehabil*. 2015;12(1):112. doi:10.1186/s12984-015-0090-9

23. Flash T, Hogan N. The coordination of arm movements: an experimentally confirmed mathematical model. *J Neurosci*. 1985;5(7):1688-1703. doi:10.1523/JNEUROSCI.05-07-01688.1985

24. Gulde P, Hermsdörfer J. Smoothness Metrics in Complex Movement Tasks. *Front Neurol*. 2018;9(SEP):615. doi:10.3389/fneur.2018.00615

25. Javaux A, Bouget D, Gruijthuijsen C, et al. A mixed-reality surgical trainer with comprehensive sensing for fetal laser minimally invasive surgery. *Int J Comput Assist Radiol Surg*. July 2018:1-9. doi:10.1007/s11548-018-1822-7

26. Brydges R, Hatala R, Zendejas B, Erwin PJ, Cook DA. Linking simulation-based educational assessments and patient-related outcomes: A systematic review and meta-analysis. *Acad Med*. 2015. doi:10.1097/ACM.0000000000000549

27. Tolsgaard MG. A multiple-perspective approach for the assessment and learning of ultrasound skills. *Perspect Med Educ*. 2018;7(3):211-213. doi:10.1007/s40037-018-0419-8

28. Bell CR, McKaigney CJ, Holden M, Fichtinger G, Rang L. Sonographic Accuracy as a Novel Tool for Point-of-care Ultrasound Competency Assessment. Uijtdehaage S, ed. *AEM Educ Train*. 2017;1(4):316-324. doi:10.1002/aet2.10064

29. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Ann Intern Med*. 2009;151(4):264. doi:10.7326/0003-4819-151-4-200908180-00135

30. Ruano J, Gómez-García F, Gay-Mimbrera J, et al. Evaluating characteristics of PROSPERO records as predictors of eventual publication of non-Cochrane systematic reviews: a meta-epidemiological study protocol. *Syst Rev*. 2018;7(1):43. doi:10.1186/s13643-018-0709-6

31. Huang X, Lin J, Demner-Fushman D. *Evaluation of PICO as a Knowledge Representation for Clinical Questions*. http://www.fpin.org/. Accessed October 29, 2018.

32. Burden C, Preshaw J, White P, Draycott TJ, Grant S, Fox R. Usability of virtual-reality simulation training in obstetric ultrasonography: a prospective cohort study. *Ultrasound Obstet Gynecol*. 2013;42(2):213-217. doi:10.1002/uog.12394

33. Todsen T, Jensen ML, Tolsgaard MG, et al. Transfer from point-of-care Ultrasonography training to diagnostic performance on patients—a randomized controlled trial. *Am J Surg*. 2016;211(1):40-45. doi:10.1016/J.AMJSURG.2015.05.025

34. Pittini R, Oepkes D, Macrury K, Reznick R, Beyene J, Windrim R. Teaching invasive perinatal procedures: assessment of a high fidelity simulator-based curriculum. *Ultrasound Obstet Gynecol*. 2002;19(5):478-483. doi:10.1046/j.1469-0705.2002.00701.x

35. Jensen JK, Dyre L, Jørgensen ME, Andreasen LA, Tolsgaard MG. Simulation-based point-of-care ultrasound training: a matter of competency rather than volume. *Acta Anaesthesiol Scand*. 2018;62(6):811-819. doi:10.1111/aas.13083

36. Madsen ME, Nørgaard LN, Tabor A, Konge L, Ringsted C, Tolsgaard MG. The Predictive Value of Ultrasound Learning Curves Across Simulated and Clinical Settings. *J Ultrasound Med*. 2017;36(1):201-208. doi:10.7863/ultra.16.01037

37. Monsky WL, Levine D, Mehta TS, et al. Using a Sonographic Simulator to Assess Residents Before Overnight Call. *Am J Roentgenol*. 2002;178(1):35-39. doi:10.2214/ajr.178.1.1780035

38. Maul H, Scharf A, Baier P, et al. Ultrasound simulators: experience with the SonoTrainer and comparative review of other training systems. 2004;24(5):581-585. doi:10.1002/uog.1119

39. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association Between Funding and Quality of Published Medical Education Research. *JAMA*. 2007;298(9):1002. doi:10.1001/jama.298.9.1002

40. The Nordic Cochrane Centre, The Cochrane Collaboration C. Review Manager (RevMan) [Computer program], version 5.3. 2014.

41. Suurmond R, van Rhee H, Hak T. Introduction, comparison, and validation of *Meta-Essentials*: A free and simple tool for meta-analysis. *Res Synth Methods*. 2017;8(4):537-553. doi:10.1002/jrsm.1260

42. Chalouhi GE, Bernardi V, Gueneuc A, Houssin I, Stirnemann JJ, Ville Y. Evaluation of trainees' ability to perform obstetrical ultrasound using simulation: challenges

and opportunities. *Am J Obstet Gynecol*. 2016;214(4):525.e1-525.e8. doi:10.1016/j.ajog.2015.10.932

43. Pajkrt E, Mol BWJWJ, Boer K, Drogtrop APP, Bossuyt PMMMM, Bilardo CMM. Intra- and interoperator repeatability of the nuchal translucency measurement. *Ultrasound Obstet Gynecol*. 2000;15(4):297-301. doi:10.1046/j.1469-0705.2000.00088.x

44. Campo R, Puga M, Meier Furst R, Wattiez A, De Wilde RL. Excellence needs training "Certified programme in endoscopic surgery". *Facts, views {\&} Vis ObGyn*. 2014;6(4):240-244. http://www.ncbi.nlm.nih.gov/pubmed/25593700. Accessed February 13, 2020.

45. Kholinne E, Gandhi MJ, Adikrishna A, et al. The Dimensionless Squared Jerk: An Objective Parameter That Improves Assessment of Hand Motion Analysis during Simulated Shoulder Arthroscopy. *Biomed Res Int*. 2018;2018:1-8. doi:10.1155/2018/7816160

46. Wyatt HJ. Detecting saccades with jerk. *Vision Res*. 1998;38(14):2147-2153. doi:10.1016/S0042-6989(97)00410-0

47. *European Training Requirements in Obstetrics and Gynaecology UEMS Section Obstetrics and Gynaecology / European Board and College of Obstetrics and Gynaecology Standing Committee on Training and Assessment*.

48. Westerway SC. Estimating fetal weight for best clinical outcome. *Australas J ultrasound Med*. 2012;15(1):13-17. doi:10.1002/j.2205-0140.2012.tb00136.x

49. Salomon LJ, Winer N, Bernard JP, Ville Y. A score-based method for quality control of fetal images at routine second-trimester ultrasound examination. *Prenat Diagn*. 2008;28(9):822-827. doi:10.1002/pd.2016

50. Takada K, Yashiro K, Takagi M. Reliability and sensitivity of jerk-cost measurement for evaluating irregularity of chewing jaw movements. *Physiol Meas*. 2006;27(7):609-622. doi:10.1088/0967-3334/27/7/005

51. Dromey BP, Peebles DM, Stoyanov D V. A Systematic Review and Meta-analysis of the Use of High-Fidelity Simulation in Obstetric Ultrasound. *Simul Healthc*. July 2020. doi:10.1097/sih.0000000000000485

52. Flin R, Maran N. Identifying and training non-technical skills for teams in acute medicine. *Qual Saf Heal Care*. 2004;13(SUPPL. 1). doi:10.1136/qshc.2004.009993

53. Eich C, Timmermann A, Russo SG, et al. Simulator-based training in paediatric anaesthesia and emergency medicine – Thrills, skills and attitudes. *Br J Anaesth*. 2007;98(4):417-419. doi:10.1093/bja/aem051

54. Brunyé TT, Gardony AL. Eye tracking measures of uncertainty during perceptual decision making. *Int J Psychophysiol*. 2017;120:60-68. doi:10.1016/J.IJPSYCHO.2017.07.008

55. Dromey BP, Ahmed S, Vasconcelos F, et al. Dimensionless squared jerk: An objective differential to assess experienced and novice probe movement in obstetric ultrasound. *Prenat Diagn*. November 2020:pd.5855. doi:10.1002/pd.5855

56. Judy AD. *FireScholars A STUDY OF FLIGHT SIMULATION TRAINING TIME,*

AIRCRAFT TRAINING TIME, AND PILOT COMPETENCE AS MEASURED BY THE NAVAL STANDARD SCORE. https://firescholars.seu.edu/coe/22. Accessed June 25, 2020.

57. Taylor JL, Kennedy Q, Noda A, Yesavage JA. Pilot age and expertise predict flight simulator performance: A 3-year longitudinal study. *Neurology*. 2007;68(9):648-654. doi:10.1212/01.wnl.0000255943.10045.c0

58. Burden C, Preshaw J, White P, Draycott TJ, Grant S, Fox R. Validation of Virtual Reality Simulation for Obstetric Ultrasonography. *Simul Healthc J Soc Simul Healthc*. 2012;7(5):269-273. doi:10.1097/SIH.0b013e3182611844

59. Salomon LJ, Alfirevic Z, Berghella V, et al. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound Obstet {\&} Gynecol*. 2011;37(1):116-126. doi:10.1002/uog.8831

60. Hopper AN, Jamison MH, Lewis WG. Learning curves in surgical practice. *Postgrad Med J*. 2007;83(986):777. doi:10.1136/PGMJ.2007.057190

61. Madsen ME, Konge L, Nørgaard LN, et al. Assessment of performance measures and learning curves for use of a virtual-reality ultrasound simulator in transvaginal ultrasound examination. *Ultrasound Obstet Gynecol*. 2014;44(6):693-699. doi:10.1002/uog.13400

62. *OSATS*.; 2013.

63. Villar J, Ismail LC, Victora CG, et al. International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet*. 2014;384(9946):857-868. doi:10.1016/S0140-6736(14)60932-6

64. *2018 Training Data Analysis Topic: Ultrasound Training Background*.

65. Baruch Y, Holtom BC. Survey response rate levels and trends in organizational research: *https://doi.org/101177/0018726708094863*. 2008;61(8):1139-1160. doi:10.1177/0018726708094863

66. OP20.05: Challenges to ultrasound training in obstetrics and gynecology: a survey across three Scandinavian countries - Tolsgaard - 2013 - Ultrasound in Obstetrics & Gynecology - Wiley Online Library. https://obgyn.onlinelibrary.wiley.com/doi/full/10.1002/uog.12894. Accessed August 24, 2020.

67. Henwood PC, Mackenzie DC, Liteplo AS, et al. Point-of-Care Ultrasound Use, Accuracy, and Impact on Clinical Decision Making in Rwanda Hospitals. *J Ultrasound Med*. 2017;36(6):1189-1194. doi:10.7863/ULTRA.16.05073

68. Novak K, Tanyingoh D, Petersen F, et al. Clinic-based Point of Care Transabdominal Ultrasound for Monitoring Crohn's Disease: Impact on Clinical Decision Making. *J Crohn's Colitis*. 2015;9(9):795-801. doi:10.1093/ECCO-JCC/JJV105

69. Herron J. Augmented Reality in Medical Education and Training. *http://dx.doi.org/101080/1542406520161175987*. 2016;13(2):51-55. doi:10.1080/15424065.2016.1175987

70. We found 7 critical HoloLens details that Microsoft hid inside its developer docs. https://www.pcworld.com/article/419869/we-found-7-critical-hololens-details-that-microsoft-hid-inside-its-developer-docs.html. Accessed January 4, 2022.

71. Microsoft HoloLens May Cause Discomfort As It Gets Extremely Hot. https://infinityleap.com/microsoft-hololens-may-cause-discomfort-as-it-gets-extremely-hot/. Accessed January 4, 2022.

72. Chu MWA, Moore J, Peters T, et al. Augmented reality image guidance improves navigation for beating heart mitral valve repair. *Innovations (Phila)*. 2012;7(4):274-281. doi:10.1097/IMI.0B013E31827439EA

73. Rahmatullah B, Papageorghiou AT, Noble JA. *LNCS 7512 - Integration of Local and Global Features for Anatomical Object Detection in Ultrasound*. Vol 7512.; 2012. https://link.springer.com/content/pdf/10.1007%2F978-3-642-33454-2_50.pdf. Accessed October 23, 2018.

74. Barsom EZ, Graafland M, Schijven MP. Systematic review on the effectiveness of augmented reality applications in medical training. *Surg Endosc 2016 3010*. 2016;30(10):4174-4183. doi:10.1007/S00464-016-4800-6

75. Blum T, Heining SM, Kutter O, Navab N. Advanced training methods using an augmented reality ultrasound simulator. *Sci Technol Proc - IEEE 2009 Int Symp Mix Augment Reality, ISMAR 2009*. 2009:177-178. doi:10.1109/ISMAR.2009.5336476

76. Tang KS, Cheng DL, Mi E, Greenberg PB. Augmented reality in medical education: a systematic review. *Can Med Educ J*. 2020;11(1):e81. doi:10.36834/CMEJ.61705

77. Ebner F, De Gregorio A, Schochter F, Bekes I, Janni W, Lato K. Effect of an Augmented Reality Ultrasound Trainer App on the Motor Skills Needed for a Kidney Ultrasound: Prospective Trial. *JMIR Serious Games*. 2019;7(2). doi:10.2196/12713

78. Huff M, Maurer AE, Merkt M. Producing gestures establishes a motor context for procedural learning tasks. *Learn Instr*. 2018;58:245-254. doi:10.1016/J.LEARNINSTRUC.2018.07.008

79. Singer MA, Goldin-Meadow S. Children learn when their teacher's gestures and speech differ. *Psychol Sci*. 2005;16(2):85-89. doi:10.1111/j.0956-7976.2005.00786.x

80. Watson HA, Carter J, Seed PT, Tribe RM, Shennan AH. The QUiPP App: a safe alternative to a treat-all strategy for threatened preterm labor. *Ultrasound Obstet Gynecol*. 2017;50(3):342-346. doi:10.1002/UOG.17499

81. Schlosser PD, Grundgeiger T, Sanderson PM, Happel O. An exploratory clinical evaluation of a head-worn display based multiple-patient monitoring application: impact on supervising anesthesiologists' situation awareness. *J Clin Monit Comput*. 2019;33(6):1119-1127. doi:10.1007/S10877-019-00265-4

82. Stanton NA, Plant KL, Roberts AP, Allison CK, Howell M. Seeing through the mist: an evaluation of an iteratively designed head-up display, using a simulated degraded visual environment, to facilitate rotary-wing pilot situation awareness and workload. *Cogn Technol Work*. 2020;22(3):549-563. doi:10.1007/S10111-019-00591-2/FIGURES/13

83. Tepper OM, Rudy HL, Lefkowitz A, et al. Mixed Reality with HoloLens: Where Virtual

Reality Meets Augmented Reality in the Operating Room. 2017;140(5):1066-1070. doi:10.1097/PRS.0000000000003802

84. Mahmood F, Mahmood E, Dorfman RG, et al. Augmented Reality and Ultrasound Education: Initial Experience. *J Cardiothorac Vasc Anesth*. 2018;32(3):1363-1367. doi:10.1053/J.JVCA.2017.12.006

85. Kaneko N, Sato M, Takeshima T, Sehara Y, Watanabe E. Ultrasound-guided central venous catheterization using an optical see-through head-mounted display: A pilot study. *J Clin Ultrasound*. 2016;44(8):487-491. doi:10.1002/JCU.22374

86. Tolsgaard MG, Ringsted C, Dreisler E, et al. Sustained effect of simulation-based ultrasound training on clinical performance: a randomized trial. *Ultrasound Obstet Gynecol*. 2015;46(3):312-318. doi:10.1002/uog.14780

87. *Fetal Anomaly Fetal Anomaly Screening Programme Programme Handbook*.; 2015. www.gov.uk/topic/population-screening-programmes. Accessed August 14, 2018.

88. What's new in Unity 2020.3.14 - Unity. https://unity3d.com/unity/whats-new/2020.3.14. Accessed November 16, 2021.

89. Releases · microsoft/MixedRealityToolkit-Unity · GitHub. https://github.com/Microsoft/MixedRealityToolkit-Unity/releases. Accessed November 16, 2021.

90. Qian L, Deguet A, Kazanzides P. ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery. *Healthc Technol Lett*. 2018;5(5):194-200. doi:10.1049/HTL.2018.5065

91. Kogan JR, Hatala R, Hauer KE, Holmboe E. Guidelines: The do's, don'ts and don't knows of direct observation of clinical skills in medical education. *Perspect Med Educ*. 2017;6(5):286. doi:10.1007/S40037-017-0376-7

92. Org WA, Ellaway R, Masters K, Ellaway RH. AMEE Guides in Medical Education e-Learning in Medical Education Teaching and Learning e-Learning in Medical Education Part 1: Learning, teaching and assessment Part 2: Technology, management and design The Authors e-Learning in Medical Education. 2008. www.almondtds.com. Accessed November 15, 2021.

93. Bano S, Dromey B, Vasconcelos F, et al. AutoFB: Automating Fetal Biometry Estimation from Standard Ultrasound Planes. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2021;12907 LNCS:228-238. doi:10.1007/978-3-030-87234-2_22/COVER

94. Di Vece C, Dromey B, Vasconcelos F, David AL, Peebles D, Stoyanov D. Deep learning-based plane pose regression in obstetric ultrasound. *Int J Comput Assist Radiol Surg*. 2022;17(5):833-839. doi:10.1007/S11548-022-02609-Z/FIGURES/5

95. Dromey B, Vasconcelos F, Ourselin S, David ALL, Stoyanov D, Peebles D. VP34.01: Dimensionless jerk: an objective differential to assess experienced and novice ultrasound operators in a clinical setting. *Ultrasound Obstet {\&} Gynecol*. 2020;56(S1):196. doi:10.1002/uog.22834

96. World Congress 2022 Programme – RCOG Events. https://rcogevents.com/world-congress-2022-programme/. Accessed June 17, 2022.

97. Program. https://sites.google.com/view/ipcai2022/program. Accessed June 17, 2022.

98. Black RE. Global Prevalence of Small for Gestational Age Births. *Nestle Nutr Inst Workshop Ser*. 2015;81:1-7. doi:10.1159/000365790